

# DIFFERENTIAL TOPOLOGY

**Victor Guillemin**

**Alan Pollack**

*Massachusetts Institute of Technology*

Prentice-Hall, Inc., *Englewood Cliffs, New Jersey*

*Library of Congress Cataloging in Publication Data*

GUILLEMIN, V      date

Differential topology.

Bibliography: p. 212.

I. Differential topology. I. Pollack, Alan, joint  
author. II. Title.

QA613.6.G84      514'.7      74-4115

ISBN 0-13-212605-2

© 1974 by Prentice-Hall, Inc., Englewood Cliffs, New Jersey  
All rights reserved

Prentice-Hall International, Inc., *London*  
Prentice-Hall of Australia, Pty. Ltd., *Sydney*  
Prentice-Hall of Canada, Ltd., *Toronto*  
Prentice-Hall of India Private Limited, *New Delhi*  
Prentice-Hall of Japan, Inc., *Tokyo*

# Contents

Preface ix

Straight Forward to the Student xiii

Table of Symbols xv

## CHAPTER 1

Manifolds and Smooth Maps 1

- §1 DEFINITIONS 1
- §2 DERIVATIVES AND TANGENTS 8
- §3 THE INVERSE FUNCTION THEOREM AND, IMMERSIONS 13
- §4 SUBMERSIONS 20
- §5 TRANSVERSALITY 27
- §6 HOMOTOPY AND STABILITY 33
- §7 SARD'S THEOREM AND MORSE FUNCTIONS 39
- §8 EMBEDDING MANIFOLDS IN EUCLIDEAN SPACE 48

## CHAPTER 2

### Transversality and Intersection 57

- §1 MANIFOLDS WITH BOUNDARY 57
- §2 ONE-MANIFOLDS AND SOME CONSEQUENCES 64
- §3 TRANSVERSALITY 67
- §4 INTERSECTION THEORY MOD 2 77
- §5 WINDING NUMBERS AND THE JORDAN-BROUWER SEPARATION THEOREM 85
- §6 THE BORSUK-ULAM THEOREM 91

## CHAPTER 3

### Oriented Intersection Theory 94

- §1 MOTIVATION 94
- §2 ORIENTATION 95
- §3 ORIENTED INTERSECTION NUMBER 107
- §4 LEFSCHETZ FIXED-POINT THEORY 119
- §5 VECTOR FIELDS AND THE POINCARÉ-HOPF THEOREM 132
- §6 THE HOPF DEGREE THEOREM 141
- §7 THE EULER CHARACTERISTIC AND TRIANGULATIONS 148

## CHAPTER 4

### Integration on Manifolds 151

- §1 INTRODUCTION 151
- §2 EXTERIOR ALGEBRA 153
- §3 DIFFERENTIAL FORMS 162
- §4 INTEGRATION ON MANIFOLDS 165
- §5 EXTERIOR DERIVATIVE 174
- §6 COHOMOLOGY WITH FORMS 178
- §7 STOKES THEOREM 182
- §8 INTEGRATION AND MAPPINGS 188
- §9 THE GAUSS-BONNET THEOREM 194



APPENDIX 1

Measure Zero and Sard's Theorem 202

APPENDIX 2

Classification of Compact One-Manifolds 208

Bibliography 212

Index 217



# Preface

The intent of this book is to provide an elementary and intuitive approach to differential topology. The topics covered are nowadays usually discussed in graduate algebraic topology courses as by-products of the big machinery, the homology and cohomology functors. For example, the Borsuk-Ulam theorem drops out of the multiplicative structure on the cohomology ring of projective space; the Lefschetz theorem comes from Poincaré duality and the Künneth theorem; the Jordan-Brouwer separation theorem follows from Alexander duality; and so on. We have two objections to this big-machinery approach: it obscures the elegant intuitive content of the subject matter, and it gives the student the impression that only big machines can do mathematics. We have attempted to make the results mentioned above and results like them (such as the Gauss-Bonnet theorem, the degree theorem, the Hopf theorem on vector fields) the main topic of our book rather than a mixed bag of interesting examples. In doing so we have abandoned algebraic topology altogether. Our point of view is that these theorems belong in a much more geometric realm of topology, namely intersection theory. Of course, intersection theory, properly done, requires its own apparatus: the transversality theorem. We must confess our sense of vulnerability to the charge that we have replaced one machine with another. Perhaps; *chacun à son goût*. The transversality arguments, it seems to us, can be visualized by the student, something which we feel cannot be honestly said of the singular homology functor.

This book is appropriate for a leisurely first year graduate course. We have also successfully taught a course based on the book to juniors and seniors. For undergraduates we suggest that certain topics be deleted—for example, the discussion of De Rham theory—and that the main emphasis be placed on mod 2 intersection theory, with some of the subjects from Chapter 3 presented mod 2 rather than with orientations. (The reader will notice, incidentally, that the section on De Rham theory is completely independent of the rest of the book. We have avoided using it even in our proof of the degree formula in Chapter 4, Section 8. We prove this formula using Stokes theorem rather than the theory of the top cohomology class, as in the usual treatment.)

The book is divided into four chapters. Chapter 1 contains the elementary theory of manifolds and smooth mappings. We define manifolds as subsets of Euclidean space. This has the advantage that manifolds appear as objects already familiar to the student who has studied calculus in  $\mathbf{R}^2$  and  $\mathbf{R}^3$ ; they are simply curves and surfaces generalized to higher dimensions. We also avoid confusing the student at the start with the abstract paraphernalia of charts and atlases. The most serious objection to working in Euclidean space is that it obscures the difference between properties intrinsic to the manifold and properties of its embedding. We have endeavored to make the student aware of this distinction, yet we have not scrupled to use the ambient space to make proofs more comprehensible. (See, for example, our use of the tubular neighborhood theorem in Chapter 2, Section 3). To provide cohesiveness to the elementary material, we have tried to emphasize the “stable” and “generic” quality of our definitions; whether this succeeds in making the basics more palatable, we leave to the reader’s judgment.

The last two sections of Chapter 1 deal with Sard’s theorem and some applications. Our most important use of Sard is in proving the transversality theorem in Chapter 2, but before doing so we use it to deduce the existence of Morse functions and to establish Whitney’s embedding theorem. (Incidentally, it may seem pointless to prove the embedding theorem since our manifolds already sit in Euclidean space. We feel we have just placed the emphasis elsewhere: does there exist a  $k$ -dimensional manifold so pathological that one cannot find a diffeomorphic copy of it inside Euclidean space of specified dimension  $N$ ? Answer: no, provided  $N \geq 2k + 1$ .)

Chapter 2 begins by adding boundaries to manifolds. We classify one-manifolds and present Hirsch’s proof of the Brouwer fixed-point theorem. Then the transversality theorem is derived, implying that transversal intersections are generic. Transversality permits us to define intersection numbers, and the one-manifold classification shows that they are homotopy invariants. At first we do intersection theorem mod 2, so that the student can become familiar with the topology without worrying about orientations. Moreover, mod 2 theory is the natural setting for the last two theorems of the chapter:

the Jordan-Brouwer separation theorem and the Borsuk-Ulam theorem. In each of the last three chapters we have included a section in which the student himself proves major theorems, with detailed guidance from the text. The Jordan-Brouwer separation theorem is the first of these. We found that our students received this enthusiastically, deriving real satisfaction from applying the techniques they had learned to significantly extend the theory.

In Chapter 3 we reconstruct intersection theory to include orientations. The Euler characteristic is defined as a self-intersection number and shown to vanish in odd dimensions. Next a primitive Lefschetz fixed-point theorem is proved and its use illustrated by an informal derivation of the Euler characteristics of compact surfaces. Translated into the vector field context, Lefschetz implies the Poincaré-Hopf index theorem. In an exercise section using the apparatus of the preceding discussions, the student proves the Hopf degree theorem and derives a converse to the index theorem. Finally, we relate the differential Euler characteristic to the combinatorial one.

Chapter 4 concerns forms and integration. The central result is Stokes theorem, which we do essentially as M. Spivak does in his *Calculus on Manifolds*. Stokes is used to prove an elementary but, we believe, largely underrated theorem: the degree formula relating integration to mappings. Finally, from this degree formula we derive the Gauss-Bonnet theorem for hypersurfaces in Euclidean space. Chapter 4 also includes an exercise section in which the student can construct De Rham cohomology and proves homotopy invariance. Although other problems relate cohomology to integration and intersection theory, the subject is treated essentially as an interesting aside to our primary discussion. (In particular, cohomology is not referred to anywhere else in the text.)

The original inspiration for this book was J. Milnor's lovely *Topology from the Differential Viewpoint*. Although our book in its present form involves a larger inventory of topics than Milnor's book, our debt to him remains clear.

We are indebted to Dan Quillen and John Mather for the elegant formulation of the transversality theorem given in Chapter 2, Section 3. We are also grateful to Jim King, Isadore Singer, Frank Warner, and Mike Cowan for valuable criticism, to Dennis Sullivan, Shlomo Sternberg, and Jim Munkres for many helpful conversations, and to Rena Themistocles and Phyllis Ruby for converting illegible scribble into typed manuscript. Most particularly, we are indebted to Barret O'Neill for an invaluable, detailed review of our first draft.

Cambridge, Massachusetts

VICTOR GUILLEMIN  
ALAN POLLACK



# Straight Forward to the Student

This book is written for mathematics students who have had one year of analysis and one semester of linear algebra. Included in the analysis background should be familiarity with basic topological concepts in Euclidean space: openness, connectedness, compactness, etc. We borrow two theorems from analysis which some readers may not have studied: the inverse function theorem, which is used throughout the text; and the change of variable formula for multiple integration, which is needed only for Chapter 4. An excellent reference source for these facts is Spivak [2] (pages 34 and 67 respectively).

The exercises are not only indispensable to understanding the material; they are freely referred to by the text. Those felt to be particularly fundamental—including the ones later referenced—are indicated by asterisks (\*). (In particular, asterisks are not signals to warn students of hard problems.)

Here is a miscellany of terms which are used but not defined in the text.

For maps  $f: X \rightarrow Y$  of two sets: *injective* = one-to-one; *surjective* = onto; *bijective* = injective + surjective = one-to-one and onto.

A collection of sets  $\{U_\alpha\}$  *covers* a set  $X$  if  $X$  is contained in the union  $\bigcup_\alpha U_\alpha$ . An *open cover* (or open covering) of  $X$  is a collection of open sets  $\{U_\alpha\}$  which covers  $X$ . One cover  $\{V_\beta\}$  is a *refinement* of another,  $\{U_\alpha\}$ , if every set  $V_\beta$  is contained in at least one  $U_\alpha$ . By the *second axiom of countability*, every open cover  $\{U_\alpha\}$  in  $\mathbb{R}^n$  has a countable refinement. (*Proof*: Take the

collection of all open balls which are contained in some  $U_\alpha$ , which have rational radii, and which are centered at points having only rational coordinates.)

If  $X$  is a subset contained in  $\mathbf{R}^n$ , then a subset  $V$  of  $X$  is (*relatively*) *open* in  $X$  if it can be written as the intersection of  $X$  with an open subset of  $\mathbf{R}^n$ :  $V = \tilde{V} \cap X$ ,  $\tilde{V}$  open in  $\mathbf{R}^n$ . If  $Z$  is a subset of  $X$ , we can also speak of open covers of  $Z$  in  $X$ , meaning coverings of  $Z$  by relatively open subsets of  $X$ . Every such cover of  $Z$  may be written as the intersection of  $X$  with a covering of  $Z$  by open subsets of  $\mathbf{R}^n$ . Since the second axiom of countability is valid for  $\mathbf{R}^n$ , every open cover of  $Z$  relative to  $X$  has a countable refinement. (Given  $\{U_\alpha\}$ , relatively open in  $X$ , write  $U_\alpha = \tilde{U}_\alpha \cap X$ . Then let  $\tilde{V}_\beta$  be a countable refinement of  $\{\tilde{U}_\alpha\}$  in  $\mathbf{R}^n$ , and define  $V_\beta = \tilde{V}_\beta \cap X$ ).



# Table of Symbols

$\dim$	dimension, p. 4	$H^k$	upper half space (in $\mathbf{R}^k$ ), p. 57
$S^n$	$n$ -sphere, p. 4	$\partial$	boundary, p. 57
$  $	absolute value, p. 4	Int	interior, p. 57
$\times$	product sign, pp. 4, 7	$H_x(X)$	upper half space (in $T_x$ ), p. 63
$\Delta$	diagonal, p. 7	$\vec{n}$	normal vector, pp. 64, 106
graph( )	graph (of a mapping), p. 7	$Y^\epsilon$	$\epsilon$ neighborhood, p. 69
$df_x$	derivative (of $f$ at $x$ ), p. 8	$N(Y)$	normal bundle (of $Y$ in $\mathbf{R}^n$ ), p. 71
$T_x(X)$	tangent space to $X$ at $x$ , p. 9	$N(Z; Y)$	relative normal bundle (of $Z$ in $Y$ ), p. 76
$f^{-1}( )$	preimage (with respect to $f$ ), pp. 21, 27	$\#$	number of points, p. 77
$O(n)$	orthogonal group, p. 22	$I_2$	mod 2 intersection number, pp. 78, 79
$\overline{\cap}$	"is transversal to," p. 28	$\deg_2$	mod 2 degree, p. 80
$\sim$	"is homotopic to," p. 33	$W_2$	mod 2 winding number, pp. 86, 88
$T(X)$	tangent bundle, p. 50	$n_x$	normal vector at $x$ , p. 98
$df$	derivative (on the tangent bundle), p. 50		

$-X$	$X$ with reversed orientation, p. 99	$T^\pi$	$T$ permuted by $\pi$ , p. 155
$X_0, X_1$	boundary components of $I \times X$ , p. 99	Alt	alternation operation, p. 155
$I$	intersection number (oriented) pp. 107, 112, 113	$\Lambda^p$	space of alternating $p$ -tensors, p. 156
deg	degree (of a mapping), p. 108	$\wedge$	wedge product, p. 156
$\chi$	Euler characteristic, p. 116	$\Lambda$	algebra of alternating tensors, p. 159
$L$	global Lefschetz number, p. 119	$A^*$	pull-back (of a tensor), p. 159
$L_x$	local Lefschetz number, p. 121	$f^*\omega$	pull-back (of a form), p. 163
$\vec{v}$	vector field, p. 132	$dx_i$	basis vectors for $p$ -forms, p. 163
ind	index (of a vector field), pp. 133, 134	$\int_X$	integral over $X$ , p. 167
$\phi^*\vec{v}$	pull-back (of a vector field), p. 134	$\oint$	line integral, p. 172
$\overrightarrow{\text{grad}}$	gradient, pp. 140, 177	$d\omega$	exterior derivative, p. 174
$W$	winding number, p. 144	$\sim$	is cohomologous to, p. 179
arg	argument, p. 151	$H^p(X)$	$p$ -th cohomology group, p. 179
$V^p$	$p$ -fold Cartesian product (of $V$ ), p. 153	$f^\#$	induced map on cohomology, p. 179
$\mathfrak{F}^p$	space of all $p$ -tensors, p. 154	$P$	$P$ operator, p. 179
$\otimes$	tensor product operation, p. 154	$d \arg$	argument form, pp. 188, 192
det	determinant, p. 154	$d\theta$	angle form, p. 192
$\phi_i$	basis vectors for the $p$ -fold tensor product, p. 154	$\nu_x$	volume form, p. 194
$S_p$	group of permutations of $\{1, \dots, p\}$ , p. 155	$J_f$	Jacobian of $f$ p. 195
$(-1)^\pi$	sign (of the permutation $\pi$ ), p. 155	$\kappa(x)$	curvature at $x$ , p. 195
		$S(a, b)$	rectangular solid, p. 202

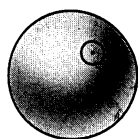
# Manifolds and Smooth Maps

## §1 Definitions

Beautiful, deep insights into the structure and properties of many geometric spaces can be developed intuitively with the aid of a few tools from elementary calculus. Because calculus is built on the local geometry of Euclidean space, it most naturally adapts itself to spaces that locally “look the same” as some Euclidean space. We call such objects *manifolds*, spaces in which the environment of each point is “just like” a small piece of Euclidean space.

The most familiar examples of manifolds are smooth surfaces like the sphere or torus (the crust of a doughnut), where each point lies in a little curved disk that may be gently flattened into a disk in the plane (Figure 1-1). An old friend that does not qualify as a manifold is the cone. Every point but one has a nice Euclidean environment, but no neighborhood of the vertex point looks like a simple piece of the plate. (See Figure 1-2.)

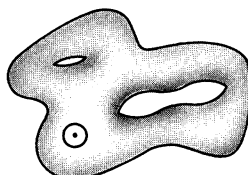
To translate our idea into a mathematical definition, we need to make precise the criterion of “sameness.” We do so in terms of mappings. A mapping  $f$  of an open set  $U \subset \mathbf{R}^n$  into  $\mathbf{R}^m$  is called *smooth* if it has continuous partial derivatives of all orders. However, when the domain of  $f$  is not open, one usually cannot speak of partial derivatives. (Why?) So we adapt the open situation to more general spaces. A map  $f: X \rightarrow \mathbf{R}^m$  defined on an



Sphere



Torus



Smooth surface

Figure 1-1



Cone

Figure 1-2

arbitrary subset  $X$  in  $\mathbf{R}^n$  is called *smooth* if it may be locally extended to a smooth map on open sets; that is, if around each point  $x \in X$  there is an open set  $U \subset \mathbf{R}^n$  and a smooth map  $F : U \rightarrow \mathbf{R}^m$  such that  $F$  equals  $f$  on  $U \cap X$  (Figure 1-3). The shorthand term *local*, referring to behavior only in some neighborhood of a point, is very convenient. Recall that the (relatively) open subsets of  $X$  are exactly those that can be written in the form  $U \cap X$ ,

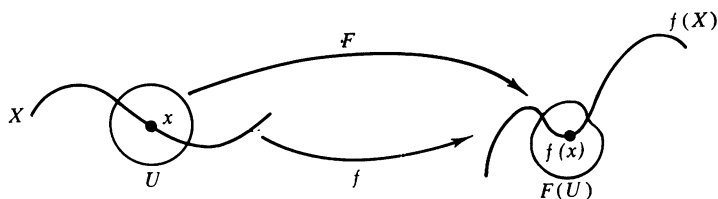


Figure 1-3

where  $U$  is an open set of  $\mathbf{R}^n$ . So smoothness is a local property;  $f: X \rightarrow \mathbf{R}^m$  is smooth if it is smooth in a neighborhood of each point of  $X$ . (Contrasting with “local” is the term *global*, which refers to the whole space  $X$  as a unified object.)

A smooth map  $f: X \rightarrow Y$  of subsets of two Euclidean spaces is a *diffeomorphism* if it is one to one and onto, and if the inverse map  $f^{-1}: Y \rightarrow X$  is also smooth.  $X$  and  $Y$  are *diffeomorphic* if such a map exists. In our philosophy, two diffeomorphic sets are intrinsically equivalent. They may be considered as two copies of a single abstract space, which may happen to be differently situated in their surrounding Euclidean spaces. Soon you will develop sufficient intuition to recognize easily many spaces as diffeomorphic. Perhaps you can begin by thinking about a few pictorial examples (Figure 1-4).

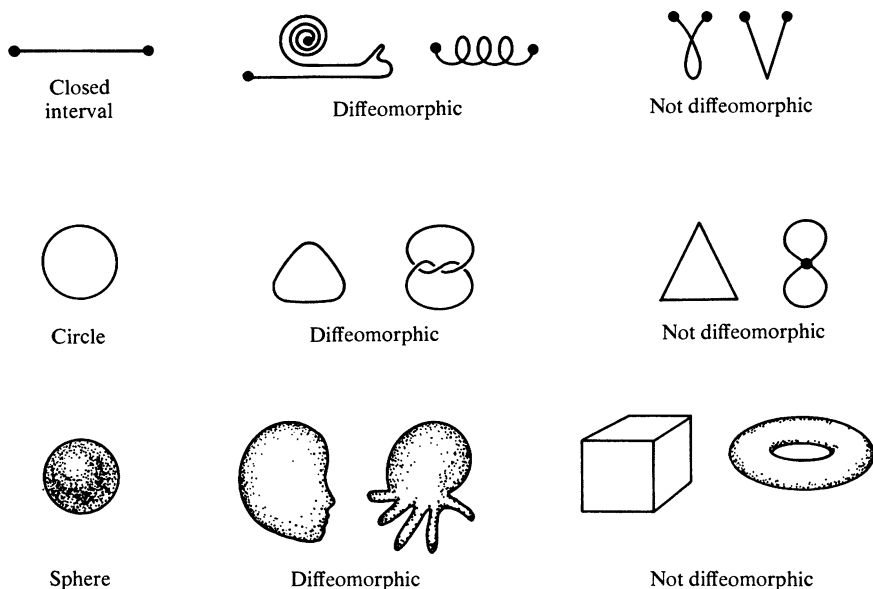


Figure 1-4

With the proper concept of equivalence, we can now define manifolds. Suppose that  $X$  is a subset of some big, ambient Euclidean space  $\mathbf{R}^N$ . Then  $X$  is a  $k$ -dimensional manifold if it is locally diffeomorphic to  $\mathbf{R}^k$ , meaning that each point  $x$  possesses a neighborhood  $V$  in  $X$  which is diffeomorphic to an open set  $U$  of  $\mathbf{R}^k$ . A diffeomorphism  $\phi: U \rightarrow V$  is called a *parametrization* of the neighborhood  $V$ . (Remember,  $V$  must be expressible as  $\tilde{V} \cap X$  for some open set  $\tilde{V} \subset \mathbf{R}^N$ .) The inverse diffeomorphism  $\phi^{-1}: V \rightarrow U$  is called a *coordinate system* on  $V$ . When we write the map  $\phi^{-1}$  in coordinates  $\phi^{-1} =$

$(x_1, \dots, x_k)$ , the  $k$  smooth functions  $x_1, \dots, x_k$  on  $V$  are called *coordinate functions*. (People tend to be a bit sloppy, calling  $x_1, \dots, x_k$  “local coordinates” on  $V$  and writing a typical point of  $V$  as  $(x_1, \dots, x_k)$ . What is really happening is that the  $k$ -tuple of coordinate functions  $(x_1, \dots, x_k)$  is used to identify  $V$  with  $U$  implicitly, and a point  $v \in V$  is identified with its coordinates  $(x_1(v), \dots, x_k(v)) \in U$ . But, after all, we already encounter such ambiguities when speaking of coordinates in Euclidean space.) The dimension  $k$  of  $X$  is often written  $\dim X$ .

As an example, let us show that the circle

$$S^1 = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$$

is a one-dimensional manifold. First, suppose that  $(x, y)$  lies in the upper semicircle where  $y > 0$ . Then  $\phi_1(x) = (x, \sqrt{1 - x^2})$  maps the open interval  $W = (-1, 1)$  bijectively onto the upper semicircle (Fig. 1-5). Its inverse  $(x, y) \rightarrow x$  defined on the upper semicircle is certainly smooth, for it extends to a smooth map of all of  $\mathbf{R}^2$  into  $\mathbf{R}^1$ . Therefore  $\phi_1$  is a parametrization. A parametrization of the lower semicircle where  $y < 0$  is similarly defined by  $\phi_2(x) = (x, -\sqrt{1 - x^2})$ . These two maps give local parametrizations of  $S^1$  around any point except the two axis points  $(1, 0)$  and  $(-1, 0)$ . To cover these points, use  $\phi_3(y) = (\sqrt{1 - y^2}, y)$  and  $\phi_4(y) = (-\sqrt{1 - y^2}, y)$  mapping  $W$  to the right and left semicircles. So we have shown the circle to be a one-dimensional manifold by covering it with four parametrizations. It is not hard to do it with only two. (Can you show that the whole circle cannot be parametrized by a single map?) Use this argument to show more generally that the  $n$ -sphere in  $\mathbf{R}^{n+1}$ ,  $S^n = \{x \in \mathbf{R}^{n+1} : |x| = 1\}$ , is an  $n$ -dimensional manifold. (Here  $|x|$  denotes the usual norm  $\sqrt{x_1^2 + \dots + x_{n+1}^2}$ .)

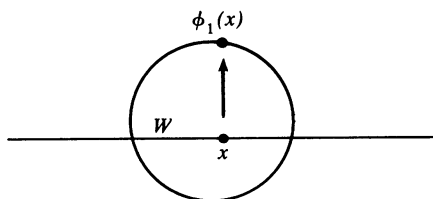


Figure 1-5

A useful method of manufacturing new manifolds from known ones is the cartesian product. Suppose that  $X$  and  $Y$  are manifolds inside  $\mathbf{R}^N$  and  $\mathbf{R}^M$  respectively, so that  $X \times Y$  is a subset of  $\mathbf{R}^M \times \mathbf{R}^N = \mathbf{R}^{M+N}$ . If  $\dim X = k$  and  $x \in X$ , we can find an open set  $W \subset \mathbf{R}^k$  and a local parametrization  $\phi : W \rightarrow X$  around  $x$ . Similarly, if  $\dim Y = l$  and  $y \in Y$ , there is an open set  $U \subset \mathbf{R}^l$  and a local parametrization  $\psi : U \rightarrow Y$  around  $y$ . Define the

map  $\phi \times \psi : W \times U \rightarrow X \times Y$  by the formula

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

Of course,  $W \times U$  is an open set in  $\mathbf{R}^k \times \mathbf{R}^l = \mathbf{R}^{k+l}$ , and it is easy to check that  $\phi \times \psi$  is a local parametrization of  $X \times Y$  around  $(x, y)$ . (Check this point carefully, especially verifying that  $(\phi \times \psi)^{-1}$  is a smooth map on the not-necessarily-open set  $X \times Y \subset \mathbf{R}^{M+N}$ ). Since this map is a local parametrization around an arbitrary point  $(x, y) \in X \times Y$ , we have proved:

**Theorem.** If  $X$  and  $Y$  are manifolds, so is  $X \times Y$ , and  $\dim X \times Y = \dim X + \dim Y$ .

We mention another useful term here. If  $X$  and  $Z$  are both manifolds in  $\mathbf{R}^N$  and  $Z \subset X$ , then  $Z$  is a *submanifold* of  $X$ . In particular,  $X$  is itself a submanifold of  $\mathbf{R}^N$ . Any open set of  $X$  is a submanifold of  $X$ .

The reader should be warned that the slothful authors will often omit the adjective “smooth” when speaking of mappings; nevertheless, smoothness is virtually always intended.

## EXERCISES

1. If  $k < l$  we can consider  $\mathbf{R}^k$  to be the subset  $\{(a_1, \dots, a_k, 0, \dots, 0)\}$  in  $\mathbf{R}^l$ . Show that smooth functions on  $\mathbf{R}^k$ , considered as a subset of  $\mathbf{R}^l$ , are the same as usual.
- \*2. Suppose that  $X$  is a subset of  $\mathbf{R}^N$  and  $Z$  is a subset of  $X$ . Show that the restriction to  $Z$  of any smooth map on  $X$  is a smooth map on  $Z$ .
- \*3. Let  $X \subset \mathbf{R}^N$ ,  $Y \subset \mathbf{R}^M$ ,  $Z \subset \mathbf{R}^L$  be arbitrary subsets, and let  $f : X \rightarrow Y$ ,  $g : Y \rightarrow Z$  be smooth maps. Then the composite  $g \circ f : X \rightarrow Z$  is smooth. If  $f$  and  $g$  are diffeomorphisms, so is  $g \circ f$ .
4. (a) Let  $B_a$  be the open ball  $\{x : |x|^2 < a\}$  in  $\mathbf{R}^k$ . ( $|x|^2 = \sum x_i^2$ ) Show that the map

$$x \rightarrow \frac{ax}{\sqrt{a^2 - |x|^2}}$$

is a diffeomorphism of  $B_a$  onto  $\mathbf{R}^k$ . [HINT: Compute its inverse directly.]

- (b) Suppose that  $X$  is a  $k$ -dimensional manifold. Show that every point in  $X$  has a neighborhood diffeomorphic to all of  $\mathbf{R}^k$ . Thus local parametrizations may always be chosen with all of  $\mathbf{R}^k$  for their domains.

- \*5. Show that every  $k$ -dimensional vector subspace  $V$  of  $\mathbf{R}^n$  is a manifold diffeomorphic to  $\mathbf{R}^k$ , and that all linear maps on  $V$  are smooth. If  $\phi: \mathbf{R}^k \rightarrow V$  is a linear isomorphism, then the corresponding coordinate functions are linear functionals on  $V$ , called *linear coordinates*.
6. A smooth bijective map of manifolds need not be a diffeomorphism. In fact, show that  $f: \mathbf{R}^1 \rightarrow \mathbf{R}^1, f(x) = x^3$ , is an example.
7. Prove that the union of the two coordinate axes in  $\mathbf{R}^2$  is not a manifold. (HINT: What happens to a neighborhood of 0 when 0 is removed?)
8. Prove that the paraboloid in  $\mathbf{R}^3$ , defined by  $x^2 + y^2 - z^2 = a$ , is a manifold if  $a > 0$ . Why doesn't  $x^2 + y^2 - z^2 = 0$  define a manifold?
9. Explicitly exhibit enough parametrizations to cover  $S^1 \times S^1 \subset \mathbf{R}^4$ .
10. "The" *torus* is the set of points in  $\mathbf{R}^3$  at distance  $b$  from the circle of radius  $a$  in the  $xy$  plane, where  $0 < b < a$ . Prove that these tori are all diffeomorphic to  $S^1 \times S^1$ . Also draw the cases  $b = a$  and  $b > a$ ; why are these not manifolds?
11. Show that one cannot parametrize the  $k$  sphere  $S^k$  by a single parametrization. [HINT:  $S^k$  is compact.]
- \*12. Stereographic projection is a map  $\pi$  from the punctured sphere  $S^2 - \{N\}$  onto  $\mathbf{R}^2$ , where  $N$  is the north pole  $(0, 0, 1)$ . For any  $p \in S^2 - \{N\}$ ,  $\pi(p)$  is defined to be the point at which the line through  $N$  and  $p$  intersects the  $xy$  plane (Figure 1-6). Prove that  $\pi: S^2 - \{N\} \rightarrow \mathbf{R}^2$  is a diffeomorphism. (To do so, write  $\pi$  explicitly in coordinates and solve for  $\pi^{-1}$ .)
- Note that if  $p$  is near  $N$ , then  $|\pi(p)|$  is large. Thus  $\pi$  allows us to think of  $S^2$  a copy of  $\mathbf{R}^2$  compactified by the addition of one point "at infinity." Since we can define stereographic projection by using the

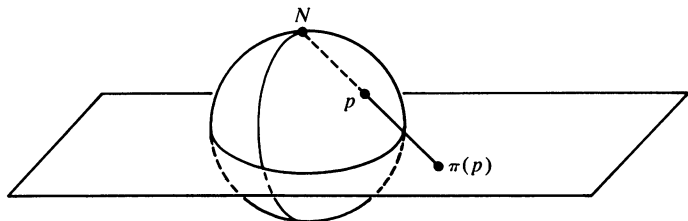


Figure 1-6



south pole instead of the north,  $S^2$  may be covered by two local parametrizations.

**\*13.** By generalizing stereographic projection define a diffeomorphism  $S^k - \{N\} \rightarrow \mathbf{R}^k$ .

**\*14.** If  $f: X \rightarrow X'$  and  $g: Y \rightarrow Y'$  are smooth maps, define a *product map*  $f \times g: X \times Y \rightarrow X' \times Y'$  by

$$(f \times g)(x, y) = (f(x), g(y)).$$

Show that  $f \times g$  is smooth.

**15.** Show that the projection map  $X \times Y \rightarrow X$ , carrying  $(x, y)$  to  $x$ , is smooth.

**\*16.** The *diagonal*  $\Delta$  in  $X \times X$  is the set of points of the form  $(x, x)$ . Show that  $\Delta$  is diffeomorphic to  $X$ , so  $\Delta$  is a manifold if  $X$  is.

**\*17.** The *graph* of a map  $f: X \rightarrow Y$  is the subset of  $X \times Y$  defined by

$$\text{graph}(f) = \{(x, f(x)): x \in X\}.$$

Define  $F: X \rightarrow \text{graph}(f)$  by  $F(x) = (x, f(x))$ . Show that if  $f$  is smooth,  $F$  is a diffeomorphism; thus  $\text{graph}(f)$  is a manifold if  $X$  is. (Note that  $\Delta = \text{graph}(\text{identity})$ .)

**\*18.** (a) An extremely useful function  $f: \mathbf{R}^1 \rightarrow \mathbf{R}^1$  is

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

Prove that  $f$  is smooth.

(b) Show that  $g(x) = f(x - a)g(b - x)$  is a smooth function, positive on  $(a, b)$  and zero elsewhere. (Here  $a < b$ .) Then

$$h(x) = \frac{\int_{-\infty}^x g \, dx}{\int_{-\infty}^{\infty} g \, dx}$$

is a smooth function satisfying  $h(x) = 0$  for  $x < a$ ,  $h(x) = 1$  for  $x > b$ , and  $0 < h(x) < 1$  for  $x \in (a, b)$ .

(c) Now construct a smooth function on  $\mathbf{R}^k$  that equals 1 on the ball of radius  $a$ , zero outside the ball of radius  $b$ , and is strictly between 0 and 1 at intermediate points. (Here  $0 < a < b$ .)

## §2 Derivatives and Tangents

We begin by recalling some facts from calculus. Suppose that  $f$  is a smooth map of an open set in  $\mathbf{R}^n$  into  $\mathbf{R}^m$  and  $x$  is any point in its domain. Then for any vector  $h \in \mathbf{R}^n$ , the derivative of  $f$  in the direction  $h$ , taken at the point  $x$ , is defined by the conventional limit

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

With  $x$  fixed, we define a mapping  $df_x: \mathbf{R}^n \rightarrow \mathbf{R}^m$  by assigning to each vector  $h \in \mathbf{R}^n$  the directional derivative  $df_x(h) \in \mathbf{R}^m$ . Note that this map, which we call the *derivative of  $f$  at  $x$* , is defined on all of  $\mathbf{R}^n$ , even though  $f$  need not be.

In calculus courses one proves that the derivative mapping  $df_x: \mathbf{R}^n \rightarrow \mathbf{R}^m$  is linear. (See, for example, Spivak [2].) And as a linear map,  $df_x$  may be represented as a matrix in terms of the standard bases. In fact, if we write  $f$  as  $f(y) = (f_1(y), \dots, f_m(y))$ , then this matrix is just the Jacobian matrix of  $f$  at  $x$ :

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x), & \dots, & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x), & \dots, & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

But although the derivative of  $f$  may be expressed in terms of its  $m \times n$  partial derivatives, it is much more natural and elegant to think of  $df_x$  as a linear map,  $df_x: \mathbf{R}^n \rightarrow \mathbf{R}^m$ .

One way in which the naturality of the definition exhibits itself is in the simplicity of the chain rule. Suppose that  $U \subset \mathbf{R}^n$  and  $V \subset \mathbf{R}^m$  are open sets, while  $f: U \rightarrow V$  and  $g: V \rightarrow \mathbf{R}^l$  are smooth maps. Then the *Chain Rule* says that for each  $x \in U$ ,

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

Thus if we have a sequence

$$\begin{array}{ccccc} U & \xrightarrow{f} & V & \xrightarrow{g} & \mathbf{R}^l, \\ & & \text{---} & \text{---} & \\ & & g \circ f & & \end{array}$$

We get a commutative diagram of derivative maps

$$\begin{array}{ccccc} \mathbf{R}^n & \xrightarrow{df_x} & \mathbf{R}^m & \xrightarrow{dg_{f(x)}} & \mathbf{R}^l. \\ & & \text{---} & \text{---} & \\ & & d(g \circ f)_x & & \end{array}$$

(A diagram of mappings is *commutative* if any two sequences of maps that start and end at the same sets produce the same composite map.)

Note that if  $f: U \rightarrow \mathbf{R}^m$  is itself a linear map  $L$ , then  $df_x = L$  for all  $x \in U$ . In particular, the derivative of the inclusion map of  $U$  into  $\mathbf{R}^n$  at any point  $x \in U$  is the identity transformation of  $\mathbf{R}^n$ .

The derivative of a mapping is its best linear approximation. So we can use derivatives to identify the linear space that best approximates a manifold  $X$  at the point  $x$ . Suppose that  $X$  sits in  $\mathbf{R}^N$  and that  $\phi: U \rightarrow X$  is a local parametrization around  $x$ , where  $U$  is an open set in  $\mathbf{R}^k$ , and assume that  $\phi(0) = x$  for convenience. The best linear approximation to  $\phi: U \rightarrow X$  at 0 is the map

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

Define the *tangent space* of  $X$  at  $x$  to be the image of the map  $d\phi_0: \mathbf{R}^k \rightarrow \mathbf{R}^N$ . So the tangent space, which we denote  $T_x(X)$ , is a vector subspace of  $\mathbf{R}^N$  whose parallel translate  $x + T_x(X)$  is the closest flat approximation to  $X$  through  $x$  (Figure 1-7). By this definition, a *tangent vector* to  $X \subset \mathbf{R}^N$  at  $x \in X$  is a point  $v$  of  $\mathbf{R}^N$  that lies in the vector subspace  $T_x(X)$  of  $\mathbf{R}^N$ . However, it is natural to picture  $v$  geometrically as the arrow running from  $x$  to  $x + v$ .

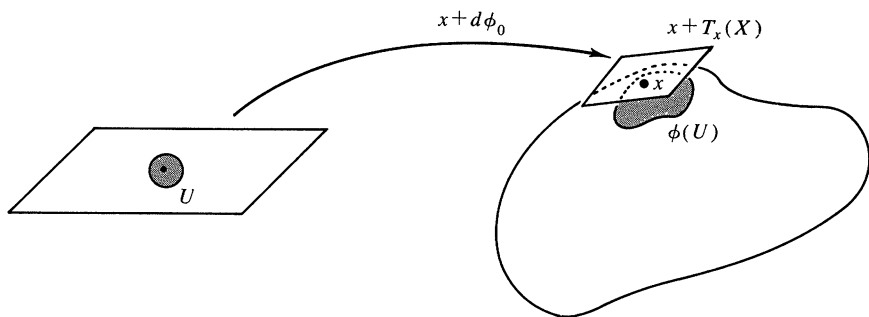


Figure 1-7

Before proceeding, we must resolve an ambiguity in the definition of  $T_x(X)$ ; will another choice of local parametrization produce the same tangent space? Suppose that  $\psi: V \rightarrow X$  is another choice, with  $\psi(0) = x$  as well. By shrinking both  $U$  and  $V$ , we may assume that  $\phi(U) = \psi(V)$ . Then the map  $h = \psi^{-1} \circ \phi: U \rightarrow V$  is a diffeomorphism. Write  $\phi = \psi \circ h$  and differentiate:  $d\phi_0 = d\psi_0 \circ dh_0$ . This relation implies that the image of  $d\phi_0$  is contained in the image of  $d\psi_0$ . The converse follows by switching the roles of  $\phi$  and  $\psi$ , so  $d\phi_0(\mathbf{R}^k) = d\psi_0(\mathbf{R}^k)$  and  $T_x(X)$  is well defined. (For another view of tangent spaces, see Exercise 12 in the exercises for this section.)

The dimension of the vector space  $T_x(X)$  is, as you expect, the dimension  $k$  of  $X$ . To prove this, we use the smoothness of the inverse  $\phi^{-1}$ . Choose an

open set  $W$  in  $\mathbf{R}^N$  and a smooth map  $\Phi': \mathbf{R}^N \rightarrow \mathbf{R}^k$  that extends  $\phi^{-1}$ . Then  $\Phi' \circ \phi$  is the identity map of  $U$ , so the chain rule implies that the sequence of linear transformations

$$\mathbf{R}^k \xrightarrow{d\phi_0} T_x(X) \xrightarrow{d\Phi'_x} \mathbf{R}^k$$

is the identity map of  $\mathbf{R}^k$ . It follows that  $d\phi_0: \mathbf{R}^k \rightarrow T_x(X)$  is an isomorphism, so  $\dim T_x(X) = k$ .

We can now construct the best linear approximation of a smooth map of arbitrary manifolds  $f: X \rightarrow Y$  at a point  $x$ . If  $f(x) = y$ , this *derivative* should be a linear transformation of tangent spaces,  $df_x: T_x(X) \rightarrow T_y(Y)$ . We require two items of our generalized definition of derivative. First, for maps in Euclidean space, we expect the new derivative to be the same as the usual one. Second, we demand the chain rule. It is easy to convince yourself that there is only one possible definition with these requisites. Suppose that  $\phi: U \rightarrow X$  parametrizes  $X$  about  $x$  and  $\psi: V \rightarrow Y$  parametrizes  $Y$  about  $y$ , where  $U \subset \mathbf{R}^k$ ,  $V \subset \mathbf{R}^l$ , and, say,  $\phi(0) = x$ ,  $\psi(0) = y$ . If  $U$  is small enough, then we can draw the following commutative square:

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

We know what  $d\phi_0$ ,  $d\psi_0$ , and  $dh_0$  must be, and the chain rule specifies that taking derivatives converts the preceding diagram into a commutative square of linear transformations:

$$\begin{array}{ccc} T_x(X) & \xrightarrow{df_x} & T_y(Y) \\ d\phi_0 \uparrow & & \uparrow d\psi_0 \\ \mathbf{R}^k & \xrightarrow{dh_0} & \mathbf{R}^l \end{array}$$

As  $d\phi_0$  is an isomorphism, there is only one acceptable definition for  $df_x$ , namely,

$$df_x = d\psi_0 \circ dh_0 \circ d\phi_0^{-1}.$$

Of course, in order to use this definition of  $df_x$ , we must verify that it does not depend on the particular parametrizations  $\phi$  and  $\psi$  used. This verification is just like the one given to establish that the definition of  $T_x(X)$  does not de-

pend on parametrization. Do write down the calculation yourself to be sure that you understand.

Does our extension of the derivative to maps of arbitrary manifolds truly conform to the chain rule as we intended? (It had better, for we showed that it is the only possibility.) Let  $g: Y \rightarrow Z$  be another smooth map, and let  $\eta: W \rightarrow Z$  parametrize  $Z$  about  $z = g(y)$ . Here  $W \subset \mathbf{R}^m$  and  $\eta(0) = z$ . Then from the commutative diagram

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

we derive the square

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

Thus, by definition,

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

By the chain rule for maps of open subsets of Euclidean spaces,  $d(j \circ h)_0 = (dj)_0 \circ (dh)_0$ . Thus

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

We have proved the

**Chain Rule.** If  $X \xrightarrow{f} Y \xrightarrow{g} Z$  are smooth maps of manifolds, then

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

## EXERCISES

- \*1. For a submanifold  $X$  of  $Y$ , let  $i: X \rightarrow Y$  be the inclusion map. Check that  $di_x$  is the inclusion map of  $T_x(X)$  into  $T_x(Y)$ .
- \*2. If  $U$  is an open subset of the manifold  $X$ , check that

$$T_x(U) = T_x(X) \quad \text{for } x \in U.$$

3. Let  $V$  be a vector subspace of  $\mathbf{R}^N$ . Show that  $T_x(V) = V$  if  $x \in V$ .
- \*4. Suppose that  $f: X \rightarrow Y$  is a diffeomorphism, and prove that at each  $x$  its derivative  $df_x$  is an isomorphism of tangent spaces.
5. Prove that  $\mathbf{R}^k$  and  $\mathbf{R}^l$  are not diffeomorphic if  $k \neq l$ .
6. The tangent space to  $S^1$  at a point  $(a, b)$  is a one-dimensional subspace of  $\mathbf{R}^2$ . Explicitly calculate the subspace in terms of  $a$  and  $b$ . [The answer is obviously the space spanned by  $(-b, a)$ , but prove it.]
7. Similarly exhibit a basis for  $T_p(S^2)$  at an arbitrary point  $p = (a, b, c)$ .
8. What is the tangent space to the paraboloid defined by  $x^2 + y^2 - z^2 = a$  at  $(\sqrt{a}, 0, 0)$ , where  $(a > 0)$ ?

- \*9. (a) Show that for any manifolds  $X$  and  $Y$ ,

$$T_{(x,y)}(X \times Y) = T_x(X) \times T_y(Y).$$

- (b) Let  $f: X \times Y \rightarrow X$  be the projection map  $(x, y) \rightarrow x$ . Show that

$$df_{(x,y)}: T_x(X) \times T_y(Y) \rightarrow T_x(X)$$

is the analogous projection  $(v, w) \rightarrow v$ .

- (c) Fixing any  $y \in Y$  gives an injection mapping  $f: X \rightarrow X \times Y$  by  $f(x) = (x, y)$ . Show that  $df_x(v) = (v, 0)$ .
- (d) Let  $f: X \rightarrow X', g: Y \rightarrow Y'$  be any smooth maps. Prove that

$$d(f \times g)_{(x,y)} = df_x \times dg_y.$$

- \*10. (a) Let  $f: X \rightarrow X \times X$  be the mapping  $f(x) = (x, x)$ . Check that  $df_x(v) = (v, v)$ .
- (b) If  $\Delta$  is the diagonal of  $X \times X$ , show that its tangent space  $T_{(x,x)}(\Delta)$  is the diagonal of  $T_x(X) \times T_x(X)$ .
- \*11. (a) Suppose that  $f: X \rightarrow Y$  is a smooth map, and let  $F: X \rightarrow X \times Y$  be  $F(x) = (x, f(x))$ . Show that

$$dF_x(v) = (v, df_x(v)).$$

- (b) Prove that the tangent space to  $\text{graph}(f)$  at the point  $(x, f(x))$  is the graph of  $df_x: T_x(X) \rightarrow T_{f(x)}(Y)$ .

- \*12. A *curve* in a manifold  $X$  is a smooth map  $t \rightarrow c(t)$  of an interval of  $\mathbf{R}^1$  into  $X$ . The *velocity vector* of the curve  $c$  at time  $t_0$ —denoted simply

$dc/dt(t_0)$ —is defined to be the vector  $dc_{t_0}(1) \in T_{x_0}(X)$ , where  $x_0 = c(t_0)$  and  $dc_{t_0}: \mathbf{R}^1 \rightarrow T_{x_0}(X)$ . In case  $X = \mathbf{R}^k$  and  $c(t) = (c_1(t), \dots, c_k(t))$  in coordinates, check that

$$\frac{dc}{dt}(t_0) = (c'_1(t_0), \dots, c'_k(t_0)).$$

Prove that every vector in  $T_x(X)$  is the velocity vector of some curve in  $X$ , and conversely. [HINT: It's easy if  $X = \mathbf{R}^k$ . Now parametrize.]

### §3 The Inverse Function Theorem and Immersions

Before we really begin to discuss the topology of manifolds, we must study the local behavior of smooth maps. Perhaps the best reason for always working with smooth maps (rather than continuous maps, as in non-differential topology) is that local behavior is often entirely specified, up to diffeomorphism, by the derivative. The elucidation of this remark is the primary objective of the first chapter.

If  $X$  and  $Y$  are smooth manifolds of the same dimension, then the simplest behavior a smooth map  $f: X \rightarrow Y$  can possibly exhibit around a point  $x$  is to carry a neighborhood of  $x$  diffeomorphically onto a neighborhood of  $y = f(x)$ . In such an instance, we call  $f$  a *local diffeomorphism* at  $x$ . A necessary condition for  $f$  to be a local diffeomorphism at  $x$  is that its derivative mapping  $df_x: T_x(X) \rightarrow T_x(Y)$  be an isomorphism. (See Exercise 4 in Section 2). The fact that this linear condition is also sufficient is the key to understanding the remark above.

**The Inverse Function Theorem.** Suppose that  $f: X \rightarrow Y$  is a smooth map whose derivative  $df_x$  at the point  $x$  is an isomorphism. Then  $f$  is a local diffeomorphism at  $x$ .

The Inverse Function Theorem is a truly remarkable and valuable fact. The derivative  $df_x$  is simply a single linear map, which we may represent by a matrix of numbers. This linear map is nonsingular precisely when the determinant of its matrix is nonzero. Thus the Inverse Function Theorem tells us that the seemingly quite subtle question of whether  $f$  maps a neighborhood of  $x$  diffeomorphically onto a neighborhood of  $y$  reduces to a trivial matter of checking if a single number—the determinant of  $df_x$ —is nonzero!

You have probably seen a proof of the Inverse Function Theorem for the special case when  $X$  and  $Y$  are open subsets of Euclidean space. One may be found in any text on calculus of several variables—for example, Spivak [2]. You should easily be able to translate the Euclidean result to the manifold setting by using local parametrizations.

Remember that the Inverse Function Theorem is purely a local result; it only tells us about the behavior of  $f$  near the point  $x$ . Even if  $df_x$  is nonsingular for every  $x \in X$ , one cannot conclude that  $f$  is globally a diffeomorphism of the spaces  $X$  and  $Y$ . Such a map is, however, a local diffeomorphism at every point  $x \in X$ , and so it will simply be called a *local diffeomorphism*. A typical example of a local diffeomorphism that is not a global diffeomorphism is the map  $f: \mathbf{R}^1 \rightarrow S^1$ , defined by  $f(t) = (\cos t, \sin t)$ .

We can suggestively reformulate the Inverse Function Theorem by using local coordinates: if  $df_x$  is an isomorphism, then one can choose local coordinates around  $x$  and  $y$  so that  $f$  appears to be the identity,  $f(x_1, \dots, x_k) = (x_1, \dots, x_k)$ . In more sophisticated language, there exist local parametrizations  $\phi: U \rightarrow X, \psi: U \rightarrow Y$  with the same open domain in  $\mathbf{R}^k$ , such that the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \phi \uparrow & & \uparrow \psi \\ U & \xrightarrow{\text{Identity}} & U \end{array}$$

In general, we shall say that two maps  $f: X \rightarrow Y$  and  $f': X' \rightarrow Y'$  are *equivalent* if there exist diffeomorphisms  $\alpha$  and  $\beta$  completing a commutative square

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \alpha \uparrow & & \uparrow \beta \\ X' & \xrightarrow{f'} & Y' \end{array}$$

(Another locution with the same meaning is “ $f$  and  $f'$  are the same up to diffeomorphism.”) In this terminology, the Inverse Function Theorem says that if  $df_x$  is an isomorphism, then  $f$  is locally equivalent, at  $x$ , to the identity. A linear map is obviously equivalent to the identity if and only if it is an isomorphism. Therefore still another version of the theorem states that  $f$  is locally equivalent to the identity precisely when  $df_x$  is.

Of course, for the Inverse Function Theorem to apply, the dimensions of  $X$  and  $Y$  must be equal. What is the best local behavior a map can have when  $\dim X < \dim Y$ ? Differentially, the best we can demand is that  $df_x: T_x(X) \rightarrow T_x(Y)$  be injective. If so,  $f$  is said to be an *immersion at  $x$* . If  $f$  is an immersion at every point, it is simply called an *immersion*. The *canonical immersion* is the standard inclusion map of  $\mathbf{R}^k$  into  $\mathbf{R}^l$  for  $l \geq k$ , where  $(a_1, \dots, a_k)$  maps to  $(a_1, \dots, a_k, 0, \dots, 0)$ . In fact, up to diffeomorphism, this is locally the only immersion.



**Local Immersion Theorem.** Suppose that  $f: X \rightarrow Y$  is an immersion at  $x$ , and  $y = f(x)$ . Then there exist local coordinates around  $x$  and  $y$  such that

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

In other words,  $f$  is locally equivalent to the canonical immersion near  $x$ .

*Proof.* Begin by choosing any local parametrizations yielding a commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \phi \uparrow & & \uparrow \psi \\ U & \xrightarrow{g} & V \end{array} \quad \begin{array}{l} \phi(0) = x \\ \psi(0) = y \end{array}$$

Now we try to augment  $g$  so that the Inverse Function Theorem may be applied. As  $dg_0: \mathbf{R}^k \rightarrow \mathbf{R}^l$  is injective, by a change of basis in  $\mathbf{R}^l$  we may assume that it has an  $l \times k$  matrix

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

where  $I_k$  is the  $k \times k$  identity matrix. Now define a map  $G: U \times \mathbf{R}^{l-k} \rightarrow \mathbf{R}^l$  by

$$G(x, z) = g(x) + (0, z).$$

$G$  maps an open set of  $\mathbf{R}^l$  into  $\mathbf{R}^l$ , and the matrix of  $dG_0$  is  $I_l$ . Therefore the Inverse Function Theorem implies that  $G$  is a local diffeomorphism of  $\mathbf{R}^l$  at 0. Note that we have defined  $G$  so that  $g = G \circ (\text{canonical immersion})$ . Since  $\psi$  and  $G$  are local diffeomorphisms at 0, so is  $\psi \circ G$ ; thus  $\psi \circ G$  may be used as a local parametrization of  $Y$  around the point  $y$ . Moreover, if we shrink  $U$  and  $V$  sufficiently, the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{\quad} & Y \\ \phi \uparrow & & \uparrow \psi \circ G \\ U & \xrightarrow{\text{Canonical immersion}} & V \end{array} \quad \text{Q.E.D.}$$

One useful corollary is obvious: if  $f$  is an immersion at  $x$ , then it is an immersion in a neighborhood of  $x$ .

It is important to recognize that the immersion condition is strictly local in nature. For example, when  $X$  and  $Y$  have the same dimension, then im-

mersions are the same as local diffeomorphisms. In contrast, actual diffeomorphisms must satisfy both a local differential condition plus a global topological one: they must be local diffeomorphisms, and they must be one-to-one and onto. So, in order to force an immersion to exhibit desirable global properties, we must append topological conditions to the local differential data.

Of particular interest is the image of an immersion. The image of the canonical immersion  $\mathbf{R}^k \rightarrow \mathbf{R}^l$  is the nicest conceivable example of a submanifold. Nevertheless, the image of an arbitrary immersion  $f: X \rightarrow Y$  need not be a submanifold of  $Y$ . Let us see why. From the Local Immersion Theorem it is evident that  $f$  maps any sufficiently small neighborhood  $W$  of an arbitrary point  $x$  diffeomorphically onto its image  $f(W)$  in  $Y$ . So every point in the image of  $f$  does sit within a parametrizable subset of  $f(X)$ . Isn't this the definition of a manifold? Not quite. For  $f(X)$  to be a manifold, points must have parametrizable *neighborhoods*, but the subsets  $f(W)$  need not be open in  $Y$ .

Consider, for example, a map that twists the circle into a figure eight (Figure 1-8). This is an immersion of  $S^1$  into  $\mathbf{R}^2$ , but its image is not a manifold. Here, of course, the trouble arises because the immersion is not one-to-one, but even the image of an injective immersion need not be a manifold. For example, the same figure eight can be realized as the image of the injective immersion of  $\mathbf{R}^1$  into  $\mathbf{R}^2$  shown in Figure 1-9.

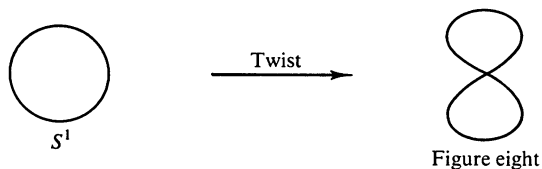


Figure 1-8

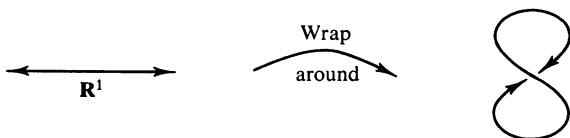


Figure 1-9

Much more anomalous instances may be cited. Let  $g: \mathbf{R}^1 \rightarrow S^1$  be the local diffeomorphism  $g(t) = (\cos 2\pi t, \sin 2\pi t)$ . Define  $G: \mathbf{R}^2 \rightarrow S^1 \times S^1$  by  $G(x, y) = (g(x), g(y))$ .  $G$  is a local diffeomorphism of the plane onto the torus. In fact, looking at  $G$  on the fundamental unit square, one may consider it to be a construction of the torus by gluing opposite sides of the square together (Figure 1-10). Now define a map of  $\mathbf{R}^1$  into the torus by restricting

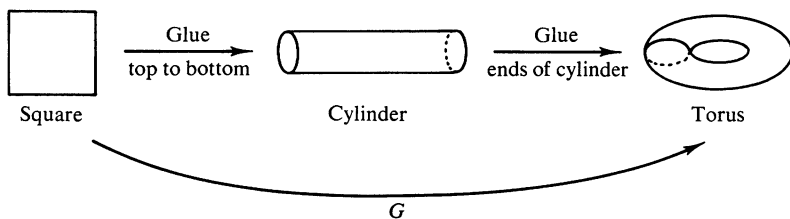


Figure 1-10

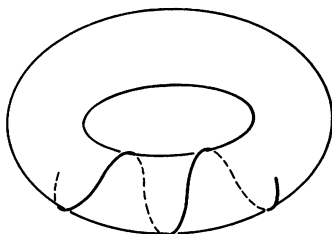


Figure 1-11

$G$  to a straight line through the origin in  $\mathbf{R}^2$  with irrational slope. Since  $G$  is a local diffeomorphism, this is an immersion that wraps  $\mathbf{R}^1$  around the torus (Figure 1-11). Moreover, the irrationality of the slope implies that the immersion is one-to-one and that its image is a dense subset of the torus!

Do these pathologies indicate that we have no hope of drawing global conclusions from our local classification theorem for immersions? Notice that the two injective immersions behave strangely because they map too many points “near infinity” in  $\mathbf{R}^1$  into small regions of the image. Perhaps prohibiting this behavior will sufficiently tame immersions. The general topological analog of points “near infinity” is the exterior of a compact subset in a given space, the compact set being thought of as very large. A map  $f: X \rightarrow Y$  is called *proper* if the preimage of every compact set in  $Y$  is compact in  $X$ . Intuitively, a proper map is one that maps points “near infinity” in  $X$  to points “near infinity” in  $Y$ . An immersion that is injective and proper is called an *embedding*. Having subjoined appropriate global topological constraints to the local immersion condition, we may now prove a reasonable global extension of the Local Immersion Theorem.

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

*Proof.* We have seen that in order to prove  $f(X)$  is a manifold, it suffices to show that the image of any open set  $W$  of  $X$  is an open subset of  $f(X)$ . If  $f(W)$  is not open in  $f(X)$ , then there exists a sequence of points  $y_i \in f(X)$  that do not belong to  $f(W)$  but that converge to a point  $y$  in  $f(W)$ . As the set  $\{y, y_i\}$  is

compact, its preimage in  $X$  must be compact. Each point  $y_i$  has precisely one preimage point  $x_i$  in  $X$ , and  $y$  possesses precisely one preimage point  $x$ , which must belong to  $W$ . Since  $\{x, x_i\}$  is compact, by passage to a subsequence we may assume that  $x_i$  converges to a point  $z \in X$ . Then  $f(x_i) \rightarrow f(z)$ ; therefore since  $f(x_i) \rightarrow f(x)$ , the injectivity of  $f$  implies that  $z = x$ . Now  $W$  is open; so since  $x_i \rightarrow x$ , we conclude that, for large  $i$ ,  $x_i \in W$ . This result contradicts the assumption  $y_i \notin f(W)$ . So  $f(X)$  is indeed a manifold. It is now trivial to check that  $f: X \rightarrow f(X)$  is a diffeomorphism, for we now know  $f$  to be a local diffeomorphism from  $X$  to  $f(X)$ . Since it is bijective, the inverse  $f^{-1}: f(X) \rightarrow X$  is well defined as a set map. But locally  $f^{-1}$  is already known to be smooth. Q.E.D.

Of course, when  $X$  itself is a compact manifold, every map  $f: X \rightarrow Y$  is proper. Thus for compact manifolds, embeddings are just one-to-one immersions.

### EXERCISES

1. Let  $A$  be a linear map of  $\mathbf{R}^n$ , and  $b \in \mathbf{R}^n$ . Show that the mapping  $x \mapsto Ax + b$  is a diffeomorphism of  $\mathbf{R}^n$  if and only if  $A$  is nonsingular.
- \*2. Suppose that  $Z$  is an  $l$ -dimensional submanifold of  $X$  and that  $z \in Z$ . Show that there exists a local coordinate system  $\{x_1, \dots, x_k\}$  defined in a neighborhood  $U$  of  $z$  in  $X$  such that  $Z \cap U$  is defined by the equations  $x_{l+1} = 0, \dots, x_k = 0$ .
3. Let  $f: \mathbf{R}^1 \rightarrow \mathbf{R}^1$  be a local diffeomorphism. Prove that the image of  $f$  is an open interval and that, in fact,  $f$  maps  $\mathbf{R}^1$  diffeomorphically onto this interval.
4. To contrast with Exercise 3, construct a local diffeomorphism  $f: \mathbf{R}^2 \rightarrow \mathbf{R}^2$  that is not a diffeomorphism onto its image. [HINT: Start with our example for  $\mathbf{R}^1 \rightarrow S^1$ .]
5. Prove that a local diffeomorphism  $f: X \rightarrow Y$  is actually a diffeomorphism of  $X$  onto an open subset of  $Y$ , provided that  $f$  is one-to-one.
6. (a) If  $f$  and  $g$  are immersions, show that  $f \times g$  is.  
 (b) If  $f$  and  $g$  are immersions, show that  $g \circ f$  is.  
 (c) If  $f$  is an immersion, show that its restriction to any submanifold of its domain is an immersion.  
 (d) When  $\dim X = \dim Y$ , show that immersions  $f: X \rightarrow Y$  are the same as local diffeomorphisms.
7. (a) Check that  $g: \mathbf{R}^1 \rightarrow S^1$ ,  $g(t) = (\cos 2\pi t, \sin 2\pi t)$ , is, in fact, a local diffeomorphism.

- (b) From Exercise 6, it follows that  $G: \mathbf{R}^2 \rightarrow S^1 \times S^1$ ,  $G = g \times g$ , is a local diffeomorphism. Also, if  $L$  is a line in  $\mathbf{R}^2$ , the restriction  $G: L \rightarrow S^1 \times S^1$  is an immersion. Prove that if  $L$  has irrational slope,  $G$  is one-to-one on  $L$ .

8. Check that the map

$$\mathbf{R}^1 \rightarrow \mathbf{R}^2, \quad t \mapsto \left( \frac{e^t + e^{-t}}{2}, \frac{e^t - e^{-t}}{2} \right),$$

is an embedding. Prove that its image is one nappe of the hyperbola  $x^2 - y^2 = 1$ .

- \*9. (a) Let  $x_1, \dots, x_N$  be the standard coordinate functions on  $\mathbf{R}^N$ , and let  $X$  be a  $k$ -dimensional submanifold of  $\mathbf{R}^N$ . Prove that every point  $x \in X$  has a neighborhood on which the restrictions of some  $k$ -coordinate functions  $x_{i_1}, \dots, x_{i_k}$  form a local coordinate system. [HINT: Let  $e_1, \dots, e_N$  be the usual basis for  $\mathbf{R}^N$ . As a linear algebra lemma, prove that the projection of  $T_x(X)$  onto the subspace spanned by  $e_{i_1}, \dots, e_{i_k}$  is bijective for some choice of  $i_1, \dots, i_k$ . Show that this implies that  $(x_{i_1}, \dots, x_{i_k})$  defines a local diffeomorphism of  $X$  into  $\mathbf{R}^k$  at the point  $x$ .]
- (b) For simplicity, assume that  $x_1, \dots, x_k$  form a local coordinate system on a neighborhood  $V$  of  $x$  in  $X$ . Prove that there are smooth functions  $g_{k+1}, \dots, g_N$  on an open set  $U$  in  $\mathbf{R}^k$  such that  $V$  may be taken to be the set

$$\{(a_1, \dots, a_k, g_{k+1}(a), \dots, g_N(a)) \in \mathbf{R}^N : a = (a_1, \dots, a_k) \in U\}.$$

That is, if we define  $g: U \rightarrow \mathbf{R}^{N-k}$  by  $g = (g_{k+1}, \dots, g_N)$ , then  $V$  equals the graph of  $g$ . Thus every manifold is locally expressible as a graph.

- \*10. *Generalization of the Inverse Function Theorem:* Let  $f: X \rightarrow Y$  be a smooth map that is one-to-one on a compact submanifold  $Z$  of  $X$ . Suppose that for all  $x \in Z$ ,

$$df_x: T_x(X) \rightarrow T_{f(x)}(Y)$$

is an isomorphism. Then  $f$  maps  $Z$  diffeomorphically onto  $f(Z)$ . (Why?) Prove that  $f$ , in fact, maps an open neighborhood of  $Z$  in  $X$  diffeomorphically onto an open neighborhood of  $f(Z)$  in  $Y$ . Note that when  $Z$  is a single point, this specializes to the Inverse Function Theorem. [HINT: Prove that, by Exercise 5, you need only show  $f$  to be one-to-one on some neighborhood of  $Z$ . Now if  $f$  isn't so, construct sequences  $\{a_i\}$  and  $\{b_i\}$  in  $X$  both converging to a point  $z \in Z$ , with  $a_i \neq b_i$  but  $f(a_i) = f(b_i)$ . Show that this contradicts the nonsingularity of  $df_z$ .]

## §4 Submersions

We continue our local analysis with the dimensional case that will be most important for us,  $\dim X \geq \dim Y$ . If  $f: X \rightarrow Y$  carries  $x$  to  $y$ , the strongest condition we can impose on its derivative  $df_x: T_x(X) \rightarrow T_y(Y)$  is surjectivity. If  $df_x$  is surjective,  $f$  is called a *submersion at  $x$* . A map that is a submersion at every point is simply called a *submersion*.

The *canonical submersion* is the standard projection of  $\mathbf{R}^k$  onto  $\mathbf{R}^l$  for  $k \geq l$ , in which  $(a_1, \dots, a_k) \rightarrow (a_1, \dots, a_l)$ . As in the case of immersions, every submersion is locally canonical, up to diffeomorphism.

**Local Submersion Theorem.** Suppose that  $f: X \rightarrow Y$  is a submersion at  $x$ , and  $y = f(x)$ . Then there exist local coordinates around  $x$  and  $y$  such that  $f(x_1, \dots, x_k) = (x_1, \dots, x_l)$ . That is,  $f$  is locally equivalent to the canonical submersion near  $x$ .

*Proof.* The demonstration is just like the local classification of immersions. Given any local parametrization diagram

$$\begin{array}{ccc}
 X & \xrightarrow{f} & Y \\
 \uparrow \phi & & \uparrow \psi \\
 U & \xrightarrow{g} & V
 \end{array}
 \quad
 \begin{array}{l}
 \phi(0) = x \\
 \psi(0) = y
 \end{array}$$

we seek to modify  $g$  and apply the Inverse Function Theorem. As  $dg_0: \mathbf{R}^k \rightarrow \mathbf{R}^l$  is surjective, by linear change of coordinates in  $\mathbf{R}^k$  we may assume it has  $l \times k$  matrix  $(I_l | 0)$ . Now define  $G: U \rightarrow \mathbf{R}^k$  by

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

where  $a = (a_1, \dots, a_k)$ . The matrix of  $dG_0$  is then  $I_k$ , so  $G$  is a local diffeomorphism at 0. Thus  $G^{-1}$  exists as a diffeomorphism of some open neighborhood  $U'$  of 0 into  $U$ . By construction,  $g = (\text{canonical submersion}) \circ G$ , so  $g \circ G^{-1}$  is the canonical submersion. Then the following square commutes:

$$\begin{array}{ccc}
 X & \xrightarrow{f} & Y \\
 \uparrow \phi \circ G^{-1} & & \uparrow \psi \\
 U' & \xrightarrow{\text{Canonical submersion}} & V
 \end{array}
 \quad \text{Q.E.D.}$$

An obvious corollary worth noting is that if  $f$  is a submersion at  $x$ , then it is actually a submersion in a whole neighborhood of  $x$ .

One of the most valuable implications of the local classification theorem concerns the geometric nature of solutions of functional equations. If  $y$  is a point of  $Y$  and  $f: X \rightarrow Y$ , the solutions of the equation  $f(x) = y$  comprise a subset of  $X$  that is called the *preimage* of  $y$ , denoted  $f^{-1}(y)$ . For general maps, the set  $f^{-1}(y)$  need not be at all reasonable as a geometric object. But suppose that  $f$  is a submersion at a point  $x \in f^{-1}(y)$ . Select local coordinates around  $x$  and  $y$  so that

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

and  $y$  corresponds to  $(0, \dots, 0)$ . Thus near  $x$ ,  $f^{-1}(y)$  is just the set of points  $(0, \dots, 0, x_{l+1}, \dots, x_k)$ . More precisely, let  $V$  denote the neighborhood of  $x$  on which the coordinate system  $(x_1, \dots, x_k)$  is defined. Then  $f^{-1}(y) \cap V$  is the set of points where  $x_1 = 0, \dots, x_l = 0$ . The functions  $x_{k+1}, \dots, x_l$  therefore form a coordinate system on the set  $f^{-1}(y) \cap V$ , which is a (relatively) open subset of  $f^{-1}(y)$ . (See Figure 1-12.)

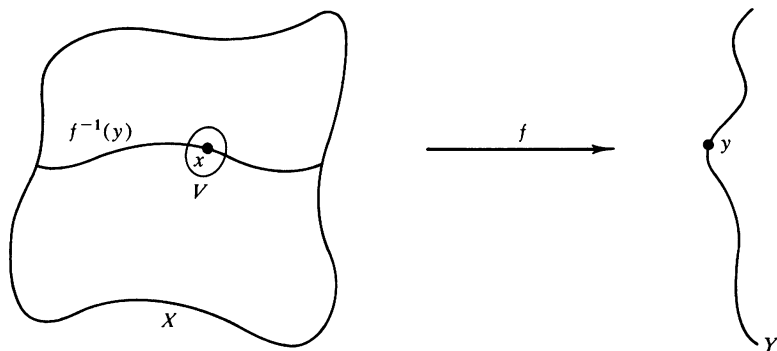


Figure 1-12

We are led to another definition. For a smooth map of manifolds  $f: X \rightarrow Y$ , a point  $y \in Y$  is called a *regular value* for  $f$  if  $df_x: T_x(X) \rightarrow T_y(Y)$  is surjective at every point  $x$  such that  $f(x) = y$ . The argument just given proves the

**Preimage Theorem.** If  $y$  is a regular value of  $f: X \rightarrow Y$ , then the preimage  $f^{-1}(y)$  is a submanifold of  $X$ , with  $\dim f^{-1}(y) = \dim X - \dim Y$ .

A point  $y \in Y$  that is not a regular value of  $f$  is called a *critical value*. The solution set  $\{x: f(x) = y\}$  may be rather complicated when  $y$  is a critical value.

One absurd punctilio must be noted about the definition of regular values. Any point  $y \in Y$  that does not belong to the image of  $f$  automatically qualifies as a regular value. This is actually a strict logical consequence of the definition, but if stupidities like this annoy you, simply consider it as a special

definition. In fact, we shall occasionally want to find some point not belonging to an image  $f(x)$ , and information about regular values will be helpful.

For emphasis, let us reiterate the definition of regular values for each dimensional possibility. When  $\dim X > \dim Y$ , the regularity of a value  $y$  means that  $f$  is a submersion at each preimage point  $x \in f^{-1}(y)$ . When  $\dim X = \dim Y$ , it means that  $f$  is a local diffeomorphism at each preimage point. Finally, if  $\dim X < \dim Y$ , then every point in  $f(X)$  is a critical value, and the regular values are those never hit by  $f$ . The case where  $\dim X = \dim Y$  is especially important; we urge you to work Exercise 7 to develop your intuition.

How much simpler it is to create submanifolds by the use of the Preimage Theorem than tediously to construct local parametrizations. For example, consider the map  $f: \mathbf{R}^k \rightarrow \mathbf{R}$  defined by

$$f(x) = |x|^2 = x_1^2 + \cdots + x_k^2.$$

The derivative  $df_a$  at the point  $a = (a_1, \dots, a_k)$  has matrix  $(2a_1, \dots, 2a_k)$ . Thus  $df_a: \mathbf{R}^k \rightarrow \mathbf{R}$  is surjective unless  $f(a) = 0$ , so every nonzero real number is a regular value of  $f$ . In particular, we prove effortlessly that the sphere  $S^{k-1} = f^{-1}(1)$  is a  $k - 1$  dimensional manifold.

A more powerful application of the theorem is provided by the orthogonal group  $O(n)$ , the group of linear transformations of  $\mathbf{R}^n$  that preserve distance. Now the space  $M(n)$  of all  $n \times n$  matrices is a manifold; in fact (by rearranging entries along a single line), it is nothing but  $\mathbf{R}^{n^2}$ .  $O(n)$  is the group of matrices that satisfy the equation  $AA^t = I$ , where  $A^t$  is the transpose of  $A$  and  $I$  is the identity matrix. We will now show that  $O(n)$  is a manifold. Note first, that for any matrix  $A$ , the matrix  $AA^t$  is symmetric, for it equals its own transpose. The vector space  $S(n)$  of all symmetric  $n \times n$  matrices is obviously a submanifold of  $M(n)$  diffeomorphic to  $\mathbf{R}^k$ , where  $k = n(n + 1)/2$ , and the map  $f: M(n) \rightarrow S(n)$ , defined by  $f(A) = AA^t$ , is smooth.  $O(n) = f^{-1}(I)$ , so, in order to establish that the orthogonal group is a manifold, we need only show  $I$  to be a regular value of  $f$ . Hence we compute the derivative of  $f$  at a matrix  $A$ . By definition,

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



Is  $df_A: T_A M(n) \rightarrow T_{f(A)} S(n)$  surjective when  $A$  belongs to  $f^{-1}(I) = O(n)$ ? By our identifications of  $M(n)$  and  $S(n)$  with Euclidean spaces, we have that  $T_A M(n) = M(n)$  and  $T_{f(A)} S(n) = S(n)$ . The matrix  $I$  is a regular value of  $f$  if and only if  $df_A: M(n) \rightarrow S(n)$  is surjective for all  $A \in O(n)$ . That is, for any  $C \in S(n)$ , there must exist a  $B \in M(n)$  solving the equation  $df_A(B) = C$ , or  $BA' + AB' = C$ . As  $C$  is symmetric, we may write  $C = \frac{1}{2}C + \frac{1}{2}C'$ , and then note that we can solve for  $B$  in the equation  $BA' = \frac{1}{2}C$ . Because  $A'A = I$ , multiplication on the right by  $A$  yields  $B = \frac{1}{2}CA$ . Then, in fact,

$$df_A(B) = (\tfrac{1}{2}CA)A' + A(\tfrac{1}{2}CA)' = \tfrac{1}{2}C(AA') + \tfrac{1}{2}(AA')C' = \tfrac{1}{2}C + \tfrac{1}{2}C' = C,$$

so we have found the desired matrix  $B$ . Thus  $f$  is a submersion at every  $A$  in  $f^{-1}(I)$ , and  $I$  is a regular value of  $f$ . Hence  $O(n)$  is a submanifold of  $M(n)$ ; moreover,

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

Note that  $O(n)$  is a group as well as a manifold. As an exercise, show that the group operations are smooth in the following sense: the multiplication map  $O(n) \times O(n) \rightarrow O(n)$  defined by  $(A, B) \mapsto AB$  and the inversion map  $O(n) \rightarrow O(n)$  defined by  $A \mapsto A^{-1}$  are both smooth maps of manifolds. (HINT: Here  $A^{-1} = A'$ .) A group that is a manifold, and whose group operations are smooth, is called a *Lie group*.

Let us now consider some variants of the preceding material that will be useful in the next section. Here is a practical sort of problem: suppose that  $g_1, \dots, g_l$  are smooth, real-valued functions on a manifold  $X$  of dimension  $k \geq l$ . Under what conditions is the set  $Z$  of common zeros a reasonable geometric object? We can answer this question easily by considering the map

$$g = (g_1, \dots, g_l): X \rightarrow \mathbf{R}^l.$$

Since  $Z = g^{-1}(0)$ ,  $Z$  is a submanifold of  $X$  if  $0$  is a regular value of  $g$ .

We can reformulate the regularity condition for  $0$  directly in terms of the functions  $g_i$ . Since each  $g_i$  is a smooth map of  $X$  into  $\mathbf{R}$ , its derivative at a point  $x$  is a linear map  $d(g_i)_x: T_x(X) \rightarrow \mathbf{R}$ ; that is,  $d(g_i)_x$  is a linear functional on the vector space  $T_x(X)$ . Furthermore, you can quickly verify that  $dg_x: T_x(X) \rightarrow \mathbf{R}^l$  is surjective if and only if the  $l$  functionals  $d(g_1)_x, \dots, d(g_l)_x$  are linearly independent on  $T_x(X)$ . We express this condition by saying the  $l$  functions  $g_1, \dots, g_l$  are *independent at  $x$* . So now we can “translate” the Preimage Theorem into this language:

**Proposition.** If the smooth, real-valued functions  $g_1, \dots, g_l$  on  $X$  are independent at each point where they all vanish, then the set  $Z$  of common zeros is a submanifold of  $X$  with dimension equal to  $\dim X - l$ .

It is convenient here to define the *codimension* of an arbitrary submanifold  $Z$  of  $X$  by the formula  $\text{codim } Z = \dim X - \dim Z$ . (Note that this depends not only on  $Z$  but on the surrounding manifold  $X$  as well.) Thus  $l$  independent functions on  $X$  cut out a submanifold of codimension  $l$ . Historically, the study of zero sets of collections of functions has been of considerable mathematical interest. Classical algebraic geometry, for example, concerns the nature of sets cut out in Euclidean space by polynomials.

Is the converse of the preceding proposition true? That is, can every submanifold  $Z$  of  $X$  be “cut out” by independent functions? We shall see that the answer is *no* (Exercise 20, Chapter 2, Section 3). However, we can get two useful partial converses:

**Partial Converse 1.** If  $y$  is a regular value of a smooth map  $f: X \rightarrow Y$ , then the preimage submanifold  $f^{-1}(y)$  can be cut out by independent functions.

*Proof.* Just choose a diffeomorphism  $h$  of a neighborhood  $W$  of  $y$  with a neighborhood of the origin in  $\mathbf{R}^l$ , where  $h(y) = 0$ . Now set  $g = h \circ f$  and check that 0 is a regular value for  $g$ . Therefore the coordinate functions  $g_1, \dots, g_l$  will work.

**Partial Converse 2.** Every submanifold of  $X$  is *locally* cut out by independent functions.

*Proof.* More specifically, let  $Z$  be a submanifold of codimension  $l$ , and let  $z$  be any point of  $Z$ . Then we claim that there exist  $l$  independent functions  $g_1, \dots, g_l$  defined on some open neighborhood  $W$  of  $z$  in  $X$  such that  $Z \cap W$  is the common vanishing set of the  $g_i$ . (Here we are stating the converse to the proposition for the submanifold  $Z \cap W$  in the manifold  $W$ .) This converse follows immediately from the Local Immersion Theorem for the immersion  $Z \rightarrow W$ ; in fact, it is just Exercise 2 of the last section.

In particular, taking  $Z$  to be an arbitrary Euclidean space, we note that every manifold is locally definable by a collection of independent functions in Euclidean space. (Again, however, it is not true that all manifolds can be *globally* cut out by independent functions. We shall see why later.)

It is useful to note the following

**Proposition.** Let  $Z$  be the preimage of a regular value  $y \in Y$  under the smooth map  $f: X \rightarrow Y$ . Then the kernel of the derivative  $df_x: T_x(X) \rightarrow T_y(Y)$  at any point  $x \in Z$  is precisely the tangent space to  $Z$ ,  $T_x(Z)$ .

*Proof.* Since  $f$  is constant on  $Z$ ,  $df_x$  is zero on  $T_x(Z)$ . But  $df_x: T_x(X) \rightarrow T_y(Y)$  is surjective, so the dimension of the kernel of  $df_x$  must be

$$\dim T_x(X) - \dim T_y(Y) = \dim X - \dim Y = \dim Z.$$

Thus  $T_x(Z)$  is a subspace of the kernel that has the same dimension as the complete kernel; hence  $T_x(Z)$  must be the kernel. Q.E.D.

## EXERCISES

- \*1. If  $f: X \rightarrow Y$  is a submersion and  $U$  is an open set of  $X$ , show that  $f(U)$  is open in  $Y$ .
- \*2. (a) If  $X$  is compact and  $Y$  connected, show every submersion  $f: X \rightarrow Y$  is surjective.  
 (b) Show that there exist no submersions of compact manifolds into Euclidean spaces.
3. Show that the curve  $t \rightarrow (t, t^2, t^3)$  embeds  $\mathbf{R}^1$  into  $\mathbf{R}^3$ . Find two independent functions that globally define the image. Are your functions independent on all of  $\mathbf{R}^3$ , or just on an open neighborhood of the image?
4. Prove the following extension of Partial Converse 2. Suppose that  $Z \subset X \subset Y$  are manifolds, and  $z \in Z$ . Then there exist independent functions  $g_1, \dots, g_l$  on a neighborhood  $W$  of  $z$  in  $Y$  such that

$$Z \cap W = \{y \in W : g_1(y) = 0, \dots, g_l(y) = 0\}$$

$$\text{and} \quad X \cap W = \{y \in W : g_1(y) = 0, \dots, g_m(y) = 0\},$$

where  $l - m$  is the codimension of  $Z$  in  $X$ .

5. Check that 0 is the only critical value of the map  $f: \mathbf{R}^3 \rightarrow \mathbf{R}^1$  defined by

$$f(x, y, z) = x^2 + y^2 - z^2.$$

Prove that if  $a$  and  $b$  are either both positive or both negative, then  $f^{-1}(a)$  and  $f^{-1}(b)$  are diffeomorphic. [HINT: Consider scalar multiplication by  $\sqrt{b/a}$  on  $\mathbf{R}^3$ .] Pictorially examine the catastrophic change in the topology of  $f^{-1}(c)$  as  $c$  passes through the critical value.

6. More generally, let  $p$  be any homogeneous polynomial in  $k$ -variables. Homogeneity means

$$p(tx_1, \dots, tx_k) = t^m p(x_1, \dots, x_k).$$

Prove that the set of points  $x$ , where  $p(x) = a$ , is a  $k - 1$  dimensional submanifold of  $\mathbf{R}^k$ , provided that  $a \neq 0$ . Show that the manifolds obtained with  $a > 0$  are all diffeomorphic, as are those with  $a < 0$ . [HINT:

Use Euler's identity for homogeneous polynomials

$$\sum_{i=1}^k x_i \frac{\partial p}{\partial x_i} = m \cdot p$$

to prove that 0 is the only critical value of  $p$ .]

- \*7. (*Stack of Records Theorem.*) Suppose that  $y$  is a regular value of  $f: X \rightarrow Y$ , where  $X$  is compact and has the same dimension as  $Y$ . Show that  $f^{-1}(y)$  is a finite set  $\{x_1, \dots, x_N\}$ . Prove there exists a neighborhood  $U$  of  $y$  in  $Y$  such that  $f^{-1}(U)$  is a disjoint union  $V_1 \cup \dots \cup V_N$ , where  $V_i$  is an open neighborhood of  $x_i$  and  $f$  maps each  $V_i$  diffeomorphically onto  $U$ . [HINT: Pick disjoint neighborhoods  $W_i$  of  $x_i$  that are mapped diffeomorphically. Show that  $f(X - \cup W_i)$  is compact and does not contain  $y$ .] See Figure 1-13.

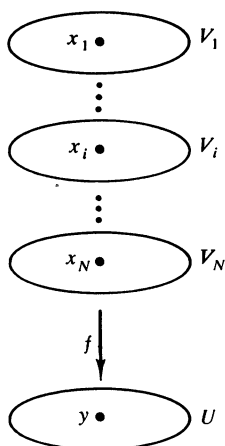


Figure 1-13

8. Let

$$p(z) = z^m + a_1 z^{m-1} + \dots + a_m$$

be a polynomial with complex coefficients, and consider the associated map  $z \rightarrow p(z)$  of the complex plane  $\mathbb{C} \rightarrow \mathbb{C}$ . Prove that this is a submersion except at finitely many points.

9. Show that the orthogonal group  $O(n)$  is compact. [HINT: Show that if  $A = (a_{ij})$  is orthogonal, then for each  $i$ ,  $\sum_j a_{ij}^2 = 1$ .]

10. Verify that the tangent space to  $O(n)$  at the identity matrix  $I$  is the vector space of skew symmetric  $n \times n$  matrices—that is, matrices  $A$  satisfying  $A^t = -A$ .
11. (a) The  $n \times n$  matrices with determinant  $+1$  form a group denoted  $SL(n)$ . Prove that  $SL(n)$  is a submanifold of  $M(n)$  and thus is a Lie group. [HINT: Prove that 0 is the only critical value of  $\det: M(n) \rightarrow \mathbf{R}$ . In fact, if  $\det(A) \neq 0$ , then show that  $\det$  is already a submersion when restricted to the set  $\{tA, t > 0\}$ . Remark: This is really a special case of Exercise 5.]
- (b) Check that the tangent space to  $SL(n)$  at the identity matrix consists of all matrices with trace equal to zero.
12. Prove that the set of all  $2 \times 2$  matrices of rank 1 is a three-dimensional submanifold of  $\mathbf{R}^4 = M(2)$ . [HINT: Show that the determinant function is a submersion on the manifold of nonzero  $2 \times 2$  matrices  $M(2) - \{0\}$ .]
13. Prove that the set of  $m \times n$  matrices of rank  $r$  is a submanifold of  $\mathbf{R}^{mn}$  of codimension  $(m-r)(n-r)$ . [HINT: Suppose, for simplicity, that an  $m \times n$  matrix  $A$  has the form

$$A = \begin{pmatrix} \overset{r}{B} & \overset{n-r}{C} \\ \underset{m-r}{D} & E \end{pmatrix},$$

where the  $r \times r$  matrix  $B$  is nonsingular. Postmultiply by the nonsingular matrix

$$\begin{pmatrix} I & -B^{-1}C \\ 0 & I \end{pmatrix}$$

to prove that  $\text{rank}(A) = r$  if and only if  $E - DB^{-1}C = 0$ .]

## §5 Transversality

We have observed that the solutions of an equation  $f(x) = y$  form a smooth manifold, provided that  $y$  is a regular value of the map  $f: X \rightarrow Y$ . Consider, now, sets of points in  $X$  whose functional values are constrained, not necessarily to be a constant  $y$ , but to satisfy an arbitrary smooth condition. Thus assume  $Z$  to be a submanifold of  $Y$ , and examine the set of solutions of the relation  $f(x) \in Z$ . When can we be assured that this solution set, the preimage  $f^{-1}(Z)$ , is a tractable geometric object? This question will lead us to define a new differential property, an extension of the notion of regularity, which will become the major theme of the book.

Whether  $f^{-1}(Z)$  is a manifold is a local matter. That is, it is a manifold if and only if every point  $x \in f^{-1}(Z)$  has a neighborhood  $U$  in  $X$  such that  $f^{-1}(Z) \cap U$  is a manifold. This observation allows us to reduce study of the relation  $f(x) \in Z$  to the simpler case we have already examined, where  $Z$  is a single point. For if  $y = f(x)$ , we may write  $Z$  in a neighborhood of  $y$  as the zero set of a collection of independent functions  $g_1, \dots, g_l$ ,  $l$  being the codimension of  $Z$  in  $Y$ . Then near  $x$ , the preimage  $f^{-1}(Z)$  is the zero set of the functions  $g_1 \circ f, \dots, g_l \circ f$ . Let  $g$  denote the submersion  $(g_1, \dots, g_l)$  defined around  $y$  (Figure 1-14). Now to the map  $g \circ f: W \rightarrow \mathbf{R}^l$  we may apply the results already obtained;  $(g \circ f)^{-1}(0)$  is guaranteed to be a manifold when 0 is a regular value of  $g \circ f$ .

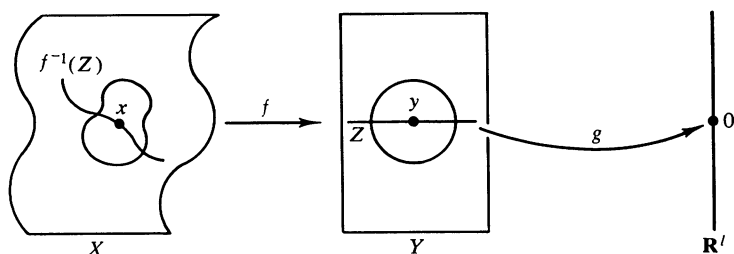


Figure 1-14

Although the map  $g$  is rather arbitrary, the condition that 0 be a regular value of  $g \circ f$  may easily be reformulated in terms of  $f$  and  $Z$  alone. Since

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

the linear map  $d(g \circ f)_x: T_x(X) \rightarrow \mathbf{R}^l$  is surjective if and only if  $dg_y$  carries the image of  $df_x$  onto  $\mathbf{R}^l$ . But  $dg_y: T_y(Y) \rightarrow \mathbf{R}^l$  is a surjective linear transformation whose kernel is the subspace  $T_y(Z)$ . Thus  $dg_y$  carries a subspace of  $T_y(Y)$  onto  $\mathbf{R}^l$  precisely if that subspace and  $T_y(Z)$  together span all of  $T_y(Y)$ . We conclude that  $g \circ f$  is a submersion at the point  $x \in f^{-1}(Z)$  if and only if

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

We have been led inexorably to this equation by our investigation. The map  $f$  is said to be *transversal* to the submanifold  $Z$ , abbreviated  $f \bar{\cap} Z$ , if the equation holds true at each point  $x$  in the preimage of  $Z$ . We have proved

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

With respect to the assertion about codimension, note that locally we have written  $f^{-1}(Z)$  as the zero set of  $l$  independent functions  $g_1 \circ f, \dots, g_l \circ f$ . Therefore the codimension of  $f^{-1}(Z)$  in  $X$  is  $l$ , which was originally set to be the codimension of  $Z$  in  $Y$ .

When  $Z$  is just a single point  $y$ , its tangent space is the zero subspace of  $T_y(Y)$ . Thus  $f$  is transversal to  $y$  if  $df_x[T_x(X)] = T_y(Y)$  for all  $x \in f^{-1}(y)$ , which is to say that  $y$  is a regular value of  $f$ . So transversality includes the notion of regularity as a special case.

For a very simple example of transversality, consider the map  $f: \mathbf{R}^1 \rightarrow \mathbf{R}^2$  defined by  $f(t) = (0, t)$ , and let  $Z$  be the  $x$  axis in  $\mathbf{R}^2$  (Figure 1-15). The map  $g: \mathbf{R}^1 \rightarrow \mathbf{R}^2$  defined by  $g(t) = (t, t^2)$ , however, fails to be transversal to  $Z$  (Figure 1-16).

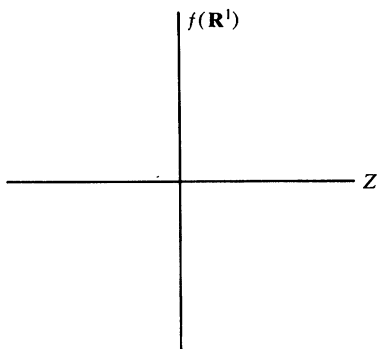


Figure 1-15

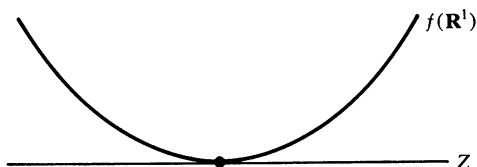


Figure 1-16

The most important and readily visualized special situation concerns the transversality of the inclusion map  $i$  of one submanifold  $X \subset Y$  with another submanifold  $Z \subset Y$ . To say a point  $x \in X$  belongs to the preimage  $i^{-1}(Z)$  simply means that  $x$  belongs to the intersection  $X \cap Z$ . Also, the derivative  $di_x: T_x(X) \rightarrow T_x(Y)$  is merely the inclusion map of  $T_x(X)$  into  $T_x(Y)$ . So  $i \pitchfork Z$  if and only if, for every  $x \in X \cap Z$ ,

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

Notice that this equation is symmetric in  $X$  and  $Z$ . When it holds, we shall

say that the two submanifolds  $X$  and  $Z$  are *transversal*, and write  $X \pitchfork Z$ . The theorem above specializes to

**Theorem.** The intersection of two transversal submanifolds of  $Y$  is again a submanifold. Moreover,

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

The additivity of codimension (which is derived by trivial arithmetic from the codimension assertion of the previous theorem) is absolutely natural. Around a point  $x$  belonging to  $X \cap Z$ , the submanifold  $X$  is cut out of  $Y$  by  $k = \text{codim } X$  independent functions, and  $Z$  is cut out by  $l = \text{codim } Z$  independent functions. Then  $X \cap Z$  is locally just the vanishing set of the combined collection of  $k + l$  functions; that these  $k + l$  functions are together independent is exactly the transversality condition.

One must remember that the transversality of  $X$  and  $Z$  also depends on the ambient space  $Y$ . For example, the two coordinate axes intersect transversally in  $\mathbf{R}^2$ , but not when considered to be submanifolds of  $\mathbf{R}^3$ . In general, if the dimensions of  $X$  and  $Z$  do not add up to at least the dimension of  $Y$ , then they can only intersect transversally by not intersecting at all. For example if  $X$  and  $Y$  are curves in  $\mathbf{R}^3$ , then  $X \pitchfork Y$  implies  $X \cap Y$  is empty. For some example see Figures 1-17 to 1-19. In each case, examine the intersection with the theorem in mind.

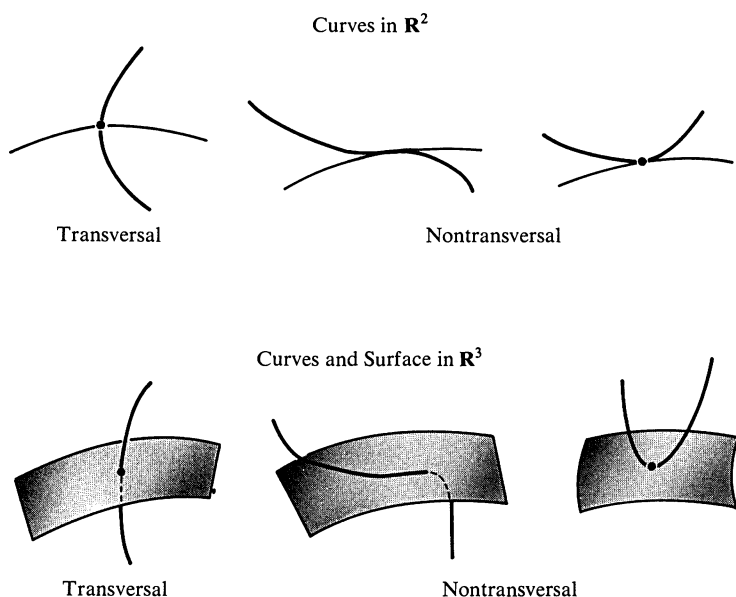
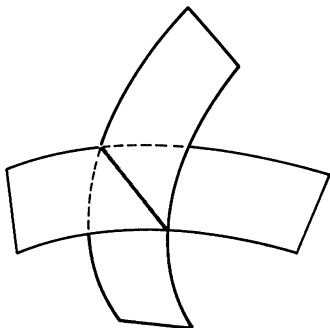
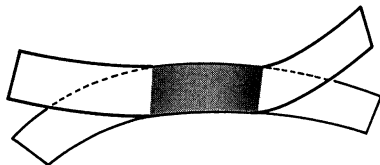


Figure 1-17

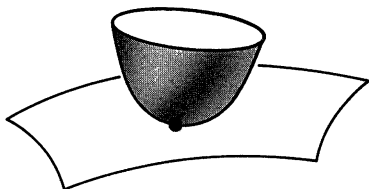




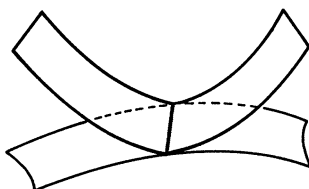
Transversal



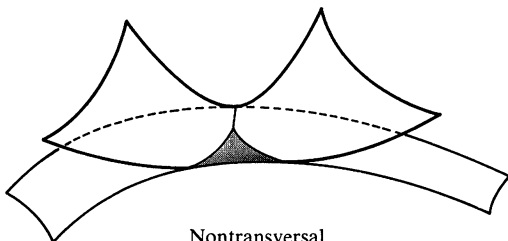
Nontransversal



Nontransversal

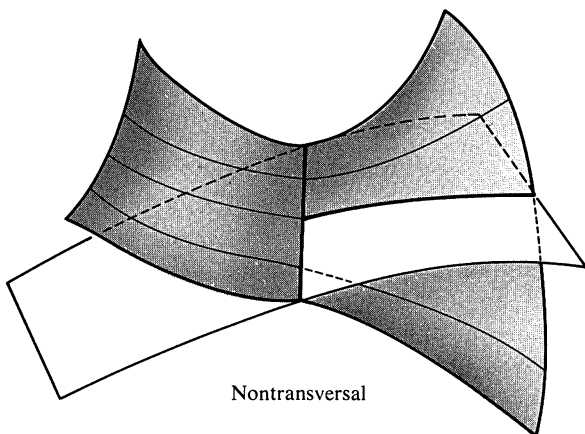


Nontransversal



Nontransversal

**Figure 1-18**



Nontransversal

**Figure 1-19**

Transversality will be fundamental to the topological theory of subsequent chapters, so be sure to devote some time now toward developing your intuition. The best procedure is probably just to draw several dozen pictures of intersecting manifolds in 2- and 3-space. You will find that transversal intersections are rather prosaic but very dependable and clean. But see how utterly bizarre nontransversal intersections can be—and examine your constructions in order to understand precisely where transversality is violated. We have started you off with a series of tame, but typical, examples.

### EXERCISES

1. (a) Suppose that  $A: \mathbf{R}^k \rightarrow \mathbf{R}^n$  is a linear map and  $V$  is a vector subspace of  $\mathbf{R}^n$ . Check that  $A \pitchfork V$  means just  $A(\mathbf{R}^k) + V = \mathbf{R}^n$ .  
 (b) If  $V$  and  $W$  are linear subspaces of  $\mathbf{R}^n$ , then  $V \pitchfork W$  means just  $V + W = \mathbf{R}^n$ .
2. Which of the following linear spaces intersect transversally?  
 (a) The  $xy$  plane and the  $z$  axis in  $\mathbf{R}^3$ .  
 (b) The  $xy$  plane and the plane spanned by  $\{(3, 2, 0), (0, 4, -1)\}$  in  $\mathbf{R}^3$ .  
 (c) The plane spanned by  $\{(1, 0, 0), (2, 1, 0)\}$  and the  $y$  axis in  $\mathbf{R}^3$ .  
 (d)  $\mathbf{R}^k \times \{0\}$  and  $\{0\} \times \mathbf{R}^l$  in  $\mathbf{R}^n$ . (Depends on  $k, l, n$ .)  
 (e)  $\mathbf{R}^k \times \{0\}$  and  $\mathbf{R}^l \times \{0\}$  in  $\mathbf{R}^n$ . (Depends on  $k, l, n$ .)  
 (f)  $V \times \{0\}$  and the diagonal in  $V \times V$ .  
 (g) The symmetric ( $A' = A$ ) and skew symmetric ( $A' = -A$ ) matrices in  $M(n)$ .
3. Let  $V_1, V_2, V_3$  be linear subspaces of  $\mathbf{R}^n$ . One says they have “normal intersection” if  $V_i \pitchfork (V_j \cap V_k)$  whenever  $i \neq j$  and  $i \neq k$ . Prove that this holds if and only if

$$\text{codim}(V_1 \cap V_2 \cap V_3) = \text{codim } V_1 + \text{codim } V_2 + \text{codim } V_3$$

- \*4. Let  $X$  and  $Z$  be transversal submanifolds of  $Y$ . Prove that if  $y \in X \cap Z$ , then

$$T_y(X \cap Z) = T_y(X) \cap T_y(Z).$$

(“The tangent space to the intersection is the intersection of the tangent spaces.”)

- \*5. More generally, let  $f: X \rightarrow Y$  be a map transversal to a submanifold  $Z$  in  $Y$ . Then  $W = f^{-1}(Z)$  is a submanifold of  $X$ . Prove that  $T_x(W)$  is the preimage of  $T_{f(x)}(Z)$  under the linear map  $df_x: T_x(X) \rightarrow T_{f(x)}(Y)$ .

("The tangent space to the preimage of  $Z$  is the preimage of the tangent space of  $Z$ .") (Why does this imply Exercise 4?)

6. Suppose that  $X$  and  $Z$  do not intersect transversally in  $Y$ . May  $X \cap Z$  still be a manifold? If so, must its codimension still be  $\text{codim } X + \text{codim } Z$ ? (Can it be?) Answer with drawings.
- \*7. Let  $X \xrightarrow{f} Y \xrightarrow{g} Z$  be a sequence of smooth maps of manifolds, and assume that  $g$  is transversal to a submanifold  $W$  of  $Z$ . Show  $f \pitchfork g^{-1}(W)$  if and only if  $g \circ f \pitchfork W$ .
8. For which values of  $a$  does the hyperboloid defined by  $x^2 + y^2 - z^2 = 1$  intersect the sphere  $x^2 + y^2 + z^2 = a$  transversally? What does the intersection look like for different values of  $a$ ?
9. Let  $V$  be a vector space, and let  $\Delta$  be the diagonal of  $V \times V$ . For a linear map  $A: V \rightarrow V$ , consider the graph  $W = \{(v, Av) : v \in V\}$ . Show that  $W \pitchfork \Delta$  if and only if  $+1$  is not an eigenvalue of  $A$ .
10. Let  $f: X \rightarrow X$  be a map with fixed point  $x$ ; that is,  $f(x) = x$ . If  $+1$  is not an eigenvalue of  $df_x: T_x(X) \rightarrow T_x(X)$ , then  $x$  is called a *Lefschetz fixed point* of  $f$ .  $f$  is called a *Lefschetz map* if all its fixed points are Lefschetz. Prove that if  $X$  is compact and  $f$  is Lefschetz, then  $f$  has only finitely many fixed points.
11. A theorem of analysis states that every closed subset of  $\mathbf{R}^k$  is the zero set of some smooth function  $f: \mathbf{R}^k \rightarrow \mathbf{R}$ . Use this theorem to show that if  $C$  is any closed subset of  $\mathbf{R}^k$ , then there is a submanifold  $X$  of  $\mathbf{R}^{k+1}$  such that  $X \cap \mathbf{R}^k = C$ . [Here we consider  $\mathbf{R}^k$  as a submanifold of  $\mathbf{R}^{k+1}$  via the usual inclusion  $(a_1, \dots, a_k) \mapsto (a_1, \dots, a_k, 0)$ .] Because closed sets may be extremely bizarre, this shows how bad nontransversal intersections can be.

## §6 Homotopy and Stability

A great many properties of a map are not altered if the map is deformed in a smooth manner. Intuitively, one smooth map  $f_1: X \rightarrow Y$  is a deformation of another  $f_0: X \rightarrow Y$  if they may be joined by a smoothly evolving family of maps  $f_t: X \rightarrow Y$ . (See Figure 1-20.) The precisely formulated definition is one of the fundamental concepts of topology. Let  $I$  denote the unit interval  $[0, 1]$  in  $\mathbf{R}$ . We say that  $f_0$  and  $f_1$  are *homotopic*, abbreviated  $f_0 \sim f_1$ , if there exists a smooth map  $F: X \times I \rightarrow Y$  such that  $F(x, 0) = f_0(x)$  and  $F(x, 1) = f_1(x)$ .  $F$  is called a *homotopy* between  $f_0$  and  $f_1$ . Homotopy is

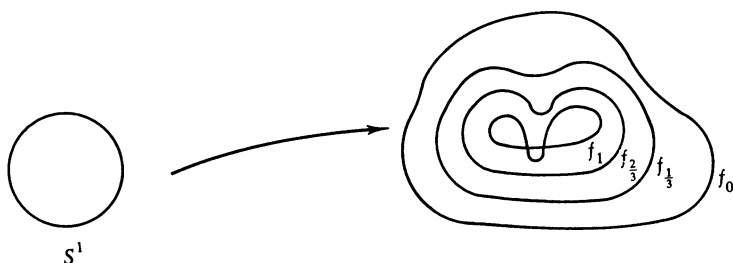


Figure 1-20

an equivalence relation on smooth maps from  $X$  to  $Y$  (see Exercise 1), and the equivalence class to which a mapping belongs is its *homotopy class*. In order to recover the smoothly evolving family of maps joining  $f_0$  and  $f_1$ , just define  $f_t: X \rightarrow Y$  by  $f_t(x) = F(x, t)$ .

In the real world of sense perceptions and physical measurements, no continuous quantity or functional relationship is ever perfectly determined. The only physically meaningful properties of a mapping, consequently, are those that remain valid when the map is slightly deformed. Such properties are *stable* properties, and the collection of maps that possess a particular stable property may be referred to as a *stable class* of maps. Specifically, a property is stable provided that whenever  $f_0: X \rightarrow Y$  possesses the property and  $f_t: X \rightarrow Y$  is a homotopy of  $f_0$ , then, for some  $\epsilon > 0$ , each  $f_t$  with  $t < \epsilon$  also possesses the property.

Consider, for example, curves in the plane, smooth maps of  $\mathbf{R}^1$  into  $\mathbf{R}^2$ . The property that a curve pass through the origin is not stable, since a small wiggle can immediately distort any such curve to avoid 0. (See Figure 1-21). Nor is the property of intersecting the  $x$  axis stable (Figure 1-22). However, transversal intersection with the  $x$ -axis is a stable property, as you can easily

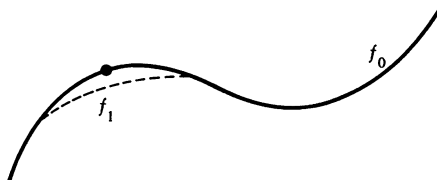


Figure 1-21

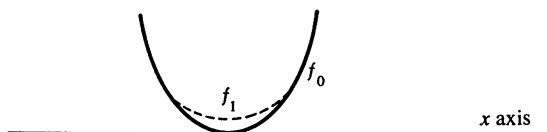


Figure 1-22

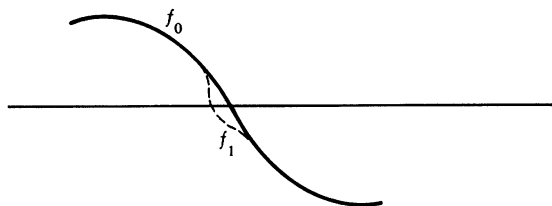


Figure 1-23

see (Figure 1-23). This situation is quite general. The naive point-set condition of intersection is seldom stable, and therefore it is meaningless in the physical world. Transversality, a notion that at first appears unintuitively formal, is all we can really experience.

We prove that all the differential properties of maps  $X \rightarrow Y$  discussed so far are stable, provided that  $X$  is compact.

**Stability theorem.** The following classes of smooth maps of a compact manifold  $X$  into a manifold  $Y$  are stable classes:

- (a) local diffeomorphisms.
- (b) immersions.
- (c) submersions.
- (d) maps transversal to any specified submanifold  $Z \subset Y$ .
- (e) embeddings.
- (f) diffeomorphisms.

The notion of stability provides insight into transversality. For example, why can two curves in  $\mathbf{R}^3$  never intersect transversally except when they do not intersect at all? The formal answer is that  $1 + 1 < 3$ , but there is a more geometric reason. By a small deformation of either curve, one can abruptly pull the two entirely apart; their intersection is not stable (Figure 1-24).

The same principle explains the real reason for all automatic dimensional exclusions of the transversality definition. If  $\dim X + \dim Z < \dim Y$ , and

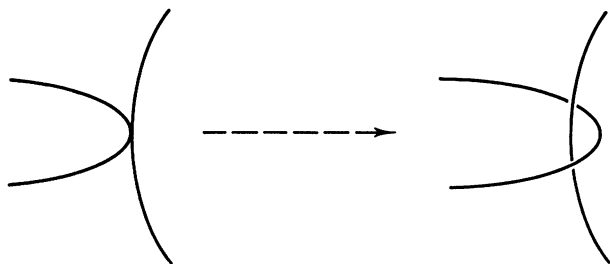


Figure 1-24

$f: X \rightarrow Y$  hits  $Z$ , then one can perturb  $f$  so as to pull it off  $Z$  immediately. Thus no map  $f$  can stably intersect  $Z$ .

Furthermore, other nontransversal intersections, in dimensions not arithmetically prohibited, can be understood from the same point of view, as the diagrams in Figure 1-25 illustrate.

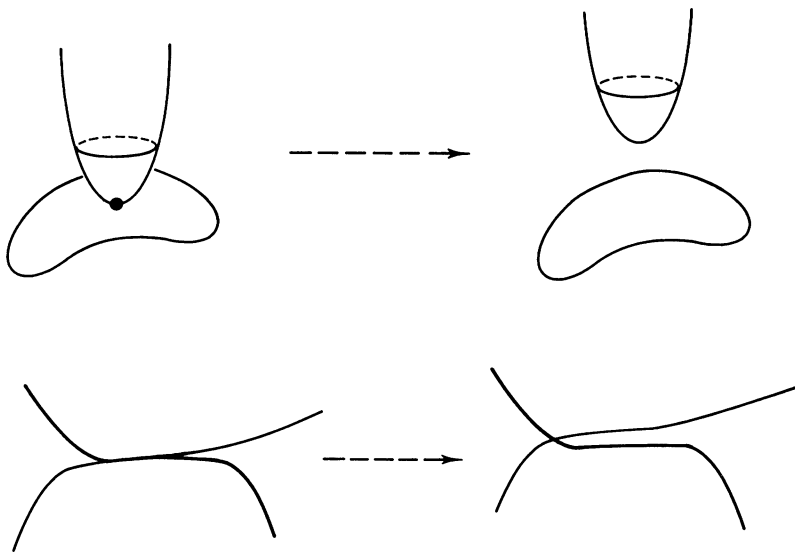


Figure 1-25

**Proof of Stability Theorem.** The stability of the first four classes is proven in the same way. Local diffeomorphisms are just immersions in the special case when  $\dim X = \dim Y$ , so we start with class (b). If  $f_t$  is a homotopy of the immersion  $f_0$ , we must produce an  $\epsilon > 0$  such that  $d(f_t)_x$  is injective for all points  $(x, t)$  in  $X \times [0, \epsilon] \subset X \times I$ . The compactness of  $X$  implies that any open neighborhood of  $X \times \{0\}$  in  $X \times I$  contains  $X \times [0, \epsilon]$  if  $\epsilon$  is sufficiently small. Therefore we need only prove that each point  $(x_0, 0)$  has a neighborhood  $U$  in  $X \times I$  such that  $d(f_t)_x$  is injective for  $(x, t) \in U$ . Since this assertion is local, we need only prove it when  $X$  is an open piece of  $\mathbf{R}^k$  and  $Y$  an open piece of  $\mathbf{R}^l$ . The injectivity of  $d(f_0)_{x_0}$  implies that its  $l \times k$  Jacobian matrix

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

contains a  $k \times k$  submatrix with nonzero determinant. But each partial

$$\frac{\partial (f_t)_i}{\partial x_j} (x),$$

as a function on  $X \times I$ , is continuous. Because the determinant function is also continuous, that same  $k \times k$  submatrix must be nonsingular for all points  $(x, t)$  in a neighborhood of  $(x_0, 0)$  as claimed.

The proof of class (c) is virtually identical. For (d), recall that transversality with respect to  $Z$  may be locally translated into a submersion condition, so the proof of (d) is quite similar.

For (e), we need now only show that if  $f_0$  is one-to-one, so is  $f_t$  if  $t$  is small enough. This proof is a relative of Exercise 10, Section 3. Define a smooth map  $G: X \times I \rightarrow Y \times I$  by  $G(x, t) = (f_t(x), t)$ . Then if (e) is false, there is a sequence  $t_i \rightarrow 0$  and distinct points  $x_i, y_i \in X$  such that  $G(x_i, t_i) = G(y_i, t_i)$ . As  $X$  is compact, we may pass to a subsequence to obtain convergence  $x_i \rightarrow x_0, y_i \rightarrow y_0$ . Then

$$G(x_0, 0) = \lim G(x_i, t_i) = \lim G(y_i, t_i) = G(y_0, 0).$$

But  $G(x_0, 0) = f_0(x_0)$  and  $G(y_0, 0) = f_0(y_0)$ , so if  $f_0$  is injective,  $x_0$  must equal  $y_0$ . Now, locally, we may work in Euclidean space. The matrix of  $dG_{(x_0, 0)}$  is just

$$\left( \begin{array}{c|c} d(f_0)_{x_0} & \begin{matrix} a_1 \\ \vdots \\ \vdots \\ a_l \end{matrix} \\ \hline \begin{matrix} 0 \dots 0 \end{matrix} & 1 \end{array} \right),$$

where the numbers  $a_j$  are not of interest. Since  $d(f_0)_{x_0}$  is injective, its matrix must have  $k$  independent rows. Thus the matrix of  $dG_{(x_0, 0)}$  has  $k+1$  independent rows, so  $dG_{(x_0, 0)}$  must be an injective linear map. Consequently,  $G$  is an immersion around  $(x_0, 0)$  and thus must be one-to-one on some neighborhood of  $(x_0, 0)$ . But for large  $i$ , both  $(x_i, t_i)$  and  $(y_i, t_i)$  belong to this neighborhood, a contradiction.

Finally, we leave (f) for Exercise 8. Q.E.D.

### EXERCISES

- \*1. Suppose that  $f_0, f_1: X \rightarrow Y$  are homotopic. Show that there exists a homotopy  $\tilde{F}: X \times I \rightarrow Y$  such that  $\tilde{F}(x, t) = f_0(x)$  for all  $t \in [0, \frac{1}{4}]$  and  $\tilde{F}(x, t) = f_1(x)$  for all  $t \in [\frac{3}{4}, 1]$ . [HINT: Find a smooth function  $\rho: \mathbf{R} \rightarrow \mathbf{R}$  such that  $\rho(t) = 0$  if  $t \leq \frac{1}{4}$ ,  $\rho(t) = 1$  if  $t \geq \frac{3}{4}$ . Now let  $F$  be any homotopy and set  $\tilde{F}(x, t) = F(x, \rho(t))$ .]
- \*2. Prove that homotopy is an equivalence relation: if  $f \sim g$  and  $g \sim h$ , then  $f \sim h$ . [HINT: To join the homotopies together, you need Exercise 1. Why?]

- \*3. Show that every connected manifold  $X$  is *arcwise connected*: given any two points  $x_0, x_1 \in X$ , there exists a smooth curve  $f: I \rightarrow X$  with  $f(0) = x_0, f(1) = x_1$ . [HINT: Use Exercise 2 to show that the relation " $x_0$  and  $x_1$  can be joined by a smooth curve" is an equivalence relation on  $X$ , and check that the equivalence classes are open.]
4. A manifold  $X$  is *contractible* if its identity map is homotopic to some constant map  $X \rightarrow \{x\}$ ,  $x$  being a point of  $X$ . Check that if  $X$  is contractible, then all maps of an arbitrary manifold  $Y$  into  $X$  are homotopic. (And conversely.)
5. Show that  $\mathbf{R}^k$  is contractible.
6. A manifold  $X$  is *simply connected* if it is connected and if every map of the circle  $S^1$  into  $X$  is homotopic to a constant. Check that all contractible spaces are simply connected, but convince yourself that the converse is false. (Soon we shall develop tools that easily prove the converse false.)
- \*7. Show that the *antipodal map*  $x \rightarrow -x$  of  $S^k \rightarrow S^k$  is homotopic to the identity if  $k$  is odd. HINT: Start off with  $k = 1$  by using the linear maps defined by

$$\begin{pmatrix} \cos \pi t & -\sin \pi t \\ \sin \pi t & \cos \pi t \end{pmatrix}.$$

8. Prove that diffeomorphisms constitute a stable class of mappings of compact manifolds; that is, prove part (f) of the Stability Theorem. [HINT: Reduce to the connected case. Then use the fact that local diffeomorphisms map open sets into open sets, plus part (e) of the theorem.]
9. Prove that the Stability Theorem is false on noncompact domains. Here's one counterexample, but find others yourself to understand what goes wrong. Let  $\rho: \mathbf{R} \rightarrow \mathbf{R}$  be a function with  $\rho(s) = 1$  if  $|s| < 1$ ,  $\rho(s) = 0$  if  $|s| > 2$ . Define  $f_t: \mathbf{R} \rightarrow \mathbf{R}$  by  $f_t(x) = x\rho(tx)$ . Verify that this is a counterexample to all six parts of the theorem. [For part (d), use  $Z = \{0\}$ .]
- \*10. A *deformation* of a submanifold  $Z$  in  $Y$  is a smooth homotopy  $i_t: Z \rightarrow Y$  where  $i_0$  is the inclusion map  $Z \rightarrow Y$  and each  $i_t$  is an embedding. Thus  $Z_t = i_t(Z)$  is a smoothly varying submanifold of  $Y$  with  $Z_0 = Z$ . Show that if  $Z$  is compact, then any homotopy  $i_t$  of its inclusion map is a deformation for small  $t$ . Give a counterexample in the noncompact case (other than the triviality where  $\dim Z = \dim Y$ ).



11. We shall have use for a direct generalization of the notion of homotopy. Suppose that  $f_s: X \rightarrow Y$  is a family of smooth maps, indexed by a parameter  $s$  that varies over a subset  $S$  in some Euclidean space. We say that  $\{f_s\}$  is a *smooth family* of mappings if the map  $F: X \times S \rightarrow Y$ , defined by  $F(x, s) = f_s(x)$ , is smooth. Check that the Stability Theorem generalizes immediately to the following: if  $f_0$  belongs to any of the classes listed, then there exists an  $\epsilon > 0$  such that  $f_s$  belongs to the same class if  $|s_0 - s| < \epsilon$ .

## §7 Sard's Theorem and Morse Functions

The preimage of a regular value of the smooth map  $f: X \rightarrow Y$  is a nice submanifold of  $X$ . This simple fact has led us to a generalization of the notion of regularity—namely, transversality—which we hope will be a key to decipher some of the secrets of the topology of manifolds. But the regularity condition on values of  $f$  is a strong one. Perhaps the condition is so strong that regular values occur too rarely for our Preimage Theorem to be of much use. In fact, precisely the opposite is true, as guaranteed by the second deep theorem to be borrowed from advanced calculus.

**Sard's Theorem.** If  $f: X \rightarrow Y$  is any smooth map of manifolds, then almost every point in  $Y$  is a regular value of  $f$ .

The statement may sound vague but that will be rectified. First, we declare an arbitrary set  $A$  in  $\mathbf{R}^l$  to have *measure zero* if it can be covered by a countable number of rectangular solids with arbitrary small total volume. Of course, a rectangular solid in  $\mathbf{R}^l$  is just a cartesian product of  $l$  intervals in  $\mathbf{R}^1$ , and its volume is the product of the lengths of the  $l$  intervals. Thus  $A$  has measure zero if, for every  $\epsilon > 0$ , there exists a countable collection  $\{S_1, S_2, \dots\}$  of rectangular solids in  $\mathbf{R}^l$ , such that  $A$  is contained in the union of the  $S_i$  and

$$\sum_{i=1}^{\infty} \text{vol}(S_i) < \epsilon.$$

The concept of measure zero is extended to manifolds via local parametrizations. An arbitrary subset  $C \subset Y$  has *measure zero* if, for every local parametrization  $\psi$  of  $Y$ , the preimage  $\psi^{-1}(C)$  has measure zero in Euclidean space. The condition really need not be verified for every parametrization, for it is not difficult to show that if  $A \subset \mathbf{R}^l$  has measure zero and  $g: \mathbf{R}^l \rightarrow \mathbf{R}^l$  is a smooth map, then  $g(A)$  has measure zero. (Proof in Appendix A.) It follows that  $C$  has measure zero, provided that it can be covered by the images of some collection of local parametrizations  $\psi_\alpha$  satisfying the condition that  $\psi_\alpha^{-1}(C)$  has measure zero for each  $\alpha$ .

The assertion in Sard's theorem that "almost every point" of  $Y$  is a regular value for  $X$  means that the points that are not regular values constitute a set of measure zero. Since the complement of the regular values are the critical values, Sard's theorem may be restated:

**Sard's Theorem (Restated).** The set of critical values of a smooth map of manifolds  $f: X \rightarrow Y$  has measure zero.

The proof of Sard's theorem has been relegated to Appendix A, because it has a distinctly different flavor from the topological matters to which the theorem will be applied.

Not surprisingly, no rectangular solid in  $\mathbf{R}^l$  has measure zero, and thus none can be contained in a set of measure zero. (A proof of this remark may also be found in Appendix A.) Consequently, no set of measure zero in a manifold  $Y$  can contain a non-vacuous open set. So we observe the following corollary:

**Corollary.** The regular values of any smooth map  $f: X \rightarrow Y$  are dense in  $Y$ . In fact, if  $f_i: X_i \rightarrow Y$  are any countable number of smooth maps, then the points of  $Y$  that are simultaneously regular values for all of the  $f_i$  are dense.

The second statement is true because the union of any countable collection  $\{C_1, C_2, \dots\}$  of sets of measure zero still has measure zero. Given  $\epsilon > 0$ , choose for each  $i$  a sequence of rectangular solids  $\{S_1^i, S_2^i, \dots\}$  that cover  $C_i$ , and such that

$$\sum_{j=1}^{\infty} \text{vol}(S_j^i) < \frac{\epsilon}{2^i}.$$

Then the countable collection  $\{S_j^i\}$  covers  $\bigcup C_i$  and has total volume less than

$$\sum_i \frac{\epsilon}{2^i} = \epsilon.$$

Now, for the corollary, just take  $C_i$  to be the set of critical points of  $f_i$ .

We introduce two new words for concepts already defined. If  $f: X \rightarrow Y$  is a smooth map, a point  $x \in X$  is a *regular point* of  $f$  if  $df_x: T_x(X) \rightarrow T_y(Y)$  is surjective; one also says that  $f$  is *regular* at  $x$  (which thus means the same as " $f$  is a submersion at  $x$ "). If  $df_x$  is not surjective,  $x$  is a *critical point* of  $f$ . These terms are convenient and quite standard, but, unfortunately, they are easily confused with the related terms "regular value" and "critical value." *Caveat:* keep them straight! Regular and critical *values* live in  $Y$ ; regular and critical *points* live in  $X$ . [Note:  $y$  is a regular value if every  $x \in f^{-1}(y)$  is a regular point.  $y$  is a critical value if so much as one  $x \in f^{-1}(y)$  is a critical point.]

In particular, remember that Sard's theorem says that the set of critical *values* of a map  $f: X \rightarrow Y$  has measure zero in  $Y$ , not that the set of critical *points* has measure zero in  $X$ . For a trivial reminder, examine any constant map of  $X$  into  $Y$ ; if  $\dim Y > 0$ , the set of critical points is all of  $X$  and thus is certainly not of measure zero.

Let us illustrate the use of Sard's theorem with a typical application, the final installment in our discussion of optimal local behavior of mappings. Here we consider only smooth functions on a manifold  $X$ —that is, maps  $f: X \rightarrow \mathbf{R}$ . At a particular  $x \in X$ ,  $f$  is either regular or  $df_x = 0$ . If it is regular, then we can choose a coordinate system around  $x$  so that  $f$  is simply the first coordinate function. Thus we really know all about the local behavior of  $f$  at regular points, at least up to diffeomorphism. But what can we say at critical points? Topological considerations often compel maps to have critical points. For example, if  $X$  is compact, any function on  $X$  must have a maximum and minimum. But if  $f(x)$  is an extreme value, then obviously  $f$  cannot be a coordinate function near  $x$ , so  $df_x$  must be zero. Thus on compact domains every function has at least two critical points (except for the case  $X = \{\text{one point}\}$ ).

At least there is a “best form” for functions around critical points. Let us first work in  $\mathbf{R}^k$ , supposing  $f$  to have a critical point at  $x$ . You probably recall from calculus that when  $df_x = 0$ —that is, when all the partials  $(\partial f / \partial x_1), \dots, (\partial f / \partial x_k)$  vanish at  $x$ —then rather straightforward tests exist for determining whether  $f$  has a maximum, a minimum, or a saddle at the point  $x$ . These tests involve the second derivatives of  $f$ , and they provide definite information as long as the *Hessian matrix* of second partials

$$H = \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)$$

is nonsingular at  $x$ . If the Hessian is nonsingular at the critical point  $x$ , one says that  $x$  is a *nondegenerate critical point* of  $f$ .

Nondegenerate critical points are reasonable in at least one sense: they are isolated from the other critical points of  $f$ . To see this, define a map  $g: \mathbf{R}^k \rightarrow \mathbf{R}^k$  by the formula

$$g = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_k} \right).$$

Then  $df_x = 0$  if and only if  $g(x) = 0$ . Moreover, in the standard matrix representation, the derivative  $dg_x$  is nothing but the Hessian of  $f$  at  $x$ . So if  $x$  is nondegenerate, then not only does  $g(x) = 0$  but  $g$  maps a neighborhood of  $x$  diffeomorphically onto a neighborhood of 0 as well. Thus  $g$  can be zero at no other points in this neighborhood, and so  $f$  has no other critical points there.

The local behavior of a function at a nondegenerate critical point is completely determined, up to diffeomorphism, by a theorem called the *Morse Lemma*. This “canonical form” theorem is analogous in spirit to the local classification theorems for immersions and submersions; it explicitly specifies the function  $f$ , in an appropriate coordinate system, in terms of the Hessian matrix of second derivatives at the critical point.

**Morse Lemma.** Suppose that the point  $a \in \mathbf{R}^k$  is a nondegenerate critical point of the function  $f$ , and

$$(h_{ij}) = \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)$$

is the Hessian of  $f$  at  $a$ . Then there exists a local coordinate system  $(x_1, \dots, x_k)$  around  $a$  such that

$$f = f(a) + \sum h_{ij} x_i x_j$$

near  $a$ .

Thus every function near a nondegenerate critical point is locally equivalent to a quadratic polynomial, the coefficients of which constitute the Hessian. Obviously, the Morse Lemma, which we will not prove or use, is much stronger than the assertion that the Hessian determines whether  $f$  has a maximum or a minimum at  $a$ . (See Exercise 11.)

The concept of nondegeneracy makes sense on manifolds, via local parametrizations. Suppose that  $f: X \rightarrow \mathbf{R}$  has a critical point at  $x$  and that  $\phi$  is a local parametrization carrying the origin to  $x$ . Then 0 is a critical point for the function  $f \circ \phi$ , for  $d(f \circ \phi)_0 = df_x \circ d\phi_0 = 0$ . We shall declare  $x$  to be *nondegenerate* for  $f$  if 0 is nondegenerate for  $f \circ \phi$ . The difficulty with such local definitions is that one must always prove the choice of parametrization to be unimportant. In this case, if  $\phi_1$  and  $\phi_2$  are two choices, then  $f \circ \phi_1 = (f \circ \phi_2) \circ \psi$ , where  $\psi = \phi_2^{-1} \circ \phi_1$ . Thus we must prove

**Lemma.** Suppose that  $f$  is a function on  $\mathbf{R}^k$  with a nondegenerate critical point at 0, and  $\psi$  is a diffeomorphism with  $\psi(0) = 0$ . Then  $f \circ \psi$  also has a nondegenerate critical point at 0.

*Proof.* We simply compute with the explicit form of the chain rule in  $\mathbf{R}^k$ . Let  $f' = f \circ \psi$ , and let  $H$  and  $H'$  be the Hessians at 0 of  $f$  and  $f'$ , respectively. We must show that  $\det(H) \neq 0$  implies  $\det(H') \neq 0$ . Now the chain rule says

$$\frac{\partial f'}{\partial x_i}(x) = \sum_{\alpha} \frac{\partial f}{\partial x_{\alpha}}[\psi(x)] \frac{\partial \psi_{\alpha}}{\partial x_i}(x).$$

Thus

$$\frac{\partial^2 f'}{\partial x_i \partial x_j}(0) = \sum_{\alpha} \sum_{\beta} \frac{\partial^2 f}{\partial x_{\alpha} \partial x_{\beta}}(0) \frac{\partial \psi_{\alpha}}{\partial x_i}(0) \frac{\partial \psi_{\beta}}{\partial x_j}(0) + \sum_{\alpha} \frac{\partial f}{\partial x_{\alpha}}(0) \frac{\partial^2 \psi_{\alpha}}{\partial x_i \partial x_j}(0).$$

Since 0 is a critical point of  $f$ , each term in the second sum is zero. So in the notation of matrix multiplication,  $H' = (d\psi_0)'H(d\psi_0)$ , where  $t$  means "transpose." Now  $\det(d\psi_0) \neq 0$ , since  $\psi$  is a diffeomorphism, and  $\det(d\psi_0)' = \det(d\psi_0)$ . Thus the multiplicative rule for determinants shows that  $\det(H') \neq 0$  if  $\det(H) \neq 0$ . Q.E.D.

Although the verification that nondegeneracy is defined on manifolds was less than elegant, the concept is of great value. There are several closely interrelated reasons for its significance. First, the Morse Lemma completely describes the behavior of functions at nondegenerate points, making detailed constructions possible. Second, functions whose critical points are all nondegenerate—called *Morse functions*—tell a great deal about the topology of their domain manifolds. We shall only be able to illustrate this marvelous subject in two comparatively pedestrian instances (Appendix A, to be read in conjunction with the second chapter, and an exercise in Chapter 3, Section 5). But Marston Morse has made a delightful film [4], describing the two-dimensional situation intuitively. Another fine reference is John Milnor's book *Morse Theory* [3]; you will certainly want to read this book at some point, but we especially urge you to read soon the informal introduction in the first three pages. (Milnor also presents a simple proof of the Morse Lemma.)

A third reason for the importance of nondegeneracy is that it is the common situation; the occurrence of degenerate critical points is really quite rare. In fact, we will deduce from Sard that, in a reasonable sense, the vast majority of functions are actually Morse functions. Suppose that the manifold  $X$  sits in  $\mathbf{R}^N$ , and let  $x_1, \dots, x_N$  be the usual coordinate functions on  $\mathbf{R}^N$ . If  $f$  is a function on  $X$  and  $a = (a_1, \dots, a_N)$  is an  $N$ -tuple of numbers, we define a new function  $f_a$  on  $X$  by

$$f_a = f + a_1x_1 + \cdots + a_Nx_N.$$

The result we allude to is the following

**Theorem.** No matter what the function  $f: X \rightarrow \mathbf{R}$  is, for almost every  $a \in \mathbf{R}^n$  the function  $f_a$  is a Morse function on  $X$ .

As earlier, the phrase "almost every" means that the values  $a$  for which the assertion is false only constitute a set of measure zero.

First we prove the result in  $\mathbf{R}^k$ .

**Lemma.** Let  $f$  be a smooth function on an open set  $U$  of  $\mathbf{R}^k$ . Then for almost all  $k$ -tuples  $a = (a_1, \dots, a_k)$  in  $\mathbf{R}^k$ , the function

$$f_a = f + a_1x_1 + \cdots + a_kx_k$$

is a Morse function on  $U$ .

*Proof.* We again use the map  $g: U \rightarrow \mathbf{R}^k$  defined by

$$g = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_k} \right).$$

Now the derivative of  $f_a$  at a point  $p$  is represented in coordinates by

$$(df_a)_p = \left( \frac{\partial f_a}{\partial x_1}(p), \dots, \frac{\partial f_a}{\partial x_k}(p) \right) = g(p) + a.$$

Thus  $p$  is a critical point of  $f_a$  if and only if  $g(p) = -a$ . Moreover, since  $f_a$  and  $f$  have the same second partials, the Hessian of  $f$  at  $p$  is the matrix  $(dg)_p$ . Now assume that  $-a$  is a regular value for  $g$ . Then whenever  $g(p) = -a$ ,  $(dg)_p$  is nonsingular. Consequently, every critical point of  $f_a$  is nondegenerate. But Sard implies that  $-a$  is a regular value of  $g$  for almost all  $a \in \mathbf{R}^k$ .

Q.E.D.

*Proof of Theorem.* Suppose that  $x$  is any point in  $X$  and that  $x_1, \dots, x_N$  are the standard coordinate functions on  $\mathbf{R}^N$ . Then the restrictions of some  $k$  of these coordinate functions  $x_{i_1}, \dots, x_{i_k}$  to  $X$  constitute a coordinate system in a neighborhood of  $x$ . This was an exercise, but here is a proof. If  $\phi_1, \dots, \phi_N$  is the standard basis of linear functionals on  $\mathbf{R}^N$ , then some  $k$  of these  $\phi_{i_1}, \dots, \phi_{i_k}$  are linearly independent when restricted to the vector subspace  $T_x(X)$ . Since the derivative of the coordinate function  $x_i: \mathbf{R}^N \rightarrow \mathbf{R}$  is just the linear functional  $\phi_i: \mathbf{R}^N \rightarrow \mathbf{R}$ , the derivative at  $x$  of the restriction of  $x_i$  to  $X$  is the restriction of  $\phi_i$  to  $T_x(X)$ . Therefore the linear independence of  $\phi_{i_1}, \dots, \phi_{i_k}$  on  $T_x(X)$  implies that the map  $(x_{i_1}, \dots, x_{i_k}): X \rightarrow \mathbf{R}^k$  is a local diffeomorphism at  $x$ .

Therefore we can cover  $X$  with open subsets  $U_\alpha$  such that on each  $\alpha$  some  $k$  of the functions  $x_1, \dots, x_N$  form a coordinate system. Moreover, by the second axiom of countability (see the “Straight Forward”), we may assume there are only countably many  $U_\alpha$ .

Now suppose that  $U_\alpha$  is one of these sets, and, for convenience, assume that  $(x_1, \dots, x_k)$  is a coordinate system on  $U_\alpha$ . For each  $N - k$  tuple  $c = (c_{k+1}, \dots, c_N)$ , consider the function

$$f_{(0,c)} = f + c_{k+1}x_{k+1} + \dots + c_N x_N.$$

The lemma implies that for almost all  $b \in \mathbf{R}^k$ , the function

$$f_{(b,c)} = f_{(0,c)} + b_1 x_1 + \dots + b_k x_k$$

is a Morse function on  $U_\alpha$ . Now let  $S_\alpha$  be the set of points  $a$  in  $\mathbf{R}^N$  such that  $f_a$  is not a Morse function on  $U_\alpha$ . We have just shown that each “horizontal slice”  $S_\alpha \cap \mathbf{R}^k \times \{c\}$  has measure zero (considered as a subset of  $\mathbf{R}^k$ ). It

should seem believable to you that any reasonable subset of  $\mathbf{R}^N$  whose horizontal slices all have  $\mathbf{R}^k$ -measure zero must itself have measure zero in  $\mathbf{R}^N$ ; this is a form of “Fubini’s theorem,” proved in Appendix A. Granting it for the moment, we conclude that every  $S_\alpha$  has measure zero in  $\mathbf{R}^N$ .

Obviously, a function has a degenerate critical point on  $X$  if and only if it has one in some  $U_\alpha$ . Thus the set of  $N$ -tuples  $a$  for which  $f_a$  is not a Morse function on  $X$  is the union of the  $S_\alpha$ . Since a countable union of sets of measure zero still has measure zero, we are finished. Q.E.D.

### EXERCISES

1. Show that  $\mathbf{R}^k$  is of measure zero in  $\mathbf{R}^l$ ,  $k < l$ .
2. Let  $A$  be a measure zero subset of  $\mathbf{R}^k$ . Show that  $A \times \mathbf{R}^l$  is of measure zero in  $\mathbf{R}^{k+l}$ . (This implies Exercise 1. How?)
3. Suppose that  $Z$  is a submanifold of  $X$  with  $\dim Z < \dim X$ . Prove that  $Z$  has measure zero in  $X$  (without using Sard!).
4. Prove that the rational numbers have measure zero in  $\mathbf{R}^1$ , even though they are dense.
5. Exhibit a smooth map  $f: \mathbf{R} \rightarrow \mathbf{R}$  whose set of critical values is dense. [HINT: Write the rationals in a sequence  $r_0, r_1, \dots$ . Now construct a smooth function on  $[i, i+1]$  that is zero near the endpoints and that has  $r_i$  as a critical value (Figure 1-26).]

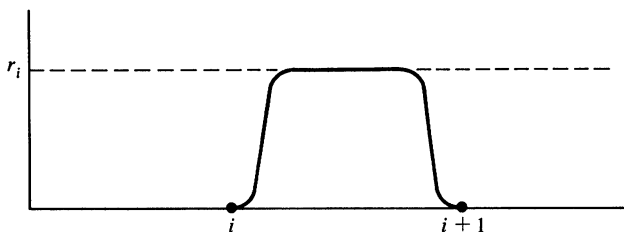


Figure 1-26

- \*6. Prove that the sphere  $S^k$  is simply connected if  $k > 1$ . [HINT: If  $f: S^1 \rightarrow S^k$  and  $k > 1$ , Sard gives you a point  $p \notin f(S^1)$ . Now use stereographic projection.]
7. When  $\dim X < \dim Y$ , Sard says that the image of any smooth map  $f: X \rightarrow Y$  has measure zero in  $Y$ . Prove this “mini-Sard” yourself, assuming the fact that if  $A$  has measure zero in  $\mathbf{R}^l$  and  $g: \mathbf{R}^l \rightarrow \mathbf{R}^l$  is

smooth, then  $g(A)$  also has measure zero. [HINT: Reduce to the case of a map  $f$  of an open set  $U \subset \mathbf{R}^k$  into  $\mathbf{R}^l$ . Consider  $F: U \times \mathbf{R}^{l-k} \rightarrow \mathbf{R}^l$  defined by  $F(x, t) = f(x)$ .]

8. Analyze the critical behavior of the following functions at the origin. Is the critical point nondegenerate? Is it isolated? Is it a local maximum or minimum?
  - (a)  $f(x, y) = x^2 + 4y^3$
  - (b)  $f(x, y) = x^2 - 2xy + y^2$
  - (c)  $f(x, y) = x^2 + y^4$
  - (d)  $f(x, y) = x^2 + 11xy + y^2/2 + x^6$
  - (e)  $f(x, y) = 10xy + y^2 + 75y^3$
9. Prove the Morse Lemma in  $\mathbf{R}^1$ . [HINT: Use this elementary calculus lemma: for any function  $f$  on  $R$  and any point  $a \in R$ , there is another function  $g$  such that

$$f(x) = f(a) + (x - a)f'(a) + (x - a)^2g(x).$$

This result is proven on page 135.]

10. Suppose that  $f = \sum a_{ij}x_ix_j$  in  $\mathbf{R}^k$ . Check that its Hessian matrix is  $H = (a_{ij})$ . Considering  $\mathbf{R}^k$  as the vector space of column vectors,  $H$  operates as a linear map by left multiplication, as usual. Show that if  $Hv = 0$ , then  $f$  is critical all along the line through  $v$  and 0. Thus the origin is an isolated critical point if and only if  $H$  is nonsingular.
11. Using the Morse Lemma, prove that if  $a$  is a nondegenerate critical point of a function  $f$ , there exists a local coordinate system  $(x_1, \dots, x_n)$  around  $a$  such that

$$f = f(a) + \sum_{i=1}^n \epsilon_i x_i^2, \quad \epsilon_i = \pm 1.$$

[HINT: Diagonalize  $(h_{ij})$ .]

12. Prove that the function  $f$  in Exercise 11 has a maximum at  $a$  if all the  $\epsilon_i$ 's are negative and a minimum if they are all positive. Show that if neither is the case, then  $a$  is neither a maximum nor a minimum.
13. Show that the determinant function on  $M(n)$  is Morse if  $n = 2$ , but not if  $n > 2$ .
14. Show that the "height function"  $(x_1, \dots, x_k) \rightarrow x_k$  on the sphere  $S^{k-1}$  is a Morse function with two critical points, the poles. Note that one pole is a maximum and the other a minimum.



15. Let  $X$  be a submanifold of  $\mathbf{R}^N$ . Prove that there exists a linear map  $l: \mathbf{R}^N \rightarrow \mathbf{R}$  whose restriction to  $X$  is a Morse function. (Exercise 14 is a special case.)
16. Let  $f$  be a smooth function on an open set  $U \subset \mathbf{R}^k$ . For each  $x \in U$ , let  $H(x)$  be the Hessian matrix of  $f$ , whether  $x$  is critical or not. Prove that  $f$  is Morse if and only if

$$\det(H)^2 + \sum_{i=1}^k \left( \frac{\partial f}{\partial x_i} \right)^2 > 0 \quad \text{on } U.$$

17. Suppose that  $f_t$  is a homotopic family of functions on  $\mathbf{R}^k$ . Show that if  $f_0$  is Morse in some neighborhood of a compact set  $K$ , then so is every  $f_t$  for  $t$  sufficiently small. [HINT: Show that the sum in Exercise 16 is bounded away from 0 on a neighborhood of  $K$  as long as  $t$  is small.]
18. (Stability of Morse Functions) Let  $f$  be a Morse function on the compact manifold  $X$ , and let  $f_t$  be a homotopic family of functions with  $f_0 = f$ . Show that each  $f_t$  is Morse if  $t$  is sufficiently small. [HINT: Exercise 17.]
19. Let  $X$  be a compact manifold. Prove that there exist Morse functions on  $X$  which take distinct values at distinct critical points. [HINT: Let  $f$  be Morse, and let  $x_1, \dots, x_N$  be its critical points. Let  $\rho_i$  be a smooth function on  $X$  that is one on a small neighborhood of  $x_i$  and zero outside a slightly larger neighborhood. Choose numbers  $a_1, \dots, a_N$  such that

$$f(x_i) + a_i \neq f(x_j) + a_j \quad \text{if } i \neq j.$$

Prove that if the  $a_i$  are small enough, then

$$f + \sum_{i=1}^N a_i \rho_i$$

has the same critical points as  $f$  and is even arbitrarily close to  $f$ .]

20. (a) Suppose that  $X$  is a compact manifold in  $\mathbf{R}^N$  and  $f$  is a function on  $X$ . Show that the  $N$ -tuples  $(a_1, \dots, a_N)$  for which

$$f_a = f + a_1 x_1 + \dots + a_N x_N$$

is a Morse function constitute an open set.

- (b) Remove the compactness assumption on  $X$ , and show that the set  $\{a: f_a \text{ Morse}\}$  is a countable intersection of open sets. [HINT: Use (a), plus the second axiom of countability.]
- (c) The set  $\{a: f_a \text{ not Morse}\}$  is therefore a countable union of closed sets. Show that this is enough to justify the use of Fubini in our proof of the existence of Morse functions. (See Appendix A.)

21. Let  $\phi: X \rightarrow \mathbf{R}^N$  be an immersion. Show for “almost every”  $a_1, \dots, a_N$ ,  $a_1\phi_1 + \dots + a_N\phi_N$  is a Morse function on  $X$  where  $\phi_1, \dots, \phi_N$  are the coordinate functions of  $\phi$ . [HINT: Show that the proof we gave for the existence of Morse functions only requires  $X$  to be immersed in  $\mathbf{R}^N$ , not embedded.]
22. Here is an application of Morse theory to electrostatics. Let  $x_1, \dots, x_4$  be points in general position in  $\mathbf{R}^3$  (that is, they don't all lie in a plane.) Let  $q_1, \dots, q_4$  be electric charges placed at these points. The potential function of the resulting electric field is

$$V_q = \frac{q_1}{r_1} + \dots + \frac{q_4}{r_4}$$

where  $r_i = |x - x_i|$ . The critical points of  $V_q$  are called *equilibrium points* of the electric field, and an equilibrium point is *non-degenerate* if the critical point is. Prove that for “almost every”  $q$  the equilibrium points of  $V_q$  are non-degenerate and finite in number. [HINT: Show that the map:  $\mathbf{R}^3 - \{x_1, \dots, x_4\} \rightarrow \mathbf{R}^4$  with coordinates  $r_1, r_2, r_3, r_4$  is an immersion and apply Exercise 21.]

## §8 Embedding Manifolds in Euclidean Space

The second application we shall give for Sard's theorem is a proof of the Whitney embedding theorem. A  $k$ -dimensional manifold  $X$  has been defined as a subset of some Euclidean space  $\mathbf{R}^N$  that may be enormous compared to  $X$ . This ambient Euclidean space is rather arbitrary when we consider the manifold  $X$  as an abstract object. For example, if  $M > N$ , then  $\mathbf{R}^N$  naturally embeds in  $\mathbf{R}^M$ , so one might have constructed the same manifold  $X$  in  $\mathbf{R}^M$  instead. Whitney inquired how large  $N$  must be in order that  $\mathbf{R}^N$  contain a diffeomorphic copy of every  $k$ -dimensional manifold. His preliminary answer was that  $N = 2k + 1$  suffices; this is the result we shall prove. After a great deal of hard work, Whitney improved his result by one, establishing that every  $k$ -dimensional manifold actually embeds in  $\mathbf{R}^{2k}$ .

One way to interpret the Whitney theorem is as a limit to the possible complexity of manifolds. Any manifold that may be defined in  $\mathbf{R}^N$  may also be defined in  $\mathbf{R}^{N+1}$ ; but perhaps the extra room for twisting in  $\mathbf{R}^{N+1}$  allows the construction of manifolds there that cannot exist inside the smaller space  $\mathbf{R}^N$ . (In fact, it is not a priori obvious that any single Euclidean space is large enough to contain all manifolds of a given dimension.) A classic example is the Klein bottle, a surface that can be constructed in  $\mathbf{R}^4$  by attaching the

circular ends of a cylinder to one another, but with the opposite orientations. (When ends are identified with the same orientation, the resulting manifold is just the torus.) In  $\mathbf{R}^3$ , an immersion of the Klein bottle exists, but self-intersection is unavoidable; there is not enough twisting room in  $\mathbf{R}^3$  for an embedding (Figure 1-27). (To envision an embedding in  $\mathbf{R}^4$ , represent the fourth dimension by density of red coloration and allow the bottle in the drawing to blush as it passes through itself.)

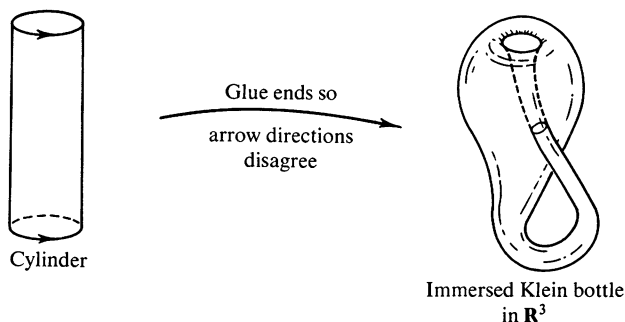


Figure 1-27

The Whitney theorem implies that after twice the dimension  $k$  of the manifolds being constructed, extra twisting room in huge Euclidean spaces is wasted; everything can be done in  $\mathbf{R}^{2k}$ . The Klein bottle illustrates that Whitney's result is optimal, for it is a  $k = 2$  dimensional manifold that does not embed in  $3 = 2k - 1$  space. (The circle is another example.)

Why  $N = 2k + 1$ ? As we see it, the underlying geometrical reason why  $2k + 1$  space can accommodate all  $k$ -dimensional manifolds is a basic property of transversality that will be proved in the next chapter. To wit, if  $X$  and  $Z$  are submanifolds of  $Y$  with  $\dim X + \dim Z < \dim Y$ , then one can pull  $X$  and  $Z$  apart by stretching either one an arbitrarily small amount (Exercise 6, Chapter 2, Section 3). For the moment, think loosely of a manifold as an intricate piece of plumbing, a construction made by joining  $k$ -dimensional pipes and various other standard fixtures (like  $k$ -hemispheres) according to some combinatorial prescription. Suppose that we follow instructions for fitting the pieces together inside  $\mathbf{R}^N$ . As the pipes are bent, they may occasionally pass through one another, like the immersed Klein bottle in  $\mathbf{R}^3$ . But once the linkages are complete, we can return to remove the self-intersections. As long as  $N \geq 2k + 1$ , a minute stretching of the intersecting sections will produce a manifold conforming to prescription. (This technique does not explain Whitney's optimal result  $N = 2k$ . For instance, intersecting curves can be pulled apart by local deformations in  $\mathbf{R}^3$  but not in  $\mathbf{R}^2$ .) See Figure 1-28.

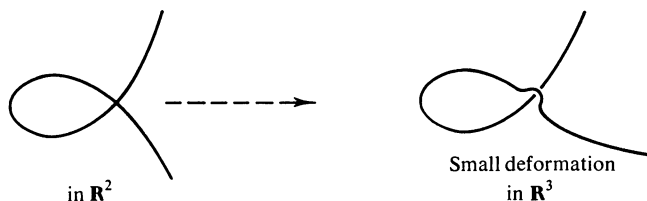


Figure 1-28

Modeling an acceptable proof on the heuristic argument above would be quite difficult. Fortunately, a straightforward, if less visual, proof may be derived directly from Sard's theorem.

A useful object in proving the Whitney theorem is the *tangent bundle* of a manifold  $X$  in  $\mathbf{R}^N$ . The tangent spaces to  $X$  at various points are vector subspaces of  $\mathbf{R}^N$  that will generally overlap one another. The tangent bundle  $T(X)$  is an artifice used to pull them apart. Specifically,  $T(X)$  is the subset of  $X \times \mathbf{R}^N$  defined by

$$T(X) = \{(x, v) \in X \times \mathbf{R}^N : v \in T_x(X)\}.$$

$T(X)$  contains a natural copy  $X_0$  of  $X$ , consisting of the points  $(x, 0)$ . In the direction perpendicular to  $X_0$ , it contains copies of each tangent space  $T_x(X)$ , embedded as the sets  $\{(x, v) : \text{with } x \text{ fixed}\}$ .

Any smooth map  $f: X \rightarrow Y$  induces a global *derivative map*  $df: T(X) \rightarrow T(Y)$ , defined by  $df(x, v) = (f(x), df_x(v))$ . Note that  $T(X)$  is a subset of Euclidean space; i.e.,  $X \subset \mathbf{R}^N$ , so  $T(X) \subset \mathbf{R}^N \times \mathbf{R}^N$ . Therefore if  $Y \subset \mathbf{R}^M$ , then  $df$  maps a subset of  $\mathbf{R}^{2N}$  into  $\mathbf{R}^{2M}$ . We claim that  $df$  is smooth. For since  $f: X \rightarrow \mathbf{R}^M$  is smooth, it extends around any point to a smooth map  $F: U \rightarrow \mathbf{R}^M$ , where  $U$  is an open set of  $\mathbf{R}^N$ . Then  $dF: T(U) \rightarrow \mathbf{R}^{2M}$  locally extends  $df$ . But  $T(U)$  is all of  $U \times \mathbf{R}^N$ , an open set in  $\mathbf{R}^{2N}$ , and as a map of this open set,  $dF$  is obviously defined by a smooth formula. This shows that  $df: T(X) \rightarrow \mathbf{R}^{2M}$  may be locally extended to a smooth map on an open subset of  $\mathbf{R}^{2N}$ , meaning that  $df$  is smooth.

The chain rule says that for smooth maps  $f: X \rightarrow Y$  and  $g: Y \rightarrow Z$ , the composite  $dg \circ df: T(X) \rightarrow T(Z)$  equals  $d(g \circ f)$ . Consequently, if  $f: X \rightarrow Y$  is a diffeomorphism, so is  $df: T(X) \rightarrow T(Y)$ , for the chain rule implies that  $df^{-1} \circ df$  is the identity map of  $T(X)$  and  $df \circ df^{-1}$  is the identity map of  $T(Y)$ . Thus diffeomorphic manifolds have diffeomorphic tangent bundles. As a result,  $T(X)$  is an object intrinsically associated to  $X$ ; it does not depend on the ambient Euclidean space.

Note that if  $W$  is an open set of  $X$ , and hence also a manifold, then  $T(W)$  is the subset  $T(X) \cap (W \times \mathbf{R}^N)$  in  $T(X)$ . Since  $W \times \mathbf{R}^N$  is open in  $X \times \mathbf{R}^N$ ,  $T(W)$  is open in the topology of  $T(X)$ . Now suppose that  $W$  is the image of a

local parametrization  $\phi: U \rightarrow W$ ,  $U$  being an open set in  $\mathbf{R}^k$ . Then  $d\phi: T(U) \rightarrow T(W)$  is a diffeomorphism. But  $T(U) = U \times \mathbf{R}^k$  is an open subset of  $\mathbf{R}^{2k}$ , so  $d\phi$  serves to parametrize the open set  $T(W)$  in  $T(X)$ . Since every point of  $T(X)$  sits in such a neighborhood, we have proved

**Proposition.** The tangent bundle of a manifold is another manifold, and  $\dim T(X) = 2 \dim X$ .

Now we prove a version of Whitney's result.

**Theorem.** Every  $k$ -dimensional manifold admits a one-to-one immersion in  $\mathbf{R}^{2k+1}$ .

*Proof.* In fact, if  $X \subset \mathbf{R}^N$  is  $k$ -dimensional and  $N > 2k + 1$ , we shall produce a linear projection  $\mathbf{R}^N \rightarrow \mathbf{R}^{2k+1}$  that restricts to a one-to-one immersion of  $X$ . Proceeding inductively, we prove that if  $f: X \rightarrow \mathbf{R}^M$  is an injective immersion with  $M > 2k + 1$ , then there exists a unit vector  $a \in \mathbf{R}^M$  such that the composition of  $f$  with the projection map carrying  $\mathbf{R}^M$  onto the orthogonal complement of  $a$  is still an injective immersion. Now the complement  $H = \{b \in \mathbf{R}^M : b \perp a\}$  is an  $M - 1$  dimensional vector subspace of  $\mathbf{R}^M$ , hence isomorphic to  $\mathbf{R}^{M-1}$ ; thus we obtain an injective immersion into  $\mathbf{R}^{M-1}$ .

Define a map  $h: X \times X \times \mathbf{R} \rightarrow \mathbf{R}^M$  by  $h(x, y, t) = t[f(x) - f(y)]$ . Also, define a map  $g: T(X) \rightarrow \mathbf{R}^M$  by  $g(x, v) = df_x(v)$ . Since  $M > 2k + 1$ , Sard's theorem implies that there exists a point  $a \in \mathbf{R}^M$  belonging to neither image; note that  $a \neq 0$ , since  $0$  belongs to both images.

Let  $\pi$  be the projection of  $\mathbf{R}^M$  onto the orthogonal complement  $H$  of  $a$ . Certainly  $\pi \circ f: X \rightarrow H$  is injective. For suppose that  $\pi \circ f(x) = \pi \circ f(y)$ . Then the definition of  $\pi$  implies that  $f(x) - f(y) = ta$  for some scalar  $t$ . If  $x \neq y$  then  $t \neq 0$ , because  $f$  is injective. But then  $h(x, y, 1/t) = a$ , contradicting the choice of  $a$ .

Similarly,  $\pi \circ f: X \rightarrow H$  is an immersion. For suppose that  $v$  is a nonzero vector in  $T_x(X)$  for which  $d(\pi \circ f)_x(v) = 0$ . Because  $\pi$  is linear, the chain rule yields  $d(\pi \circ f)_x = \pi \circ df_x$ . Thus  $\pi \circ df_x(v) = 0$ , so  $df_x(v) = ta$  for some scalar  $t$ . Because  $f$  is an immersion,  $t \neq 0$ . Thus  $g(x, 1/t) = a$ , again contradicting the choice of  $a$ . Q.E.D.

For compact manifolds, one-to-one immersions are the same as embeddings, so we have just proved the embedding theorem in the compact case. In general, we must modify the immersion to make it proper—a topological, not a differential problem. The situation is typical of differential topology; quite often fundamental differential concepts are most naturally and intuitively developed for compact manifolds, then extended by technical tricks to arbitrary manifolds. Rather than allow such technicalities to divert you now

from the primary conceptual development of our subject, you may choose to skip the remainder of this section for the present. Later you can return to learn the technique, which we shall need again.

The fundamental trick used for such generalizations is the following theorem.

**Theorem.** Let  $X$  be an arbitrary subset of  $\mathbf{R}^N$ . For any covering of  $X$  by (relatively) open subsets  $\{U_\alpha\}$ , there exists a sequence of smooth functions  $\{\theta_i\}$  on  $X$ , called a *partition of unity* subordinate to the open cover  $\{U_\alpha\}$ , with the following properties:

- (a)  $0 \leq \theta_i(x) \leq 1$  for all  $x \in X$  and all  $i$ .
- (b) Each  $x \in X$  has a neighborhood on which all but finitely many functions  $\theta_i$  are identically zero.
- (c) Each function  $\theta_i$  is identically zero except on some closed set contained in one of the  $U_\alpha$ .
- (d) For each  $x \in X$ ,  $\sum_i \theta_i(x) = 1$ . (Note that according to (b), this sum is always finite.)

*Proof.* Each  $U_\alpha$  may be written as  $X \cap W_\alpha$  for some open set  $W_\alpha$  in the ambient Euclidean space  $\mathbf{R}^N$ . Set  $W = \bigcup_\alpha W_\alpha$ , and let  $\{K_j\}$  be any nested sequence of compact sets exhausting the open set  $W$ . That is,

$$\bigcup_{j=1}^{\infty} K_j = W$$

and  $K_j \subset \text{Interior}(K_{j+1})$ . (For instance, let  $K_j = \{z \in W : |z| < j \text{ and distance of } z \text{ to } \mathbf{R}^N - W \geq 1/j\}$ .) The collection of all open balls of  $\mathbf{R}^N$  whose closures belong to at least one  $W_\alpha$  is an open cover of  $W$ . Select a finite number of such balls that cover the set  $K_2$ . According to Exercise 9 of Section 1, for each ball selected we may find a smooth nonnegative function on  $\mathbf{R}^N$  that is identically one on that ball and zero outside a closed set contained in one of the  $W_\alpha$ . Call these functions  $\eta_1, \eta_2, \dots, \eta_r$ . (See Figure 1-29.)

We continue building a sequence of functions inductively. For each  $j \geq 3$ , the compact set  $K_j - \text{Int}(K_{j-1})$  is contained inside the open set  $W - K_{j-2}$ . The collection of all open balls small enough to have their closures contained both in  $W - K_{j-2}$  and in some  $W_\alpha$  forms an open cover of  $K_j - \text{Int}(K_{j-1})$ . Extract a finite subcover, and then add to our sequence  $\{\eta_i\}$  one function for each ball; the function is to be equal to one on the ball and zero outside a closed set contained in both  $W - K_{j-2}$  and in one of the  $W_\alpha$ .

By construction, for each  $j$  only finitely many functions  $\eta_i$  fail to vanish on  $K_j$ . Thus since every point of  $W$  belongs to the interior of some  $K_j$ , the sum

$$\sum_{j=1}^{\infty} \eta_j$$

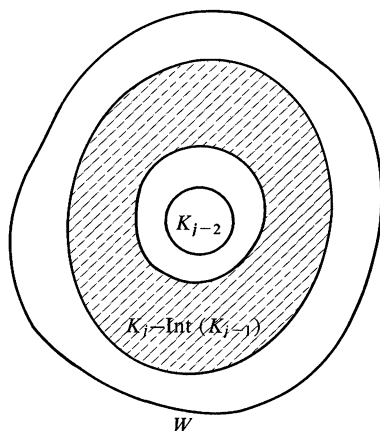


Figure 1-29

is actually finite in a neighborhood of every point of  $W$ . Furthermore, at least one term is nonzero at any point of  $W$ . Therefore

$$\frac{\eta_i}{\sum_{j=1}^{\infty} \eta_j}$$

is a well-defined smooth function. If we let  $\theta_i$  be the restriction of this to  $X$ , then we are done. Q.E.D.

**Corollary.** On any manifold  $X$  there exists a proper map  $\rho: X \rightarrow \mathbf{R}$

*Proof.* Let  $\{U_\alpha\}$  be the collection of open subsets of  $X$  that have compact closure, and let  $\theta_i$  be a subordinate partition of unity. Then

$$\rho = \sum_{i=1}^{\infty} i\theta_i$$

is a well-defined smooth function. If  $\rho(x) \leq j$ , then obviously at least one of the first  $j$  functions  $\theta_1, \dots, \theta_j$  must be nonzero at  $x$ . Therefore  $\rho^{-1}([ -j, j ])$  is contained in

$$\bigcup_{i=1}^j \{x : \theta_i(x) \neq 0\},$$

a set with compact closure. But every compact set in  $\mathbf{R}$  is contained in some interval  $[ -j, j ]$ . Q.E.D.

**Whitney Theorem.** Every  $k$ -dimensional manifold embeds in  $\mathbf{R}^{2k+1}$ .

*Proof.* Begin with a one-to-one immersion of  $X$  into  $\mathbf{R}^{2k+1}$ . Composing with any diffeomorphism of  $\mathbf{R}^{2k+1}$  into its unit ball—say,  $z \rightarrow z/(1 + |z|^2)$ —

we obtain an injective immersion  $f: X \rightarrow \mathbf{R}^{2k+1}$  such that  $|f(x)| < 1$  for all  $x \in X$ . Let  $\rho: X \rightarrow \mathbf{R}$  be a proper function, and define a new injective immersion  $F: X \rightarrow \mathbf{R}^{2k+2}$  by  $F(x) = (f(x), \rho(x))$ . Now drop back down to  $\mathbf{R}^{2k+1}$  as in the earlier theorem by composing  $F$  with an orthogonal projection  $\pi: \mathbf{R}^{2k+2} \rightarrow H$ , where  $H$  is the linear space perpendicular to a suitable unit vector,  $a$  in  $\mathbf{R}^{2k+2}$ .

Recall that the map  $\pi \circ F: X \rightarrow H$  is still an injective immersion for almost every  $a \in S^{2k+1}$ , so we may pick an  $a$  that happens to be neither of the sphere's two poles. But now  $\pi \circ F$  is easily seen to be proper. In fact, given any bound  $c$ , we claim that there exists another number  $d$  such that the set of points  $x \in X$  where  $|\pi \circ F(x)| \leq c$  is contained in the set where  $|\rho(x)| \leq d$ . As  $\rho$  is proper, the latter is a compact subset of  $X$ . Thus the claim implies that the preimage under  $\pi \circ F$  of every closed ball in  $H$  is a compact subset of  $X$ , showing that  $\pi \circ F$  is proper. If the claim is false, then there exists a sequence of points  $\{x_i\}$  in  $X$  for which  $|\pi \circ F(x_i)| < c$  but  $\rho(x_i) \rightarrow \infty$ . Remember that, by definition, for every  $z \in \mathbf{R}^{2k+2}$  the vector  $\pi(z)$  is the one point in  $H$  for which  $z - \pi(z)$  is a multiple of  $a$ . Thus  $F(x_i) - \pi \circ F(x_i)$  is a multiple of  $a$  for each  $i$ , and hence so is the vector

$$w_i = \frac{1}{\rho(x_i)}[F(x_i) - \pi \circ F(x_i)].$$

Consider what happens as  $i \rightarrow \infty$ .

$$\frac{F(x_i)}{\rho(x_i)} = \left( \frac{f(x_i)}{\rho(x_i)}, 1 \right) \rightarrow (0, \dots, 0, 1),$$

because  $|f(x_i)| < 1$  for all  $i$ . The quotient

$$\frac{\pi \circ F(x_i)}{\rho(x_i)}$$

has norm  $\leq c/\rho(x_i)$ , so it converges to zero. Thus  $w_i \rightarrow (0, \dots, 0, 1)$ . But each  $w_i$  is a multiple of  $a$ ; therefore so is the limit. We conclude that  $a$  must be either the north or south pole of  $S^{2k+1}$ , a contradiction. This proves the claim and the theorem. Q.E.D.

## EXERCISES

1. Show that  $T(\mathbf{R}^k) = \mathbf{R}^k \times \mathbf{R}^k$ .
2. Let  $g$  be a smooth, everywhere-positive function on  $X$ . Check that the multiplication map  $T(X) \rightarrow T(X)$ ,  $(x, v) \rightarrow (x, g(x)v)$ , is smooth.



3. Show that  $T(X \times Y)$  is diffeomorphic to  $T(X) \times T(Y)$ .
4. Show that the tangent bundle to  $S^1$  is diffeomorphic to the cylinder  $S^1 \times \mathbf{R}^1$ .
5. Prove that the *projection* map  $p: T(X) \rightarrow X$ ,  $p(x, v) = x$ , is a submersion.
- \*6. A *vector field*  $\vec{v}$  on a manifold  $X$  in  $\mathbf{R}^N$  is a smooth map  $\vec{v}: X \rightarrow \mathbf{R}^N$  such that  $\vec{v}(x)$  is always tangent to  $X$  at  $x$ . Verify that the following definition (which does not explicitly mention the ambient  $\mathbf{R}^N$ ) is equivalent: a vector field  $\vec{v}$  on  $X$  is a *cross section* of  $T(X)$ —that is, a smooth map  $\vec{v}: X \rightarrow T(X)$  such that  $p \circ \vec{v}$  equals the identity map of  $X$ . ( $p$  as in Exercise 5.)
- \*7. A point  $x \in X$  is a *zero* of the vector field  $\vec{v}$  if  $\vec{v}(x) = 0$ . Show that if  $k$  is odd, there exists a vector field  $\vec{v}$  on  $S^k$  having no zeros. [HINT: For  $k = 1$ , use  $(x_1, x_2) \rightarrow (-x_2, x_1)$ .] It is a rather deep topological fact that nonvanishing vector fields do not exist on the even spheres. We will see why in Chapter 3.
- \*8. Prove that if  $S^k$  has a nonvanishing vector field, then its antipodal map is homotopic to the identity (Compare Section 6, Exercise 7.) [HINT: Show that you may take  $|\vec{v}(x)| = 1$  everywhere. Now rotate  $x$  to  $-x$  in the direction indicated by  $\vec{v}(x)$ .]
9. Let  $S(X)$  be the set of points  $(x, v) \in T(X)$  with  $|v| = 1$ . Prove that  $S(X)$  is a  $2k - 1$  dimensional submanifold of  $T(X)$ ; it is called the *sphere bundle* of  $X$ . [HINT: Consider the map  $(x, v) \rightarrow |v|^2$ .]
10. *The Whitney Immersion Theorem.* Prove that every  $k$ -dimensional manifold  $X$  may be immersed in  $\mathbf{R}^{2k}$ .
11. Show that if  $X$  is a compact  $k$ -dimensional manifold, then there exists a map  $X \rightarrow \mathbf{R}^{2k-1}$  that is an immersion except at finitely many points of  $X$ . Do so by showing that if  $f: X \rightarrow \mathbf{R}^{2k}$  is an immersion and  $a$  is a regular value for the map  $F: T(X) \rightarrow \mathbf{R}^{2k}$ ,  $F(x, v) = df_x(v)$ , then  $F^{-1}(a)$  is a finite set. Show that  $\pi \circ f$  is an immersion except on  $f^{-1}(a)$ , where  $\pi$  is an orthogonal projection perpendicular to  $a$ . The exceptional points, in  $f^{-1}(a)$ , are called *cross caps*. [HINT: Show that there are only finitely many preimages of  $a$  under  $F$  in the compact set  $\{(x, v): |v| \leq 1\} \subset T(X)$ . For if  $(x_i, v_i)$  are infinitely many preimages, pick a subsequence so that  $x_i \rightarrow x$ ,  $v_i/|v_i| \rightarrow w$ . Now show that  $df_x(w) = 0$ .]

12. Whitney showed<sup>†</sup> that for maps of two-manifolds into  $\mathbf{R}^3$ , a typical cross cap looks like the map  $(x, y) \rightarrow (x, xy, y^2)$ . Check that this is an immersion except at the origin. What does its image look like?
13. An open cover  $\{V_\alpha\}$  of a manifold  $X$  is *locally finite* if each point of  $X$  possesses a neighborhood that intersects only finitely many of the sets  $V_\alpha$ . Show that any open cover  $\{U_\alpha\}$  admits a locally finite refinement  $\{V_\alpha\}$ . [HINT: Partition of unity.]
- \*14. *Inverse Function Theorem Revisited.* Use a partition-of-unity technique to prove a noncompact version of Exercise 10, Section 3. Suppose that the derivative of  $f: X \rightarrow Y$  is an isomorphism whenever  $x$  lies in the submanifold  $Z \subset X$ , and assume that  $f$  maps  $Z$  diffeomorphically onto  $f(Z)$ . Prove that  $f$  maps a neighborhood of  $Z$  diffeomorphically onto a neighborhood of  $f(Z)$ . [Outline: Find local inverses  $g_i: U_i \rightarrow X$ , where  $\{U_i\}$  is a locally finite collection of open subsets of  $Y$  covering  $f(Z)$ . Define  $W = \{y \in U_i: g_i(y) = g_j(y) \text{ whenever } y \in U_i \cap U_j\}$ . The maps  $g_i$  “patch together” to define a smooth inverse  $g: W \rightarrow X$ . Finish by proving that  $W$  contains an open neighborhood of  $f(Z)$ ; this is where local finiteness is needed.]
15. *The Smooth Urysohn Theorem.* If  $A$  and  $B$  are disjoint, smooth, closed subsets of a manifold  $X$ , prove that there is a smooth function  $\phi$  on  $X$  such that  $0 \leq \phi \leq 1$  with  $\phi = 0$  on  $A$  and  $\phi = 1$  on  $B$ . [HINT: Partition of unity.]

<sup>†</sup>H. Whitney, “The General Type of Singularity of a Set of  $2n - 1$  Functions of  $n$  Variables,” *Duke Math. Journal*, **10** (1943), 161–172.

# Transversality and Intersection

## §1 Manifolds with Boundary

We now enlarge the class of geometric objects under study by allowing our manifolds to possess boundaries. For example, we wish to consider spaces like the closed unit ball in  $\mathbf{R}^n$ , whose boundary is  $S^{n-1}$ , or the compact cylindrical surface  $S^1 \times [0, 1]$  in  $\mathbf{R}^3$ , which is bounded by two copies of the circle. (See Figure 2-1.) These spaces fail to be manifolds because neighborhoods of points in their boundaries are not diffeomorphic to open sets in Euclidean space. (See Exercise 1.) The simplest example of all is the upper half-space  $H^k$  in  $\mathbf{R}^k$ , consisting of all points with nonnegative final coordinate. The boundary of  $H^k$  is  $\mathbf{R}^{k-1}$  under its usual embedding in  $\mathbf{R}^k$ . Because of its simplicity, we adopt  $H^k$  as our model. (See Figure 2-2.)

**Definition.** A subset  $X$  of  $\mathbf{R}^N$  is a  $k$ -dimensional manifold with boundary if every point of  $X$  possesses a neighborhood diffeomorphic to an open set in the space  $H^k$ . As before, such a diffeomorphism is called a *local parametrization* of  $X$ . The *boundary* of  $X$ , denoted  $\partial X$ , consists of those points that belong to the image of the boundary of  $H^k$  under some local parametrization. Its complement is called the interior of  $X$ ,  $\text{Int}(X) = X - \partial X$ .

Do not confuse the boundary or interior of  $X$ , as we have defined them, with the topological boundary or interior of  $X$  as a subset of  $\mathbf{R}^N$ . Although

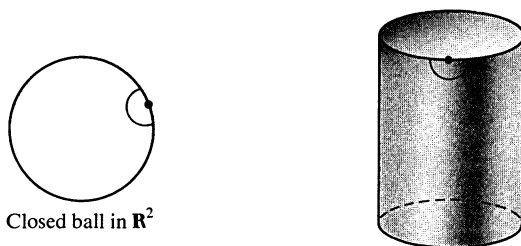


Figure 2-1

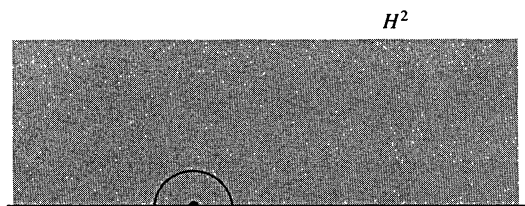


Figure 2-2

the notions often agree when  $\dim X = N$ , they definitely do not correspond when  $\dim X < N$ . We shall always use the words in the manifold sense. Note that manifolds as defined earlier also qualify as “manifolds with boundary,” although their boundaries are empty. We shall still reserve the unmodified term “manifold” for these, even though, for emphasis, we will sometimes call them “boundaryless.”

The product of two manifolds with boundary, unfortunately, is not generally another manifold with boundary, as the square  $[0, 1] \times [0, 1]$  illustrates. But at least the following proposition is true.

**Proposition.** The product of a manifold without boundary  $X$  and a manifold with boundary  $Y$  is another manifold with boundary. Furthermore,

$$\partial(X \times Y) = X \times \partial Y,$$

and

$$\dim(X \times Y) = \dim X + \dim Y.$$

*Proof.* If  $U \subset \mathbb{R}^k$  and  $V \subset H^l$  are open, then

$$U \times V \subset \mathbb{R}^k \times H^l = H^{k+l}$$

is open. Moreover, if  $\phi: U \rightarrow X$  and  $\psi: V \rightarrow Y$  are local parametrizations, so is  $\phi \times \psi: U \times V \rightarrow X \times Y$ . Q.E.D.

The most important application of this observation will be to the homotopy mapping space  $X \times I$  of a boundaryless manifold  $X$ .

Tangent spaces and derivatives are still defined in the setting of manifolds with boundary. First, suppose that  $g$  is a smooth map of an open set  $U$  of  $H^k$  into  $\mathbf{R}^l$ . If  $u$  is an interior point of  $U$ , then the derivative  $dg_u$  is already defined. If  $u \in \partial U$ , the smoothness of  $g$  means that it may be extended to a smooth map  $\tilde{g}$  defined in an open neighborhood of  $u$  in  $\mathbf{R}^k$ ; define  $d\tilde{g}_u$  to be the derivative  $d\tilde{g}_u: \mathbf{R}^k \rightarrow \mathbf{R}^l$ . We must show that if  $\tilde{g}$  is another local extension of  $g$ , then  $d\tilde{g}_u = d\tilde{g}_u$ . Let  $u_i$  be any sequence of points in  $\text{Int}(U)$  that converge to  $u$ . Because  $\tilde{g}$  and  $\tilde{g}$  both agree with  $g$  on  $\text{Int}(U)$ ,

$$d\tilde{g}_{u_i} = d\tilde{g}_{u_i}.$$

Now use the continuity of these derivatives with respect to  $u_i$ ; letting  $u_i \rightarrow u$ , we obtain  $d\tilde{g}_u = d\tilde{g}_u$ , as desired. Remember: even at boundary points, the derivative  $dg_u$  is still a linear map of all of  $\mathbf{R}^k$  into  $\mathbf{R}^l$ .

It is trivial to verify that differentiation of smooth maps defined on half-spaces still follows the chain rule. This allows us to extend differentiation to manifolds with boundary exactly as we did for boundaryless manifolds. If  $X \subset \mathbf{R}^N$  is a  $k$ -dimensional manifold with boundary, define its *tangent space*  $T_x(X)$  at a point  $x \in X$  to be the image of the derivative of any local parametrization around  $x$ . Check that  $T_x(X)$  is a  $k$ -dimensional linear subspace of  $\mathbf{R}^N$ —even when  $x$  is a boundary point!—whose definition is independent of the choice of parametrization. Also verify that for a smooth map  $f: X \rightarrow Y$  of two manifolds with boundary, the derivative at any point  $x$  may be defined just as before as a linear transformation  $df_x: T_x(X) \rightarrow T_{f(x)}(Y)$ . Moreover, the chain rule remains valid.

If  $X$  is a manifold with boundary,  $\text{Int}(X)$  is automatically a boundaryless manifold of the same dimension as  $X$ . The reason is that any interior point is in the range of a local parametrization whose domain is an open set of  $H^k$  contained entirely in  $\text{Int}(H^k)$  and therefore is an open set of  $\mathbf{R}^k$ . More interesting is

**Proposition.** If  $X$  is a  $k$ -dimensional manifold with boundary, then  $\partial X$  is a  $(k - 1)$  dimensional manifold without boundary.

*Proof.* The essential point is to show that if  $x$  is in the boundary with respect to one system of local coordinates, then it is in the boundary with respect to any other system. If  $x \in \partial X$ , there is a local parametrization  $\phi: U \rightarrow V$ , where  $U$  is an open subset of  $H^k$  and  $V$  is an open subset of  $X$ . We need only show that  $\phi(\partial U) = \partial V$ , for then  $\phi$  restricts to a diffeomorphism of  $\partial U = U \cap \partial H^k$ , an open set in  $\mathbf{R}^{k-1}$ , with  $\partial V = \partial X \cap V$ , a neighborhood of  $x$  in  $\partial X$ . By definition,  $\phi(\partial U) \subset \partial V$ , so we need to demonstrate that  $\phi(\partial U) \supset \partial V$ . That is, if  $\psi$  is any local parametrization mapping an open set  $W$  of  $H^k$  into  $V$ , we must show that  $\phi(\partial U) \supset \psi(\partial W)$ , or, equivalently,  $\partial U \supset \phi^{-1} \circ \psi(\partial W)$ . So let  $g = \phi^{-1} \circ \psi: W \rightarrow U$ , and suppose that some

$w \in \partial W$  maps to an interior point  $u = g(w)$  of  $U$ . As  $\phi$  and  $\psi$  are both diffeomorphisms,  $g$  must be a diffeomorphism of  $W$  onto an open subset  $g(W)$  of  $U$ . As usual, the chain rule implies that the derivative of its inverse,  $d(g^{-1})_u$ , is bijective. But since  $u \in \text{Int}(U)$ ,  $g(W)$  contains a neighborhood of  $u$  that is open in  $\mathbf{R}^k$ . Thus the Inverse Function Theorem, applied to the map  $g^{-1}$  defined on this open subset of  $\mathbf{R}^k$ , implies that the image of  $g^{-1}$  contains a neighborhood of  $w$  that is open in  $\mathbf{R}^k$ . This contradicts the assumption  $w \in \partial W$ . Q.E.D.

Observe that if  $x \in \partial X$ , then the tangent space to the boundary  $T_x(\partial X)$  is a linear subspace of  $T_x(X)$  with codimension 1. (See Figure 2-3.) For any smooth map  $f$  defined on  $X$ , let us introduce the notation  $\partial f$  for the restriction of  $f$  to  $\partial X$ . The derivative of  $\partial f$  at  $x$  is just the restriction of  $df_x$  to the subspace  $T_x(\partial X)$ .

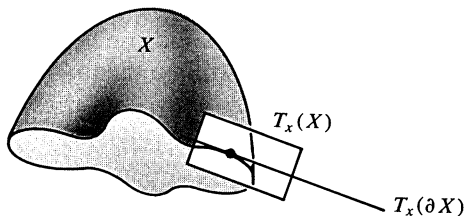


Figure 2-3

All definitions that have been formulated in terms of derivatives of mappings make sense verbatim when the manifolds have boundaries. However, in seeking extensions of the fundamental theorems of the first chapter, we must impose additional restrictions on our maps. We would like conditions that would guarantee that if  $f: X \rightarrow Y$  encounters a submanifold  $Z$  of  $Y$ , then  $f^{-1}(Z)$  is a manifold with boundary. We also want  $\partial f^{-1}(Z) = f^{-1}(Z) \cap \partial X$ . Unfortunately, the transversality of  $f$  alone doesn't guarantee this. (For example, let  $f: H^2 \rightarrow \mathbf{R}$  be the map  $(x_1, x_2) \rightarrow x_2$ , and let  $Z$  be  $\{0\}$ . Then  $f^{-1}(Z) = \partial H^2$ .) The right condition turns out to be an additional transversality assumption along the boundary.

**Theorem.** Let  $f$  be a smooth map of a manifold  $X$  with boundary onto a boundaryless manifold  $Y$ , and suppose that both  $f: X \rightarrow Y$  and  $\partial f: \partial X \rightarrow Y$  are transversal with respect to a boundaryless submanifold  $Z$  in  $Y$ . Then the preimage  $f^{-1}(Z)$  is a manifold with boundary

$$\partial\{f^{-1}(Z)\} = f^{-1}(Z) \cap \partial X,$$

and the codimension of  $f^{-1}(Z)$  in  $X$  equals the codimension of  $Z$  in  $Y$ .

*Proof.* The restriction of  $f$  to the boundaryless manifold  $\text{Int}(X)$  is transversal to  $Z$ ; so, by the earlier theorem,  $f^{-1}(Z) \cap \text{Int}(X)$  is a boundaryless manifold of correct codimension. Now we need only examine  $f^{-1}(Z)$  in a neighborhood of a point  $x \in f^{-1}(Z) \cap \partial X$ . As usual, reduce to the case where  $Z$  is a single point by introducing a submersion  $\phi$  of a neighborhood of  $f(x)$  in  $Y$  onto  $\mathbf{R}^l$ , such that in this neighborhood  $Z = g^{-1}(0)$ . Here  $l = \text{codim } Z$ . Then  $\phi \circ f$  is defined in a neighborhood of  $x$  in  $X$ , and the intersection of  $f^{-1}(Z)$  with that neighborhood is  $(\phi \circ f)^{-1}(0)$ . We now refer  $\phi \circ f$  back to Euclidean space by choosing a local parametrization  $h: U \rightarrow X$  around  $x$ , where  $U$  is an open set of  $H^k$ , and then setting  $g = \phi \circ f \circ h$ . As  $h: U \rightarrow h(U)$  is a diffeomorphism, the set  $f^{-1}(Z)$  is a manifold with boundary in a neighborhood of  $x$  if and only if  $(f \circ h)^{-1}(Z) = g^{-1}(0)$  is a manifold with boundary near  $u = h^{-1}(x) \in \partial U$ .

Just as in the boundaryless case, the transversality assumption

$$df_x(T_x(X)) + T_{f(x)}(Z) = T_{f(x)}(Y)$$

translates into the fact that  $x$  is a regular point of  $\phi \circ f$  or, equivalently, that  $g$  is regular at  $u$ . By definition, the smoothness of  $g$  means that it extends to a smooth map  $\tilde{g}$  defined on a neighborhood  $\tilde{U}$  of  $u$  in  $\mathbf{R}^k$ . As  $d\tilde{g}_u = dg_u$ ,  $\tilde{g}$  is also regular at  $u$ . Since  $\tilde{g}$  is a map of boundaryless manifolds, the preimage  $\tilde{g}^{-1}(0)$ , intersected with some neighborhood of its regular point  $u$ , is a boundaryless submanifold  $S$  of  $\mathbf{R}^k$ .

As  $g^{-1}(0) = S \cap H^k$  in a neighborhood of  $u$ , we must show that  $S \cap H^k$  is a manifold with boundary. This is where the transversality assumption on  $\partial f$  is essential. Denote by  $\pi$  the restriction to  $S$  of the last coordinate function on  $\mathbf{R}^k$ ,  $\pi: S \rightarrow \mathbf{R}$ . Then

$$S \cap H^k = \{s \in S: \pi(s) \geq 0\}.$$

We claim that 0 is a regular value for  $\pi$ , for, if not, there is a point  $s \in S$  such that  $\pi(s) = 0$  and  $d\pi_s = 0$ . Of course,  $\pi(s) = 0$  simply means that  $s \in S \cap \partial H^k$ . Also, as  $\pi: \mathbf{R}^k \rightarrow \mathbf{R}$  is linear,  $d\pi_s$  equals  $\pi$ . Thus the fact that  $d\pi_s$  is zero on  $T_s(S)$  means only that the last coordinate of every vector in  $T_s(S)$  is zero or, equivalently, that

$$T_s(S) \subset T_s(\partial H^k) = \mathbf{R}^{k-1}.$$

But since  $S = \tilde{g}^{-1}(0)$ , we know that the kernel of  $dg_s = d\tilde{g}_s: \mathbf{R}^k \rightarrow \mathbf{R}$  is just  $T_s(S)$ . Now the derivative of  $\partial g$  at  $s$  is the restriction of  $dg_s: \mathbf{R}^k \rightarrow \mathbf{R}$  to  $\mathbf{R}^{k-1}$ . Therefore if the kernel of  $dg_s$  is contained in  $\mathbf{R}^{k-1}$ , then the two linear maps  $dg_s: \mathbf{R}^k \rightarrow \mathbf{R}$  and  $d(\partial g)_s: \mathbf{R}^{k-1} \rightarrow \mathbf{R}$  must have the same kernel. But the transversality conditions imply that both maps are surjective; so the standard dimension relation for linear maps tells us that the kernel of  $dg_s$

has dimension  $k - 1$ , whereas the kernel of  $d(\partial g)_s$  has dimension  $k - 2$ . Because this fact contradicts the conclusion we just deduced, it must be that 0 is a regular value for  $g$ . Finally, the following lemma completes the proof.

**Lemma.** Suppose that  $S$  is a manifold without boundary and that  $\pi: S \rightarrow \mathbf{R}$  is a smooth function with regular value 0. Then the subset  $\{s \in S: \pi(s) \geq 0\}$  is a manifold with boundary, and the boundary is  $\pi^{-1}(0)$ .

*Proof.* The set where  $\pi > 0$  is open in  $S$  and is therefore a submanifold of the same dimension as  $S$ . So suppose that  $\pi(s) = 0$ . Because  $\pi$  is regular at  $s$ , it is locally equivalent to the canonical submersion near  $s$ . But the lemma is obvious for the canonical submersion. Q.E.D.

The lemma is not without independent interest. For example, setting  $S = \mathbf{R}^n$  and  $\pi(s) = 1 - |s|^2$ , it proves that the closed unit ball  $\{s \in \mathbf{R}^n: |s| \leq 1\}$  is a manifold with boundary.

The generalization of Sard's theorem to manifolds with boundary is more straightforward.

**Sard's theorem.** For any smooth map  $f$  of a manifold  $X$  with boundary into a boundaryless manifold  $Y$ , almost every point of  $Y$  is a regular value of both  $f: X \rightarrow Y$  and  $\partial f: \partial X \rightarrow Y$ .

*Proof.* Because the derivative of  $\partial f$  at a point  $x \in \partial X$  is just the restriction of  $df_x$  to the subspace  $T_x(\partial X) \subset T_x(X)$ , it is obvious that if  $\partial f$  is regular at  $x$ , so is  $f$ . Thus a point  $y \in Y$  fails to be a regular value of both  $f: X \rightarrow Y$  and  $\partial f: \partial X \rightarrow Y$  only when it is a critical value of  $f: \text{Int}(X) \rightarrow Y$  or  $\partial f: \partial X \rightarrow Y$ . But since  $\text{Int}(X)$  and  $\partial X$  are both boundaryless manifolds, both sets of critical values have measure zero. Thus the complement of the set of common regular values for  $f$  and  $\partial f$ , being the union of two sets of measure zero, itself has measure zero. Q.E.D.

## EXERCISES

1. If  $U \subset \mathbf{R}^k$  and  $V \subset H^k$  are neighborhoods of 0, prove that there exists no diffeomorphism of  $V$  with  $U$ .
2. Prove that if  $f: X \rightarrow Y$  is a diffeomorphism of manifolds with boundary, then  $\partial f$  maps  $\partial X$  diffeomorphically onto  $\partial Y$ .
3. Show that the square  $S = [0, 1] \times [0, 1]$  is not a manifold with boundary. [HINT: If  $f$  maps a neighborhood of the corner  $s$  into  $H^2$ , and carries boundary to boundary, show that two independent vectors  $v_1$  and  $v_2$  in  $T_s(S)$  are mapped to dependent vectors  $df_s(v_1), df_s(v_2)$ .] See Figure 2-4.



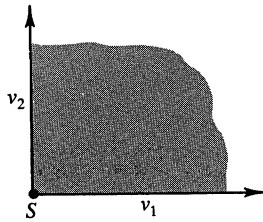


Figure 2-4

4. Show that the solid hyperboloid  $x^2 + y^2 - z^2 \leq a$  is a manifold with boundary ( $a > 0$ ).
5. Indicate for which values of  $a$  the intersection of the solid hyperboloid  $x^2 + y^2 - z^2 \leq a$  and the unit sphere  $x^2 + y^2 + z^2 = 1$  is a manifold with boundary? What does it look like?
6. There are two standard ways of making manifolds with boundary out of the unit square by gluing a pair of opposite edges (Figure 2-5). Simple gluing produces the cylinder, whereas gluing after one twist produces the closed Möbius band. Check that the boundary of the cylinder is two copies of  $S^1$ , while the boundary of the Möbius band is one copy of  $S^1$ ; consequently, the cylinder and Möbius band are not diffeomorphic. What happens if you twist  $n$  times before gluing?

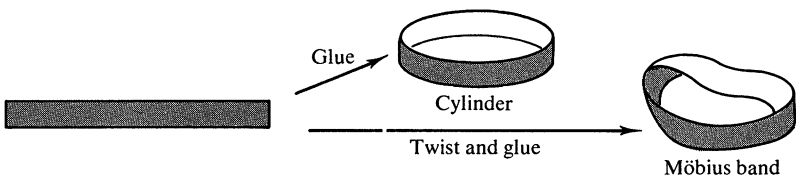


Figure 2-5

- \*7. Suppose that  $X$  is a manifold with boundary and  $x \in \partial X$ . Let  $U \xrightarrow{\phi} X$  be a local parametrization with  $\phi(0) = x$ , where  $U$  is an open subset of  $H^k$ . Then  $d\phi_0 : \mathbb{R}^k \rightarrow T_x(X)$  is an isomorphism. Define the *upper half-space*  $H_x(X)$  in  $T_x(X)$  to be the image of  $H^k$  under  $d\phi_0$ ,  $H_x(X) = d\phi_0(H^k)$ . Prove that  $H_x(X)$  does not depend on the choice of local parametrization.
- \*8. Show that there are precisely two unit vectors in  $T_x(X)$  that are perpendicular to  $T_x(\partial X)$  and that one lies inside  $H_x(X)$ , the other outside. The one in  $H_x(X)$  is called the *inward unit normal vector* to the boundary,

and the other is the *outward unit normal vector* to the boundary. Denote the outward unit normal by  $\vec{n}(x)$ . Note that if  $X$  sits in  $\mathbf{R}^N$ ,  $\vec{n}$  may be considered to be a map of  $\partial X$  into  $\mathbf{R}^N$ . Prove that  $\vec{n}$  is smooth. (Specifically, what is  $\vec{n}(x)$  for  $X = H^k$ ?) See Figure 2-6.

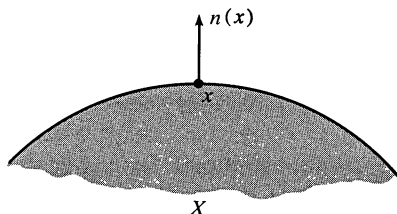


Figure 2-6

9. (a) Show that  $\partial X$  is a closed subset of  $X$ . (In particular,  $\partial X$  is compact if  $X$  is.)  
 (b) Find some examples in which  $\partial X$  is compact but  $X$  is not.
10. Let  $x \in \partial X$  be a boundary point. Show that there exists a smooth non-negative function  $f$  on some open neighborhood  $U$  of  $x$ , such that  $f(z) = 0$  if and only if  $z \in \partial U$ , and if  $z \in \partial U$ , then  $df_z(\vec{n}(z)) > 0$ .
11. (Converse to Lemma of page 62) Show that if  $X$  is any manifold with boundary, then there exists a smooth nonnegative function  $f$  on  $X$ , with a regular value at 0, such that  $\partial X = f^{-1}(0)$ . [HINT: Use a partition of unity to glue together the local functions of Exercise 10. What guarantees regularity?]

## §2 One-Manifolds and Some Consequences

Up to diffeomorphism, the only compact, connected, one-dimensional manifolds with boundary are the closed interval and the circle. This is one of those absolutely obvious statements whose proof turns out to be technically less trivial than expected. The idea is simple enough. Beginning at some particular point, just run along the curve at constant speed. Since the manifold is compact, you cannot run forever over new territory; either you arrive again at your starting point and the curve must be a circle, or else you run into a boundary point and it is an interval. A careful proof has been provided in an appendix, so for the moment we simply assert

**The Classification of One-Manifolds.** Every compact, connected, one-dimensional manifold with boundary is diffeomorphic to  $[0, 1]$  or  $S^1$ .

As every compact one-manifold with boundary is the disjoint union of finitely many connected components, we obtain a trivial corollary that will have surprisingly nontrivial applications.

**Corollary.** The boundary of any compact one-dimensional manifold with boundary consists of an even number of points.

The first use to which we put this corollary is in proving:

**Theorem.** If  $X$  is any compact manifold with boundary, then there exists no smooth map  $g: X \rightarrow \partial X$  such that  $\partial g: \partial X \rightarrow \partial X$  is the identity. That is, there is no “retraction” of  $X$  onto its boundary.

*Proof.* Suppose that such a  $g$  exists, and let  $z \in \partial X$  be a regular value. ( $z$  exists thanks to Sard.) Then  $g^{-1}(z)$  is a submanifold of  $X$  with boundary. As the codimension of  $g^{-1}(z)$  in  $X$  equals the codimension of  $\{z\}$  in  $\partial X$ , namely  $\dim X - 1$ ,  $g^{-1}(z)$  is one dimensional and compact. But since  $\partial g = \text{identity}$ ,

$$\partial g^{-1}(z) = g^{-1}(z) \cap \partial X = \{z\},$$

contradicting the corollary. Q.E.D.

We can now prove a famous theorem of Brouwer, usually proved either by using the sophisticated machinery of algebraic topology or through elaborate combinatorics. (The following “transversality” proof is due to M. Hirsch.)

**Brouwer Fixed-Point Theorem.** Any smooth map  $f$  of the closed unit ball  $B^n \subset \mathbf{R}^n$  into itself must have a fixed point; that is,  $f(x) = x$  for some  $x \in B^n$ .

*Proof.* Suppose that there exists an  $f$  without fixed points, and we shall construct a retraction  $g: B^n \rightarrow \partial B^n$ . Since  $f(x) \neq x$ , the two points  $x$  and  $f(x)$  determine a line. Let  $g(x)$  be the point where the line segment starting at  $f(x)$  and passing through  $x$  hits the boundary (Figure 2-7). If  $x \in \partial B^n$  already,

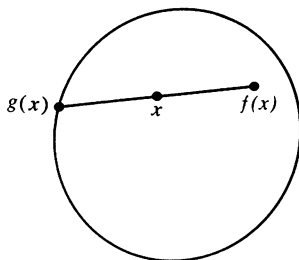


Figure 2-7

$g(x) = x$ . Thus  $g: B^n \rightarrow B^n$  is the identity on  $\partial B^n$ . We only need show that  $g$  is smooth to obtain a contradiction to the retraction theorem and thereby complete the proof. Since  $x$  is in the line segment between  $f(x)$  and  $g(x)$ , we may write the vector  $g(x) - f(x)$  as a multiple  $t$  times the vector  $x - f(x)$ , where  $t \geq 1$ . Thus  $g(x) = tx + (1 - t)f(x)$ . If  $t$  depends smoothly on  $x$ , then  $g$  is smooth. Take the dot product of both sides of this formula. Because  $|g(x)| = 1$ , you obtain the formula

$$t^2 |x - f(x)|^2 + 2tf(x) \cdot [x - f(x)] + |f(x)|^2 - 1 = 0.$$

The latter may look ugly, but it has the redeeming virtue of being a quadratic polynomial with a unique positive root. (There is also a root with  $t \leq 0$ , corresponding to the point where the line from  $x$  through  $f(x)$  hits the boundary.) Now we need only substitute into the quadratic formula of high school to obtain an expression for  $t$  in terms of smooth functions of  $x$ . Q.E.D.

### EXERCISES

- Any one-dimensional, compact, connected submanifold of  $\mathbf{R}^3$  is diffeomorphic to a circle. But can it be deformed into a circle within  $\mathbf{R}^3$ ? Draw some pictures, or try with string.
- Show that the fixed point in the Brouwer theorem need not be an interior point.
- Find maps of the solid torus into itself having no fixed points. Where does the proof of the Brouwer theorem fail?
- Prove that the Brouwer theorem is false for the open ball  $|x|^2 < a$ . [HINT: See Chapter 1, Section 1, Exercise 4.]
- Prove the Brouwer theorem for maps of  $[0, 1]$  directly, without using regular values.
- Prove the Brouwer theorem for continuous maps  $f: B^n \rightarrow B^n$ . Use the Weierstrass Approximation Theorem, which says that for any  $\epsilon > 0$  there exists a polynomial mapping  $p: \mathbf{R}^n \rightarrow \mathbf{R}^n$  such that  $|f - p| < \epsilon$  on  $B^n$ . (One reference: Rudin's *Principles of Mathematical Analysis*, "The Stone-Weierstrass Theorem.") [HINT: First show that given  $\delta > 0$ , you can find  $p$  so that  $|f - p| < \delta$  and  $p: B^n \rightarrow B^n$ . Now use the fact that if  $f$  has no fixed points,  $|f(x) - x| > c > 0$  on  $B^n$ .]
- As a surprisingly concrete application of Brouwer, prove this theorem of Frobenius: if the entries in an  $n \times n$  real matrix  $A$  are all nonnegative,

then  $A$  has a real nonnegative eigenvalue. [HINT: It suffices to assume  $A$  nonsingular; otherwise 0 is an eigenvalue. Let  $A$  also denote the associated linear map of  $\mathbf{R}^n$ , and consider the map  $v \rightarrow Av/|Av|$  restricted to  $S^{n-1} \rightarrow S^{n-1}$ . Show that this maps the “first quadrant”

$$Q = \{(x_1, \dots, x_n) \in S^{n-1} : \text{all } x_i \geq 0\}$$

into itself. It is not hard to prove, although we don't expect you to bother, that  $Q$  is homeomorphic with  $B^{n-1}$ ; that is, there exists a continuous bijection  $Q \rightarrow B^{n-1}$  having a continuous inverse. Now use Exercise 6.] See Figure 2-8.

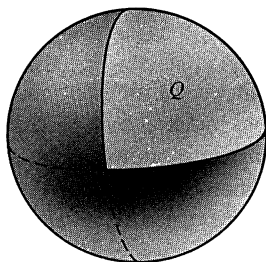
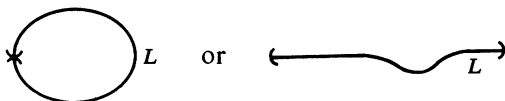


Figure 2-8

8. Suppose that  $\dim X = 1$  and  $L$  is a subset diffeomorphic to an open interval in  $\mathbf{R}^1$ . Prove that  $\bar{L} - L$  consists of at most two points:



This is needed for the classification of one-manifolds. [HINT: Given  $g: (a, b) \xrightarrow{\sim} L$ , let  $p \in \bar{L} - L$ . Let  $J$  be a closed subset of  $X$  diffeomorphic to  $[0, 1]$ , such that 1 corresponds to  $p$  and 0 corresponds to some  $g(t) \in L$ . Prove that  $J$  contains either  $g(a, t)$  or  $g(t, b)$ , by showing that the set  $\{s \in (a, t) : g(s) \in J\}$  is open and closed in  $(a, b)$ .]

### §3 Transversality

Earlier we proved that transversality is a property that is stable under small perturbations, at least for maps with compact domains. From Sard's theorem we shall deduce the much more subtle and valuable fact that transversality is a generic quality: any smooth map  $f: X \rightarrow Y$ , no matter how bizarre its behavior with respect to a given submanifold  $Z$  in  $Y$ , may be de-

formed by an arbitrary small amount into a map that is transversal to  $Z$ . Physically, stability means that transversal maps are actually observable. The fact that transversality is generic says that *only* transversal maps are observable. In this sense, almost all mappings are transversal. (See Figure 2-9.).

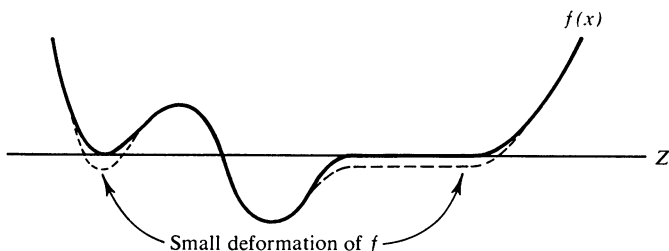


Figure 2-9

The key to transversality is families of mappings. Suppose  $f_s: X \rightarrow Y$  is a family of smooth maps, indexed by a parameter  $s$  that ranges over some set  $S$ . As we did with homotopies, consider the map  $F: X \times S \rightarrow Y$  defined by  $F(x, s) = f_s(x)$ . We require that the family vary smoothly by assuming  $S$  to be a manifold and  $F$  to be smooth. The central theorem is

**The Transversality Theorem.** Suppose that  $F: X \times S \rightarrow Y$  is a smooth map of manifolds, where only  $X$  has boundary, and let  $Z$  be any boundaryless submanifold of  $Y$ . If both  $F$  and  $\partial F$  are transversal to  $Z$ , then for almost every  $s \in S$ , both  $f_s$  and  $\partial f_s$  are transversal to  $Z$ .

*Proof.* The preimage  $W = F^{-1}(Z)$  is a submanifold of  $X \times S$  with boundary  $\partial W = W \cap \partial(X \times S)$ . Let  $\pi: X \times S \rightarrow S$  be the natural projection map. We shall show that whenever  $s \in S$  is a regular value for the restriction map  $\pi: W \rightarrow S$ , then  $f_s \pitchfork Z$ , and whenever  $s$  is a regular value for  $\partial\pi: \partial W \rightarrow S$ , then  $\partial f_s \pitchfork Z$ . By Sard's theorem, almost every  $s \in S$  is a regular value for both maps, so the theorem follows.

In order to show that  $f_s \pitchfork Z$ , suppose that  $f_s(x) = z \in Z$ . Because  $F(x, s) = z$  and  $F \pitchfork Z$ , we know that

$$dF_{(x,s)}T_{(x,s)}(X \times S) + T_z(Z) = T_z(Y);$$

that is, given any vector  $a \in T_z(Y)$ , there exists a vector  $b \in T_{(x,s)}(X \times S)$  such that

$$dF_{(x,s)}(b) - a \in T_z(Z).$$

We want to exhibit a vector  $v \in T_x(X)$  such that  $df_s(v) - a \in T_z(Z)$ . Now

$$T_{(x,s)}(X \times S) = T_x(X) \times T_s(S),$$

so  $b = (w, e)$  for vectors  $w \in T_x(X)$  and  $e \in T_s(S)$ . If  $e$  were zero we would be done, for since the restriction of  $F$  to  $X \times \{s\}$  is  $f_s$ , it follows that

$$dF_{(x,s)}(w, 0) = df_s(w).$$

Although  $e$  need not be zero, we may use the projection  $\pi$  to kill it off. As

$$d\pi_{(x,s)}: T_x(X) \times T_s(S) \longrightarrow T_s(S)$$

is just projection onto the second factor, the regularity assumption that  $d\pi_{(x,s)}$  map  $T_{(x,s)}(W)$  onto  $T_s(S)$  tells us that there is some vector of the form  $(u, e)$  in  $T_{(x,s)}(W)$ . But  $F: W \rightarrow Z$ , so  $dF_{(x,s)}(u, e) \in T_z(Z)$ . Consequently, the vector  $v = w - u \in T_x(X)$  is our solution. For

$$df_s(v) - a = dF_{(x,s)}[(w, e) - (u, e)] - a = [dF_{(x,s)}(w, e) - a] - dF_s(u, e),$$

and both of the latter vectors belong to  $T_z(Z)$ .

Precisely the same argument shows that  $\partial f_s \nparallel Z$  when  $s$  is a regular value of  $\partial\pi$ . (In fact, this is a special instance of what we just proved, for the case of the boundaryless manifold  $\partial X$  and the map  $\partial F: (\partial X) \times S \rightarrow Y$ .) Q.E.D.

The Transversality Theorem easily implies that transversal maps are generic when the target manifold  $Y$  is a Euclidean space  $\mathbf{R}^M$ . If  $f: X \rightarrow \mathbf{R}^M$  is any smooth map, take  $S$  to be an open ball of  $\mathbf{R}^M$  itself and then define  $F: X \times S \rightarrow \mathbf{R}^M$  by  $F_{(x,s)} = f(x) + s$ . For any fixed  $x \in X$ ,  $F$  is a translation of the ball  $S$ , obviously a submersion. So, of course,  $F$  is a submersion of  $X \times S$  and therefore transversal to any submanifold  $Z$  of  $\mathbf{R}^M$ . According to the Transversality Theorem, for almost every  $s \in S$ , the map  $f_s(x) = f(x) + s$  is transversal to  $Z$ . Thus  $f$  may be deformed into a transversal map by the simple addition of an arbitrarily small quantity  $s$ .

For an arbitrary, boundaryless, target manifold  $Y$ , the proof that transversality is generic follows the same course.  $Y$  sits inside some Euclidean space  $\mathbf{R}^M$ , and we have just shown how to vary a given map  $f: X \rightarrow Y$  through a family of maps carrying  $X$  into  $\mathbf{R}^M$ . All we need to do now is somehow to project these maps down onto  $Y$ , thereby obtaining a suitable family mapping  $X$  into  $Y$ . To do so, we must understand a little of the geometry of  $Y$  with respect to its environment. As usual, the compact case is clearest.

**$\epsilon$ -Neighborhood Theorem.** For a compact boundaryless manifold  $Y$  in  $\mathbf{R}^M$  and a positive number  $\epsilon$ , let  $Y^\epsilon$  be the open set of points in  $\mathbf{R}^M$  with distance less than  $\epsilon$  from  $Y$ . If  $\epsilon$  is sufficiently small, then each point  $w \in Y^\epsilon$  possesses a unique closest point in  $Y$ , denoted  $\pi(w)$ . Moreover, the map  $\pi: Y^\epsilon \rightarrow Y$  is a submersion. When  $Y$  is not compact, there still exists a submersion  $\pi: Y^\epsilon \rightarrow Y$  that is the identity on  $Y$ , but now  $\epsilon$  must be allowed to be a

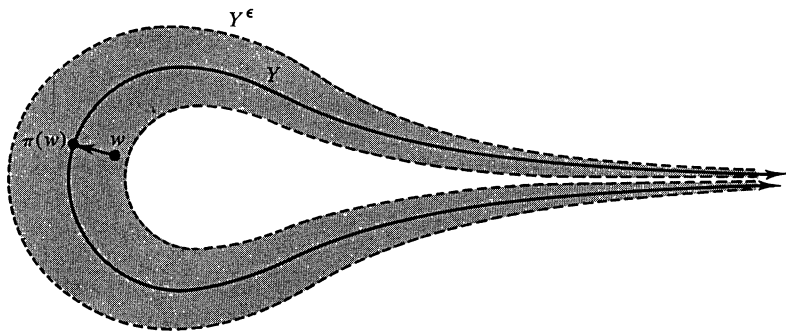


Figure 2-10

smooth positive function on  $Y$ , and  $Y^\epsilon$  is defined as  $\{w \in \mathbf{R}^M : |w - y| < \epsilon(y) \text{ for some } y \in Y\}$ . See Figure 2-10.

We postpone the proof for a moment.

**Corollary.** Let  $f: X \rightarrow Y$  be a smooth map,  $Y$  being boundaryless. Then there is an open ball  $S$  in some Euclidean space and a smooth map  $F: X \times S \rightarrow Y$  such that  $F(x, 0) = f(x)$ , and for any fixed  $x \in X$  the map  $s \rightarrow F(x, s)$  is a submersion  $S \rightarrow Y$ . In particular, both  $F$  and  $\partial F$  are submersions.

*Proof.* Let  $S$  be the unit ball in  $\mathbf{R}^M$ , the ambient Euclidean space of  $Y$ , and define

$$F(x, s) = \pi[f(x) + \epsilon(f(x))s].$$

Since  $\pi: Y^\epsilon \rightarrow Y$  restricts to the identity on  $Y$ ,  $F(x, 0) = f(x)$ . For fixed  $x$ ,

$$s \rightarrow f(x) + \epsilon(f(x))s$$

is certainly a submersion of  $S \rightarrow Y$ . As the composition of two submersions is another,  $s \rightarrow F(x, s)$  is a submersion.  $F$  and  $\partial F$  must obviously be submersions, for they are submersions even when restricted to the submanifolds  $\{x\} \times S$ , and one such submanifold passes through each point of  $X \times S$  and each point of  $(\partial X) \times S$ . Q.E.D.

That transversality is generic follows directly. The form we shall require is the

**Transversality Homotopy Theorem.** For any smooth map  $f: X \rightarrow Y$  and any boundaryless submanifold  $Z$  of the boundaryless manifold  $Y$ , there exists a smooth map  $g: X \rightarrow Y$  homotopic to  $f$  such that  $g \pitchfork Z$  and  $\partial g \pitchfork Z$ .



*Proof.* For the family of mappings  $F$  of the corollary, the Transversality Theorem implies that  $f_s \bar{\cap} Z$  and  $\partial f_s \bar{\cap} Z$  for almost all  $s \in S$ . But each  $f_s$  is homotopic to  $f$ , the homotopy  $X \times I \rightarrow Y$  being  $(x, t) \rightarrow F(x, ts)$ .

Q.E.D.

To prove the  $\epsilon$ -Neighborhood Theorem, we introduce an artifice similar to the tangent bundle. For each  $y \in Y$ , define  $N_y(Y)$ , the *normal space* of  $Y$  at  $y$ , to be the orthogonal complement of  $T_y(Y)$  in  $\mathbf{R}^M$ . The *normal bundle*  $N(Y)$  is then defined to be the set

$$\{(y, v) \in Y \times \mathbf{R}^M : v \in N_y(Y)\}.$$

Note that unlike  $T(Y)$ ,  $N(Y)$  is not intrinsic to the manifold  $Y$  but depends on the specific relationship between  $Y$  and the surrounding  $\mathbf{R}^M$ . There is a natural projection map  $\sigma : N(Y) \rightarrow Y$  defined by  $\sigma(y, v) = y$ .

In order to show that  $N(Y)$  is a manifold, we must recall an elementary fact from linear algebra. Suppose that  $A : \mathbf{R}^M \rightarrow \mathbf{R}^k$  is a linear map. Its *transpose* is a linear map  $A' : \mathbf{R}^k \rightarrow \mathbf{R}^M$  characterized by the dot product equation  $Av \cdot w = v \cdot A'w$  for all  $v \in \mathbf{R}^M$ ,  $w \in \mathbf{R}^k$ . If the matrix of  $A$  is  $(a_{ij})$  with respect to the standard bases, then the matrix of  $A'$  is  $(a_{ji})$ . Note that if  $A$  is surjective, then  $A'$  maps  $\mathbf{R}^k$  isomorphically onto the orthogonal complement of the kernel of  $A$ . For if  $A'w = 0$ , then  $Av \cdot w = v \cdot A'w = 0$ , so that  $w \perp A(\mathbf{R}^M)$ ; since  $A$  is surjective,  $w$  must be zero and thus  $A'$  is injective. Similarly, if  $Av = 0$  then  $0 = Av \cdot w = v \cdot A'w$ , so  $A'(\mathbf{R}^k) \perp \text{Kernel}(A)$ . Thus  $A'$  maps  $\mathbf{R}^k$  injectively into the orthogonal complement of  $\text{Kernel}(A)$ . As  $\text{Kernel}(A)$  has dimension  $M - k$ , its complement has dimension  $k$ , so  $A'$  is surjective, too.

**Proposition.** If  $Y \subset \mathbf{R}^M$ , then  $N(Y)$  is a manifold of dimension  $M$  and the projection  $\sigma : N(Y) \rightarrow Y$  is a submersion.

*Proof.* Define  $Y$  locally by equations: around any given point of  $Y$ , find an open set  $\tilde{U}$  of  $\mathbf{R}^M$  and a submersion  $\phi : \tilde{U} \rightarrow \mathbf{R}^k$  ( $k = \text{codim } Y$ ) such that  $U = Y \cap \tilde{U} = \phi^{-1}(0)$ . The set  $N(U)$  equals  $N(Y) \cap (U \times \mathbf{R}^M)$ , thus is open in  $N(Y)$ . For each  $y \in U$ ,  $d\phi_y : \mathbf{R}^M \rightarrow \mathbf{R}^k$  is surjective and has kernel  $T_y(Y)$ . Therefore its transpose maps  $\mathbf{R}^k$  isomorphically onto  $N_y(Y)$ . The map  $\psi : U \times \mathbf{R}^k \rightarrow N(U)$ , defined by  $\psi(y, v) = (y, d\phi_y'v)$ , is thus bijective, and it is easy to check that it is an embedding of  $U \times \mathbf{R}^k$  into  $Y \times \mathbf{R}^M$ . Consequently,  $N(U)$  is a manifold parametrized by  $\psi$ , with dimension  $= \dim U + k = \dim Y + \text{codim } Y = M$ . Since every point of  $N(Y)$  has such a neighborhood,  $N(Y)$  is a manifold. Since  $\sigma \circ \psi : U \times \mathbf{R}^k \rightarrow U$  is just the standard submersion,  $\sigma$  is a submersion. Q.E.D.

*Proof of the  $\epsilon$ -Neighborhood Theorem.* Let  $h : N(Y) \rightarrow \mathbf{R}^M$  be  $h(y, v) = y + v$ . Notice that  $h$  is regular at every point of  $Y \times \{0\}$  in  $N(Y)$ , for through  $(y, 0)$

there pass two natural complementary manifolds of  $N(Y)$ , i.e.,  $Y \times \{0\}$  and  $\{y\} \times N_y(Y)$ . The derivative of  $h$  at  $(y, 0)$  maps the tangent space of  $Y \times \{0\}$  at  $(y, 0)$  onto  $T_y(Y)$  and maps the tangent space of  $\{y\} \times N_y(Y)$  at  $(y, 0)$  onto  $N_y(Y)$ . Therefore it maps onto  $T_y(Y) + N_y(Y) = \mathbf{R}^M$ .

Since  $h$  maps  $Y \times \{0\}$  diffeomorphically onto  $Y$  and is regular at each  $(y, 0)$ , it must map a neighborhood of  $Y \times \{0\}$  diffeomorphically onto a neighborhood of  $Y$  in  $\mathbf{R}^M$  (Exercise 14 of Chapter 1, Section 8). Now any neighborhood of  $Y$  contains some  $Y^\epsilon$ ; this point is obvious when  $Y$  is compact and easy to show in general (Exercise 1). Thus  $h^{-1}: Y^\epsilon \rightarrow N(Y)$  is defined, and  $\pi = \sigma \circ h^{-1}: Y^\epsilon \rightarrow Y$  is the desired submersion. The geometric description of  $\pi$  for compact manifolds, which we never use, will be derived in Exercise 3. Q.E.D.

We will need a somewhat stronger form of the Transversality Homotopy Theorem. Let us say that a map  $f: X \rightarrow Y$  is *transversal* to  $Z$  on a subset  $C$  of  $X$  if the transversality condition

$$df_x T_x(X) + T_{f(x)}(Z) = T_{f(x)}(Y)$$

is satisfied at every point  $x \in C \cap f^{-1}(Z)$ .

**Extension Theorem.** Suppose that  $Z$  is a closed submanifold of  $Y$ , both boundaryless, and  $C$  is a closed subset of  $X$ . Let  $f: X \rightarrow Y$  be a smooth map with  $f \bar{\cap} Z$  on  $C$  and  $\partial f \bar{\cap} Z$  on  $C \cap \partial X$ . Then there exists a smooth map  $g: X \rightarrow Y$  homotopic to  $f$ , such that  $g \bar{\cap} Z$ ,  $\partial g \bar{\cap} Z$ , and on a neighborhood of  $C$  we have  $g = f$ .

**Lemma.** If  $U$  is an open neighborhood of the closed set  $C$  in  $X$ , then there exists a smooth function  $\gamma: X \rightarrow [0, 1]$  that is identically equal to one outside  $U$  but that is zero on a neighborhood of  $C$ .

*Proof.* Let  $C'$  be any closed set contained in  $U$  that contains  $C$  in its interior, and let  $\{\theta_i\}$  be a partition of unity subordinate to the open cover  $\{U, X - C'\}$  of  $X$ . (The proof of the Partition of Unity Theorem is still valid when  $X$  has boundary.) Then just take  $\gamma$  to be the sum of those  $\theta_i$  that vanish outside of  $X - C'$ . Q.E.D.

**Proof of theorem.** First we show that  $f \bar{\cap} Z$  on a neighborhood of  $C$ . If  $x \in C$  but  $x \notin f^{-1}(Z)$ , then since  $Z$  is closed,  $X - f^{-1}(Z)$  is a neighborhood of  $x$  on which  $f \bar{\cap} Z$  automatically. If  $x \in f^{-1}(Z)$ , then there is a neighborhood  $W$  of  $f(x)$  in  $Y$  and a submersion  $\phi: W \rightarrow \mathbf{R}^k$  such that  $f \bar{\cap} Z$  at a point  $x' \in f^{-1}(Z \cap W)$  precisely when  $\phi \circ f$  is regular at  $x'$ . But  $\phi \circ f$  is regular at  $x$ , so it is regular in a neighborhood of  $x$ . Thus  $f \bar{\cap} Z$  on a neighborhood of every point  $x \in C$ , and so  $f \bar{\cap} Z$  on a neighborhood  $U$  of  $X$ .

Let  $\gamma$  be the function in the lemma, and set  $\tau = \gamma^2$ . Since  $d\tau_x = 2\gamma(x) d\gamma_x$ ,  $d\tau_x = 0$  wherever  $\tau(x) = 0$ . Now we modify the map  $F: X \times S \rightarrow Y$  which we used in proving the Homotopy Theorem, defining  $G: X \times S \rightarrow Y$  by  $G(x, s) = F(x, \tau(x)s)$ . We claim  $G \nparallel Z$ . For suppose that  $(x, s) \in G^{-1}(Z)$ , and say,  $\tau(x) \neq 0$ . Then the map  $S \rightarrow Y$ , defined by  $r \rightarrow G(x, r)$ , is a submersion, being the composition of the diffeomorphism  $r \rightarrow \tau(x)r$  with the submersion  $r \rightarrow F(x, r)$ . Consequently,  $G$  is regular at  $(x, s)$ , so certainly  $G \nparallel Z$  at  $(x, s)$ . When  $\tau(x) = 0$  we evaluate  $dG_{(x, s)}$  at any element

$$(v, w) \in T_x(X) \times T_s(S) = T_x(X) \times \mathbf{R}^M.$$

If, for clarity, we define  $m: X \times S \rightarrow X \times S$  by  $m(x, s) = (x, \tau(x)s)$ , then its derivative is

$$dm_{(x, s)}(v, w) = (v, \tau(x) \cdot w + d\tau_x(v) \cdot s),$$

where  $\tau(x), d\tau_x(v) \in \mathbf{R}$  are scalars multiplying the vectors  $w, s \in \mathbf{R}^M$ . Now apply the chain rule to  $G = F \circ m$ , substituting  $\tau(x) = 0$  and  $d\tau_x = 0$ :

$$dG_{(x, s)}(v, w) = dF_{(x, 0)}(v, 0).$$

Since  $F$  equals  $f$  when restricted to  $X \times \{0\}$

$$dG_{(x, s)}(v, w) = df_x(v).$$

But if  $\tau(x) = 0$ , then  $x \in U$  and  $f \nparallel Z$  at  $x$ , so the fact that  $df_x$  and  $dG_{(x, s)}$  have the same images implies that  $G \nparallel Z$  at  $(x, s)$ .

A similar analysis shows that  $\partial G \nparallel Z$ . Now, by the Transversality Theorem, we can pick an  $s$  for which the map  $g(x) = G(x, s)$  satisfies  $g \nparallel Z$  and  $\partial g \nparallel Z$ . As before,  $g$  is homotopic to  $f$ . Furthermore, if  $x$  belongs to the neighborhood of  $C$  on which  $\tau = 0$ , then  $g(x) = G(x, s) = F(x, 0) = f(x)$ .

Q.E.D.

Since  $\partial X$  is always closed in  $X$ , we obtain the special case

**Corollary.** If, for  $f: X \rightarrow Y$ , the boundary map  $\partial f: \partial X \rightarrow Y$  is transversal to  $Z$ , then there exists a map  $g: X \rightarrow Y$  homotopic to  $f$  such that  $\partial g = \partial f$  and  $g \nparallel Z$ .

Focusing attention on the boundary, the corollary may be interpreted in another useful form. Suppose that  $h: \partial X \rightarrow Y$  is a map transversal to  $Z$ . Then if  $h$  extends to any map of the whole manifold  $X \rightarrow Y$ , it extends to a map that is transversal to  $Z$  on all of  $X$ .

## EXERCISES

1. (For the  $\epsilon$ -Neighborhood Theorem) Show that any neighborhood  $\tilde{U}$  of  $Y$  in  $\mathbf{R}^M$  contains some  $Y^\epsilon$ ; moreover, if  $Y$  is compact,  $\epsilon$  may be taken constant. [HINT: Find covering open sets  $U_\alpha \subset Y$  and  $\epsilon_\alpha > 0$ , such that  $U_\alpha^{\epsilon_\alpha} \subset \tilde{U}$ . Let  $\{\theta_i\}$  be a subordinate partition of unity, and show that  $\epsilon = \sum \theta_i \epsilon_i$  works.]
2. Let  $Y$  be a compact submanifold of  $\mathbf{R}^M$ , and let  $w \in \mathbf{R}^M$ . Show that there exists a (not necessarily unique) point  $y \in Y$  closest to  $w$ , and prove that  $w - y \in N_y(Y)$ . [HINT: If  $c(t)$  is a curve on  $Y$  with  $c(0) = y$ , then the smooth function  $|w - c(t)|^2$  has a minimum at 0. Use Exercise 12, Chapter 1, Section 2.]
3. Use Exercise 2 to verify the geometric characterization of  $\pi: Y^\epsilon \rightarrow Y$ , for compact  $Y$ . Assume that  $h: N(Y) \rightarrow \mathbf{R}^M$  carries a neighborhood of  $Y$  in  $N(Y)$  diffeomorphically onto  $Y^\epsilon$ ,  $\epsilon$  constant. Prove that if  $w \in Y^\epsilon$ , then  $\pi(w)$  is the *unique* point of  $Y$  closest to  $w$ .
4. (General Position Lemma) Let  $X$  and  $Y$  be submanifolds of  $\mathbf{R}^N$ . Show that for almost every  $a \in \mathbf{R}^N$  the translate  $X + a$  intersects  $Y$  transversally.
- \*5. Suppose that the compact submanifold  $X$  in  $Y$  intersects another submanifold  $Z$ , but  $\dim X + \dim Z < \dim Y$ . Prove that  $X$  may be pulled away from  $Z$  by an arbitrarily small deformation: given  $\epsilon > 0$  there exists a deformation  $X_t = i_t(X)$  such that  $X_1$  does not intersect  $Z$  and  $|x - i_1(x)| < \epsilon$  for all  $x \in X$ . (Note: You need Exercise 11, Chapter 1, Section 6. The point here is to make  $X_t$  a manifold.)
6. Sharpen Exercise 5. Assume that  $Z$  is closed in  $Y$  and let  $U$  be any open set in  $X$  containing  $Z \cap X$ . Show that the deformation  $X_t$  may be chosen to be constant outside of  $U$  (Figure 2-11).

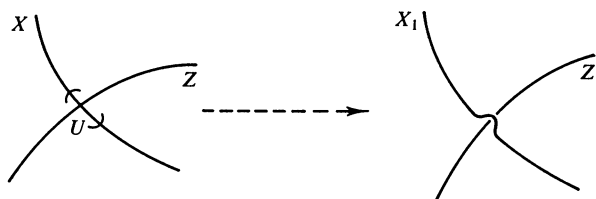


Figure 2-11

- \*7. Suppose that  $X$  is a submanifold of  $\mathbf{R}^N$ . Show that “almost every” vector space  $V$  of any fixed dimension  $l$  in  $\mathbf{R}^N$  intersects  $X$  transversally. [HINT: The set  $S \subset (\mathbf{R}^N)^l$  consisting of all linearly independent  $l$ -tuples of vectors in  $\mathbf{R}^N$  is open in  $\mathbf{R}^{Nl}$ , and the map  $\mathbf{R}^l \times S \rightarrow \mathbf{R}^N$  defined by

$$[(t_1, \dots, t_l), v_1, \dots, v_l] \longrightarrow t_1 v_1 + \dots + t_l v_l$$

is a submersion.]

8. Suppose that  $f: \mathbf{R}^n \rightarrow \mathbf{R}^n$  is a smooth map,  $n > 1$ , and let  $K \subset \mathbf{R}^n$  be compact and  $\epsilon > 0$ . Show that there exists a map  $f': \mathbf{R}^n \rightarrow \mathbf{R}^n$  such that  $df'_x$  is never zero, but  $|f - f'| < \epsilon$  on  $K$ . Prove that this result is false for  $n = 1$ . [HINT: Let  $M(n) = \{n \times n \text{ matrices}\}$ , and show that the map  $F: \mathbf{R}^n \times M(n) \rightarrow M(n)$ , defined by  $F(x, A) = df_x \cdot A$ , is a submersion. Pick  $A$  so that  $F_A \not\cap \{0\}$ ; where is  $n > 1$  used?]

9. Let  $f: \mathbf{R}^k \rightarrow \mathbf{R}^k$ , and, for each  $a \in \mathbf{R}^k$ , define

$$f_a(x) = f(x) + a_1 x_1 + \dots + a_k x_k.$$

Prove that for almost all  $a \in \mathbf{R}^k$ ,  $f_a$  is a Morse function. [HINT: Consider

$$\mathbf{R}^k \times \mathbf{R}^k \longrightarrow \mathbf{R}^k, \quad (x, a) \longrightarrow \frac{\partial f}{\partial x_i} + a_1, \dots, \frac{\partial f}{\partial x_k} + a_k.$$

Show that this is a submersion, hence  $\not\cap \{0\}$ .]

10. Let  $X$  be an  $n - 1$  dimensional submanifold of  $\mathbf{R}^n$ , a “hypersurface.” A point in  $\mathbf{R}^n$  is called a *focal point* of  $X$  if it is a critical value of the normal bundle map  $h: N(X) \rightarrow \mathbf{R}^n$ ,  $h(x, v) = x + v$ . Locate the focal points of the parabola  $y = x^2$  in  $\mathbf{R}^2$ . [Answer: You get a curve with a cusp at  $(0, \frac{1}{2})$ .]
11. Let  $X$  be a one-dimensional submanifold of  $\mathbf{R}^2$ , and let  $p \in X$ . Choose linear coordinates in  $\mathbf{R}^2$  so that  $p$  is the origin, the  $x$  axis is the tangent line to  $X$  at  $p$ , and the  $y$  axis is the normal line. Show that in a neighborhood of  $p = 0$ ,  $X$  is the graph of a function  $y = f(x)$  with  $f(0) = 0$  and  $f'(0) = 0$ . The quantity  $f''(0)$  is called the *curvature* of  $X$  at  $p$ , denoted  $\kappa(p)$ . Show that if  $\kappa(p) \neq 0$  then  $X$  has a focal point along the normal line at distance  $1/\kappa(p)$  from  $p$ . [HINT: Show that the normal space to  $X$  at a point  $x$  near  $p$  is spanned by  $(-f'(x), 1)$ . Now compute the normal bundle map  $h: N(X) \rightarrow \mathbf{R}^2$ .]

- \*12. Let  $Z$  be a submanifold of  $Y$ , where  $Y \subset \mathbf{R}^M$ . Define the *normal bundle to  $Z$  in  $Y$*  to be the set  $N(Z; Y) = \{(z, v) : z \in Z, v \in T_z(Y) \text{ and } v \perp T_z(Z)\}$ . Prove that  $N(Z; Y)$  is a manifold with the same dimension as  $Y$ . [HINT: Let  $g_1, \dots, g_l$  be independent functions in a neighborhood  $\tilde{U}$  of  $z$  in  $\mathbf{R}^M$ , with

$$U = Z \cap \tilde{U} = \{g_1 = 0, \dots, g_l = 0\}$$

and

$$Y \cap \tilde{U} = \{g_{k+1} = 0, \dots, g_l = 0\}$$

(Exercise 4, Chapter 1, Section 4). Show that the associated parametrization  $U \times \mathbf{R}^l \rightarrow N(Z; \mathbf{R}^M)$  as constructed in the text restricts to a parametrization  $U \times \mathbf{R}^k \rightarrow N(Z; Y)$ .]

13. Consider  $S^{k-1}$  as a submanifold of  $S^k$  via the usual embedding mapping  $(x_1, \dots, x_k) \rightarrow (x_1, \dots, x_k, 0)$ . Show that at  $p \in S^{k-1}$  the orthogonal complement to  $T_p(S^{k-1})$  in  $T_p(S^k)$  is spanned by the vector  $(0, \dots, 0, 1)$ . Prove that  $N(S^{k-1}; S^k)$  is diffeomorphic to  $S^{k-1} \times \mathbf{R}$ .
14. Prove that the map  $\sigma : N(Z; Y) \rightarrow Z$ ,  $\sigma(z, v) = z$ , is a submersion. What specifically is the preimage  $\sigma^{-1}(z)$ , which we denote by  $N_z(Z; Y)$ ?
15. Show that the map  $z \rightarrow (z, 0)$  embeds  $Z$  as a submanifold of  $N(Z; Y)$ .
- \*16. *Tubular Neighborhood Theorem.* Prove that there exists a diffeomorphism from an open neighborhood of  $Z$  in  $N(Z; Y)$  onto an open neighborhood of  $Z$  in  $Y$ . [HINT: Let  $Y^\epsilon \xrightarrow{\pi} Y$  be as in the  $\epsilon$ -Neighborhood Theorem. Consider the map  $h : N(Z; Y) \rightarrow \mathbf{R}^M$ ,  $h(z, v) = z + v$ . Then  $W = h^{-1}(Y^\epsilon)$  is an open neighborhood of  $Z$  in  $N(Z; Y)$ . The sequence of maps  $W \xrightarrow{h} Y^\epsilon \xrightarrow{\pi} Y$  is the identity on  $Z$ , so use Exercise 14 of Chapter 1, Section 8.]
17. Let  $\Delta$  be the diagonal in  $X \times X$ . Show that the orthogonal complement to  $T_{(x,x)}(\Delta)$  in  $T_{(x,x)}(X \times X)$  is the collection of vectors  $\{(v, -v) : v \in T_x(X)\}$ . [See Exercise 10, Chapter 1, Section 2.]
- \*18. Prove that the map  $T(X) \rightarrow N(\Delta; X \times X)$ , defined by sending  $(x, v)$  to  $((x, x), (v, -v))$ , is a diffeomorphism. Use the Tubular Neighborhood Theorem to conclude that there is a diffeomorphism of a neighborhood of  $X$  in  $T(X)$  with a neighborhood of  $\Delta$  in  $X \times X$ , extending the usual diffeomorphism  $X \rightarrow \Delta$ ,  $x \rightarrow (x, x)$ .
- \*19. Let  $Z$  be a submanifold of codimension  $k$  in  $Y$ . We say that the normal bundle  $N(Z; Y)$  is *trivial* if there exists a diffeomorphism  $\Phi : N(Z; Y)$

$\rightarrow Z \times \mathbf{R}^k$  that restricts to a linear isomorphism  $N_z(Z; Y) \rightarrow \{z\} \times \mathbf{R}^k$  for each point  $z \in Z$ . As a check on your grasp of the construction, prove that  $N(Z; Y)$  is always *locally trivial*. That is, each point  $z \in Z$  has a neighborhood  $V$  in  $Z$  such that  $N(V; Y)$  is trivial.

- \*20. Prove that  $N(Z; Y)$  is trivial if and only if there exists a set of  $k$  independent global defining functions  $g_1, \dots, g_k$  for  $Z$  on some set  $U$  in  $Y$ . That is,

$$Z = \{y \in U : g_1(y) = 0, \dots, g_k(y) = 0\}.$$

[HINT: If  $N(Z; Y)$  is trivial, then there obviously exist global defining functions for  $Z$  in  $N(Z; Y)$ . Transfer these functions to an open set in  $Y$  via the Tubular Neighborhood Theorem, Exercise 16. Conversely, if there exists a submersion  $g : U \rightarrow \mathbf{R}^k$  with  $g^{-1}(0) = Z$ , check that, for each  $z \in Z$ , the transpose map  $dg'_z : \mathbf{R}^k \rightarrow T_z(Y)$  carries  $\mathbf{R}^k$  isomorphically onto the orthogonal complement of  $T_z(Z)$  in  $T_z(Y)$ ; thus  $\Phi^{-1} : Z \times \mathbf{R}^k \rightarrow N(Z; Y)$  is defined by  $\Phi^{-1}(z, a) = (z, dg'_z a)$ .]

## §4 Intersection Theory Mod 2

The previous section was technical and rather difficult. We now hope to convince you that the effort was worth it. In this section we will use the transversality lemma and the other results of Section 3 to develop a simple intuitive invariant for intersecting manifolds, from which we will be able to obtain many nice geometric consequences.

Two submanifolds  $X$  and  $Z$  inside  $Y$  have *complementary dimension* if  $\dim X + \dim Z = \dim Y$ . If  $X \bar{\cap} Z$ , this dimension condition makes their intersection  $X \cap Z$  a zero-dimensional manifold. (We are working now without boundaries.) If we further assume that both  $X$  and  $Z$  are closed and that at least one of them, say  $X$ , is compact, then  $X \cap Z$  must be a finite set of points. Provisionally, we might refer to the number of points in  $X \cap Z$  as the “intersection number” of  $X$  and  $Z$ , indicated by  $\#(X \cap Z)$ . See Figure 2-12.

How can we generalize this discussion to define the intersection number of the compact  $X$  with an arbitrary closed  $Z$  of complementary dimension?

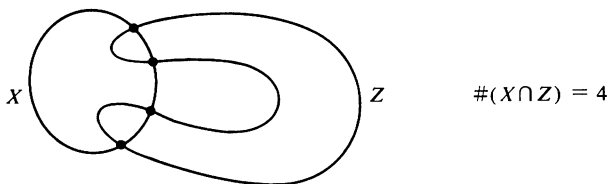


Figure 2-12

Without transversality,  $X \cap Z$  may be some frowzy, useless conglomeration. But we can wiggle  $X$ , deforming it ever so slightly, and make it transversal to  $Z$ . Then we might simply define the intersection number of  $X$  and  $Z$  to be the intersection numbers we obtain after the plastic surgery. The difficulty is that different alterations of  $X$  can produce different intersection numbers (Figure 2-13). Happily the idea is savable nonetheless, because as we shall see, the intersection numbers obtained with different deformations at least agree mod 2. All deformations either give even intersection numbers, or they all give odd ones. Thus a mod 2 intersection number of  $X$  and  $Z$  can be meaningfully defined.

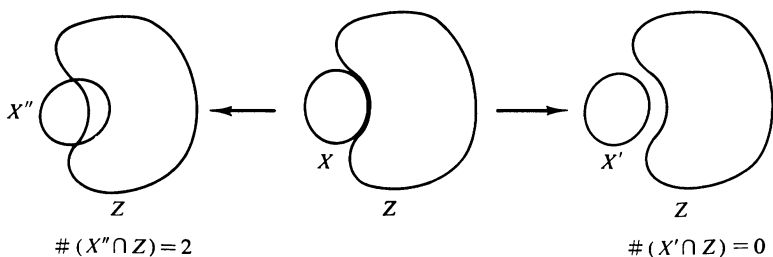


Figure 2-13

We must deal with the necessity of deforming  $X$  in a mathematically precise manner. Attempting to define deformations of arbitrary point sets in  $Y$  is hopeless, so we shift our point of view somewhat. Considering  $X$  as an abstract manifold and its inclusion mapping  $i: X \hookrightarrow Y$  simply as an embedding, we know how to deform  $i$ , namely by homotopy. Since embeddings form a stable class of mappings, any small homotopy of  $i$  gives us another embedding  $X \rightarrow Y$  and thus produces an image manifold that is a diffeomorphic copy of  $X$  adjacent to the original.

We are led to the following more general situation. Suppose that  $X$  is any compact manifold, not necessarily inside  $Y$ , and  $f: X \rightarrow Y$  is a smooth map transversal to the closed manifold  $Z$  in  $Y$ , where  $\dim X + \dim Z = \dim Y$ . Then  $f^{-1}(Z)$  is a closed zero-dimensional submanifold of  $X$ , hence a finite set. Define the *mod 2 intersection number* of the map  $f$  with  $Z$ ,  $I_2(f, Z)$ , to be the number of points in  $f^{-1}(Z)$  modulo 2. For an *arbitrary* smooth map  $g: X \rightarrow Y$ , select any map  $f$  that is homotopic to  $g$  and transversal to  $Z$ , and define  $I_2(g, Z) = I_2(f, Z)$ . The ambiguity in the definition is remedied by:

**Theorem.** If  $f_0, f_1: X \rightarrow Y$  are homotopic and both transversal to  $Z$ , then  $I_2(f_0, Z) = I_2(f_1, Z)$ .

*Proof.* Let  $F: X \times I \rightarrow Y$  be a homotopy of  $f_0$  and  $f_1$ . By the Extension Theorem, we may assume that  $F \pitchfork Z$ . Since  $\partial(X \times I) = X \times \{0\} \cup X \times \{1\}$



and  $\partial F$  is  $f_0$  on  $X \times \{0\}$  and  $f_1$  on  $X \times \{1\}$ ,  $\partial F \cap Z$ . Then  $F^{-1}(Z)$  is a one-dimensional submanifold of  $X \times I$  with boundary

$$\partial F^{-1}(Z) = F^{-1}(Z) \cap \partial(X \times I) = f_0^{-1}(Z) \times \{0\} \cup f_1^{-1}(Z) \times \{1\}.$$

From the classification of one-manifolds,  $\partial F^{-1}(Z)$  must have an even number of points, so  $\#f_0^{-1}(Z) = \#f_1^{-1}(Z) \bmod 2$ . Q.E.D.

Because homotopy is an equivalence relation, we immediately obtain

**Corollary.** If  $g_0, g_1: X \rightarrow Y$  are arbitrary homotopic maps, then we have  $I_2(g_0, Z) = I_2(g_1, Z)$ .

Return now to the motivating special case. If  $X$  is a compact submanifold of  $Y$  and  $Z$  a closed submanifold of complementary dimension, we define the mod 2 intersection number of  $X$  with  $Z$  by  $I_2(X, Z) = I_2(i, Z)$ , where  $i: X \hookrightarrow Y$  is the inclusion. When  $X \cap Z$ , then  $I_2(X, Z)$  is just  $\#X \cap Z \bmod 2$ . If  $I_2(X, Z) \neq 0$ , then no matter how  $X$  is deformed it cannot be pulled entirely away from  $Z$ . For example, on the torus  $Y = S^1 \times S^1$  the two complementary circles  $S^1 \times \{1\}$  and  $\{1\} \times S^1$  have nonzero mod 2 intersection numbers (Figure 2-14).

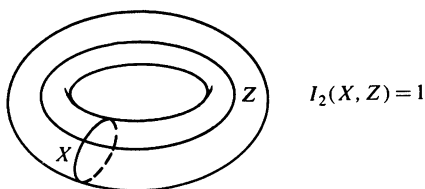


Figure 2-14

A curious situation prevails when  $\dim X = \frac{1}{2} \dim Y$ , for then we may consider  $I_2(X, X)$  the *mod 2 self-intersection number* of  $X$ . An illustrative example is the central curve on the open Möbius band. Make a model of the Möbius band by taping together the ends of a twisted paper strip, and show that

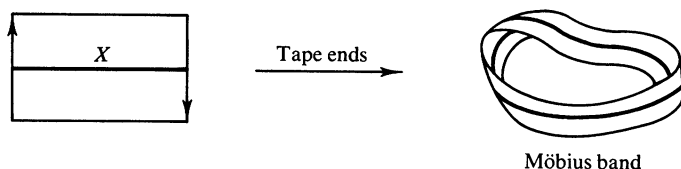


Figure 2-15

$I_2(X, X) = 1$ . Convince yourself that no matter how  $X$  is deformed, it can never be pulled entirely away from its original position (Figure 2-15).

If  $X$  happens to be the boundary of some  $W$  in  $Y$ , then  $I_2(X, Z) = 0$ . For if  $Z \cap X$ , then, roughly speaking,  $Z$  must “pass out” of  $W$  as often as it “enters”; hence  $\#(X \cap Z)$  is even (Figure 2-16).

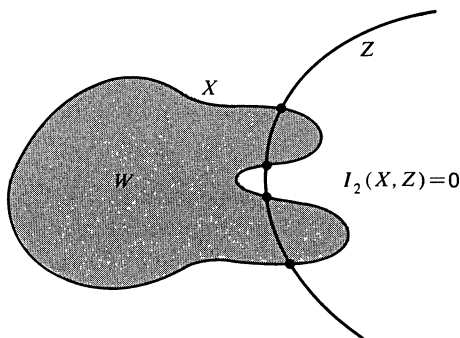


Figure 2-16

In the more general setting, we formulate this observation as follows:

**Boundary Theorem.** Suppose that  $X$  is the boundary of some compact manifold  $W$  and  $g: X \rightarrow Y$  is a smooth map. If  $g$  may be extended to all of  $W$ , then  $I_2(g, Z) = 0$  for any closed submanifold  $Z$  in  $Y$  of complementary dimension.

*Proof.* Let  $G: W \rightarrow Y$  extend  $g$ ; that is,  $\partial G = g$ . From the Transversality Homotopy Theorem, we obtain a homotopic map  $F: W \rightarrow Y$ , with  $F \cap Z$  and  $f = \partial F \cap Z$ . Then  $f \sim g$ , so  $I_2(g, Z) = \#f^{-1}(Z) \bmod 2$ . But  $F^{-1}(Z)$  is a compact one-dimensional manifold with boundary, so  $\#\partial F^{-1}(Z) = \#f^{-1}(Z)$  is even. Q.E.D.

Intersection theory gives us an interesting “homotopy invariant” attached to maps between manifolds of the same dimension. The definition depends on the following fact.

**Theorem.** If  $f: X \rightarrow Y$  is a smooth map of a compact manifold  $X$  into a connected manifold  $Y$  and  $\dim X = \dim Y$ , then  $I_2(f, \{y\})$  is the same for all points  $y \in Y$ . This common value is called the *mod 2 degree* of  $f$ , denoted  $\deg_2(f)$ .

We emphasize that degree mod 2 is defined only when the range manifold  $Y$  is connected and the domain  $X$  is compact; whenever  $\deg_2$  is written, these assumptions are implicit.

*Proof.* Given any  $y \in Y$ , alter  $f$  homotopically, if necessary, to make it transversal to  $\{y\}$ . Now by the “Stack of Records Theorem” (Chapter 1, Section 4, Exercise 7), we can find a neighborhood  $U$  of  $y$  such that the preimage  $f^{-1}(U)$  is a disjoint union  $V_1 \cup \cdots \cup V_m$ , where each  $V_i$  is an open set in  $X$  mapped by  $f$  diffeomorphically onto  $U$ . We conclude that  $I_2(f, \{z\}) = n \bmod 2$  for all points  $z \in U$ . Consequently, the function defined on  $Y$  by  $y \mapsto I_2(f, \{y\})$  is locally constant. Since  $Y$  is connected, it must be globally constant. Q.E.D.

Note that calculating the mod 2 degree of  $f$  is a very easy matter. Just pick any regular value  $y$  for  $f$ , and count preimage points:  $\deg_2(f) = \#f^{-1}(y) \bmod 2$ . For example, the map  $z \mapsto z^n$  (complex multiplication), which wraps  $S^1$  smoothly around  $S^1$   $n$  times, has mod 2 degree zero if  $n$  is even, one if  $n$  is odd.

Since  $\deg_2$  is defined as an intersection number, we immediately obtain

**Theorem.** Homotopic maps have the same mod 2 degree.

**Theorem.** If  $X = \partial W$  and  $f: X \rightarrow Y$  may be extended to all of  $W$ , then  $\deg_2(f) = 0$ .

We can apply this topological observation to the ancient problem of investigating the zeros of functions. Suppose, for example, that  $p: \mathbb{C} \rightarrow \mathbb{C}$  is a smooth, complex function and  $W$  is a smooth compact region in the plane, a two-dimensional manifold with boundary. Does  $p(z) = 0$  for some  $z \in W$ ? We assume that  $p$  has no zeros on  $\partial W$ , so that  $p/|p|: \partial W \rightarrow S^1$  is defined and smooth as a map of compact one-manifolds. Now if  $p$  has no zeros inside  $W$ , then  $p/|p|$  is defined on all of  $W$ , and the last theorem implies

**Proposition.** If the mod 2 degree of  $p/|p|: \partial W \rightarrow S^1$  is nonzero, then the function  $p$  has a zero inside  $W$ .

How remarkable that by simply counting the number of times  $p(z)$  points in a given direction while  $z$  is on the boundary, we may be able to tell whether the equation  $p(z) = 0$  is solvable inside  $W$ ! In order to indicate both the power and the limitation of this theory, let us prove half of the Fundamental Theorem of Algebra. Suppose that

$$p(z) = z^m + a_1 z^{m-1} + \cdots + a_m$$

is a monic complex polynomial, and define a homotopy by

$$\begin{aligned} p_t(z) &= tp(z) + (1-t)z^m \\ &= z^m + t(a_1 z^{m-1} + \cdots + a_m). \end{aligned}$$

It is easily seen that if  $W$  is a closed ball of sufficiently large radius, then none of the  $p_i$  have zeros on  $\partial W$ . For

$$\frac{p_i(z)}{z^m} = 1 + t \left( a_1 \frac{1}{z} + \cdots + a_m \frac{1}{z^m} \right)$$

and the term in parenthesis  $\rightarrow 0$  as  $z \rightarrow \infty$ . Thus the homotopy  $p_i/|p_i|: \partial W \rightarrow S^1$  is defined for all  $t$ , so we conclude that  $\deg_2(p/|p|) = \deg_2(p_0/|p_0|)$ . The polynomial  $p_0$  is just  $z^m$ ; therefore every point in  $S^1$  has precisely  $m$  preimage points in  $\partial W$  under  $p_0/|p_0|$ . Since we compute  $\deg_2$  by counting preimages of regular values,  $\deg_2(p_0/|p_0|) = m \bmod 2$ . The preceding proposition now gives

**One-half Fundamental Theorem of Algebra.** Every complex polynomial of odd degree has a root.

This famous result was effortlessly obtained, but it also demonstrates the insufficiency of mod 2 intersection theory: too much information is wasted. The  $\deg_2$  invariant is too crude to distinguish the polynomial  $z^2$  from a constant, so it cannot prove the even half of the Fundamental Theorem of Algebra. In the next chapter we shall refine our whole approach in order to construct a theory of considerably greater power. Yet the naive methods already developed continue to produce surprisingly deep insights with a minimum of strain, as illustrated by the following two sections.

### EXERCISES

1. Prove that there exists a complex number  $z$  such that

$$z^7 + \cos(|z|^2)(1 + 93z^4) = 0.$$

(For heaven's sake, don't try to compute it!)

2. Let  $X \xrightarrow{f} Y \xrightarrow{g} Z$  be a sequence of smooth maps of manifolds, with  $X$  compact. Assume that  $g$  is transversal to a closed submanifold  $W$  of  $Z$ , so  $g^{-1}(W)$  is a submanifold of  $Y$ . Verify that

$$I_2(f, g^{-1}(W)) = I_2(g \circ f, W).$$

(See Exercise 7, Chapter 1, Section 5.) In particular, check that if the conditions are such that one of these intersection numbers is defined, then so is the other.

3. Suppose that  $X$  and  $Z$  are compact manifolds and that  $f: X \rightarrow Y$ , and  $g: Z \rightarrow Y$  are smooth maps into the manifold  $Y$ . If  $\dim X + \dim Z =$

$\dim Y$ , we can define the mod 2 intersection number of  $f$  and  $g$  by  $I_2(f, g) = I_2(f \times g, \Delta)$ , where  $\Delta$  is the diagonal of  $Y \times Y$ .

- (a) Prove that  $I_2(f, g)$  is unaltered if either  $f$  or  $g$  is varied by a homotopy.
- (b) Check that  $I_2(f, g) = I_2(g, f)$  [HINT: Use Exercise 2 with the “switching diffeomorphism”  $(a, b) \rightarrow (b, a)$  of  $Y \times Y \rightarrow Y \times Y$ .]
- (c) If  $Z$  is actually a submanifold of  $Y$  and  $i: Z \hookrightarrow Y$  is its inclusion, show that

$$\widetilde{I_2}(f, i) = I_2(f, Z).$$

- (d) Prove that for two compact submanifolds  $X$  and  $Z$  in  $Y$ ,

$$I_2(X, Z) = I_2(Z, X).$$

(Note: This is trivial when  $X \cap Z = \emptyset$ . So why did we use this approach?)

- \*4. If  $f: X \rightarrow Y$  is homotopic to a constant map, show that  $I_2(f, Z) = 0$  for all complementary dimensional closed  $Z$  in  $Y$ , except perhaps if  $\dim X = 0$ . [HINT: Show that if  $\dim Z < \dim Y$ , then  $f$  is homotopic to a constant  $X \rightarrow \{y\}$ , where  $y \notin Z$ . If  $X$  is one point, for which  $Z$  will  $I_2(f, Z) \neq 0$ ?]
- \*5. Prove that intersection theory is vacuous in contractible manifolds: if  $Y$  is contractible and  $\dim Y > 0$ , then  $I_2(f, Z) = 0$  for every  $f: X \rightarrow Y$ ,  $X$  compact and  $Z$  closed,  $\dim X + \dim Z = \dim Y$ . (No dimension-zero anomalies here.) In particular, intersection theory is vacuous in Euclidean space.
- \*6. Prove that no compact manifold—other than the one-point space—is contractible. [HINT: Apply Exercise 5 to the identity map.]
- \*7. Prove that  $S^1$  is not simply connected. [HINT: Consider the identity map.]
- \*8. (a) Let  $f: S^1 \rightarrow S^1$  be any smooth map. Prove that there exists a smooth map  $g: \mathbf{R} \rightarrow \mathbf{R}$  such that  $f(\cos t, \sin t) = (\cos g(t), \sin g(t))$ , and satisfying  $g(2\pi) = g(0) + 2\pi q$  for some integer  $q$ . [HINT: First define  $g$  on  $[0, 2\pi]$ , and show that  $g(2\pi) = g(0) + 2\pi q$ . Now extend  $g$  by demanding  $g(t + 2\pi) = g(t) + 2\pi q$ .]  
 (b) Prove that  $\deg_2(f) = q \bmod 2$ .
- 9. (Degrees are the only interesting intersection numbers on spheres.) Suppose  $f: X \rightarrow S^k$  is smooth, where  $X$  is compact and  $0 < \dim X < k$ .

Then for all closed  $Z \subset S^k$  of dimension complementary to  $X$ ,  $I_2(X, Z) = 0$ . [HINT: By Sard, there exists  $p \notin f(X) \cap Z$ . Use stereographic projection, plus Exercise 5.]

10. Prove that  $S^2$  and the torus are not diffeomorphic.
11. Suppose that  $f: X \rightarrow Y$  has  $\deg_2(f) \neq 0$ . Prove that  $f$  is onto. [Remember that  $X$  must be compact,  $Y$  connected, and  $\dim X = \dim Y$  for  $\deg_2(f)$  to make sense.]
12. If  $Y$  is not compact, then  $\deg_2(f) = 0$  for all maps  $f: X \rightarrow Y$  ( $X$  compact).
13. If  $f: X \rightarrow Y$  is transversal to  $Z$  and  $\dim X + \dim Z = \dim Y$ , then we can at least define  $I_2(f, Z)$  as  $\#f^{-1}(Z) \bmod 2$ , as long as  $f^{-1}(Z)$  is finite. Let us explore how useful this definition is without the two assumptions made in the text:  $X$  compact and  $Z$  closed. Find examples to show:
  - (a)  $I_2(f, Z)$  may not be a homotopy invariant if  $Z$  is not closed.
  - (b)  $I_2(f, Z)$  may not be a homotopy invariant if  $X$  is not compact.
  - (c) The Boundary Theorem is false without the requirement that  $Z$  be closed.
  - (d) The Boundary Theorem is false without the requirement that  $X$  be compact.
  - (e) The Boundary Theorem is false without the requirement that  $W$  be compact, even if  $X = \partial W$  is compact and  $Z$  is closed. [HINT: Look at the cylinder  $S^1 \times R$ .]
- \*14. Two compact submanifolds  $X$  and  $Z$  in  $Y$  are *cobordant* if there exists a compact manifold with boundary,  $W$ , in  $Y \times I$  such that  $\partial W = X \times \{0\} \cup Z \times \{1\}$ . Show that if  $X$  may be deformed into  $Z$ , then  $X$  and  $Z$  are cobordant. However, the “trousers example” in Figure 2-17 shows that the converse is false.

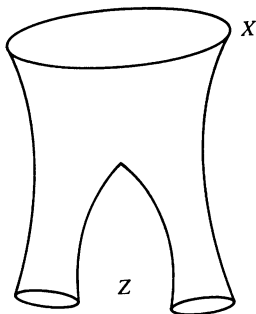


Figure 2-17

- \*15. Prove that if  $X$  and  $Z$  are cobordant in  $Y$ , then for every compact manifold  $C$  in  $Y$  with dimension complementary to  $X$  and  $Z$ ,  $I_2(X, C) = I_2(Z, C)$ . [HINT: Let  $f$  be the restriction to  $W$  of the projection map  $Y \times I \rightarrow Y$ , and use the Boundary Theorem.]
16. Prove that  $\deg_2(f)$  is well defined by direct application of the Boundary Theorem. [HINT: If  $y_0, y_1 \in Y$ , alter  $f$  homotopically to get  $f \bar{\cap} \{y_0, y_1\}$ . Now let  $c: I \rightarrow Y$  be a curve with  $c(0) = y_0$ ,  $c(1) = y_1$ , and define  $F: X \times I \rightarrow Y \times Y$  by  $f \times c$ . Examine  $\partial F$ .]
17. Derive the Nonretraction Theorem of Section 2 from the Boundary Theorem.
18. Suppose that  $Z$  is a compact submanifold of  $Y$  with  $\dim Z = \frac{1}{2} \dim Y$ . Prove that if  $Z$  is globally definable by independent functions, then  $I_2(Z, Z) = 0$ . [HINT: By Exercise 20, Section 3,  $N(Z; Y) = Z \times \mathbb{R}^k$ . Certainly  $I_2(Z, Z) = 0$  in  $Z \times \mathbb{R}^k$ , since  $Z \times \{a\} \cap Z$  is empty. Now use the Tubular Neighborhood Theorem, Exercise 16, Section 3.]
19. Show that the central circle  $X$  in the open Möbius band has mod 2 intersection number  $I_2(X, X) = 1$ . [HINT: Show that when the ends of the strip in Figure 2-18 are glued together with a twist,  $X'$  becomes a manifold that is a deformation of  $X$ .]

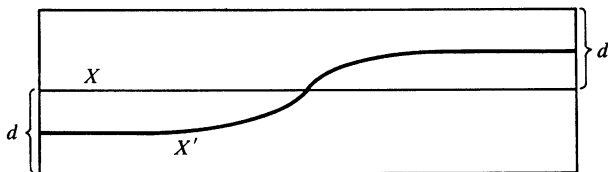


Figure 2-18

**Corollary** to Exercises 19 and 20. The central circle in the Möbius band is not definable by an independent function.

## §5 Winding Numbers and the Jordan-Brouwer Separation Theorem

The Classical Jordan Curve Theorem says that every simple closed curve in  $\mathbb{R}^2$  divides the plane into two pieces, the “inside” and “outside” of the curve. Lest the theorem appear too obvious, try your intuition on the example shown in Figure 2-19.

This section is a self-guided expedition with gun and camera into the wilds of such jungles, and in  $n$  dimensions, too! Begin with a compact, con-

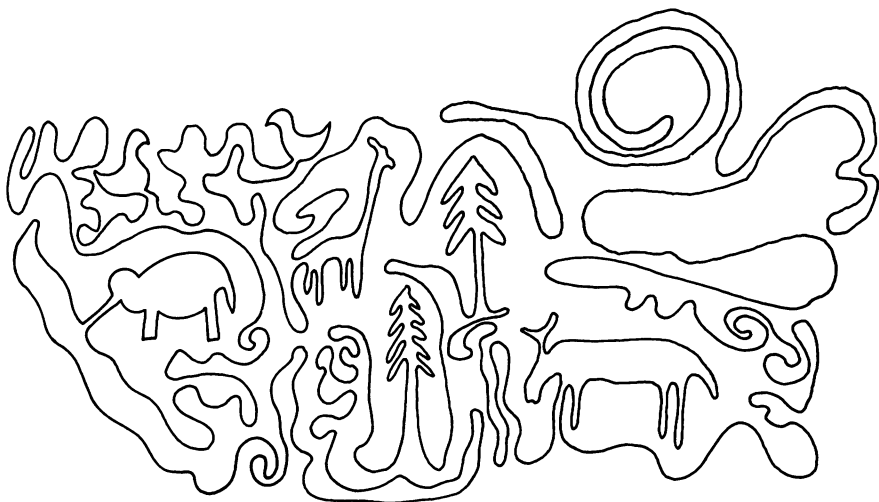


Figure 2-19

nected manifold  $x$  and a smooth map  $f: X \rightarrow \mathbf{R}^n$ . Suppose that  $\dim X = n - 1$ , so that, in particular,  $f$  might be the inclusion map of a hypersurface into  $\mathbf{R}^n$ . (In general, a *hypersurface* in a manifold is a submanifold of codimension one.) We wish to study how  $f$  wraps  $X$  around in  $\mathbf{R}^n$ , so take any point  $z$  of  $\mathbf{R}^n$  not lying in the image  $f(X)$ . To see how  $f(x)$  winds around  $z$ , we inquire how often the unit vector

$$u(x) = \frac{f(x) - z}{|f(x) - z|},$$

which indicates the direction from  $z$  to  $f(x)$ , points in a given direction. From Intersection Theory, we know that  $u: X \rightarrow S^{n-1}$  hits almost every direction vector the same number of times mod 2, namely,  $\deg_2(u)$  times. So seize this invariant and define the *mod 2 winding number* of  $f$  around  $z$  to be  $W_2(f, z) = \deg_2(u)$ . (See Figure 2-20.)

In a moment you will use this notion (i.e., mod 2 winding number) to establish a generalized version of the Jordan curve theorem, but first some exercises will develop a preliminary theorem. The proof introduces a beauti-

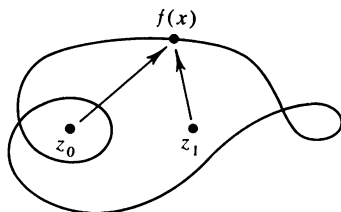


Figure 2-20



fully simple technique that appears repeatedly in later sections. Hints are provided at the end of the section, but it should be fun to fit the proof together yourself.

**Theorem.** Suppose that  $X$  is the boundary of  $D$ , a compact manifold with boundary, and let  $F: D \rightarrow \mathbf{R}^n$  be a smooth map extending  $f$ ; that is,  $\partial F = f$ . Suppose that  $z$  is a regular value of  $F$  that does not belong to the image of  $f$ . Then  $F^{-1}(z)$  is a finite set, and  $W_2(f, z) = \#F^{-1}(z) \bmod 2$ . That is,  $f$  winds  $X$  around  $z$  as often as  $F$  hits  $z$ , mod 2. (See Figure 2-21.)

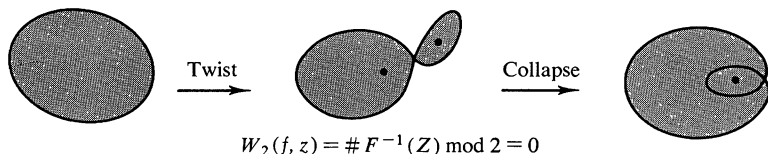


Figure 2-21

Here are some exercises to help you construct a proof:

1. Show that if  $F$  does not hit  $z$ , then  $W_2(f, z) = 0$ .
2. Suppose that  $F^{-1}(z) = \{y_1, \dots, y_l\}$ , and around each point  $y_i$  let  $B_i$  be a ball. (That is,  $B_i$  is the image of a ball in  $\mathbf{R}^n$  via some local parametrization of  $D$ .) Demand that the balls be disjoint from one another and from  $X = \partial D$ . Let  $f_i: \partial B_i \rightarrow \mathbf{R}^n$  be the restriction of  $F$ , and prove that

$$W_2(f, z) = W_2(f_1, z) + \dots + W_2(f_l, z) \bmod 2.$$

(See Figure 2-22.)

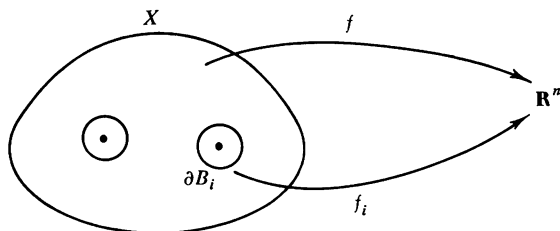


Figure 2-22

3. Use the regularity of  $z$  to choose the balls  $B_i$  so that  $W_2(f_i, z) = 1$ , and thus prove the theorem.

Now assume that  $X$  actually is a compact, connected hypersurface in  $\mathbf{R}^n$ . If  $X$  really does separate  $\mathbf{R}^n$  into an inside and an outside, then it should be

the boundary of a compact  $n$ -dimensional manifold with boundary—namely, its inside. In this case, the preceding theorem tells us that if  $z \in \mathbf{R}^n$  is any point not on  $X$ , then  $W_2(X, z)$  must be 1 or 0, depending on whether  $z$  lies inside or outside of  $X$ . [Here  $W_2(X, z)$  is used for the winding number of the inclusion map of  $X$  around  $z$ .] The next exercises help you reverse this reasoning to prove the Separation Theorem.

4. Let  $z \in \mathbf{R}^n - X$ . Prove that if  $x$  is any point of  $X$  and  $U$  any neighborhood of  $x$  in  $\mathbf{R}^n$ , then there exists a point of  $U$  that may be joined to  $z$  by a curve not intersecting  $X$  (Figure 2-23).

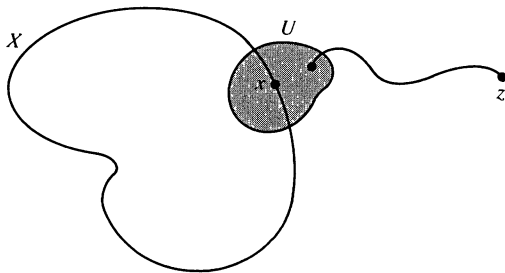


Figure 2-23

5. Show that  $\mathbf{R}^n - X$  has, at most, two connected components.
6. Show that if  $z_0$  and  $z_1$  belong to the same connected component of  $\mathbf{R}^n - X$ , then  $W_2(X, z_0) = W_2(X, z_1)$ .
7. Given a point  $z \in \mathbf{R}^n - X$  and a direction vector  $\vec{v} \in S^{n-1}$ , consider the ray  $r$  emanating from  $z$  in the direction of  $\vec{v}$ ,

$$r = \{z + t\vec{v} : t \geq 0\}.$$

Check that the ray  $r$  is transversal to  $X$  if and only if  $\vec{v}$  is a regular value of the direction map  $u: X \rightarrow S^{n-1}$ . In particular, almost every ray from  $z$  intersects  $X$  transversally.

8. Suppose that  $r$  is a ray emanating from  $z_0$  that intersects  $X$  transversally in a nonempty (necessarily finite) set. Suppose that  $z_1$  is any other point on  $r$  (but not on  $X$ ), and let  $l$  be the number of times  $r$  intersects  $X$  between  $z_0$  and  $z_1$ . Verify that  $W_2(X, z_0) = W_2(X, z_1) + l \pmod{2}$ . (See Figure 2-24.)
9. Conclude that  $\mathbf{R}^n - X$  has precisely two components,

$$D_0 = \{z : W_2(X, z) = 0\} \quad \text{and} \quad D_1 = \{z : W_2(X, z) = 1\}.$$

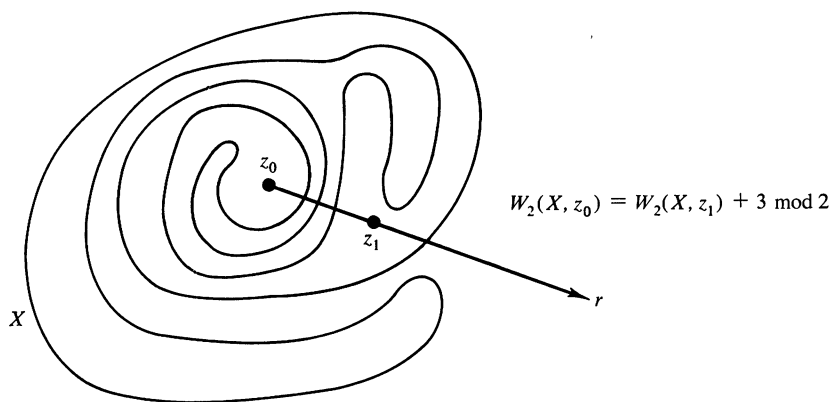


Figure 2-24

10. Show that if  $z$  is very large, then  $W_2(X, z) = 0$ .

11. Combine these observations to prove

**The Jordan-Brouwer Separation Theorem.** The complement of the compact, connected hypersurface  $X$  in  $\mathbb{R}^n$  consists of two connected open sets, the “outside”  $D_0$  and the “inside”  $D_1$ . Moreover,  $\bar{D}_1$  is a compact manifold with boundary  $\partial \bar{D}_1 = X$ .

Note that we have actually derived a simple procedure for determining whether a given point  $z$  lies inside or outside of  $X$ .

12. Given  $z \in \mathbb{R}^n - X$ , let  $r$  be any ray emanating from  $z$  that is transversal to  $X$ . Show that  $z$  is inside  $X$  if and only if  $r$  intersects  $X$  in an odd number of points (Figure 2-25).

**Hints** (listed by exercise number)

1. If  $u$  extends to  $D$ , then  $\deg_2(u) = 0$ .
2. Use 1, replacing  $D$  by

$$D' = D - \bigcup_{i=1}^l \text{Int}(B_i).$$

3. If  $f_i$  carries  $\partial B_i$  diffeomorphically onto a small sphere centered at  $z$ , then  $u_i: \partial B_i \rightarrow S^{n-1}$  is bijective. But  $f$  is a local diffeomorphism at  $y_i$ , so you can choose such  $B_i$ .
4. Show that the points  $x \in X$  for which the statement is true constitute a nonempty, open, and closed set. (Closed: easy. Open: use

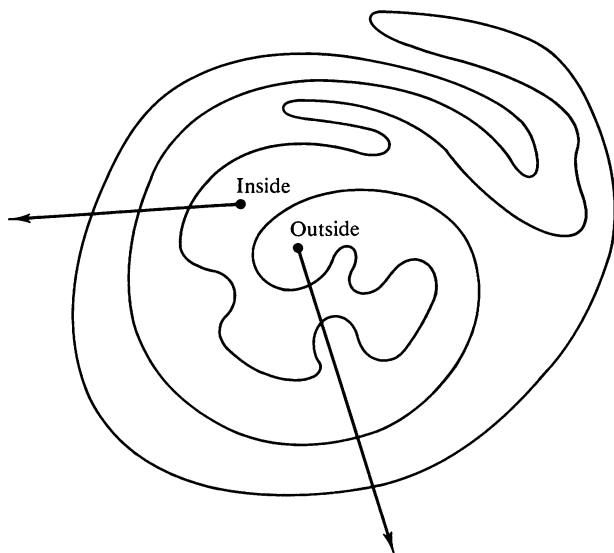


Figure 2-25

coordinates making  $X$  look locally like a piece of  $\mathbf{R}^{n-1}$  in  $\mathbf{R}^n$ . Non-empty: consider the straight line joining  $z$  to the closest point in  $X$ .)

5. Let  $B$  be a small ball such that  $B - X$  has two components, and fix points  $z_0$  and  $z_1$  in opposite components. Every point of  $\mathbf{R}^n - X$  may be joined to either  $z_0$  or  $z_1$  by a curve not intersecting  $X$ . (We have not yet excluded the possibility that  $z_0$  and  $z_1$  may be so joined!)
6. If  $z_t$  is a curve in  $\mathbf{R}^n - X$ , then the homotopy

$$u_t(x) = \frac{x - z_t}{|x - z_t|}$$

is defined for all  $t$ . Thus  $u_0$  and  $u_1$  have the same mod 2 degree.

7. Either compute directly or use Exercise 7, Chapter 1, Section 5. Note that if  $g: \mathbf{R}^n - \{z\} \rightarrow S^{n-1}$  is  $g(y) = (y - z)/|y - z|$ , then  $u: X \rightarrow S^{n-1}$  is  $g$  composed with the inclusion map of  $X$ . So by the exercise cited,  $u \cap \{\vec{v}\}$  if and only if  $X \cap g^{-1}\{\vec{v}\}$ .
8. By Exercise 7,  $\vec{v}$  is a regular value for both  $u_0$  and  $u_1$ . But

$$\#u_0^{-1}(\vec{v}) = \#u_1^{-1}(\vec{v}) + l.$$

9. Exercise 8 implies that both  $D_0$  and  $D_1$  are nonempty. Now apply Exercises 5 and 6.
10. Since  $X$  is compact, when  $|z|$  is large the image  $u(X)$  on  $S^{n-1}$  lies in a small neighborhood of  $z/|z|$ .

11. By Exercise 10,  $D_1$  is compact, and  $\bar{D}_1 = D_1 \cup X$ . Produce a local parametrization of  $\bar{D}_1$  around a point  $x \in X$  as follows: Let  $\psi: B \rightarrow \mathbb{R}^n$  map a ball  $B$  around 0 in  $\mathbb{R}^n$  diffeomorphically onto a neighborhood of  $x$  in  $\mathbb{R}^n$ , carrying  $B \cap \mathbb{R}^{n-1}$  onto  $X \cap \psi(B)$ . Use Exercises 4 and 6 to prove either that  $\psi(B \cap H^n) \subset D_1$  and  $\psi(B \cap -H^n) \subset D_0$  or the reverse. In either case,  $\psi$  restricts to a local parametrization of  $\bar{D}_1$  around  $x$ .
12. Combine Exercise 8 with Exercise 10.

## §6 The Borsuk-Ulam Theorem

We shall use our winding number apparatus to prove another famous theorem from topology, the Borsuk-Ulam theorem. One form of it is

**Borsuk-Ulam Theorem.** Let  $f: S^k \rightarrow \mathbb{R}^{k+1}$  be a smooth map whose image does not contain the origin, and suppose that  $f$  satisfies the symmetry condition

$$f(-x) = -f(x) \quad \text{for all } x \in S^k,$$

Then  $W_2(f, 0) = 1$ .

Informally, any map that is symmetric around the origin must wind around it an odd number of times.

*Proof.* Proceed by induction on  $k$ . For  $k = 1$ , see Exercise 2.

Now assume the theorem true for  $k - 1$ , and let  $f: S^k \rightarrow \mathbb{R}^{k+1} - \{0\}$  be symmetric. Consider  $S^{k-1}$  to be the equator of  $S^k$ , embedded by  $(x_1, \dots, x_k) \rightarrow (x_1, \dots, x_k, 0)$ . The idea of the proof is rather like Exercise 12 in the previous section. We will compute  $W_2(f, 0)$  by counting how often  $f$  intersects a line  $l$  in  $\mathbb{R}^{k+1}$ . By choosing  $l$  disjoint from the image of the equator, we can use the inductive hypothesis to show that the equator winds around  $l$  an odd number of times. Finally, it is easy to calculate the intersection of  $f$  with  $l$  once we know the behavior of  $f$  on the equator.

Denote the restriction of  $f$  to the equator  $S^{k-1}$  by  $g$ . In choosing a suitable line  $l$ , use Sard to select a unit vector  $\vec{a}$  that is a regular value for both maps

$$\frac{g}{|g|}: S^{k-1} \rightarrow S^k \quad \text{and} \quad \frac{f}{|f|}: S^k \rightarrow S^k.$$

From symmetry, it is clear that  $-\vec{a}$  is also a regular value for both maps. By dimensional comparison, regularity for  $g/|g|$  simply means that  $g/|g|$  never hits  $\vec{a}$  or  $-\vec{a}$ ; consequently,  $g$  never intersects the line  $l = \mathbb{R} \cdot \vec{a}$ . We let you check that regularity for  $f/|f|$  is equivalent to the condition  $f \nparallel l$ . (See Exercise 7, Chapter 1, Section 5.)

Now, by definition,

$$W_2(f, 0) = \deg_2 \left( \frac{f}{|f|} \right) = \# \left( \frac{f}{|f|} \right)^{-1}(\vec{a}) \bmod 2.$$

And  $f/|f|$  hits  $+\vec{a}$  precisely as often as it hits  $-\vec{a}$ , due to symmetry. Thus

$$\# \left( \frac{f}{|f|} \right)^{-1}(\vec{a}) = \frac{1}{2} \# f^{-1}(l).$$

We can calculate this result on the upper hemisphere alone. Let  $f_+$  be the restriction of  $f$  to the upper hemisphere  $S_+^k$ , the points where  $x_{k+1} \geq 0$ . By symmetry, plus the fact that  $f(\text{equator})$  does not intersect  $l$ , we know that

$$\# f_+^{-1}(l) = \frac{1}{2} \# f^{-1}(l).$$

We conclude that  $W_2(f, 0) = \# f_+^{-1}(l) \bmod 2$ .

The virtue of the latter expression for  $W_2(f, 0)$  is that the upper hemisphere is a manifold with boundary, and on its boundary  $\partial S_+^k = S^{k-1}$  we can use the inductive hypothesis. However, the dimensions are not correct for applying the inductive hypothesis to the map  $g: S^{k-1} \rightarrow \mathbf{R}^{k+1}$ . So let  $V$  be the orthogonal complement of  $l$ , and let  $\pi: \mathbf{R}^{k+1} \rightarrow V$  be orthogonal projection. Since  $g$  is symmetric and  $\pi$  is linear, the composite  $\pi \circ g: S^{k-1} \rightarrow V$  is symmetric; moreover,  $\pi \circ g$  is never zero, for  $g$  never intersects  $\pi^{-1}(0) = l$ . Identify the  $k$ -dimensional vector space  $V$  with  $\mathbf{R}^k$  and invoke the inductive hypothesis:  $W_2(\pi \circ g, 0) = 1$ .

Now since  $f_+ \nmid l$ ,

$$\pi \circ f_+: S_+^k \longrightarrow V$$

is transversal to  $\{0\}$ . So by the first theorem of the preceding section,

$$W_2(\pi \circ g, 0) = \#(\pi \circ f_+)^{-1}(0).$$

But

$$(\pi \circ f_+)^{-1}(0) = f_+^{-1}(l),$$

So

$$W_2(f, 0) = \# f_+^{-1}(l) = W_2(\pi \circ g, 0) = 1 \bmod 2. \quad \text{Q.E.D.}$$

One very simple geometrical observation is clear from the proof.

**Theorem.** If  $f: S^k \rightarrow \mathbf{R}^{k+1} - \{0\}$  is symmetric about the origin ( $f(-x) = -f(x)$ ), then  $f$  intersects every line through 0 at least once.

*Proof.* If  $f$  never hits  $l$ , then use this  $l$  in the proof, obtaining the contradiction

$$W_2(f, 0) = \frac{1}{2} \# f^{-1}(l) = 0. \quad \text{Q.E.D.}$$

This theorem has a number of surprising consequences, such as

**Theorem.** Any  $k$  smooth functions  $f_1, \dots, f_k$  on  $S^k$  that all satisfy the symmetry condition  $f_i(-x) = -f_i(x)$ ,  $i = 1, \dots, k$ , must possess a common zero.

*Proof.* If not, apply the corollary to the map

$$f(x) = (f_1(x), \dots, f_k(x), 0),$$

taking the  $x_{k+1}$  axis for  $l$ .

We may rephrase the statement as

**Theorem.** For any  $k$  smooth functions  $g_1, \dots, g_k$  on  $S^k$  there exists a point  $p \in S^k$  such that

$$g_1(p) = g_1(-p), \dots, g_k(p) = g_k(-p).$$

*Proof.* Convert to above by setting

$$f_i(x) = g_i(x) - g_i(-x). \quad \text{Q.E.D.}$$

A meteorological formulation of this result (for  $S^2$ ) is that at any given time there are two places in the world, at opposite ends of the earth from each other, having exactly the same weather (i.e., exactly the same temperature and pressure). Here's another verbalization of the theorem: if a balloon is deflated and laid on the floor, two antipodal points end up over the same point of the floor.

## EXERCISES

1. Show that the Borsuk-Ulam theorem is equivalent to the following assertion: if  $f: S^k \rightarrow S^k$  carries antipodal points to antipodal points, then  $\deg_2(f) = 1$ .
2. Prove that any map  $f: S^1 \rightarrow S^1$  mapping antipodal points to antipodal points has  $\deg_2(f) = 1$  by a direct computation with Exercise 8, Section 4. [HINT: If  $g: \mathbf{R} \rightarrow \mathbf{R}$  is as in Exercise 8, Section 4, show that

$$g(s + \pi) = g(s) + \pi q,$$

where  $q$  is odd.]

3. Let  $p_1, \dots, p_n$  be real homogeneous polynomials in  $n + 1$  variables, all of odd order. Prove that the associated functions on  $\mathbf{R}^{n+1}$  simultaneously vanish along some line through the origin.

# Oriented Intersection Theory

## §1 Motivation

Mod 2 intersection theory lacks the delicacy required for discriminations of any subtlety, but careful reexamination suggests a refinement. Suppose that  $f: X \rightarrow Y$  is transversal to a submanifold  $Z$  of dimension complementary to  $X$  ( $\dim X + \dim Z = \dim Y$ ) all three being boundaryless manifolds and  $X$  compact. Then  $f^{-1}(Z)$  is a finite set whose cardinality we called the “provisional intersection number” of  $f$  with  $Z$ . In order to make intersection numbers an invariant of homotopy, we discarded all the information carried by the provisional number  $\#f^{-1}(Z)$  except for one bit, its parity (even or odd).

Recall the reasoning. If  $f_0, f_1: X \rightarrow Y$  are homotopic and both transversal to  $Z$ , there exists a homotopy  $F: X \times I \rightarrow Y$  that is also transversal.  $F^{-1}(Z)$  then consists of some circles interior to  $X \times I$ , plus a few curves joining in pairs the points of

$$\partial F^{-1}(Z) = f_0^{-1}(Z) \times \{0\} \cup f_1^{-1}(Z) \times \{1\}.$$

(See Figure 3-1.) So although  $\#f_0^{-1}(Z)$  and  $\#f_1^{-1}(Z)$  need not be equal, at least they are both even or both odd; hence  $I_2(f_0, Z) = I_2(f_1, Z)$ .

But look more closely. Some curves join points in opposite ends of the cylinder  $X \times I$  and thus actually serve to compare  $\#f_0^{-1}(Z)$  and  $\#f_1^{-1}(Z)$ ,



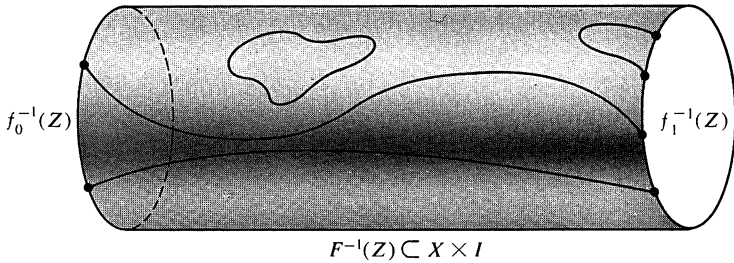


Figure 3-1

whereas other curves join points in the same end. In order to extract a homotopy invariant from the provisional data, we really only need cancel the contributions of any linked pairs of points that lie in the same end.

The difficulty is to recognize beforehand which points must be discarded. When we only have the one map  $f$ , can we tell how many points of  $f^{-1}(Z)$  will be joined by a homotopy? And is this cancellation factor the same for every possible homotopy? Both answers are affirmative, as subsequent sections will prove. The additional technique, involving some easy linear algebra, consists of little more than keeping track of a few plus and minus signs. Such bookkeeping always requires a little care and patience, but the rewards more than compensate for the minor annoyance.

## §2 Orientation

Suppose that  $V$  is a finite-dimensional real vector space and  $\beta = \{v_1, \dots, v_k\}$  is an ordered basis. If  $\beta' = \{v'_1, \dots, v'_k\}$  is another ordered basis, then there is a unique linear isomorphism  $A: V \rightarrow V$  such that  $\beta' = A\beta$  (where  $A\beta$  denotes the ordered basis  $\{Av_1, \dots, Av_k\}$ ). We shall say that  $\beta$  and  $\beta'$  are *equivalently oriented* if the determinant of this linear transformation  $A$  is positive. Thanks to the product rule for determinants, this procedure defines an equivalence relation partitioning the set of all ordered bases for  $V$  into two classes.

An *orientation* of  $V$  is an arbitrary decision to affix a positive sign to the elements of one equivalence class and a negative sign to the others. The sign given an ordered basis  $\beta$  is called its *orientation*, so  $\beta$  is either *positively oriented* or *negatively oriented*, depending on which equivalence class it belongs to. There are precisely two possible orientations for  $V$ , and in order to distinguish between them it suffices merely to specify the sign of any single ordered basis. For example, the standard orientation of Euclidean space is the one for which the standard ordered basis is positively oriented. (Notice that the ordering of basis elements is essential. Interchanging the positions of

any two basis vectors in  $\beta$  produces a different ordered basis with opposite orientation.)

A separate definition is required for the zero-dimensional vector space. Here an orientation is just a choice of sign  $+1$  or  $-1$  associated, if you like, with “the empty basis.”

If  $A: V \rightarrow W$  is an isomorphism of vector spaces, then whenever two ordered bases  $\beta$  and  $\beta'$  on  $V$  belong to the same equivalence class, so do the two ordered bases  $A\beta$  and  $A\beta'$  on  $W$ . Thus if both  $V$  and  $W$  are oriented, meaning that an orientation is specified for both, the sign of  $A\beta$  is either always the same as the sign of  $\beta$  or always opposite. That is,  $A$  either *preserves* or *reverses* orientation.

Now return to manifolds. An *orientation* of  $X$ , a manifold with boundary, is a smooth choice of orientations for all the tangent spaces  $T_x(X)$ . The smoothness condition is to be interpreted in the following sense: around each point  $x \in X$  there must exist a local parametrization  $h: U \rightarrow X$ , such that  $dh_u: \mathbf{R}^k \rightarrow T_{h(u)}(X)$  preserves orientation at each point  $u$  of the domain  $U \subset H^k$ . (The orientation on  $\mathbf{R}^k$  is implicitly assumed to be the standard one.) A map like  $h$  whose derivative preserves orientations at every point is simply called an *orientation-preserving* map. Not all manifolds possess orientations, the most famous example being the Möbius strip. You are invited to make a paper model and “prove” pictorially that no smooth orientation of the Möbius strip exists. The difficulty is that if you walk around a transparent strip pitching pennies heads up, eventually you return to the starting point to find tails up! (See Figure 3-2.)

For zero-dimensional manifolds, orientations are very simple. To each point  $x \in X$  we simply assign an *orientation number*  $+1$  or  $-1$ , there being no question of smoothness.

$X$  is *orientable* if it may be given an orientation. If so, then  $X$  obviously admits at least two distinct orientations, for if one is specified we need only

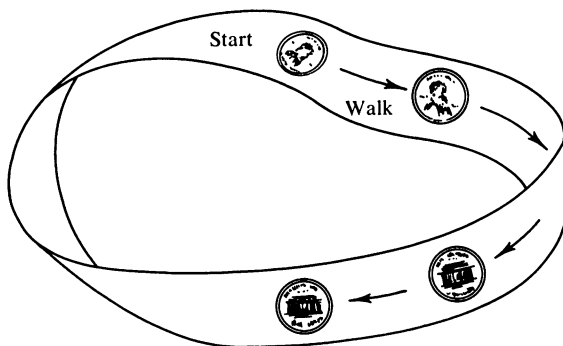


Figure 3-2

reverse the orientations of each tangent space to obtain an *opposite orientation*.

**Proposition.** A connected, orientable manifold with boundary admits exactly two orientations.

*Proof.* We show that the set of points at which two orientations agree and the set where they disagree are both open. Consequently, two orientations of a connected manifold are either identical or opposite. So let  $h: U \rightarrow X$ ,  $h': U' \rightarrow X$  be local parametrizations around  $x \in X$  such that  $dh_u$  preserves the first orientation and  $dh'_{u'}$  preserves the second, for all  $u \in U$ ,  $u' \in U'$ . We may assume  $h(0) = x = h'(0)$ , and  $h(U) = h'(U')$ . Then if the two orientations of  $T_x(X)$  agree,  $d(h^{-1} \circ h')_0: \mathbf{R}^k \rightarrow \mathbf{R}^k$  is orientation preserving. Thus  $d(h^{-1} \circ h')_{u'}$  has positive determinant at  $u' = 0$ , and the continuity of the determinant function implies that it has positive determinant for  $u'$  in a neighborhood of 0. Arguing back, this says that the two orientations agree in a neighborhood of  $x$ . Similarly, if the two orientations disagree at  $x$ , then  $d(h^{-1} \circ h')_{u'}$  is negative near 0, so the orientations must be opposite in a neighborhood of  $x$ . Q.E.D.

By an *oriented manifold*, we mean a manifold together with a specified smooth orientation. If  $X$  is oriented, we shall symbolically denote the oriented manifold obtained by reversing the orientation on  $X$  as  $-X$ . Thus the foregoing proposition says that if  $X$  is connected, the only two orientations on the underlying manifold are those of  $X$  and  $-X$ .

If  $X$  and  $Y$  are oriented and one of them is boundaryless, then  $X \times Y$  acquires a *product orientation* as follows. At each point  $(x, y) \in X \times Y$ ,

$$T_{(x,y)}(X \times Y) = T_x(X) \times T_y(Y).$$

Let  $\alpha = \{v_1, \dots, v_k\}$  and  $\beta = \{w_1, \dots, w_l\}$  be ordered bases for  $T_x(X)$  and for  $T_y(Y)$ , respectively, and denote by  $(\alpha \times 0, 0 \times \beta)$  the ordered basis  $\{(v_1, 0), \dots, (v_k, 0), (0, w_1), \dots, (0, w_l)\}$  of  $T_x(X) \times T_y(Y)$ . Define the orientation on  $T_x(X) \times T_y(Y)$  by setting

$$\text{sign}(\alpha \times 0, 0 \times \beta) = \text{sign}(\alpha) \text{sign}(\beta).$$

Check that this orientation does not depend on the choice of  $\alpha$  and  $\beta$ .

An orientation of  $X$  naturally induces a *boundary orientation* on  $\partial X$ . At every point  $x \in X$ ,  $T_x(\partial X)$  has codimension 1 in  $T_x(X)$ . Therefore there are precisely two unit vectors in  $T_x(X)$  that are perpendicular to  $T_x(\partial X)$ ; one points inward and one points outward. More precisely, if  $h: U \rightarrow X$  is a local parametrization around  $x$ ,  $U$  being open in  $H^k$  and  $h(0) = u$ , then  $(dh_0)^{-1}: T_x(X) \rightarrow \mathbf{R}^k$  carries one unit normal vector into  $H^k$ , the *inward*

unit normal, and carries the other into  $-H^k$ , the *outward* unit normal (Figure 3-3). You checked that this distinction is independent of the choice of  $h$  in Exercise 7, Chapter 2, Section 1. We denote the outward unit normal at  $x$  by  $n_x$ . Now orient  $T_x(\partial X)$  by declaring the sign of any ordered basis  $\beta = \{v_1, \dots, v_{k-1}\}$  to be the sign of the ordered basis  $\{n_x, \beta\} = \{n_x, v_1, \dots, v_{k-1}\}$  for  $T_x(X)$ . It is easy to check that this procedure smoothly orients each  $T_x(\partial X)$  and thus defines an orientation of  $\partial X$ .

As an example, the closed unit ball  $B^2$  in  $\mathbf{R}^2$  inherits the standard orientation from  $\mathbf{R}^2$ . The induced orientation on  $S^1$  is the one for which the counter-clockwise-pointing vectors are positive. (See Figure 3-4.)

A particularly important example (Figure 3-5) is the homotopy space  $X \times I$  or, more conveniently for the moment,  $I \times X$ . For each  $t \in I$ , the slice  $X_t = \{t\} \times X$  is naturally diffeomorphic to  $X$ , so let us orient  $X_t$  in order to

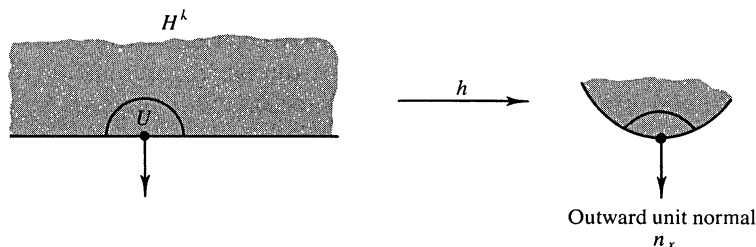


Figure 3-3

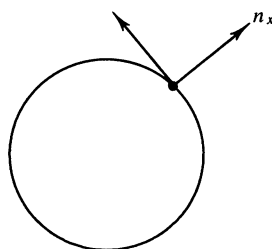


Figure 3-4

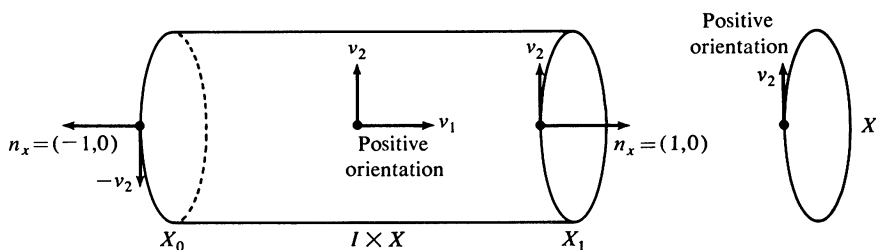


Figure 3-5

make this diffeomorphism  $x \rightarrow (t, x)$  orientation preserving. Now  $\partial(I \times X)$  is, as a set,  $X_1 \cup X_0$ ; but what is its boundary orientation? Along  $X_1$ , the outward-pointing normal vector is  $n_{(1,x)} = (1, 0) \in T_1(I) \times T_x(X)$ . Any ordered basis for  $T_{(1,x)}(X_1)$  is  $0 \times \beta$ , where  $\beta$  is an ordered basis for  $T_x(X)$ . By definition of the boundary orientation,

$$\text{sign}(0 \times \beta) = \text{sign}(n_{(1,x)}, 0 \times \beta),$$

whereas by the definition of the product orientation,

$$\text{sign}(1 \times 0, 0 \times \beta) = \text{sign}(1) \text{sign}(\beta) = \text{sign}(\beta).$$

This result shows that the boundary orientation on  $X_1$  equals its orientation as a copy of  $X$ . However, along  $X_0$ , the outward-pointing normal is  $n_{(0,x)} = (-1, 0)$ . So the sign of the basis  $0 \times \beta$  for  $T_{(0,x)}(X_0)$  in the boundary orientation on  $X_0$  equals

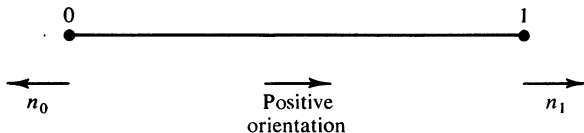
$$\text{sign}(-1 \times 0, 0 \times \beta) = \text{sign}(-1) \text{sign}(\beta) = -\text{sign}(\beta).$$

Thus the boundary orientation on  $X_0$  is the reverse of its orientation as a copy of  $X$ .

We conclude that as an oriented manifold,  $\partial(I \times X) = X_1 \cup (-X_0)$ . We shall adopt for this union the suggestive symbolism

$$\partial(I \times X) = X_1 - X_0.$$

If  $\dim X = 1$ , then  $\partial X$  is zero dimensional. The orientation of the zero-dimensional vector space  $T_x(\partial X)$  is equal to the sign of the basis  $\{n_x\}$  for  $T_x(X)$ . Consider, in particular, the compact interval  $X = [0, 1]$  with its standard orientation inherited from  $\mathbf{R}^1$ . At  $x = 1$ , the outward normal vector is  $1 \in \mathbf{R}^1 = T_1(X)$ , which is positively oriented, and at  $x = 0$  the outward normal is the negatively oriented  $-1 \in \mathbf{R}^1 = T_0(X)$ . Thus the orientation of  $T_1(\partial X)$  is  $+1$ , and the orientation of  $T_0(\partial X)$  is  $-1$ .



Reversing the orientation on  $[0, 1]$  simply reverses the orientations at each boundary point. Now let  $X$  be any compact oriented one-manifold with boundary. Since the boundary points of  $X$  are connected by diffeomorphic copies of the interval (thanks to the theorem classifying one-manifolds), we have

**Observation.** The sum of the orientation numbers at the boundary points of any compact oriented one-dimensional manifold with boundary is zero.

Thus orientation numbers are precisely the artifice needed to refine intersection theory. But first we must put orientations on transversal preimages. To do so, we shall use the following simple observation. Suppose that  $V = V_1 \oplus V_2$  is a direct sum. Then orientations on any two of these vector spaces automatically induces a *direct sum orientation* on the third, as follows. Choose ordered bases  $\beta_1$  and  $\beta_2$  for  $V_1$  and  $V_2$ , respectively, and let  $\beta = (\beta_1, \beta_2)$  be the combined ordered basis for  $V$ . Now simply demand that  $\text{sign}(\beta) = \text{sign}(\beta_1) \cdot \text{sign}(\beta_2)$ . Since the two spaces are oriented, this procedure uniquely affixes a sign to one ordered basis for their sum, and thereby determines an orientation of the sum. We let you check that different choices of  $\beta_1$  and  $\beta_2$  would induce the same orientation. But note that the order of the summands  $V_1$  and  $V_2$  is crucial, for  $(\beta_1, \beta_2)$  may not be equivalent in orientation to  $(\beta_2, \beta_1)$ .

Now let  $f: X \rightarrow Y$  be a smooth map with  $f \pitchfork Z$  and  $\partial f \pitchfork Z$ , where  $X$ ,  $Y$ , and  $Z$  are all oriented and the last two are boundaryless. We define a *preimage orientation* on the manifold-with-boundary  $S = f^{-1}(Z)$ . If  $f(x) = z \in Z$ , then  $T_x(S)$  is the preimage of  $T_z(Z)$  under the derivative map  $df_x: T_x(X) \rightarrow T_z(Y)$ . Let  $N_x(S; X)$  be the orthogonal complement to  $T_x(S)$  in  $T_x(X)$ . Then

$$N_x(S; X) \oplus T_x(S) = T_x(X),$$

so that we need only choose an orientation on  $N_x(S; X)$  to obtain a direct sum orientation on  $T_x(S)$ .

Because

$$df_x T_x(X) + T_z(Z) = T_z(Y),$$

and  $T_x(S)$  is the entire preimage of  $T_z(Z)$ , we get a direct sum

$$df_x N_x(S; X) \oplus T_z(Z) = T_z(Y).$$

Thus the orientations on  $Z$  and  $Y$  induce a direct image orientation on  $df_x N_x(S; X)$ . But  $T_x(S)$  contains the entire kernel of the linear map  $df_x$ , so  $df_x$  must map  $N_x(S; X)$  isomorphically onto its image. Therefore the induced orientation on  $df_x N_x(S; X)$  defines an orientation on  $N_x(S; X)$  via the isomorphism  $df_x$ .

All this discussion serves to define orientations on each tangent space  $T_x(S)$ ; we leave the verification of smoothness to you. (But smoothness should be rather obvious, since the construction used only some algebra plus the smoothly varying map  $df_x$ .) Unscrambling the formalities and developing a feeling for the preimage orientation are a matter of struggling through a few

calculations, so some examples are included in the exercises. For reference, here again are the two direct sum equations that define the preimage orientation. (The ordering of summands is important!)

$$\begin{aligned} df_x N_x(S; X) \oplus T_z(Z) &= T_z(Y) \\ N_x(S; X) \oplus T_x(S) &= T_x(X), \end{aligned}$$

where  $S = f^{-1}(Z)$ .

Our use of  $N_x(S; X)$ , the orthogonal complement of  $T_x(S)$  in  $T_x(X)$ , was only for convenient definiteness. In fact, if  $H$  is any other subspace of  $T_x(X)$  complementary to  $T_x(S)$ , then the two direct sums

$$\begin{aligned} df_x H \oplus T_z(Z) &= T_z(Y) \\ H \oplus T_x(S) &= T_x(X) \end{aligned}$$

define the same orientation on  $T_x(S)$ . Checking this is a matter of simple algebra, which we leave for Exercise 28.

Let us prove an important property of our definitions. Suppose that we have a map  $f: X \rightarrow Y$  as above, where  $f \pitchfork Z$ ,  $\partial f \pitchfork Z$ , the manifolds are oriented, and only  $X$  has boundary. Then the manifold  $\partial f^{-1}(Z)$  acquires two orientations—one as the preimage of  $Z$  under the map  $\partial f: \partial X \rightarrow Y$ , and a second as the boundary of the manifold  $f^{-1}(Z)$ . It turns out that these orientations differ by the sign  $(-1)^{\text{codim } Z}$ . The proof is a good exercise in understanding the definitions; in fact, just figuring out the statement is a valuable endeavor.

**Proposition.**  $\partial[f^{-1}(Z)] = (-1)^{\text{codim } Z}(\partial f)^{-1}(Z)$ .

*Proof.* Denote  $f^{-1}(Z)$  by  $S$ , as above. Let  $H$  be a subspace of  $T_x(\partial X)$  complementary to  $T_x(\partial S)$ , so

$$H \oplus T_x(\partial S) = T_x(\partial X).$$

(Figure 3-6.) Note that  $H$  is also complementary to  $T_x(S)$  in  $T_x(X)$ ; simply compare dimensions, noting that  $H$  must be disjoint from  $T_x(S)$  because

$$T_x(S) \cap T_x(\partial X) = T_x(\partial S).$$

So we may use  $H$  to define the orientation of both  $S$  and  $\partial S$  at  $x$ . (Here we need Exercise 28.) Since  $H \subset T_x(\partial X)$ , the maps  $df_x$  and  $d(\partial f)_x$  agree on  $H$ . Thus  $H$  is assigned the same orientation under the two maps, via the direct sum  $df_x H \oplus T_z(Z) = T_z(Y)$ . Now that  $H$  is oriented, the orientation of  $S$

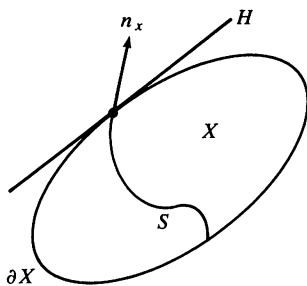


Figure 3-6

induced by  $f$  is defined by  $H \oplus T_x(S) = T_x(X)$ , and the orientation of  $\partial S$  induced by  $\partial f$  is defined by  $H \oplus T_x(\partial S) = T_x(\partial X)$ .

Let  $n_x$  be the outward unit vector to  $\partial S$  in  $S$ , and let  $\mathbf{R} \cdot n_x$  represent the one-dimensional subspace spanned by  $n_x$ , oriented so that  $\{n_x\}$  is a positively oriented basis. Unfortunately for our computation,  $n_x$  need not be perpendicular to all of  $T_x(\partial X)$ . Nonetheless, we claim that the orientations of  $T_x(\partial X)$  and  $T_x(X)$  are related by the direct sum

$$\mathbf{R} \cdot n_x \oplus T_x(\partial X) = T_x(X).$$

We leave you to check this; just note that  $n_x$  is an outward-pointing vector, and use Exercise 27.

Into the equation

$$T_x(X) = \mathbf{R} \cdot n_x \oplus T_x(\partial X)$$

insert the formulas for the preimage orientations of  $S$  and  $\partial S$ , obtaining

$$H \oplus T_x(S) = \mathbf{R} \cdot n_x \oplus H \oplus T_x(\partial S).$$

Since  $l = \dim H$  transpositions are required to move  $n_x$  from the left to the right of an ordered basis for  $H$ , the direct sum on the right has  $(-1)^l$  times the orientation of  $H \oplus \mathbf{R} \cdot n_x \oplus T_x(\partial S)$ . Thus  $T_x(S)$  must have  $(-1)^l$  times the orientation of  $\mathbf{R} \cdot n_x \oplus T_x(\partial S)$  when  $\partial S$  is given its preimage orientation. But, by definition, when  $\partial S$  is given the boundary orientation, then  $T_x(S)$  has exactly the orientation of  $\mathbf{R} \cdot n_x \oplus T_x(\partial S)$ . Conclude that the preimage and boundary orientations of  $T_x(\partial S)$  must differ by  $(-1)^l$ . Since

$$l = \dim H = \operatorname{codim} S = \operatorname{codim} Z,$$

we are done. Q.E.D.



## EXERCISES

1. Prove that the relation of being “equivalently oriented” is indeed an equivalence relation on ordered bases.
- \*2. Let  $\beta = \{v_1, \dots, v_k\}$  be an ordered basis for  $V$ . Show that:
  - (a) replacing one  $v_i$  by a multiple  $cv_i$  yields an equivalently oriented ordered basis if  $c > 0$ , an oppositely oriented one if  $c < 0$ .
  - (b) transposing two elements (i.e., interchanging the places of  $v_i$  and  $v_j$ ,  $i \neq j$ ) yields an oppositely oriented ordered basis.
  - (c) subtracting from one  $v_i$  a linear combination of the others yields an equivalently oriented ordered basis.
3. A sequence of linear maps of vector spaces

$$V_1 \xrightarrow{A_1} V_2 \xrightarrow{A_2} \cdots V_n \xrightarrow{A_n} V_{n+1}$$

is *exact* if

$$\text{Image}(A_i) = \text{Kernel}(A_{i+1}), \quad i = 1, \dots, n.$$

Thus exactness of  $0 \rightarrow U \xrightarrow{A} V \xrightarrow{B} W \rightarrow 0$  means that  $A$  is injective,  $B$  surjective, and  $\text{Image}(A) = \text{Kernel}(B)$ . Such a sequence is called a *short exact sequence*. Show how orientations of any two vector spaces in a short exact sequence automatically determine an orientation on the third. [HINT: Proceed as in the special case of direct sum orientations, in which  $V = U \oplus W$ .]

- \*4. Suppose that  $V$  is the direct sum of  $V_1$  and  $V_2$ . Prove that the direct sum orientation from  $V_1 \oplus V_2$  equals  $(-1)^{(\dim V_1)(\dim V_2)}$  times the orientation from  $V_2 \oplus V_1$ .
5. Prove that boundary orientations are smooth. (See Exercise 7, Chapter 2, Section 1.)
- \*6.  $H^k$  is oriented by the standard orientation of  $\mathbf{R}^k$ . Thus  $\partial H^k$  acquires a boundary orientation. But  $\partial H^k$  may be identified with  $\mathbf{R}^{k-1}$ . Show that the boundary orientation agrees with the standard orientation of  $\mathbf{R}^{k-1}$  if and only if  $k$  is even. Symbolically, we may express this as  $\partial H^k = (-1)^k \mathbf{R}^{k-1}$ .
7. Specifically write down the orientation of  $S^2$  (as the boundary of  $B^3$ ), by writing a positively oriented ordered basis for the tangent space at each  $(a, b, c)$ .

8. Define a function  $f: S^2 \rightarrow \mathbf{R}$  by  $f(x, y, z) = z$ . For the regular values  $-1 < t < 1$ ,  $f^{-1}(t)$  are the longitude circles at height  $t$ . What are their orientations? [Explicitly exhibit a positively oriented vector at a typical point of  $f^{-1}(t)$ .]
9. Prove that the boundary orientation of  $S^k = \partial B^{k+1}$  is the same as its preimage orientation under the map

$$g: \mathbf{R}^{k+1} \longrightarrow \mathbf{R}, \quad g(x) = |x|^2.$$

10. Suppose that  $f$  is a smooth function on  $\mathbf{R}^1$ , and let  $X \subset \mathbf{R}^2$  be its graph  $X = \{(x, f(x))\}$ . Specifically write down the orientation of  $X$  as the preimage of 0 under the map

$$F: \mathbf{R}^2 \longrightarrow \mathbf{R}, \quad F(x, y) = f(x) - y.$$

[HINT:  $T_{(x,y)}(X)$  is spanned by  $v = (1, f'(x))$ . So the normal direction to  $X$  is spanned by  $n = (-f'(x), 1)$ . Is  $v$  or  $-v$  positively oriented?]

11. Similarly, let  $f$  be a smooth function on  $\mathbf{R}^2$ , and let  $S \subset \mathbf{R}^3$  be its graph  $S = \{(x, y, f(x, y))\}$ . Explicitly determine the orientation of  $S$  under the map

$$F: \mathbf{R}^3 \longrightarrow \mathbf{R}, \quad F(x, y, z) = z - f(x, y).$$

[HINT:  $T_{(x,y,z)}(S)$  is spanned by

$$v_1 = \left(1, 0, \frac{\partial F}{\partial x}(x, y)\right) \quad \text{and} \quad v_2 = \left(0, 1, \frac{\partial F}{\partial y}(x, y)\right),$$

so the normal line is spanned by

$$n = \left(-\frac{\partial F}{\partial x}(x, y), -\frac{\partial F}{\partial y}(x, y), 1\right).$$

Is  $\{v_1, v_2\}$  or  $\{v_2, v_1\}$  positively oriented?]

12. Suppose that  $f: X \rightarrow Y$  is a diffeomorphism of connected oriented manifolds with boundary. Show that if  $df_x: T_x(X) \rightarrow T_{f(x)}(Y)$  preserves orientation at one point  $x$ , then  $f$  preserves orientation globally.
- \*13. Prove that every compact hypersurface in Euclidean space is orientable. [HINT: Jordan-Brouwer Separation Theorem.]
- \*14. Let  $X$  and  $Z$  be transversal submanifolds of  $Y$ , all three being oriented. Let  $X \cap Z$  denote the intersection manifold with the orientation pre-

scribed by the inclusion map  $X \rightarrow Y$ . Now suppose that

$$\dim X + \dim Z = \dim Y,$$

so  $X \cap Z$  is zero dimensional. Then at any point  $y \in X \cap Z$ ,

$$T_y(X) \oplus T_y(Z) = T_y(Y).$$

Check that the orientation number of  $y$  in  $X \cap Z$  is  $+1$  if the orientations of  $X$  and  $Z$  add up—in the order of the above sum—to the orientation of  $Y$ , and  $-1$  if not.

15. If  $\dim X + \dim Z = \dim Y$  and  $X \bar{\cap} Z$ , prove that

$$X \cap Z = (-1)^{(\dim X)(\dim Z)} Z \cap X.$$

16. Now drop the assumption of dimensional complementarity. Prove that whenever  $X \bar{\cap} Z$  in  $Y$ , the two orientations on the intersection manifold are related by

$$X \cap Z = (-1)^{(\operatorname{codim} X)(\operatorname{codim} Z)} Z \cap X.$$

Note that whenever  $X$  and  $Z$  do have complementary dimensions, then  $(\operatorname{codim} X)(\operatorname{codim} Z) = (\dim X)(\dim Z)$ . [HINT: Show that the orientation of  $S = X \cap Z$  is specified by the formula

$$[N_y(S; X) \oplus N_y(S; Z)] \oplus T_y(S) = T_y(Y).$$

But for  $Z \cap X$ , the first two spaces are interchanged.]

17. Compute the orientation of  $X \cap Z$  in the following examples by exhibiting positively oriented bases at every point. [By convention, we orient the three coordinate axes so that the standard basis vectors are positively oriented. Orient the  $xy$  plane so that  $\{(1, 0, 0), (0, 1, 0)\}$  is positive and the  $yz$  plane so that  $\{(0, 1, 0), (0, 0, 1)\}$  is positive. Finally, orient  $S^1$  and  $S^2$  as the boundary of  $B^2$  and  $B^3$ , respectively.]
- $X = x$  axis,  $Z = y$  axis (in  $\mathbf{R}^2$ )
  - $X = S^1$ ,  $Z = y$  axis (in  $\mathbf{R}^2$ )
  - $X = xy$ -plane,  $Z = z$  axis (in  $\mathbf{R}^3$ )
  - $X = S^2$ ,  $Z = yz$  plane (in  $\mathbf{R}^3$ )
  - $X = S^1$  in  $xy$  plane,  $Z = yz$  plane (in  $\mathbf{R}^3$ )
  - $X = xy$  plane,  $Z = yz$  plane (in  $\mathbf{R}^3$ )
  - $X =$  hyperboloid  $x^2 + y^2 - z^2 = a$  with preimage orientation ( $a > 0$ ),  $Z = xy$  plane (in  $\mathbf{R}^3$ )

**\*18.** Suppose that  $Z$  is a hypersurface in the oriented manifold  $Y$ , a submanifold of codimension 1. Prove that the following statements are equivalent:

- (a)  $Z$  is orientable.
- (b) There exists a smooth field of normal vectors  $\vec{n}(z)$  along  $Z$  in  $Y$ .
- (c) The normal bundle  $N(Z; Y)$  is trivial.
- (d)  $Z$  is globally definable by an independent function; that is, there is a smooth function  $\theta$  on a neighborhood of  $Z$  such that  $\theta^{-1}(0) = Z$ , and  $d\theta$  is nonzero at every point of  $Z$ .

[HINT: For (a)  $\langle \Rightarrow \rangle$  (b), mimic the definition of boundary orientations. (b)  $\langle \Rightarrow \rangle$  (c) is a trivial verification. For (c)  $\langle \Rightarrow \rangle$  (d), see Exercise 20, Chapter 2, Section 3.]

**\*19.** Suppose that  $Z$  is an oriented hypersurface in the oriented manifold  $Y$ , and let  $\vec{n}$  be a smooth field of unit normal vectors along  $Z$  in  $Y$ . Note that  $\vec{n}$  defines an orientation in each normal space  $N_z(Z; Y)$ —namely, the orientation in which the vector  $\vec{n}(z)$  is a positively oriented basis. Show that if the direct sum orientation on  $N_z(Z; Y) \oplus T_z(Z)$  and the orientation of  $T_z(Y)$  agree at one point  $z \in Z$ , then they agree in the entire connected component of  $Z$  containing  $z$ . Show that there is precisely one choice of  $\vec{n}$  for which the orientations always agree; we shall call this  $\vec{n}$  the *outward unit normal field* along  $Z$ . (For the case of a surface in  $\mathbf{R}^3$ , physicists say that  $\vec{n}$  conforms to the “right-hand rule.”) Check that for boundary orientations this is the usual outward normal.

**20.** Prove that the Möbius band is not orientable. [HINT: The central circle is orientable. Now use Exercise 18, plus Exercise 19 of Chapter 2, Section 4.]

**21.** Prove that the Möbius band cannot be globally defined in  $\mathbf{R}^3$  by an independent function.

**22.** Let  $V$  be a vector space. Show that both orientations on  $V$  define the same product orientation on  $V \times V$ .

**23.** Let  $X$  be an orientable manifold. Show that the product orientation on  $X \times X$  is the same for all choices of orientation on  $X$ . (See Exercise 22.)

**24.** Suppose that  $X$  is not orientable. Prove that  $X \times Y$  is never orientable, no matter what manifold  $Y$  may be. [HINT: First show that  $X \times \mathbf{R}^1$  is not orientable. But every  $Y$  has an open subset diffeomorphic to  $\mathbf{R}^1$ .]

**25.** Prove that there exists a natural orientation on some neighborhood of the diagonal  $\Delta$  in  $X \times X$ , whether or not  $X$  can be oriented. Compare

this with Exercises 23 and 24 and the fact that  $\Delta$  itself is orientable if and only if  $X$  is. [HINT: You can cover a neighborhood of  $\Delta$  by local parametrizations  $\phi \times \phi: U \times U \rightarrow X \times X$ , where  $\phi: U \rightarrow X$  is a local parametrization of  $X$ . Apply Exercise 23 to  $U \times U$ .]

26. Prove that every simply connected manifold  $X$  is orientable. [HINT: Pick an “origin”  $x \in X$ , and choose an orientation for the single space  $T_x(X)$ . If  $y \in X$ , orient  $T_y(X)$  as follows. Choose a sequence of open sets  $U_1, \dots, U_i$ , each  $U_i$  diffeomorphic to an open ball in  $\mathbf{R}^k$ , such that each  $U_i \cap U_{i+1} \neq \emptyset$ , and  $x \in U_1, y \in U_i$ . Successively orient the sets  $U_i$ , and show that the orientation induced on  $T_y(X)$  from  $U_i$  does not depend on the choice of the  $U_i$ .]
27. The definition of boundary orientation uses the outward unit normal  $n_x$  to  $\partial X$  at  $x$ . Show that perpendicularity is unnecessary. That is, if  $h_x \in T_x(X)$  is any other vector pointing outward, then it defines the same orientation as  $n_x$ . (“Outward pointing” is defined without perpendicularity assumptions; see Exercises 7 and 8 of Chapter 2, Section 1.) [HINT: Show that  $h_x = cn_x + v$ , where  $v \in T_x(\partial X)$  and  $c > 0$ .]
28. Similar to Exercise 27, show that orthogonality is not needed in defining preimage orientations. Specifically, if

$$H \oplus T_x(S) = T_x(X),$$

then this equation, plus

$$df_x H \oplus T_z(Z) = T_z(Y),$$

defines the same preimage orientation as defined by the orthogonal complement.

### §3 Oriented Intersection Number

We are prepared now to reconstruct intersection theory. The setting is the same as for mod 2 theory:  $X$ ,  $Y$ , and  $Z$  are boundaryless manifolds,  $X$  is compact,  $Z$  is a closed submanifold of  $Y$ , and  $\dim X + \dim Z = \dim Y$ . Now, however, we shall work exclusively with oriented manifolds. (Occasionally we will capsulize these hypotheses in the phrase “ $f: X \rightarrow Y$  and  $Z$  are appropriate for intersection theory.”)

If  $f: X \rightarrow Y$  is transversal to  $Z$ , then  $f^{-1}(Z)$  is a finite number of points, each with an orientation number  $\pm 1$  provided by the preimage orientation. Define the *intersection number*  $I(f, Z)$  to be the sum of these orientation numbers.

The orientation number at a point  $x \in f^{-1}(Z)$  is quite simple. For if  $f(x) = z \in Z$ , then transversality plus dimensional complementarity give a direct sum

$$df_x T_x(X) \oplus T_z(Z) = T_z(Y).$$

Now  $df_x$  must be an isomorphism onto its image, so that the orientation of  $X$  provides an orientation of  $df_x T_x(X)$ . Then the orientation number at  $x$  is  $+1$  if the orientations on  $df_x T_x(X)$  and  $T_z(Z)$  “add up” to the prescribed orientation on  $Y$  (in that order!), and  $-1$  if not.

The observation in the last section about one-manifolds is the key to the essential homotopy invariance of our definition. First, suppose that  $X$  is the boundary of a compact  $W$  and that  $f$  extends in some manner to a mapping  $F: W \rightarrow Y$ . By the Extension Theorem, we may assume  $F \pitchfork Z$ . Thus  $F^{-1}(Z)$  is a compact oriented one-manifold with boundary  $\partial F^{-1}(Z) = f^{-1}(Z)$ . Consequently, the sum of the orientation numbers at points in  $f^{-1}(Z)$  must be zero, and we have proved

**Proposition.** If  $X = \partial W$  and  $f: X \rightarrow Y$  extends to  $W$ , then  $I(f, Z) = 0$  ( $W$  compact).

In particular, suppose that  $f_0$  and  $f_1$  are homotopic and both transversal to  $Z$ . Then if  $F: I \times X \rightarrow Y$  is a homotopy between them, we know that  $I(\partial F, Z) = 0$ . But  $\partial(I \times X) = X_1 - X_0$ , and via the natural identifications of  $X_0$  and  $X_1$  with  $X$ ,  $\partial F$  equals  $f_0$  on  $X_0$  and  $f_1$  on  $X_1$ . It follows that

$$\partial F^{-1}(Z) = f_1^{-1}(Z) - f_0^{-1}(Z),$$

so

$$I(\partial F, Z) = I(f_1, Z) - I(f_0, Z).$$

This result proves

**Proposition.** Homotopic maps always have the same intersection numbers.

So far the statement only has meaning for maps transversal to  $Z$ . But as in the mod 2 theory, this homotopy invariance allows us to extend intersection theory to arbitrary maps. Given any  $g: X \rightarrow Y$ , select a homotopic map  $f$  that is transversal to  $Z$ , and define  $I(g, Z) = I(f, Z)$ . As we have just shown, any other choice of  $f$  must provide the same number. Note that the two previous propositions are automatically true for arbitrary maps as well as transversal ones. (Examples in a moment.)

When  $Y$  is connected and has the same dimension as  $X$ , we define the *degree* of an arbitrary smooth map  $f: X \rightarrow Y$  to be the intersection number of

$f$  with any point  $y$ ,  $\deg(f) = I(f, \{y\})$ . Our proof that  $I_2(f, \{y\})$  is the same for all points  $y \in Y$  works perfectly well for the oriented theory, so  $\deg(f)$  is acceptably defined. Later in the section we shall also deduce this fact from more general technical considerations. Notice that in order to calculate  $\deg(f)$ , one simply selects any regular value  $y$  and counts the preimage points  $\{x: f(x) = y\}$ , except that a point  $x$  makes a contribution of  $+1$  or  $-1$  to the sum, depending on whether the isomorphism  $df_x: T_x(X) \rightarrow T_y(Y)$  preserves or reverses orientation.

Degree is defined as an intersection number, so it must automatically be a homotopy invariant. To understand the way in which the canceling effect of orientation numbers converts the raw preimage data into a homotopy invariant, we suggest you draw a few pictures. For example, one can create many maps of  $S^1$  into itself by smoothly deforming the circle inside the plane and the projecting back onto  $S^1$  (Figure 3-7). Since such a map is homotopic to the identity, its degree must be  $+1$ . Look at the preimage of various regular values to verify this fact. (What are the critical values?)

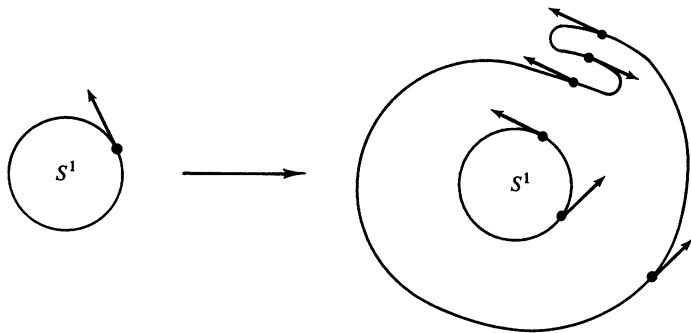


Figure 3-7

Another class of interesting maps of the circle involves the restrictions of the complex monomials  $z \rightarrow z^m$ . When  $m > 0$  this wraps the circle uniformly around itself  $m$  times preserving orientation. The map is everywhere regular and orientation preserving, so its degree is the number of preimages of any point—namely,  $m$ . (You should be able to prove these assertions easily by calculating with local parametrizations. We suggest using local parametrizations derived from the map  $\theta \rightarrow (\cos \theta, \sin \theta)$  of  $\mathbf{R}^1 \rightarrow S^1$ .) Similarly, when  $m < 0$  the map is everywhere regular but orientation reversing. As each point has  $|m|$  preimages, the degree is  $-|m| = m$ . Finally, when  $m = 0$  the map is constant, so its degree is zero. (By contrast, the mod 2 theory gives only  $m \bmod 2$ .)

One immediate consequence of this calculation (which could not have been proven with mod 2 theory) is the interesting fact that the circle admits an infinite number of homotopically distinct mappings, for since  $\deg(z^m) = m$ , none of these maps can be homotopic to another one.

The extendability theorem for intersection numbers specializes to the following statement about degrees.

**Proposition.** Suppose that  $f: X \rightarrow Y$  is a smooth map of compact oriented manifolds having the same dimension and that  $X = \partial W$  ( $W$  compact). If  $f$  can be extended to all of  $W$ , then  $\deg(f) = 0$ .

We may now complete an item left unfinished by the mod 2 theory. For any complex polynomial  $p(z)$  of order  $m$ , we showed earlier that on a circle  $S$  of sufficiently large radius  $r$  in the plane,

$$\frac{p(z)}{|p(z)|} \quad \text{and} \quad \frac{z^m}{|z^m|} = \left(\frac{z}{r}\right)^m$$

are homotopic maps of  $S \rightarrow S^1$ . Thus  $p/|p|$  must have the same degree as  $(z/r)^m$ —namely,  $m$ . When  $m > 0$  we conclude that  $p/|p|$  will not extend to the whole disk of radius  $r$ , implying that  $p$  must have a zero inside the disk. This proves

**The Fundamental Theorem of Algebra.** Every nonconstant complex polynomial has a root.

We can refine this argument to obtain more detailed information about the zeros of  $p$ . At any point  $z_0 \in \mathbb{C}$ , we may factor

$$p(z) = (z - z_0)^l q(z),$$

where  $q(z_0) \neq 0$ . Of course,  $p(z_0) = 0$  precisely when  $l > 0$ , in which case  $l$  is called the *multiplicity* of the zero  $z_0$ .

**Proposition.** Let  $W$  be a smooth compact region in  $\mathbb{C}$  whose boundary contains no zeros of the polynomial  $p$ . Then the total number of zeros of  $p$  inside  $W$ , counting multiplicities, is the degree of the map  $p/|p|: \partial W \rightarrow S^1$ .

*Proof.* By a “smooth compact region” we simply mean a compact two-submanifold with boundary. A trivial theorem from algebra says that  $p$  has only finitely many roots,  $z_0, \dots, z_n$ , in  $W$ . Around each  $z_i$  circumscribe a small closed disk  $D_i$ , making the disks all disjoint from each other and from the boundary. Then  $p/|p|$  is defined on

$$W' = W - \bigcup_{i=0}^n D_i.$$



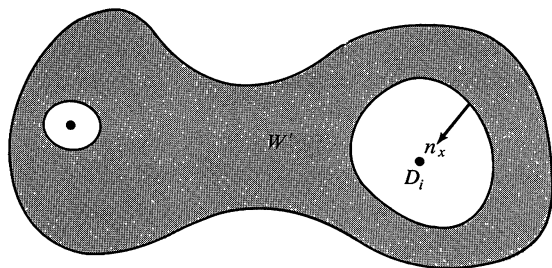


Figure 3-8

Note that the orientation of  $\partial D_i$  as part of the boundary of  $W'$  is opposite to its usual orientation as the boundary of  $D_i$ , for the outward-pointing normal vector for  $W'$  is the inward-pointing normal vector for  $D_i$  (Figure 3-8). Thus, as oriented manifolds,

$$\partial W' = \partial W - \bigcup_{i=0}^n \partial D_i.$$

Since the degree of  $p/|p|$  is zero on  $\partial W'$ , we conclude that its degree on  $W'$  equals the sum of its degrees on the oriented circles  $\partial D_i$ .

We complete the proof by showing that the degree of  $p/|p|: \partial D_i \rightarrow S^1$  is the multiplicity of the zero of  $p$  at  $z_i$ . Write

$$p(z) = (z - z_i)^l q(z),$$

where  $q(z_i) \neq 0$ . Since  $p$  has no zero in  $D_i$  other than  $z_i$ ,  $q$  can never be zero in  $D_i$ . If  $r$  is the radius of  $D_i$ , then  $g: S^1 \rightarrow \partial D_i$ , defined by  $g(z) = z_i + rz$ , is an orientation-preserving diffeomorphism of  $S^1$  onto  $\partial D_i$ . Thus the degree of  $p/|p|: \partial D_i \rightarrow S^1$  equals the degree of  $p \circ g/|p \circ g|: S^1 \rightarrow S^1$ . Define a homotopy  $h_t: S^1 \rightarrow S^1$  by

$$h_t(z) = \frac{z^l q(z_i + trz)}{|q(z_i + trz)|}.$$

(Division is admissible, since  $z_i + trz \in D_i$ .) Now

$$h_1 = \frac{p \circ g}{|p \circ g|},$$

while

$$h_0(z) = cz^l,$$

where  $c$  is the constant  $q(z_i)/|q(z_i)|$ . Since  $\deg(h_0) = l$ , conclude that  $\deg(h_1) = l$ . Q.E.D.

Let us shift our view to a different special instance of intersection theory. When  $X$  also happens to be a submanifold of  $Y$ , then, as in the mod 2 case, we define its intersection number with  $Z$ ,  $I(X, Z)$ , to be the intersection number of the inclusion map of  $X$  with  $Z$ . If  $X \bar{\cap} Z$ , then  $I(X, Z)$  is calculated by counting the points of  $X \cap Z$ , where a point  $y$  is included with a plus sign if the orientation of  $X$  and  $Z$  (in that order!) “add up” at  $y$  to the orientation of  $Y$ ; otherwise  $y$  is counted with a minus sign (Figure 3-9).

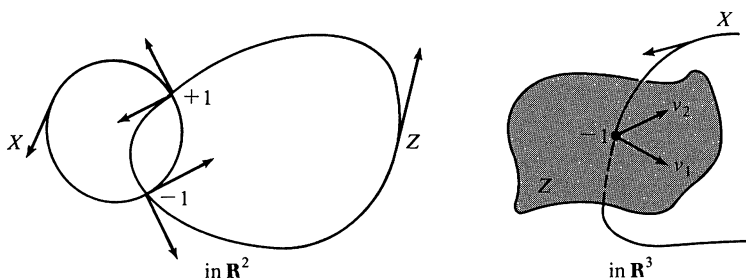
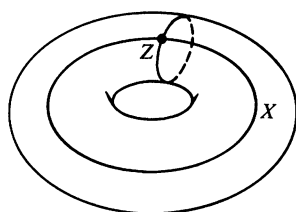


Figure 3-9



Two circles on the torus

$$I(X, Z) = -I(Z, X)$$

Figure 3-10

Be sure you understand the figure. Another suggestion for checking your grasp of this sometimes confusing matter of orientations is to calculate the importance of ordering. When both  $X$  and  $Z$  are compact, then they possess two intersection numbers,  $I(X, Z)$  and  $I(Z, X)$ , and these numbers may be different! For example, in Figure 3-10  $I(X, Z) = -I(Z, X)$ . In a few pages we shall consider this asymmetry in general, but, as an exercise now, compute the difference directly from the definition, assuming  $X \bar{\cap} Z$ .

We know that  $I(X, Z)$  remains the same no matter how we deform  $X$ . In order to prove its invariance under perturbations of  $Z$  as well, we shall generalize our approach somewhat. For the present,  $Z$ , like  $X$ , will be permitted an existence independent of  $Y$ , and we define an intersection number for a pair of arbitrary maps  $f: X \rightarrow Y$ ,  $g: Z \rightarrow Y$ . Of course, when  $Z$  happens to be a submanifold of  $Y$  and  $g$  is its inclusion map, we expect to retrieve the old number  $I(f, Z)$ . To replace the earlier condition that  $Z$  be a *closed* submanifold, we now require it to be compact, but otherwise  $Z$  may be an arbitrary,

boundaryless manifold satisfying the dimensional constraint  $\dim X + \dim Z = \dim Y$ . As usual, everything is oriented.

First, consider the transversal case. Two maps  $f$  and  $g$  are defined to be transversal,  $f \bar{\cap} g$ , if

$$df_x T_x(X) + dg_z T_z(Z) = T_y(Y)$$

whenever

$$f(x) = y = g(z).$$

The dimensional complementarity further implies that the sum is direct and that both derivatives  $df_x$  and  $dg_z$  are injective. Thus these derivatives map  $T_x(X)$  and  $T_z(Z)$  isomorphically onto their images, so that the image spaces receive orientations from  $X$  and  $Z$ , respectively. Define the local intersection number at  $(x, z)$  to be  $+1$  if the direct sum orientation of  $df_x T_x(X) \oplus dg_z T_z(Z)$  equals the given orientation on  $T_y(Y)$ , and  $-1$  otherwise. Then  $I(f, g)$  is defined as the sum of the local contributions from all pairs  $(x, z)$  at which  $f(x) = g(z)$ . (Note that when  $g: Z \rightarrow Y$  is the inclusion map of a submanifold, then  $f \bar{\cap} g$  if and only if  $f \bar{\cap} Z$ , and if so  $I(f, g) = I(f, Z)$ .)

To show that the sum is finite, we use a standard reformulation once more. If  $\Delta$  denotes the diagonal of  $Y \times Y$ , and  $f \times g: X \times Z \rightarrow Y \times Y$  is the product map, then  $f(x) = g(z)$  precisely at pairs  $(x, y)$  in  $(f \times g)^{-1}(\Delta)$ . Now  $\dim(X \times Z) = \text{codim } \Delta$ , so if  $f \times g \bar{\cap} \Delta$ , then the preimage of  $\Delta$  is a compact zero-dimensional manifold, hence a finite set. Transversality, and more, follows from a bit of linear algebra.

**Lemma.** Let  $U$  and  $W$  be subspaces of the vector space  $V$ . Then  $U \oplus W = V$  if and only if  $U \times W \oplus \Delta = V \times V$ . (Here  $\Delta$  is the diagonal of  $V \times V$ .) Assume, also, that  $U$  and  $W$  are oriented, and give  $V$  the direct sum orientation. Now assign  $\Delta$  the orientation carried from  $V$  by the natural isomorphism  $V \rightarrow \Delta$ . Then the product orientation on  $V \times V$  agrees with the direct sum orientation from  $U \times W \oplus \Delta$  if and only if  $W$  is even dimensional.

*Proof.* Note that

$$U \cap W = 0 \iff U \times W \cap \Delta = 0.$$

Furthermore,

$$\dim U + \dim W = \dim V \iff \dim U \times W + \dim \Delta = \dim V \times V.$$

The first assertion now follows immediately. Orientations are always easiest to compute yourself, but in case you like things the hard way, read on. Let  $\{u_1, \dots, u_k\}$  and  $\{w_1, \dots, w_l\}$  be positively oriented ordered bases for  $U$  and  $W$ , respectively. Then the combined ordered basis  $\{u_1, \dots, u_k, w_1, \dots, w_l\}$  is positively oriented for  $V$ , so  $\{(u_1, u_1), \dots, (u_k, u_k), (w_1, w_1), \dots, (w_l, w_l)\}$  is positively oriented for  $\Delta$ . In the product orientation of  $U \times W$ , the ordered

basis  $\{(u_1, 0), \dots, (u_k, 0), (0, w_1), \dots, (0, w_l)\}$  is positive, so for the direct sum  $U \times W \oplus \Delta$ , the following combined ordered basis is positively oriented:

$$\{(u_1, 0), \dots, (u_k, 0), (0, w_1), \dots, (0, w_l), (u_1, u_1), \dots, (u_k, u_k), (w_1, w_1), \dots, (w_l, w_l)\}.$$

Now one may subtract one basis element from another without altering orientation, so the following basis has the same sign:

$$\{(u_1, 0), \dots, (u_k, 0), (0, w_1), \dots, (0, w_l), (0, u_1), \dots, (0, u_k), (w_1, 0), \dots, (w_l, 0)\}.$$

By a sequence of  $l \times k$  transpositions of basis elements, we may convert this basis to

$$\{(u_1, 0), \dots, (u_k, 0), (0, u_1), \dots, (0, u_k), (0, w_1), \dots, (0, w_l), (w_1, 0), \dots, (w_l, 0)\};$$

first move  $(0, u_1)$   $l$  places to the left, then  $(0, u_2)$ , and so on. Similarly,  $l(l+k)$  transpositions convert the latter to

$$\{(u_1, 0), \dots, (u_k, 0), (w_1, 0), \dots, (w_l, 0), (0, u_1), \dots, (0, u_k), (0, w_1), \dots, (0, w_l)\}.$$

Finally, this ordered basis is, by definition, positively oriented for  $V \times V$ . Each transposition of basis vectors reverses orientation. Since

$$lk + l(l+k) = 2lk + l^2$$

transpositions intervened, the first and last ordered bases are equivalently oriented if and only if  $2lk + l^2$  is even, hence if and only if  $l = \dim W$  is even. Q.E.D.

Substituting

$$U = df_x T_x(X), \quad W = dg_z T_z(Z), \quad \text{and} \quad V = T_y(Y),$$

the lemma translates as

**Proposition.**  $f \frown g$  if and only if  $f \times g \frown \Delta$ , and then

$$I(f, g) = (-1)^{\dim Z} I(f \times g, \Delta).$$

One use of the proposition is to remove the transversality assumption on  $f$  and  $g$ . For arbitrary maps  $f: X \rightarrow Y, g: Z \rightarrow Y$ , we may simply define  $I(f, g)$  to be  $(-1)^{\dim Z} I(f \times g, \Delta)$ .

**Proposition.** If  $f_0$  and  $g_0$  are respectively homotopic to  $f_1$  and  $g_1$ , then  $I(f_0, g_0) = I(f_1, g_1)$ .

*Proof.* Taking homotopies  $f_t$  and  $g_t$ , we obtain a homotopy  $f_t \times g_t$  between  $f_0 \times g_0$  and  $f_1 \times g_1$ . Q.E.D.

**Corollary.** If  $Z$  is a submanifold of  $Y$  and  $i: Z \rightarrow Y$  is its inclusion map, then  $I(f, i) = I(f, Z)$  for any map  $f: X \rightarrow Y$ .

*Proof.* As noted earlier, the statement is completely obvious from the definition when  $f \cap Z$ . But if  $f$  is arbitrary, we may deform it homotopically into a map transversal to  $Z$  without altering either  $I(f, i)$  or  $I(f, Z)$ . Q.E.D.

**Corollary.** If  $\dim X = \dim Y$  and  $Y$  is connected, then  $I(f, \{y\})$  is the same for every  $y \in Y$ . Thus  $\deg(f)$  is well defined.

*Proof.* Since  $Y$  is connected, the inclusion maps  $i_0, i_1$  of any two points  $y_0, y_1 \in Y$  are homotopic. Therefore

$$I(f, \{y_0\}) = I(f, i_0) = I(f, i_1) = I(f, \{y_1\}). \quad \text{Q.E.D.}$$

**Proposition.**  $I(f, g) = (-1)^{(\dim X)(\dim Z)} I(g, f)$ .

*Proof.* We must compare the direct sum orientations of

$$T_y(Y) = df_x T_x(X) \oplus dg_z T_z(Z) \quad \text{and} \quad T_y(Y) = dg_z T_z(Z) \oplus df_x T_x(X).$$

Write down bases and observe that to convert one to the other requires  $\dim X \cdot \dim Z$  transpositions. Q.E.D.

**Corollary.** If  $X$  and  $Z$  are both compact submanifolds of  $Y$ , then

$$I(X, Z) = (-1)^{(\dim X)(\dim Z)} I(Z, X).$$

In particular, suppose that  $\dim Y = 2 \dim X$ , so that the self-intersection number  $I(X, X)$  is defined. If  $X$  is odd dimensional, then  $I(X, X) = -I(X, X)$ , so  $I(X, X) = 0$ . Consequently,  $I_2(X, X) = I(X, X) \bmod 2$  vanishes as well. This observation yields an insight into the nonorientability of some manifolds. Because mod 2 intersection theory is defined without any assumption of orientability, we may still calculate  $I_2(X, X)$  for compact orientable submanifolds  $X$  of half dimension in an *arbitrary* manifold  $Y$ . If one of these self-intersection numbers fails to vanish, then  $Y$  cannot be oriented. For example, the central circle in the Möbius strip has nonzero mod 2 self-intersection number, so the Möbius strip is nonorientable.

If  $Y$  is a compact, oriented manifold, its *Euler characteristic*  $\chi(Y)$  is defined to be the self-intersection number of the diagonal  $\Delta$  in  $Y \times Y$ :

$$\chi(Y) = I(\Delta, \Delta).$$

The Euler characteristic is a diffeomorphism invariant of compact manifolds† that plays a fundamental role in a variety of geometric and topological situations. The next few sections will illustrate its significance and provide another, more primitive, definition. For the present, we content ourselves with this definition and one consequence of our corollary.

**Proposition.** The Euler characteristic of an odd-dimensional, compact, oriented manifold is zero.

### EXERCISES

(Many of the exercises for mod 2 intersection theory are true for oriented intersection theory. It would be valuable to look back at Chapter 2, Section 4, with orientations in mind.)

- \*1. Suppose that  $f: X \rightarrow Y$  is a diffeomorphism of compact, connected manifolds. Check that  $\deg(f) = +1$  if  $f$  preserves orientation,  $\deg(f) = -1$  if  $f$  reverses orientation.
2. (a) Compute the degree of the antipodal map  $S^k \rightarrow S^k$ ,  $x \rightarrow -x$ .  
 (b) Prove that the antipodal map is homotopic to the identity if and only if  $k$  is odd. (See Exercise 7, Chapter 1, Section 6.)  
 (c) Prove that there exists a nonvanishing vector field in  $S^k$  if and only if  $k$  is odd. (See Exercises 7 and 8, Chapter 1, Section 8).  
 (d) Could mod 2 theory prove parts (b) and (c)?

3. Let

$$p(z) = z^m + a_1 z^{m-1} + \cdots + a_m$$

be a complex polynomial function. Prove that if

$$r^m > |a_1| r^{m-1} + \cdots + |a_m|,$$

then  $p$  has a root inside the disk of radius  $r$ ,  $\{|z| < r\}$ .

4. Let  $f(z) = 1/z$  on the circle of radius  $r$  in  $\mathbb{C}$ .  
 (a) Compute  $\deg(f|f)$ .

†Although we have defined Euler characteristic only for oriented manifolds, Exercise 20 shows that it is actually defined for nonorientable manifolds as well.

- (b) Why does our proof of the Fundamental Theorem of Algebra not imply that  $1/z = 0$  for some  $z \in \mathbb{C}$ ?
5. Where does the proof of the Fundamental Theorem of Algebra fail in  $\mathbb{R}$ ? (That is, why can't we use the same argument to show that every polynomial with real coefficients has a real root?)
6. Show that  $z^2 = e^{-|z|^2}$  for some complex number  $z$ .
- \*7. Prove that the map  $S^1 \rightarrow S^1$ , defined by  $z \rightarrow \bar{z}^m$ , where  $\bar{z}$  is the complex conjugate of  $z$ , has degree  $-m$ . [HINT:  $z \rightarrow \bar{z}$  is an orientation-reversing reflection diffeomorphism.]
- \*8. According to Exercise 8, Chapter 2, Section 4, for any map  $f: S^1 \rightarrow S^1$  there exists a map  $g: \mathbb{R}^1 \rightarrow \mathbb{R}^1$  such that

$$f(\cos t, \sin t) = (\cos g(t), \sin g(t)).$$

Moreover,  $g$  satisfies  $g(t + 2\pi) = g(t) + 2\pi q$  for some integer  $q$ . Show that  $\deg(f) = q$ .

- \*9. Prove that two maps of the circle  $S^1$  into itself are homotopic if and only if they have the same degree. This is a special case of a remarkable theorem of Hopf, which we will prove later. [HINT: If  $g_0, g_1: \mathbb{R}^1 \rightarrow \mathbb{R}^1$  both satisfy  $g(t + 1) = g(t) + 2\pi q$ , then so do all the maps  $g_s = sg_1 + (1 - s)g_0$ .]
10. Suppose that  $X \xrightarrow{f} Y \xrightarrow{g} Z$  are given, and prove that  $\deg(g \circ f) = \deg(f) \cdot \deg(g)$ .
- \*11. Prove that a map  $f: S^1 \rightarrow \mathbb{R}^2 - \{0\}$  extends to the whole ball  $B = \{z \mid |z| \leq 1\}$  if and only if  $\deg(f) = 0$ . [HINT: If  $\deg(f) = 0$ , use Exercise 9 to extend  $f$  to the annulus  $A = \{\frac{1}{2} \leq |z| \leq 1\}$  so that, on the inner circle  $\{|z| = \frac{1}{2}\}$ , the extended map is constant. By the trick of Exercise 1, Chapter 1, Section 6, you can make the map constant on a whole neighborhood of the inner circle. Now extend to the rest of  $B$ .]
12. Assume that  $X \overline{\cap} Z$ , both compact and oriented, and prove directly from the definition that

$$I(X, Z) = (-1)^{(\dim X)(\dim Z)} I(Z, X).$$

13. Prove that the Euler characteristic of the product of two compact, oriented manifolds is the product of their Euler characteristics.

- \*14. Suppose that  $W \xrightarrow{g} X \xrightarrow{f} Y$  is a sequence of maps with  $f \pitchfork Z$  in  $Y$ . Show that if  $f \circ g$  and  $Z$  are appropriate for intersection theory, so are  $g$  and  $f^{-1}(Z)$ . Prove that

$$I(f \circ g, Z) = I(g, f^{-1}(Z)).$$

15. Intersection numbers for transversal maps may be interpreted through the special case of submanifolds. For let  $f: X \rightarrow Y$  and  $Z \subset Y$  be appropriate for intersection theory,  $f \pitchfork Z$ . Then  $f^{-1}(Z) = \{x_1, \dots, x_N\}$ . Show that dimensional complementarity, plus transversality, makes  $f$  an immersion at each  $x_i$ . So  $f$  maps a neighborhood  $U_i$  of  $x_i$  diffeomorphically onto a submanifold  $V_i \subset Y$ , with

$$V_i \cap Z = \{f(x_i)\} \quad \text{and} \quad V_i \pitchfork Z$$

(Figure 3-11.) Orient  $V_i$  via this diffeomorphism, and check that  $I(f, Z)$  equals the sum of the intersection numbers of  $V_1 \cap Z, \dots, V_N \cap Z$ . (However, the union  $V_1 \cup \dots \cup V_N$  may not be a manifold, for the points  $f(x_i)$  need not be distinct.)

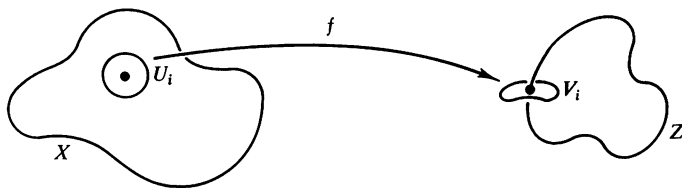


Figure 3-11

16. Let  $Z$  be a compact submanifold of  $Y$ , both oriented, with  $\dim Z = \frac{1}{2} \dim Y$ . Prove that  $I(Z, Z) = I(Z \times Z, \Delta)$ , where  $\Delta$  is the diagonal of  $Y$ . [HINT: Let  $i$  be the inclusion map of  $Z$ . Then

$$\begin{aligned} I(Z, Z) &= I(i, Z) = I(i, i) = (-1)^{\dim Z} I(i \times i, \Delta) \\ &= (-1)^{\dim Z} I(Z \times Z, \Delta). \end{aligned}$$

What happens when  $\dim Z$  is odd?]

- \*17. Prove that the Euler characteristic of an orientable manifold  $X$  is the same for all choices of orientation. [HINT: Use Exercise 16 with  $Z$  being the diagonal of  $X \times X$  and  $Y = X \times X$ . Apply Exercise 23, Section 2.]
18. There are instances when intersection numbers can be meaningfully defined even in the absence of global orientations. Let  $X$  and  $Z$  be compact submanifolds of  $Y$  with complementary dimension. Suppose that



there exists an open neighborhood  $U$  of  $X \cap Z$  in  $Y$  such that  $U$ ,  $U \cap X$ , and  $U \cap Z$  are all oriented. Prove that  $I(X, Z)$  is well defined and unaltered by small deformations of either  $X$  or  $Z$ . [HINT: Make  $X \cap Z$  by deforming it only inside  $U$ .]

19. In particular, suppose  $Z$  is a compact submanifold of  $Y$  with  $\dim Z = \frac{1}{2} \dim Y$ . Here we assume only  $Y$  oriented. Prove that the self-intersection number  $I(Z, Z)$  is still well defined and unaltered by small deformations of  $Z$ . [HINT: Define  $I(Z, Z) = I(Z \times Z, \Delta)$ , where  $\Delta$  is the diagonal of  $Y$ . According to Exercise 17, this agrees with the usual definition in the orientable case. Now invoke Exercise 25, Section 2, together with Exercise 18 above.]
20. As a special case of Exercise 19, prove that the Euler characteristic is well defined even for nonorientable manifolds and that it is still a diffeomorphism invariant. (You need Exercise 25 of Section 2 again.)

## §4 Lefschetz Fixed-Point Theory

One can use intersection theory to study the fixed points of a smooth map  $f: X \rightarrow X$  on a compact oriented manifold, the solutions of the equation  $f(x) = x$ . Perhaps the simplest question to ask is how many fixed points there are. And from an intersection theoretic point of view, a natural answer (in some sense) suggests itself. Note that  $x$  is a fixed point precisely when  $(x, f(x)) \in X \times X$  belongs to the intersection of  $\text{graph}(f)$  with the diagonal  $\Delta$ . As the latter are submanifolds of complementary dimension in  $X \times X$  (both oriented by their diffeomorphisms with  $X$ ), we may use intersection theory to “count” their common points:  $I(\Delta, \text{graph}(f))$  is called the *global Lefschetz number* of  $f$ , denoted  $L(f)$ .

Of course,  $f$  may actually have an infinite number of fixed points, the identity map being an extreme example. Thus the sense in which  $L(f)$  measures the fixed-point set is somewhat subtle. However, we shall see that when the fixed points of  $f$  do happen to be finite, then  $L(f)$  may be calculated directly in terms of the local behavior of  $f$  around its fixed points.

Before embarking on the local analysis, let us observe some immediate consequences of the intersection theory approach.

**Smooth Lefschetz Fixed-Point Theorem.**<sup>†</sup> Let  $f: X \rightarrow X$  be a smooth map on a compact orientable manifold. If  $L(f) \neq 0$ , then  $f$  has a fixed point.

<sup>†</sup>The Lefschetz number, like the Euler characteristic defined in the previous section, is a “topological” invariant—that is, it can be defined solely in terms of the open sets of  $X$  without any reference to manifold theory. From our approach it is only clear that it is a differential invariant. One can find a topological discussion of both these concepts in Greenberg’s book [12].

*Proof.* It is encouraging to find that the proof is obvious. If  $f$  has no fixed points, then  $\Delta$  and  $\text{graph}(f)$  are disjoint, hence trivially transversal. Consequently,

$$L(f) = I(\Delta, \text{graph}(f)) = 0.$$

This result gains strength as we find more efficient ways to compute the Lefschetz number; at the moment, for all we know  $L(f)$  might be zero for every  $f$ !

Since  $L(f)$  is defined as an intersection number, you can easily verify

**Proposition.**  $L(f)$  is a homotopy invariant.

The graph of the identity map is just the diagonal itself; thus  $L(\text{identity}) = I(\Delta, \Delta) = \chi(X)$ , the Euler characteristic. So

**Proposition.** If  $f$  is homotopic to the identity, then  $L(f)$  equals the Euler characteristic of  $X$ . In particular, if  $X$  admits a smooth map  $f: X \rightarrow X$  that is homotopic to the identity and has no fixed points, then  $\chi(X) = 0$ .

The obvious first candidates for study are maps  $f: X \rightarrow X$  for which  $\text{graph}(f) \bar{\cap} \Delta$ . Such maps, called *Lefschetz maps*, have only finitely many fixed points, although the converse is false. Since Lefschetz maps are defined by a transversality condition, no one should be surprised to discover that “most” maps are Lefschetz.

**Proposition.** Every map  $f: X \rightarrow X$  is homotopic to a Lefschetz map.

*Proof.* Recall from Chapter 2, Section 3, the basic lemma used to prove that transversality is generic: we can find an open ball  $S$  of some Euclidean space and a smooth map  $F: X \times S \rightarrow X$ , such that  $F(x, 0) = f(x)$  and  $s \mapsto F(x, s)$  defines a submersion  $S \rightarrow X$  for each  $x \in X$ . Note that the map  $G: X \times S \rightarrow X \times X$ , defined by  $G(x, s) = (x, F(x, s))$ , is also a submersion. For suppose that  $G(x, s) = (x, y)$ . Since  $G$  restricted to  $X \times \{s\}$  is a diffeomorphism onto  $X \times \{y\}$ , the image of  $dG_{(x,s)}$  contains  $T_x(X) \times \{0\}$ . It also contains  $\{0\} \times T_y(X)$ , as  $G$  restricted to  $\{x\} \times S$  is a submersion into  $\{x\} \times X$ . Therefore  $G$  is a submersion; hence certainly  $G \bar{\cap} \Delta$ . By the Transversality Theorem, for almost every  $s$  the map  $X \rightarrow X \times X$ , defined by  $x \mapsto G(x, s) = (x, F(x, s))$ , is  $\bar{\cap} \Delta$ . That is,  $x \mapsto F(x, s)$  is Lefschetz. Q.E.D.

What does it mean for  $f$  to be Lefschetz? Suppose that  $x$  is a fixed point of  $f$ . The tangent space of  $\text{graph}(f)$  in  $T_x(X) \times T_x(X)$  is the graph of the map  $df_x: T_x(X) \rightarrow T_x(X)$ , and the tangent space of the diagonal  $\Delta$  is the diagonal

$\Delta_x$  of  $T_x(X) \times T_x(X)$ . Thus  $\text{graph}(f) \cap \Delta$  at  $(x, x)$  if and only if

$$\text{graph}(df_x) + \Delta_x = T_x(X) \times T_x(X).$$

As  $\text{graph}(df_x)$  and  $\Delta_x$  are vector subspaces of  $T_x(X) \times T_x(X)$  with complementary dimension, they fill out everything precisely if their intersection is zero. But  $\text{graph}(df_x) \cap \Delta_x = 0$  means just that  $df_x$  has no nonzero fixed point or, in the language of linear algebra, that  $df_x$  has no eigenvector of eigenvalue  $+1$ .

We call the fixed point  $x$  a *Lefschetz fixed point* of  $f$  if  $df_x$  has no nonzero fixed point (i.e., if the eigenvalues of  $df_x$  are all unequal to  $+1$ ). So  $f$  is a Lefschetz map if and only if all its fixed points are Lefschetz. Notice that the Lefschetz condition on  $x$  is completely natural and straightforward; it is simply the infinitesimal analog of the demand that  $x$  be an isolated fixed point of  $f$ . (In fact, you proved in Exercise 10, Chapter 1, Section 5 that Lefschetz fixed points are isolated.)

If  $x$  is a Lefschetz fixed point, we denote the orientation number  $\pm 1$  of  $(x, x)$  in the intersection  $\Delta \cap \text{graph}(f)$  by  $L_x(f)$ , called the *local Lefschetz number* of  $f$  at  $x$ . Thus for Lefschetz maps,

$$L(f) = \sum_{f(x)=x} L_x(f).$$

It is quite easy to identify  $L_x(f)$  more explicitly. First, recognize that the Lefschetz condition at  $x$  is equivalent to the requirement that  $df_x - I$  be an isomorphism of  $T_x(X)$ , for the kernel of  $df_x - I$  is the fixed-point set of  $df_x$ . (Here  $I$  is the identity map of  $T_x(X)$ .) Now you should not be shocked to discover that  $L_x(f)$  simply reflects whether  $df_x - I$  preserves or reverses orientation.

**Proposition.** The local Lefschetz number  $L_x(f)$  at a Lefschetz fixed point is  $+1$  if the isomorphism  $df_x - I$  preserves orientation on  $T_x(X)$ , and it is  $-1$  if it reverses orientation. That is, the sign of  $L_x(f)$  equals the sign of the determinant of  $df_x - I$ .

*Proof.* Let  $A = df_x$  and let  $\beta = \{v_1, \dots, v_k\}$  be a positively oriented ordered basis for  $T_x(X)$ . Then

$$\{(v_1, v_1), \dots, (v_k, v_k)\} \quad \text{and} \quad \{(v_1, Av_1), \dots, (v_k, Av_k)\}$$

are positively oriented ordered bases for  $T_{(x,x)}(\Delta)$  and  $T_{(x,x)}(\text{graph}(f))$ , respectively. So the sign of  $L_x(f)$  equals the sign of the combined basis

$$\{(v_1, v_1), \dots, (v_k, v_k), (v_1, Av_1), \dots, (v_k, Av_k)\}$$

in the product orientation of  $X \times X$ . Since we may subtract a linear combination of basis vectors from another without altering orientation, the combined basis has the same sign as

$$\{(v_1, v_1), \dots, (v_k, v_k), (0, (A - I)v_1), \dots, (0, (A - I)v_k)\}.$$

Because  $A - I$  is an isomorphism, we can subtract suitable linear combinations of the last  $k$ -vectors from the first to obtain an equivalently oriented basis

$$\begin{aligned} \{(v_1, 0), \dots, (v_k, 0), (0, (A - I)v_1), \dots, (0, (A - I)v_k)\} \\ = \{\beta \times 0, 0 \times (A - I)\beta\}. \end{aligned}$$

By definition of the product orientation, the sign of the latter is equal to  $\text{sign } \beta \cdot \text{sign } (A - I)\beta$ . Q.E.D.

In order to develop a feeling for  $L_x(f)$ , let us examine the two-dimensional case. Since we are investigating a local property, assume that  $f: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ , and  $f$  fixes the origin. Set  $A = df_0$ , so that

$$f(x) = Ax + \epsilon(x),$$

where  $\epsilon(x) \rightarrow 0$  rapidly as  $x \rightarrow 0$ . Assume that  $A$  has two independent real eigenvectors; thus by the proper choice of coordinates, its matrix is diagonal,

$$A = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix}.$$

Then

$$L_0(f) = \text{sign } (\alpha_1 - 1)(\alpha_2 - 1).$$

Assume that both  $\alpha_1$  and  $\alpha_2$  are positive.

**Case 1.** Both  $\alpha_1, \alpha_2 > 1$ .  $L_0(f) = +1$ , and locally  $f$  is an “expanding map” with *source* at the origin (Figure 3-12).

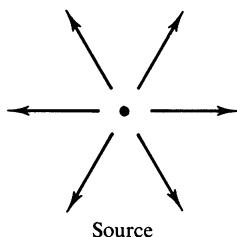


Figure 3-12

**Case 2.** Both  $\alpha_1, \alpha_2 < 1$ . Again  $L_0(f) = +1$ , and locally  $f$  is a “contracting map” with *sink* at the origin (Figure 3-13).

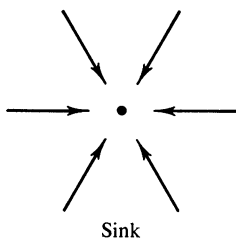


Figure 3-13

**Case 3.**  $\alpha_1 < 1 < \alpha_2$ . Here  $L_0(f) = -1$ , and the origin is a *saddlepoint* of  $f$  (Figure 3-14).

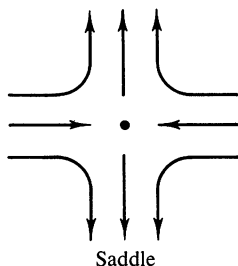


Figure 3-14

Observe how  $L_x(f)$  reports on the qualitative topological behavior of  $f$  near the fixed point  $x$ . In these two-dimensional examples, one can read off  $L_x(f)$  directly from pictorial representations of  $f$ . For example, let  $f: S^2 \rightarrow S^2$  be a smooth map of the unit sphere that moves every point except the poles due south. (Figure 3-15) If  $\pi: \mathbb{R}^3 - \{0\} \rightarrow S^2$  is the projection  $x \rightarrow x/|x|$ , then

$$f(x) = \pi(x + (0, 0, -\tfrac{1}{2}))$$

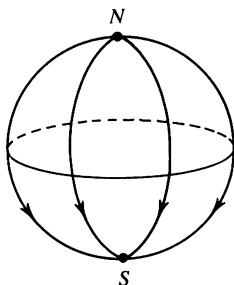


Figure 3-15

is an appropriate example. This  $f$  is a Lefschetz map with a source at the north pole and a sink at the south pole. Thus  $L_N(f) = +1 = L_S(f)$ , so

$$L(f) = L_N(f) + L_S(f) = 2.$$

As  $f$  is homotopic to the identity via

$$f_t(x) = \pi \left[ x + \left( 0, 0, -\frac{t}{2} \right) \right],$$

we have  $L(f) = \chi(S^2)$ .

**Proposition.** The Euler characteristic of  $S^2$  is 2.

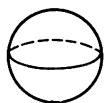
The theorem relating Euler characteristics to fixed-point theory is a double-edged sword. We have just used it to calculate  $\chi(S^2)$ , but now we may reverse direction to conclude

**Corollary.** Every map of  $S^2$  that is homotopic to the identity must possess a fixed point. In particular, the antipodal map  $x \rightarrow -x$  is not homotopic to the identity.

If you will grant a bit of poetic license, we can extend the argument to calculate the Euler characteristic of all compact oriented two-dimensional manifolds. First we must quote a theorem not to be proven here. (A classical proof may be found in Ahlfors and Sario [7], and a modern Morse theoretic proof in Gramain [6] or Wallace [5].)

**Classification of Two-Manifolds.** Every compact oriented boundaryless two-manifold is diffeomorphic to one of the following:

Surface of genus 0 (sphere):



Surface of genus 1 (torus):



Surface of genus 2:



Surface of genus 3:



etc.

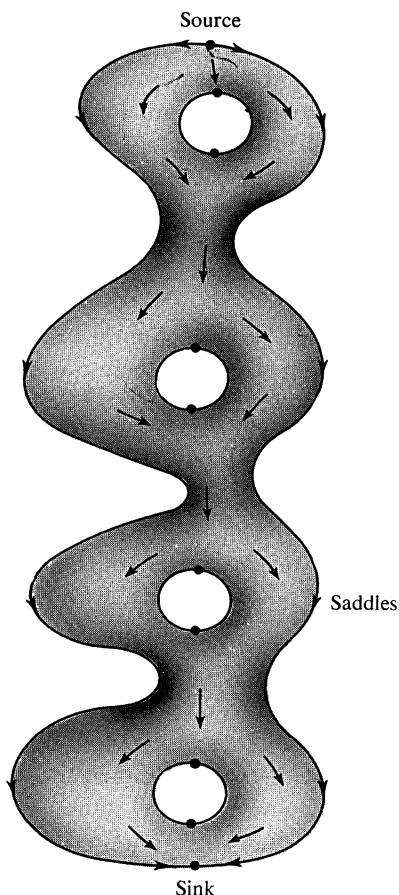


Figure 3-16

Now we “construct” a Lefschetz map on the surface of genus  $k$  as follows. Stand the surface vertically on one end, and coat it evenly with hot fudge topping. Let  $f_t(x)$  denote the oozing trajectory of the point  $x$  of fudge as time  $t$  passes. At time 0,  $f_0$  is just the identity. At time  $t > 0$ ,  $f_t$  is a Lefschetz map with one source at the top, one sink at the bottom, and saddlepoints at the top and bottom of each hole. Figure 3-16 shows the map for the surface of genus 4.

**Proposition.** The surface of genus  $k$  admits a Lefschetz map homotopic to the identity, with one source, one sink, and  $2k$  saddles. Consequently, its Euler characteristic is  $2 - 2k$ .

Note that if  $k > 1$ , the Euler characteristic of the surface of genus  $k$  is negative—rather surprising for a self-intersection number!

Despite the fact that “most” maps are Lefschetz, many common ones are not. For example, the polynomial maps  $z \rightarrow z + z^m$  on  $\mathbb{C}$  are not Lefschetz for  $m > 1$ . But we shall see that complicated fixed points are just amalgams of Lefschetz points. They behave like unstable compound particles in physics; disrupted by the slightest perturbation, they decompose into elementary constituents. That is the essence of proof of the first proposition in this section: under an arbitrarily small perturbation, the fixed points of any map  $f: X \rightarrow X$  split into Lefschetz fixed points. If  $f$  has isolated fixed points, we need only disrupt it locally.

**Splitting Proposition.** Let  $U$  be a neighborhood of the fixed point  $x$  that contains no other fixed points of  $f$ . Then there exists a homotopy  $f_t$  of  $f$  such that  $f_1$  has only Lefschetz fixed points in  $U$ , and each  $f_t$  equals  $f$  outside some compact subset of  $U$ .

*Proof.* First suppose that  $U$  is an open set in  $\mathbb{R}^k$  and that  $f: U \rightarrow \mathbb{R}^k$  fixes 0, but no other points, in  $U$ . Let  $\rho: \mathbb{R}^k \rightarrow [0, 1]$  be a smooth function that is one on a neighborhood  $V$  of the origin and zero outside a compact subset  $K \subset U$ . We shall show that there are points  $a \in \mathbb{R}^k$ , with  $|a|$  arbitrarily small, such that

$$f_t(x) = f(x) + t\rho(x)a$$

is satisfactory. Note that if  $|a|$  is small enough, then  $f_t$  has no fixed points outside of  $V$ . For since  $f$  has no fixed points in the compact  $K - V$ ,

$$|f(x) - x| > c > 0$$

there. So if  $|a| < c/2$ , we have

$$|f_t(x) - x| \geq |f(x) - x| - t\rho(x)|a| > \frac{c}{2}$$

on  $K - V$ . Of course, outside  $K$ ,

$$f_t(x) = f(x) \neq x.$$

Now use Sard to select any point  $a$ , as close as desired to 0, such that  $a$  is a regular value for the map  $x \rightarrow f(x) - x$ , and  $|a| < c/2$ . If  $x$  is a fixed point of  $f_1$ , then  $x \in V$ , so  $f_1 = f + a$  near  $x$ . Consequently,  $d(f_1)_x = df_x$ ; hence  $x$  is a Lefschetz fixed point for  $f_1$  if and only if  $df_x - I$  is nonsingular. But since  $f_1(x) = x$ ,

$$f(x) - x = a.$$



So the regularity of  $a$  implies that

$$d(f - I)_x = df_x - I$$

is nonsingular.

This proves the proposition in  $\mathbf{R}^k$ . Transferring to manifolds is easy. Choose a local parametrization  $\phi$  around  $x \in X$  carrying 0 to  $x$ , and apply the Euclidean case to  $g = \phi^{-1} \circ f \circ \phi$ . You need only check that if  $z$  is a Lefschetz fixed point for  $g$ , then  $\phi(z)$  is a Lefschetz fixed point for  $f$ ,  $f = \phi \circ g \circ \phi^{-1}$ . But the chain rule shows that  $d(f)_z$  has a nonzero fixed point if and only if  $d(g)_z$  does. Q.E.D.

The compound fixed points may decay into elementary ones in different ways. Depending on the precise perturbation, the pattern and even the number of Lefschetz “particles” produced may vary. But one factor we know is constant: the sum of the local Lefschetz numbers of the resulting fixed points. These numbers, therefore, play the role of electric charges on the particles, and the homotopy invariance of the global Lefschetz number is the analog of the conservation of charge law. In order to account for the fact that different numbers of elementary particles may be produced, think of Lefschetz fixed points with opposite signs as oppositely charged antimatter equivalents; they can annihilate each other or, conversely, they can be created in pairs.

Certainly we should be able to determine empirically the charge on a compound particle without smashing it into bits to count the resulting  $\pm 1$  charges. One experiment, using our intersection theoretic equipment, suggests itself in Euclidean space. For the fixed points of a map  $f$  on Euclidean space are the same as the zeros of the map  $x \rightarrow f(x) - x$ , and we have had experience in measuring these phenomena.

Suppose that  $x$  is an isolated fixed point of  $f$  in  $\mathbf{R}^k$ . If  $B$  is a small closed ball centered at  $x$  that contains no other fixed point, then the assignment

$$z \longrightarrow \frac{f(z) - z}{|f(z) - z|}$$

defines a smooth map  $F: \partial B \rightarrow S^{k-1}$ . We call the degree of this map the *local Lefschetz number* of  $f$  at  $x$ , denoted  $L_x(f)$ . Note that the choice of  $B$  is insignificant, for if  $B'$  is another ball with, say, smaller radius, then  $F$  is defined on the whole annulus  $\overline{B} - \overline{B'}$ . The last term is a compact manifold with boundary, and the boundary equals  $\partial B$ , with the usual orientation, union  $\partial B'$ , with the opposite orientation. Since the degree of  $F$  is zero on the boundary of  $\overline{B} - \overline{B'}$ , its degree on  $B$  must equal its degree on  $B'$ .

Our terminology requires justification.

**Proposition.** At Lefschetz fixed points, the two definitions of  $L_x(f)$  agree.

*Proof.* For simplicity, assume that  $x = 0$ . Since  $f(0) = 0$ , we may write  $f(z) = Az + \epsilon(z)$ , where  $A = df_0$  and  $\epsilon(z)/|z| \rightarrow 0$  as  $z \rightarrow 0$ . Since the linear map  $A - I$  is an isomorphism, the image of the unit ball under  $A - I$  contains some closed ball of radius  $c > 0$ . Linearity implies  $|(A - I)z| > c|z|$  for all  $z \in \mathbf{R}^k$ . Now choose the radius of the ball  $B$  small enough so that

$$\frac{|\epsilon(z)|}{|z|} < \frac{c}{2} \quad \text{on } B.$$

Define  $f_t(z) = Az + t\epsilon(z)$ .

Since  $|f_t(z) - z| \geq |(A - I)z| - t|\epsilon(z)| > \frac{1}{2}c|z|$

for  $0 \leq t \leq 1$ , the map

$$F_t(z) = \frac{f_t(z) - z}{|f_t(z) - z|}$$

defines a homotopy  $\partial B \times I \rightarrow S^{k-1}$ .

Now  $\deg(F_1)$  is precisely the new definition of  $L_0(f)$ . Since  $\deg(F_1) = \deg(F_0)$ , we have reduced to showing that the degree of

$$F_0 : z \longrightarrow \frac{(A - I)z}{|(A - I)z|}$$

is  $\pm 1$ , depending on whether  $A - I$  preserves or reverses orientation. Here we invoke an algebraic lemma, one that you will prove in the exercises (Exercise 1).

**Lemma.** Suppose that  $E$  is a linear isomorphism of  $\mathbf{R}^k$  that preserves orientation. Then there exists a homotopy  $E_t$  consisting of linear isomorphisms, such that  $E_0 = E$  and  $E_1$  is the identity. If  $E$  reverses orientation, then there exists such a homotopy with  $E_1$  equal to the reflection map

$$E_1(x_1, \dots, x_k) = (-x_1, x_2, \dots, x_k).$$

Apply this to  $E = A - I$ , obtaining a homotopy of  $F_0$  to  $z \mapsto E_1 z / |E_1 z|$ . Both possibilities for  $E_1$  preserve norm, so  $|E_1 z| = |z| = r$ , the radius of  $B$ . Thus if  $A - I$  preserves orientation,  $F_0$  is homotopic to the standard orientation-preserving diffeomorphism  $z \mapsto z/r$  of  $\partial B \rightarrow S^{k-1}$ , so  $\deg(F_0) = 1$ . If  $A - I$  reverses orientation, then  $F_0$  is homotopic to this same diffeomorphism followed by the orientation-reversing reflection map on  $S^{k-1}$ , so  $\deg(F_0) = -1$ . Q.E.D.

Not only does our new definition agree with the old for Lefschetz fixed points, it is also easily seen to be the correct “charge” invariant.

**Proposition.** Suppose that the map  $f$  in  $\mathbf{R}^k$  has an isolated fixed point at  $x$ , and let  $B$  be a closed ball around  $x$  containing no other fixed point of  $f$ . Choose any map  $f_1$  that equals  $f$  outside some compact subset of  $\text{Int}(B)$  but has only Lefschetz fixed points in  $B$ . Then

$$L_x(f) = \sum_{f_1(z)=z} L_z(f_1), \quad z \in B.$$

*Proof.*  $L_x(f)$  is the degree on  $\partial B$  of

$$F: z \longmapsto \frac{f(z) - z}{|f(z) - z|}.$$

But on  $\partial B$ ,  $F$  equals

$$F_1: z \longmapsto \frac{f_1(z) - z}{|f_1(z) - z|}.$$

Now suppose that  $z_1, \dots, z_N$  are the fixed points of  $f_1$ , and choose small balls  $B_i$  around  $z_i$  that are disjoint from each other and from  $\partial B$ . Then  $F_1$  extends to

$$B' = B - \bigcup_{i=1}^N \text{Int}(B_i),$$

so its degree as a map  $\partial B' \rightarrow S^{k-1}$  is zero. But since

$$\partial B' = \partial B - \bigcup_{i=1}^N \partial B_i,$$

this degree equals its degree on  $\partial B$ , minus its degrees on  $\partial B_i$ . Finally, the degree of  $F_1$  on  $\partial B_i$  is  $L_{z_i}(f_1)$ . Q.E.D.

Now let us extend this Euclidean definition of local Lefschetz numbers to manifolds in the most pedestrian way. If  $f: X \rightarrow X$  has an isolated fixed point at  $x$ , choose any local diffeomorphism  $\phi$  around  $x$ , and let  $g = \phi^{-1} \circ f \circ \phi$  on Euclidean space. Suppose that  $\phi(0) = x$ , and define  $L_x(f) = L_0(g)$ . Does this definition depend on the choice of  $\phi$ ? First we check Lefschetz fixed points. If  $x$  is Lefschetz, then  $L_x(f)$  is positive or negative, depending on whether  $df_x - I$  preserves or reverses orientation. But

$$dg_0 - I = (d\phi_0^{-1} \circ df_x \circ d\phi_0) - I = d\phi_0^{-1} \circ (df_x - I) \circ d\phi_0.$$

So  $dg_0 - I$  is an isomorphism if  $df_x - I$  is, and it has the same effect on orien-

tation as  $df_x - I$ . Thus 0 is also a Lefschetz fixed point for  $g$ , and  $L_0(g) = L_x(f)$ .

If  $x$  is an arbitrary fixed point, we can avoid directly checking the insignificance of coordinate choice by splitting  $x$  into Lefschetz fixed points. Given two parametrizations  $\phi$  and  $\phi'$  around  $x$ , choose a map  $f_1: X \rightarrow X$  that equals  $f$  outside some small set  $U$  contained in the images of both parametrizations and having only Lefschetz fixed points in  $U$ . Then, by the last proposition, the local Lefschetz number of  $\phi^{-1} \circ f \circ \phi$  at 0 equals the sum of the local Lefschetz numbers of  $\phi^{-1} \circ f_1 \circ \phi$  in  $\phi^{-1}(U)$ . However, we have just shown that the number computed for  $\phi^{-1} \circ f_1 \circ \phi$  at each Lefschetz fixed point equals the corresponding number for  $f$ . Thus

$$L_0(\phi^{-1} \circ f \circ \phi) = \sum_{f_1(z)=z} L_z(f_1) \quad z \in U.$$

As we obtain the same formula using  $\phi'$ ,  $L_x(f)$  is well defined.

Moreover, we have proved the

**Local Computation of the Lefschetz Number.** Let  $f: X \rightarrow X$  be any smooth map on a compact manifold, with only finitely many fixed points. Then the global Lefschetz number (which is a homotopy invariant) equals the sum of the local Lefschetz numbers:

$$L(f) = \sum_{f(x)=x} L_x(f).$$

*Proof.* Just perturb  $f$  locally around each fixed point to obtain a homotopic Lefschetz mapping  $f_1: X \rightarrow X$ . We know the theorem for  $f_1$ ; we know that  $L(f) = L(f_1)$ ; and, finally, we just showed that the sum of the local Lefschetz numbers of  $f$  equals the analogous sum for  $f_1$ . Q.E.D.

## EXERCISES

- Here is a guide to proving the lemma of page 128.
  - Show that you need only deal with the orientation-preserving case. [HINT: If  $E$  reverses orientation, compose it with the reflection.]
  - Since  $E$  has a real or complex eigenvalue, it maps some one- or two-dimensional subspace  $V \subset \mathbf{R}^k$  onto itself. Write  $\mathbf{R}^k = V \oplus W$ . Find a homotopy consisting of linear isomorphisms  $E_t$  such that  $E_0: V \rightarrow V$  and  $E_0: W \rightarrow W$ . HINT: In an appropriate basis,

$$E = \left( \begin{array}{c|c} A & B \\ \hline 0 & C \end{array} \right).$$

Set

$$E_t = \left( \begin{array}{c|c} A & tB \\ \hline 0 & C \end{array} \right).$$

- (c) For  $k = 1$ , construct the homotopy explicitly.
  - (d) Assume that  $k = 2$  and that  $E$  has only complex eigenvalues. Show that  $E_t = tI + (1 - t)E$  is always an isomorphism.
  - (e) Now induct on  $k$ .
2. Let  $A: V \rightarrow V$  be a linear map. Show that the following statements are equivalent:
    - (a) 0 is an isolated fixed point of  $A$ .
    - (b)  $A - I: V \rightarrow V$  is an isomorphism.
    - (c) 0 is a Lefschetz fixed point of  $A$ .
    - (d)  $A$  is a Lefschetz map.
  3. Prove that the following are equivalent conditions:
    - (a)  $x$  is a Lefschetz fixed point of  $f: X \rightarrow X$ .
    - (b) 0 is a Lefschetz fixed point of  $df_x: T_x(X) \rightarrow T_x(X)$ .
    - (c)  $df_x: T_x(X) \rightarrow T_x(X)$  is a Lefschetz map.
  4. Prove that the class of Lefschetz maps of a compact manifold  $X$  into itself are stable. [HINT:  $\overline{\cap}$  is a stable condition.]
  5. In the exercises of Section 2 you found some examples of submanifolds that cannot be globally defined by functions. However, these examples also involved orientation difficulties. Show that nonorientability is not the key phenomenon here, by proving that the diagonal  $\Delta$  in  $S^2 \times S^2$  is not globally definable by two independent functions. In contrast, show that the other standard copies of  $S^2$  in  $S^2 \times S^2$ —i.e.,  $S^2 \times \{a\}$ ,  $a \in S^2$ —are so definable. [HINT: Compare with Exercise 18, Chapter 2, Section 4.]
  6. Show that the map  $f(x) = 2x$  on  $\mathbf{R}^k$ , with a “source” at 0, has  $L_0(f) = 1$ . However, check that the “sink”  $f(x) = \frac{1}{2}x$  has  $L_0(f) = (-1)^k$ .
  7. Prove that

$$\chi(S^k) = \begin{cases} 2, & k \text{ even} \\ 0, & k \text{ odd} \end{cases}$$

[HINT: Use the map constructed for  $S^2$ , plus Exercise 6.]

8. Use the existence of Lefschetz maps for another proof that  $\chi(X \times Y) = \chi(X) \cdot \chi(Y)$ .

9. Show that summing local Lefschetz numbers (of maps having only finitely many fixed points) does not define a homotopy invariant without the compactness assumption. [HINT:  $\mathbf{R}^k$ , for example, is contractible.]
10. (a) Prove that map  $z \rightarrow z + z^m$  has a fixed point with local Lefschetz number  $m$  at the origin of  $\mathbf{C}$  ( $m > 0$ ).  
 (b) Show that for any  $c \neq 0$ , the homotopic map  $z \rightarrow z + z^m + c$  is Lefschetz, with  $m$  fixed points that are all close to zero if  $c$  is small.  
 (c) Show that the map  $z \rightarrow z + \bar{z}^m$  has a fixed point with local Lefschetz number  $-m$  at the origin of  $\mathbf{C}$  ( $m \geq 0$ ).
11. Show that the Euler characteristic of the orthogonal group (or any compact Lie group for that matter) is zero. [HINT: Consider left multiplication by an element  $A \neq I$  on  $O(n)$ .]

## §5 Vector Fields and the Poincaré-Hopf Theorem

A *vector field* on a manifold  $X$  in  $R^N$  is a smooth assignment of a vector tangent to  $X$  at each point  $x$ —that is, a smooth map  $\vec{v} : X \rightarrow R^N$  such that  $\vec{v}(x) \in T_x(X)$  for every  $x$ . From a local point of view, all the interesting behavior of  $\vec{v}$  occurs around its *zeros*, the points  $x \in X$  where  $\vec{v}(x) = 0$ . For if  $\vec{v}(x) \neq 0$ , then  $\vec{v}$  is nearly constant in magnitude and direction near  $x$  (Figure 3-17). However, when  $\vec{v}(x) = 0$ , the direction of  $\vec{v}$  may change radically in any small neighborhood of  $x$ . The field may circulate around  $x$ ; it may have a source, sink, or saddle; it may spiral in toward  $x$  or away; or it may form a more complicated pattern (Figure 3-18).

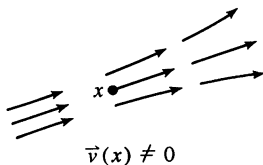


Figure 3-17

Try drawing a number of vector fields on several compact surfaces; first determine patterns around the zeros and then interpolate the remainder of the field smoothly. You will quickly discover that the topology of the manifold limits your possibilities. For example, on the sphere it is easy to create fields with exactly two zeros, as long as each zero is a sink, source, spiral, or circulation. A field with just one zero of type (f) is also readily found. None of these patterns exists on the torus. Similarly, patterns admissible on the

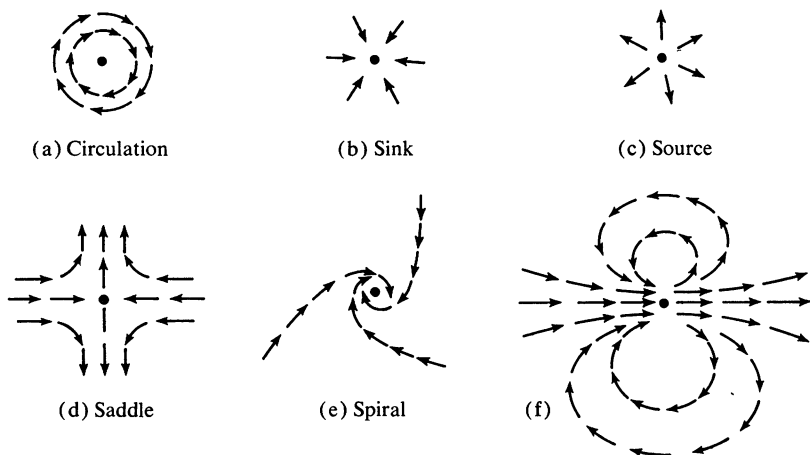


Figure 3-18

torus, like one saddle plus one source, are prohibited on the sphere. In particular, the torus has a vector field with no zeros, a situation that common experience shows is impossible on the sphere. (If your head were a doughnut, you could comb your hair without leaving a bald spot.)

In order to investigate the relation between  $\vec{v}$  and the topology of  $X$ , we must quantify the directional change of  $\vec{v}$  around its zeros. First, assume that we are in  $\mathbf{R}^k$  and that  $\vec{v}$  has an isolated zero at the origin. The direction of  $\vec{v}$  at a point  $x$  is just the unit vector  $\vec{v}(x)/|\vec{v}(x)|$ . Thus the directional variation of  $\vec{v}$  around 0 is measured by the map  $x \rightarrow \vec{v}(x)/|\vec{v}(x)|$  carrying any small sphere  $S_\epsilon$  around 0 into  $S^{k-1}$ . Choosing the radius  $\epsilon$  so small that  $\vec{v}$  has no zeros inside  $S_\epsilon$  except at the origin, we define the *index* of  $\vec{v}$  at 0,  $\text{ind}_0(\vec{v})$ , to be the degree of this directional map  $S_\epsilon \rightarrow S^{k-1}$ . As usual, the radius itself does not matter, for if  $\epsilon'$  is also suitable, then  $\vec{v}/|\vec{v}|$  extends to the annulus bounded by the two spheres.

In the two-dimensional case,  $\text{ind}_0(\vec{v})$  simply counts the number of times  $\vec{v}$  rotates completely while we walk counterclockwise around the circle; however, rotation of  $\vec{v}$  in the counterclockwise direction adds  $+1$ , whereas clockwise rotation contributes  $-1$ . Calculate the indices of the five vector fields in Figure 3-18 and compare with our answers. ( $a: +1, b: +1, c: +1, d: -1, e: +1, f: +2$ ).

In defining the index of vector fields at isolated zeros on arbitrary oriented manifolds, we use local parametrizations. Looking at the manifold locally, we see essentially a piece of Euclidean space, so we simply read off the index as if the vector field were Euclidean. The difficulty, of course, is the necessity of proving that the same index is obtained no matter which local parametrization we use for viewing.

Explicitly, suppose that  $\phi: U \rightarrow X$  is a local parametrization carrying the origin of  $\mathbf{R}^k$  to  $x$ . There is a natural way to pull back the vector field  $\vec{v}$  from  $X$  to make a vector field on the open subset  $U \subset \mathbf{R}^k$ . For each  $u \in U$ , the derivative  $d\phi_u$  is an isomorphism of  $\mathbf{R}^k$  with the tangent space of  $X$  at  $\phi(u)$ . We simply define the *pullback* vector field, denoted  $\phi^*\vec{v}$ , by assigning to  $u$  the vector that corresponds to the value of  $\vec{v}$  at  $\phi(u)$ :

$$\phi^*\vec{v}(u) = d\phi_u^{-1} \vec{v}(\phi(u)).$$

Now if  $\vec{v}$  has an isolated zero at  $x$ ,  $\phi^*\vec{v}$  has an isolated zero at the origin, and we define  $\text{ind}_x(\vec{v}) = \text{ind}_0(\phi^*\vec{v})$ . We shall avoid the tedium of proving that this definition does not depend on the choice of local parametrization with a trick, but later.

The nature of the topological limitation on  $\vec{v}$  is explained by the following celebrated theorem.

**Poincaré-Hopf Index Theorem.** If  $\vec{v}$  is a smooth vector field on the compact, oriented manifold  $X$  with only finitely many zeros, then the global sum of the indices of  $\vec{v}$  equals the Euler characteristic of  $X$ .

Knowing the Lefschetz theorem, we should have anticipated the appearance of the Euler characteristic, for vector fields are more than static collections of arrows on  $X$ —they are prescriptions for motion. Looking at a picture of a vector field, we tend to perceive a smooth fluid flow along the lines of the field. Each particle on the surface seems to travel a path that is always tangent to the field. The mathematical existence of this flow must be demonstrated by solving differential equations, technicalities that need not concern us. But, intuitively, think of allowing every particle on a compact manifold  $X$  to flow along the vector field  $\vec{v}$  for some time  $t$ . We thus obtain a transformation  $h_t$  of  $X$  whose fixed points, if  $t$  is small enough, are exactly the zeros of  $\vec{v}$ . The behavior of the flow around these points, and hence the pattern of  $\vec{v}$  near its zeros, is governed by the Lefschetz number of  $h_t$ . But as the elapsed time is shrunk to zero, the flow transformations  $h_t$  homotopically reduce to the identity transformation. Hence the Lefschetz number of the flow is nothing but the Euler characteristic of  $X$ .

We shall follow this approach in proving the Poincaré-Hopf theorem, except that we can avoid the differential equations by contenting ourselves with a crude approximation to the flow. It turns out that the only crucial property of the flow is that it is “tangent to the field at time zero,” a rather weak condition. Suppose that  $\{f_i\}$  is any homotopic family of transformations of  $X$  with  $f_0 = \text{identity}$ . We shall say that  $\{f_i\}$  is *tangent to the vector field*  $\vec{v}$  *at time zero* if, for each fixed  $x \in X$ , the vector  $\vec{v}(x)$  is tangent to the curve  $f_i(x)$  at time zero. We perform the basic local calculation in Euclidean space,



so assume that  $\vec{v}$  and the maps  $f_t$  are defined in a neighborhood  $U$  of the origin of  $\mathbf{R}^k$ .

**Proposition.** Suppose, for  $t \neq 0$ , that the maps  $f_t$  have no fixed point in the set  $U$  except the origin and that the vector field  $\vec{v}$  vanishes only at 0. If  $\{f_t\}$  is tangent to  $\vec{v}$  at time zero, then the local Lefschetz number of each  $f_t$  at 0 equals the index of  $\vec{v}$ :

$$\text{ind}_0(\vec{v}) = L_0(f_t).$$

This calculation needs nothing more than a simple application of the fundamental theorem of calculus.

**Lemma.** If  $g(t)$  is a smooth function of  $t$ , then there exists another smooth function  $r(t)$  such that

$$g(t) = g(0) + tg'(0) + t^2r(t).$$

*Proof.* First we show that  $g(t) = g(0) + tq(t)$ . Fix  $t$  and define  $h(s) = g(ts)$ . Then  $h'(s) = tg'(ts)$ , so

$$h(1) - h(0) = \int_0^1 h'(s) ds = t \int_0^1 g'(ts) ds.$$

Since

$$h(1) = g(t) \quad \text{and} \quad h(0) = g(0),$$

we need only set

$$q(t) = \int_0^1 g'(ts) ds.$$

Note that

$$q(0) = \int_0^1 g'(0) ds = g'(0).$$

Now we apply the same argument to the function  $q$ , obtaining  $q(t) = q(0) + tq(t)$ , so

$$g(t) = g(0) + tg'(0) + t^2r(t). \quad \text{Q.E.D.}$$

Note that if  $g$  depends smoothly on some other variables as well, then the function  $r$ , being defined by an integral formula, will also vary smoothly in the extra variables.

*Proof of Proposition.* For notational convenience, at any time  $t$  let  $f'_t(x)$  denote the vector whose coordinates are the derivatives at time  $t$  of the corresponding coordinates of  $f_t(x)$ ; thus the tangency of  $\{f_t\}$  to  $\vec{v}$  means  $f'_0(x) =$

$\vec{v}(x)$ . Fix  $x$  and apply the lemma to each coordinate of the function  $f_t(x)$ . We thus obtain a smooth vector-valued function  $r(t, x)$  such that

$$f_t(x) = f_0(x) + tf'_0(x) + t^2r(t, x),$$

or

$$f_t(x) = x + t\vec{v}(x) + t^2r(t, x).$$

Rearranging,

$$f_t(x) - x = t\vec{v}(x) + t^2r(t, x).$$

By assumption,

$$f_t(x) - x \neq 0 \quad \text{if } t \neq 0,$$

so we may divide each side of the equation by its norm:

$$\frac{f_t(x) - x}{|f_t(x) - x|} = \frac{\vec{v}(x) + tr(t, x)}{|\vec{v}(x) + tr(t, x)|}.$$

Now we let  $x$  vary on the sphere  $S_\epsilon$ . The degree of the map on the left is, by definition,  $L_0(f_t)$ . On the right, we have a homotopic family of maps  $S_\epsilon \rightarrow S^{k-1}$ , defined even at  $t = 0$ . In fact, at  $t = 0$  the right side becomes the map  $\vec{v}/|\vec{v}|$ , whose degree is  $\text{ind}_0(\vec{v})$ . Since degree is a homotopy invariant,

$$L_0(f_t) = \text{ind}_0(\vec{v}). \quad \text{Q.E.D.}$$

An immediate consequence of the proposition is the invariance of index under diffeomorphisms. First note that every vector field  $\vec{v}$  on  $\mathbf{R}^k$  has many suitable families  $\{f_t\}$  tangent to it at time zero. The local flow is one example, but so is its first-order approximation  $f_t(x) = x + t\vec{v}(x)$ . Now suppose that  $\vec{v}$  has an isolated zero at the origin, and let  $\phi$  be a diffeomorphism of neighborhoods of 0 with  $\phi(0) = 0$ . Then the chain rule implies that if  $\{f_t\}$  is tangent to  $\vec{v}$  at time zero,  $\{\phi^{-1} \circ f_t \circ \phi\}$  is tangent to  $\phi^*\vec{v}$ . But  $\phi$  is just a reparametrization of a neighborhood of the origin, and we know that local Lefschetz numbers do not depend on parametrization; thus

$$\text{ind}_0(\vec{v}) = L_0(f_t) = L_0(\phi^{-1} \circ f_t \circ \phi) = \text{ind}_0(\phi^*\vec{v}),$$

as claimed.

This observation also proves that the index is well defined on arbitrary manifolds. For if  $\phi_1$  and  $\phi_2$  are two local parametrizations of a neighborhood of the isolated fixed point  $x \in X$ , check that

$$\phi_1^*\vec{v} = (\phi_2^{-1} \circ \phi_1)^*[\phi_2^*\vec{v}].$$

Consequently,

$$\text{ind}_0(\phi_1^*\vec{v}) = \text{ind}_0(\phi_2^*\vec{v}).$$

*Proof of Poincaré-Hopf.* The remainder of the proof of the Poincaré-Hopf theorem is a construction. Given a vector field  $\vec{v}$  defined on the compact manifold  $X$ , we shall produce a globally defined family of maps  $f_t$  that is tangent to  $\vec{v}$  at time zero and that possesses two important properties: the fixed points of each  $f_t$  for  $t > 0$  are precisely the zeros of  $\vec{v}$ , and  $f_0$  is the identity map of  $X$ . From the proposition, it then follows that the sum of the indices of  $\vec{v}$  equals the sum of the local Lefschetz numbers of  $f_t$ —that is, the global Lefschetz number. But since  $f_0$  is the identity,  $L(f_t) = L(f_0)$  is the Euler characteristic of  $X$ .

Recall the  $\epsilon$ -Neighborhood Theorem. Let  $X^\epsilon$  denote the set of points in  $\mathbf{R}^N$  at distance less than  $\epsilon$  from  $X$ . Then if the positive number  $\epsilon$  is small enough, there is a normal projection map  $\pi: X^\epsilon \rightarrow X$  that restricts to the identity map of  $X$ . Now since  $X$  is compact, all the points  $x + t\vec{v}(x)$  lie inside  $X^\epsilon$ , as long as the parameter  $t$  is small enough and  $x \in X$ . Thus we may define

$$f_t(x) = \pi[x + t\vec{v}(x)].$$

If  $x$  is fixed, the tangent to the curve  $f_t(x)$  at time zero is  $d\pi_x \vec{v}(x)$ . But since  $\pi$  is the identity on  $X$ ,  $d\pi_x$  is the identity on  $T_x(X)$ . Thus  $d\pi_x \vec{v}(x) = \vec{v}(x)$ , and  $\{f_t\}$  is tangent to  $\vec{v}$  at  $t = 0$ .

Obviously, the zeros of  $\vec{v}$  are fixed points of  $f_t$ , and, finally, we need only show that there are no other fixed points. Remember that, by construction,  $\pi$  is a normal projection map; if  $\pi(z) = x$ , then  $z - x$  must be a vector perpendicular to  $X$ . But then if  $x$  is a fixed point of  $f_t(x)$ ,  $\pi(x + t\vec{v}(x)) = x$  so  $t\vec{v}(x) \perp T_x(X)$ . Since  $\vec{v}(x) \in T_x(X)$ , it can be perpendicular only if it is zero. Q.E.D.

From the vantage point of intersection theory, we can get a more sophisticated, and thus simpler, view of Poincaré-Hopf. We have remarked (Exercise 6, Chapter 1, Section 8) that a vector field  $\vec{v}$  on  $X$  is a cross section of the tangent bundle  $T(X)$ , so  $\text{graph}(\vec{v})$  is a copy of  $X$  in  $T(X)$ . The “identically zero” vector field gives the natural embedding  $X_0$  of  $X$  in  $T(X)$ . Thus zeros of  $\vec{v}$  correspond exactly to points of  $X_0 \cap \text{graph}(\vec{v})$ . No inspiration is now needed to decide how to measure globally the zeros of  $\vec{v}$ ; the answer has to be  $I(X_0, \text{graph}(\vec{v}))$ . See Figure 3-19. If  $X_0$  in Figure 3-19 is twisted up to the  $45^\circ$  line, you may recognize it as the infinitesimal version of Figure 3-20, where  $\{f_t\}$  is tangent to  $\vec{v}$  at time zero. In order to get a Lefschetz-free proof of the Poincaré-Hopf theorem, we must: (1) exploit the similarity of the pictures to show that  $I(X_0, \text{graph}(\vec{v}))$  equals  $I(\Delta, \text{graph}(f_t))$ , which, by homotopy invariance, is  $I(\Delta, \Delta) = \chi(X)$ ; and (2) localize  $I(X_0, \text{graph}(\vec{v}))$ —that is, once the zeros of  $\vec{v}$  are isolated, recognize  $\text{ind}_x(\vec{v})$  as the contribution of each zero  $x$ .

This scheme is carried out in Exercises 6 through 9.

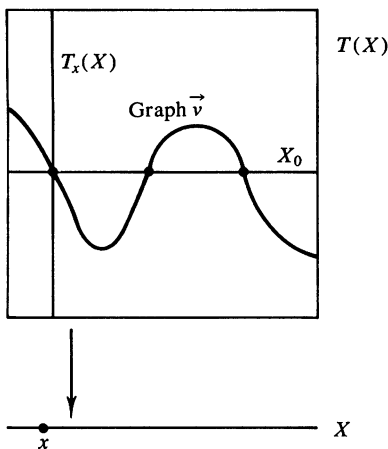


Figure 3-19

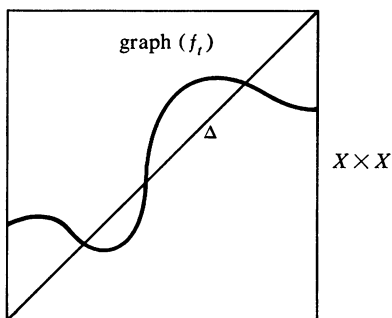


Figure 3-20

## EXERCISES

1. Let  $\vec{v}$  be the vector field on  $\mathbf{R}^2$  defined by  $\vec{v}(x, y) = (x, y)$ . Show that the family of diffeomorphisms  $h_t: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ , defined by  $h_t(z) = tz$ , is the flow corresponding to  $\vec{v}$ . That is, if we fix any  $z$ , then the curve  $t \rightarrow h_t(z)$  is always tangent to  $\vec{v}$ ; its tangent vector at any time  $t$  equals  $\vec{v}(h_t(z))$ . Draw a picture of  $\vec{v}$  and its flow curves. Compare  $\text{ind}_0(\vec{v})$  with  $L_0(h_t)$ .
2. Now let  $\vec{v}(x, y) = (-y, x)$ . Show that the flow transformations are the linear rotation maps  $h_t: \mathbf{R}^2 \rightarrow \mathbf{R}^2$  with matrix

$$\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

Draw  $\vec{v}$  and its flow curves. Also compare  $\text{ind}_0(\vec{v})$  with  $L_0(h_t)$ .

- \*3. Recall that a vector field  $\vec{v}$  on a manifold  $X$  in  $\mathbf{R}^N$  is a particular type of map  $\vec{v}: X \rightarrow \mathbf{R}^N$ . Show that at a zero  $x$ , the derivative  $d\vec{v}_x: T_x(X) \rightarrow \mathbf{R}^N$  actually carries  $T_x(X)$  into itself. [HINT: If  $X = \mathbf{R}^k \times \{0\}$ , then the claim is obvious. But you can always reduce to this case by suitable local parametrizations around  $x$  in  $\mathbf{R}^N$ .]
4. Let  $f_t: X \rightarrow X$  be the map constructed in the proof of Poincaré-Hopf,

$$f_t(x) = \pi(x + t\vec{v}(x)).$$

Prove that at a zero  $x$  of  $\vec{v}$ ,  $d(f_t)_x = I - t d\vec{v}_x$  as linear maps of  $T_x(X)$  into itself. ( $I$  is the identity.) [HINT:  $\pi$  restricted to  $X$  is the identity, so what about  $(d\pi)_x$  on  $T_x(X)$ ?]

- \*5. A zero  $x$  of  $\vec{v}$  is *nondegenerate* if  $d\vec{v}_x: T_x(X) \rightarrow T_x(X)$  is bijective. Prove that nondegenerate zeros are isolated. Furthermore, show that at a nondegenerate zero  $x$ ,  $\text{ind}_x(\vec{v}) = +1$  if the isomorphism  $d\vec{v}_x$  preserves orientation, and  $\text{ind}_x(\vec{v}) = -1$  if  $d\vec{v}_x$  reverses orientation. [HINT: Deduce from Exercise 4 that  $x$  is a nondegenerate zero of  $\vec{v}$  if and only if it is a Lefschetz fixed point of  $f_t$ .]
6. A vector field  $\vec{v}$  on  $X$  naturally defines a cross-sectional map  $f_{\vec{v}}: X \rightarrow T(X)$ , by  $f_{\vec{v}}(x) = (x, \vec{v}(x))$ .
- Show that  $f_{\vec{v}}$  is an embedding, so its image  $X_{\vec{v}}$  is a submanifold of  $T(X)$  diffeomorphic to  $X$ .
  - What is the tangent space of  $X_{\vec{v}}$  at a point  $(x, \vec{v}(x))$ ?
  - Note that the zeros of  $\vec{v}$  correspond to the intersection points of  $X_{\vec{v}}$  with  $X_0 = \{(x, 0)\}$ . Check that  $x$  is a nondegenerate zero of  $\vec{v}$  if and only if  $X_{\vec{v}} \cap X_0$  at  $(x, 0)$ .
  - If  $x$  is a nondegenerate zero of  $\vec{v}$ , show that  $\text{ind}_x(\vec{v})$  is the orientation number of the point  $(x, 0)$  in  $X_0 \cap X_{\vec{v}}$ .
7. Zeros of vector fields may be “split” by perturbations just like fixed points of maps. Suppose that  $x$  is a zero of  $\vec{v}$  and  $U$  is a neighborhood of  $x$  in  $X$  containing no other zero of  $\vec{v}$ . Prove that there exists a vector field  $\vec{v}_1$  that agrees with  $\vec{v}$  outside some compact subset of  $U$  and that has only nondegenerate fixed points inside  $U$ . [HINT: Reduce to the case where  $U$  is an open subset of  $\mathbf{R}^k$ . Pick a function  $\rho$  that is 1 near  $x$  and 0 outside a compact subset of  $U$ , and define

$$\vec{v}_1(z) = \vec{v}(z) + \rho(z)\vec{a}, \quad \vec{a} \in \mathbf{R}^k.$$

Show that if  $\vec{a}$  is small enough,  $\vec{v}_1$  can be zero only where  $\rho \equiv 1$ . Now demand that  $-\vec{a}$  be a regular value of  $\vec{v}$ .]

8. You proved in Exercise 18, Chapter 2, Section 3 that there is a diffeomorphism of a neighborhood of  $X_0$  in  $T(X)$  with a neighborhood of the diagonal  $\Delta$  in  $X \times X$ , extending the natural diffeomorphism  $X_0 \rightarrow \Delta$  defined by  $(x, 0) \rightarrow (x, x)$ . Deduce from this that  $I(X_0, X_0) = I(\Delta, \Delta)$ .
9. Show that if  $\vec{v}$  is any vector field on  $X$ , then  $X_{\vec{v}}$  may be smoothly deformed into  $X_0$ . Use this result, together with the preceding three problems, to produce an alternate proof of the Poincaré-Hopf theorem.
10. Let  $V$  be a vector subspace of  $\mathbf{R}^N$ , and let  $v \in V$ . Define a linear functional  $\psi_v$  on  $V$  by  $\psi_v(w) = v \cdot w$ . Prove that the mapping  $v \rightarrow \psi_v$  defines an isomorphism of  $V$  with its dual space  $V^*$ .
11. Let  $f$  be a real-valued function on the manifold  $X$  in  $\mathbf{R}^N$ . For each  $x \in X$ ,  $df_x : T_x(X) \rightarrow \mathbf{R}$  defines a linear functional on  $T_x(X)$ . According to Exercise 10, we may write  $df_x(w) = \vec{v}(x) \cdot w$  for some vector  $\vec{v}(x) \in T_x(X)$ . This vector field  $\vec{v}$  is called the *gradient field* of  $f$ ,  $\vec{v} = \overrightarrow{\text{grad}}(f)$ . For the special case  $X = \mathbf{R}^k$ , show that the gradient has the following simple representation in standard coordinates:

$$\overrightarrow{\text{grad}}(f) = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_k} \right).$$

12. Let  $\phi : U \rightarrow X$  be a local parametrization of the  $k$ -manifold  $X$  in  $\mathbf{R}^N$ , and we shall compute the vector field  $\phi^* \overrightarrow{\text{grad}}(f)$  on  $U$ , when  $f$  is a function on  $X$ . Let  $\{e_1, \dots, e_k\}$  be the standard basis for  $\mathbf{R}^k$ , and define smooth functions  $g_{ij}$  on  $U$  by the formula

$$g_{ij}(u) = d\phi_u(e_i) \cdot d\phi_u(e_j).$$

Now prove that

$$\phi^* \overrightarrow{\text{grad}}(f) = \sum_{i,j=1}^k \frac{\partial(f \circ \phi)}{\partial x_i} g_{ij} e_j$$

as a vector field on  $U$ .

13. Conclude from the preceding exercise that if  $f$  is a smooth function on  $X$ , then  $\overrightarrow{\text{grad}}(f)$  is a smooth vector field.
14. (a) Prove that the zeros of the vector field  $\overrightarrow{\text{grad}}(f)$  are just the critical points of  $f$ . [HINT: The functions  $g_{ij}$  in Exercise 12 never vanish on  $U$ .]  
 (b) Prove that  $x$  is a nondegenerate zero of  $\overrightarrow{\text{grad}}(f)$  if and only if it is a nondegenerate critical point of  $f$ . [HINT: Use Exercise 12 to show

that if  $\phi(u)$  is a zero of  $\overrightarrow{\text{grad}}(f)$ , then the derivative at  $u$  of  $\phi^* \overrightarrow{\text{grad}}(f)$  has as matrix the product of the Hessian of  $f \circ \phi$  at  $u$  with the matrix  $(g_{ij}(u))$ . Show that  $(g_{ij}(u))$  is positive definite to conclude that  $\det(g_{ij}(u)) > 0$ .]

15. As an exercise in overkill, use Exercise 14 to show that on every manifold  $X$  there exists a vector field with only isolated zeros. (For a more direct proof, see Exercise 11, Chapter 3, Section 6.)
16. If  $x$  is a nondegenerate critical point of a function  $f$  on  $X$ , define  $\text{ind}_x(f)$  to be the sign of the determinant of the Hessian of  $f$  at  $x$ , as computed in any coordinate system. Show that this is well defined (i.e., it is the same in all coordinate systems). [HINT: You need not compute anything if you use Exercise 14.]
17. Let  $X$  be a compact manifold and  $f$  a Morse function on  $X$ . Prove that the sum of the indices of  $f$  at its fixed points equals the Euler characteristic of  $X$ .
18. Suppose that  $\vec{v}$  is a vector field with isolated zeros in  $\mathbf{R}^k$ , and  $W$  is a compact  $k$ -dimensional submanifold of  $\mathbf{R}^k$  with boundary. Assume that  $\vec{v}$  is never zero in  $\partial W$ . Prove that the sum of the indices of  $\vec{v}$  at its zeros inside  $W$  equals the degree of the map

$$\frac{\vec{v}}{|\vec{v}|} : \partial W \longrightarrow S^k.$$

[HINT: Delete balls around the zeros and use our standard argument.]

19. (a) Vector fields on  $\mathbf{R}^k$  may be simply identified with maps of  $\mathbf{R}^k$  into itself. In particular, the map  $z \rightarrow z^m$  on  $\mathbf{C} = \mathbf{R}^2$  defines a vector field on  $\mathbf{R}^2$  with a zero at the origin. Check that this zero has index  $m$ .
- (b) Check that the index of the vector field  $z \rightarrow \bar{z}^m$  on  $\mathbf{C} = \mathbf{R}^2$ , at the origin, is  $-m$ . [HINT:  $z \rightarrow \bar{z}$  is an orientation-reversing reflection of  $\mathbf{R}^2$ .]

## §6 The Hopf Degree Theorem

Intersection theory has shown us how to make sense of the primitive notion that a map  $f: X \rightarrow Y$  be  $n$ -to-1, that it map  $n$  points of  $X$  to each point of  $Y$ . Although very few maps behave simply enough to be described by an integer  $n$  in such a naïve sense, we need only learn how to count

correctly in order to salvage the concept for general smooth maps. For if we keep tally by using orientations, then at least almost every point of  $Y$  is “hit  $n$  times.” (Here we limit ourselves to connected, compact, oriented manifolds of the same dimension.)

Now the integer  $n$ —the degree of  $f$ —has the admirable property of homotopy invariance: it is the same for all homotopic deformations of  $f$ . In this section you will prove the remarkable fact that when  $Y$  is a sphere, this simple integer completely determines the homotopy relationship. That is, two maps  $f_0, f_1 : X \rightarrow S^k$  are homotopic if and only if they have the same degree. The theorem is due to Hopf, and it will be proved through a series of exercises. Since these exercises are somewhat more ambitious than the earlier ones on Jordan-Brouwer, detailed hints are supplied at the end. (Of course, we hope you will enjoy constructing the proof without submitting to the temptation of reading the hints.) The techniques are very similar to the last few sections, so this series should serve as a good review.

For the proof, we will need a new topological term. An *isotopy* is a homotopy in which each map  $h_t$  is a diffeomorphism, and two diffeomorphisms are *isotopic* if they can be joined by an isotopy. An isotopy is *compactly supported* if the maps  $h_t$  are all equal to the identity map outside some fixed compact set.

**Isotopy Lemma.** Given any two points  $y$  and  $z$  in the connected manifold  $Y$ , there exists a diffeomorphism  $h : Y \rightarrow Y$  such that  $h(y) = z$  and  $h$  is isotopic to the identity. Moreover, the isotopy may be taken to be compactly supported.

*Proof.* If the statement is true for two points  $y$  and  $z$ , we shall call them “isotopic.” It is quite evident that this defines an equivalence relation on  $Y$ . We shall show that each equivalence class is an open set. Then  $Y$  is the disjoint union of open sets, so connectivity implies that there can be only one equivalence class, which proves the theorem.

In order to prove that the equivalence classes are open, we will construct an isotopy  $h_t$  on  $\mathbf{R}^k$  such that  $h_0$  is the identity, each  $h_t$  is the identity outside some specified small ball around 0, and  $h_1(0)$  is any desired point sufficiently close to 0. You can easily finish the proof from there: parametrize a neighborhood of  $y$  and use  $h_t$  to slide  $y$  around inside a small ball; each  $h_t$  extends smoothly to a diffeomorphism of  $Y$  by defining it to be the identity outside the ball. Thus  $y$  is isotopic to every sufficiently nearby point, proving openness.

Construct  $h_t$  first on  $\mathbf{R}^1$ . Given any  $\epsilon > 0$ , let  $\rho$  be a smooth function that vanishes outside  $(-\epsilon, \epsilon)$  and equals 1 at 0. If  $z \in \mathbf{R}^1$ , define

$$h_t(x) = x + t\rho(x)z.$$

Then  $h_t(x) = x$  if  $x \in (-\epsilon, \epsilon)$  or if  $t = 0$ , and  $h_1(0) = z$ . Is  $h_t$  an isotopy?



Compute derivatives:  $h'_t = 1 + t\rho'(x)z$ . Since  $|\rho'(x)|$  vanishes outside a compact set, it must be bounded. Hence as long as  $|z|$  is small enough,  $|t\rho'(x)z| < 1$  for all  $t \in [0, 1]$ ,  $x \in \mathbf{R}^1$ , so  $h'_t(x) > 0$ . Consequently, each  $h_t$  is strictly increasing, and the Inverse Function Theorem implies that its inverse function is also smooth. So each  $h_t$  is a diffeomorphism of  $\mathbf{R}^1$  as long as  $z$  is close enough to 0.

In  $\mathbf{R}^k$  we simply adapt this proof. Given any point in  $\mathbf{R}^k$ , we may rotate the coordinate axes so that the point sits on the first axis. So if we write  $R^k = \mathbf{R}^1 \times \mathbf{R}^{k-1}$ , we need only exhibit the desired isotopy carrying the origin to nearby points of the form  $(z, 0)$ . Choose a function  $\sigma$  on  $\mathbf{R}^{k-1}$  that is one at the origin and zero outside some small ball of radius  $\delta$ . Define  $h_t$  on  $\mathbf{R}^k$  as follows: for  $(x, y) \in \mathbf{R}^1 \times \mathbf{R}^{k-1}$ , let

$$h_t(x, y) = (x + t\sigma(y)\rho(x)z, y).$$

Then  $h_t(x, y) = (x, y)$  unless  $|x| < \epsilon$ ,  $|y| < \delta$ , and  $t > 0$ . Also,  $h_1(0, 0) = (z, 0)$ , as desired. We must show each  $h_t$  to be a diffeomorphism of  $\mathbf{R}^k$  when  $|z|$  is small. It is one-to-one and onto, for with  $|z|$  small, it induces a diffeomorphism on each line  $y = \text{constant}$ . Moreover, its derivative at the point  $(x, y) \in \mathbf{R}^1 \times \mathbf{R}^{k-1}$  has matrix

$$\left( \begin{array}{c|ccc} 1 + t\sigma(y)\rho'(x)z & a_1 & \dots & a_{k-1} \\ \hline 0 & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 0 & & & \end{array} \right),$$

$I$

where  $I$  is the  $(k-1) \times (k-1)$  identity matrix. When  $|z|$  is small, the left-corner entry is always positive, so the matrix always has positive determinant. The Inverse Function Theorem now implies that the inverse of  $h_t$  is smooth. Q.E.D.

**Corollary.** Suppose that  $Y$  is a connected manifold of dimension greater than 1, and let  $y_1, \dots, y_n$  and  $z_1, \dots, z_n$  be two sets of distinct points in  $Y$ . Then there exists a diffeomorphism  $h: Y \rightarrow Y$ , isotopic to the identity, with

$$h(y_1) = z_1, \dots, h(y_n) = z_n.$$

Moreover, the isotopy may be taken to be compactly supported.

*Proof.* Argue inductively, for the theorem gives  $n = 1$ . Assuming the corollary for  $n-1$ , obtain a compactly supported isotopy  $h'_t$  of the punctured manifold  $Y - \{y_n, z_n\}$  such that  $h'_1(y_i) = z_i$  for  $i < n$  and  $h'_0 = \text{identity}$ .

Here we use the hypothesis  $\dim Y > 1$ , which guarantees that the punctured manifold is connected. (Proof? Use arc-connectedness.) The compact-support provision implies that the  $h'_i$  are all equal to the identity near  $y_n$  and  $z_n$ , so they extend to diffeomorphisms of  $Y$  that fix those two points.

Similarly, applying the theorem to the punctured manifold

$$Y - \{y_1, \dots, y_{n-1}, z_1, \dots, z_{n-1}\},$$

we obtain a compactly supported isotopy  $h''_i$  on  $Y$  such that  $h''_1(y_n) = z_n$ ,  $h''_0 = \text{identity}$ , and all  $h''_i$  fix the points  $y_i$  and  $z_i$ ,  $i < n$ . So  $h_i = h''_i \circ h'_i$  is the required isotopy. Q.E.D.

Now we begin the proof of Hopf's theorem. It is convenient here to reintroduce explicitly the concept of winding number, which you probably noticed lurking beneath the definitions both of local Lefschetz numbers and of indices of vector fields. If  $X$  is a compact oriented,  $l$ -dimensional manifold and  $f: X \rightarrow \mathbf{R}^{l+1}$  is a smooth map, the *winding number* of  $f$  around any point  $z \in \mathbf{R}^{l+1} - f(X)$  is defined just as for mod 2 theory. Simply construct the direction map  $u: X \rightarrow S^l$  by

$$u(x) = \frac{f(x) - z}{|f(x) - z|},$$

and set  $W(f, z) = \deg(u)$ . The first exercise determines how local diffeomorphisms can wind, and the second uses this information to count preimages.

1. Let  $f: U \rightarrow \mathbf{R}^k$  be any smooth map defined on an open subset  $U$  of  $\mathbf{R}^k$ , and let  $x$  be a regular point, with  $f(x) = z$ . Let  $B$  be a sufficiently small closed ball centered at  $x$ , and define  $\partial f: \partial B \rightarrow \mathbf{R}^k$  to be the restriction of  $f$  to the boundary of  $B$ . Prove that  $W(\partial f, z) = +1$  if  $f$  preserves orientation at  $x$  and  $W(\partial f, z) = -1$  if  $f$  reverses orientation at  $x$ .
2. Let  $f: B \rightarrow \mathbf{R}^k$  be a smooth map defined on some closed ball  $B$  in  $\mathbf{R}^k$ . Suppose that  $z$  is a regular value of  $f$  that has no preimages on the boundary sphere  $\partial B$ , and consider  $\partial f: \partial B \rightarrow \mathbf{R}^k$ . Prove that the number of preimages of  $z$ , counted with our usual orientation convention, equals the winding number  $W(\partial f, z)$ .

Another necessary ingredient will be the following straightforward extension argument.

3. Let  $B$  be a closed ball in  $\mathbf{R}^k$ , and let  $f: \mathbf{R}^k - \text{Int}(B) \rightarrow Y$  be any smooth map defined outside the open ball  $\text{Int}(B)$ . Show that if the restriction

$\partial f: \partial B \rightarrow Y$  is homotopic to a constant, then  $f$  extends to a smooth map defined on all of  $\mathbf{R}^k$  into  $Y$ .

With these tools, we now direct our attention toward the Hopf theorem. The first step will be to establish a special case of the theorem. Subsequently, we shall show that the full theorem derives easily from the partial one.

**Special Case.** Any smooth map  $f: S^l \rightarrow S^l$  having degree zero is homotopic to a constant map.

4. Check that the special case implies the following corollary.

**Corollary.** Any smooth map  $f: S^l \rightarrow \mathbf{R}^{l+1} - \{0\}$  having winding number zero with respect to the origin is homotopic to a constant.

Prove the special case inductively. For  $l = 1$ , you have already established Hopf's theorem, as Exercise 9, Section 3. So assume that the special case is true for  $l = k - 1$ , and extend it to  $l = k$ . The following exercise, which is the heart of the inductive argument, uses the corollary to pull maps away from the origin.

5. Let  $f: \mathbf{R}^k \rightarrow \mathbf{R}^k$  be a smooth map with 0 as a regular value. Suppose that  $f^{-1}(0)$  is finite and that the number of preimage points in  $f^{-1}(0)$  is zero when counted with the usual orientation convention. Assuming the special case in dimension  $k - 1$ , prove that there exists a mapping  $g: \mathbf{R}^k \rightarrow \mathbf{R}^k - \{0\}$  such that  $g = f$  outside a compact set.

An obvious point worth mentioning before the next step is that because  $f$  and  $g$  are equal outside a compact set, the homotopy  $tf + (1 - t)g$  is constant outside a compact set. This fact noted, you need only devise a suitable method for reducing  $S^k$  to  $\mathbf{R}^k$  in order to complete the special case.

6. Establish the special case in dimension  $k$ .

Hopf's theorem in its full generality is essentially a particular instance of the following theorem.

**Extension Theorem.** Let  $W$  be a compact, connected, oriented  $k + 1$  dimensional manifold with boundary, and let  $f: \partial W \rightarrow S^k$  be a smooth map. Prove that  $f$  extends to a globally defined map  $F: W \rightarrow S^k$ , with  $\partial F = f$ , if and only if the degree of  $f$  is zero.

In turn, the Extension Theorem follows from the special case, but to prove it you must first step aside and establish a lemma concerning the topological triviality of Euclidean space.

7. Let  $W$  be any compact manifold with boundary, and let  $f: \partial W \rightarrow \mathbf{R}^{k+1}$  be any smooth map whatsoever. Prove that  $f$  may be extended to all of  $W$ .
8. Prove the Extension Theorem.
9. Conclude:

**The Hopf Degree Theorem.** Two maps of a compact, connected, oriented  $k$ -manifold  $X$  into  $S^k$  are homotopic if and only if they have the same degree.

As an illustration of the use of Hopf's Degree Theorem, we shall answer a classical question: which compact manifolds possess nonvanishing vector fields? The Poincaré-Hopf Index Theorem supplies a necessary condition, the nullity of the Euler characteristic. We shall show that this requirement is the only one.

**Theorem.** A compact, connected, oriented manifold  $X$  possesses a nowhere-vanishing vector field if and only if its Euler characteristic is zero.

The idea of the proof is to reverse the splitting argument introduced in the Lefschetz section and amalgamate the zeros of some arbitrary vector field. The Hopf Degree Theorem appears in the Euclidean case.

10. Let  $\vec{v}$  be a vector field on  $\mathbf{R}^n$  with finitely many zeros, and suppose that the sum of the indices of its zeros is 0. Show that there exists a vector field that has no zeros, yet equals  $\vec{v}$  outside a compact set.

To apply this on  $X$ , we merely need to produce a suitable vector field.

11. Show that on any compact manifold  $X$  there exists a vector field with only finitely many zeros.
12. In fact, let  $U$  be any open set on the compact, connected manifold  $X$ . Verify the existence of a vector field having only finitely many zeros, all of which lie in  $U$ .
13. Prove the theorem.

**Hints** (listed by exercise numbers)

1. For simplicity, take  $x = 0 = z$ , and set  $A = df_0$ . By regularity,  $A$  is bijective. Write  $f(x) = Ax + \epsilon(x)$ , where  $\epsilon(x)/|x| \rightarrow 0$  as  $|x| \rightarrow 0$ . Show that  $W(A, 0) = W(df, 0)$  if  $B$  is small enough; then use the linear isotopy lemma of Section 4 (page 128).

2. Circumscribe small balls  $B_i$  around each preimage point, and show that the degree of the directional map  $u$  on the boundary of  $B' = B - \bigcup_i B_i$  is zero (since  $u$  extends to all of  $B'$ ). Now use Exercise 1.
3. Assume that  $B$  is centered at 0, and let  $g_t: \partial B \rightarrow Y$  be a homotopy with  $g_1 = \partial f$  and  $g_0 = \text{constant}$ . Then a continuous extension of  $f$  through  $B$  is given by  $f(tx) = f_t(x)$ ,  $x \in \partial B$  and  $t \in [0, 1]$ . To extend smoothly, just modify this with our standard trick of Exercise 1, Chapter 1, Section 6.
4. The hypothesis is that the degree of  $f/|f|$  is zero, so  $f/|f|$  is homotopic to a constant. But as maps into  $\mathbf{R}^{k+1} - \{0\}$ ,  $f$  and  $f/|f|$  are homotopic.
5. Take a large ball  $B$  around the origin that contains all of  $f^{-1}(0)$ . Use Exercise 2 to show that  $\partial f: \partial B \rightarrow \mathbf{R}^k - \{0\}$  has winding number zero. The inductive hypothesis implies that  $\partial f: \partial B \rightarrow \mathbf{R}^k - \{0\}$  is homotopic to a constant, so Exercise 3 completes the proof.
6. Pick distinct regular values  $a$  and  $b$  for  $f$ . Apply the Isotopy Lemma to  $S^k - f^{-1}(b)$  to find an open neighborhood  $U$  of  $f^{-1}(a)$  that is diffeomorphic to  $\mathbf{R}^k$  and that satisfies the requirement  $b \notin f(U)$ . Let  $\alpha: \mathbf{R}^k \rightarrow U$  be a diffeomorphism, and choose another diffeomorphism  $\rho: S^k - \{b\} \rightarrow \mathbf{R}^k$  that maps  $a$  to 0. Now apply Exercise 5 to  $\beta \circ f \circ \alpha$  to find a map  $g: S^k \rightarrow S^k - \{b\}$  that is homotopic to  $f$ . But since  $S^k - \{b\}$  is diffeomorphic to  $\mathbf{R}^k$ , and thus contractible,  $g$  is homotopic to a constant.
7.  $W$  sits in some  $\mathbf{R}^N$ . Apply the  $\epsilon$ -Neighborhood Theorem for  $\partial W$  to extend  $f$  to a smooth map  $F$  defined on a neighborhood  $U$  of  $\partial W$  in  $\mathbf{R}^N$ . Let  $\rho$  be a smooth function that equals one on  $\partial W$  and zero outside a compact subset of  $U$ . Now extend  $f$  to all of  $\mathbf{R}^N$  by setting it equal to  $\rho F$  on  $U$  and 0 outside.
8. First use Exercise 7 to extend  $f$  to a map  $F: W \rightarrow \mathbf{R}^{k+1}$ . By the Transversality Extension Theorem, we may assume 0 to be a regular value of  $F$ . Use the Isotopy Lemma to place the finite point set  $F^{-1}(0)$  inside a subset  $U$  of  $\text{Int}(W)$  that is diffeomorphic to  $\mathbf{R}^{k+1}$ . Let  $B$  be a ball in  $U$  containing  $F^{-1}(0)$ , and show that  $\partial F: \partial B \rightarrow \mathbf{R}^{k+1} - \{0\}$  has winding number 0 with respect to the origin. (For this, note that  $F/|F|$  extends to the manifold  $W' = W - \text{Int}(B)$ . But we already know that its degree on  $\partial W$  is zero. Now apply the corollary of the special case.)
9. Given  $f_0, f_1: X \rightarrow S^k$ , set  $W = X \times I$ , and define the map  $f: \partial W \rightarrow S^k$  to be  $f_0$  on  $X \times \{0\}$  and  $f_1$  on  $X \times \{1\}$ . Invoke Exercise 8.
10. Compare with Exercise 5.

11. Assume that  $X \subset \mathbf{R}^N$ , and let  $T(X)$  be its tangent bundle. Define  $\rho: X \times \mathbf{R}^N \rightarrow T(X)$  by making  $\rho(x, v)$  be the orthogonal projection of the vector  $v$  into  $T_x(X)$ . Check that  $\rho$  is a submersion, and apply the Transversality Theorem with  $S = \mathbf{R}^N$ ,  $Y = T(X)$ , and  $Z = X \times \{0\}$ . Conclude that, for some  $v$ , the vector field  $x \rightarrow \rho(x, v)$  is transversal to  $X \times \{0\}$ . (An alternative proof, based on the notion of gradient vector fields, was suggested in Exercise 15, Section 5.)
12. Use the corollary to the Isotopy Lemma. (A procedure for pulling vector fields back with diffeomorphisms was demonstrated in Section 5.)
13. Use Exercises 11 and 12 to get a vector field on  $X$  with finitely many zeros, all contained in an open set  $U$  that is diffeomorphic to  $\mathbf{R}^k$ . The Poincaré-Hopf theorem implies that the sum of the indices is zero, so pull the field back to  $\mathbf{R}^k$  and apply Exercise 10.

## §7 The Euler Characteristic and Triangulations

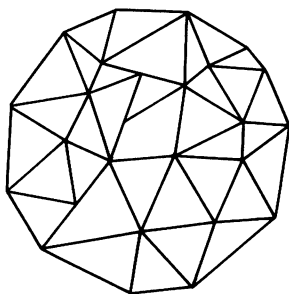
Although the original definition of the Euler characteristic as a self-intersection number may have appeared strange and unmotivated, we have already discovered that it expresses fundamental topological information. The vector fields and mappings admitted by a compact oriented manifold are strictly limited by its Euler characteristic. Yet this invariant is really far more primitive than you might expect; however, in order to see this point, we must reexamine manifolds from quite a different perspective.

One way to try to understand the anatomy of an object is to dissect it into familiar pieces, then examine the way the pieces fit together. We might, for instance, slice a surface  $X$  into “polygons.”

This section is to be strictly informal and intuitive. There is no need to prescribe the technical conditions for allowable partitions of  $X$ . Simply break up  $X$  into pieces that look like curved copies of standard plane polygons—triangles, squares, and so on. (See Figure 3-21.) The most common procedure is to use only triangles, so such dissections are called *triangulations*.

Now just count the parts. Let  $F$  denote the number of polygonal faces,  $E$  the number of edges, and  $V$  the number of vertices on our tessellated surface. If you draw a few examples on the sphere—and you are as brilliantly observant as Descartes or Euler (or if you remember your high school geometry class)—you might notice that the following relation is, in each case, satisfied:  $F - E + V = 2$ .

The constancy of this alternating sum is not a property unique to the sphere. In fact, however we triangulate a compact surface  $X$ ,  $F - E + V$  will always equal a constant characteristic of the surface, its Euler char-



A polygonal sphere

Figure 3-21

acteristic. Proving this theorem is not difficult for us, as long as we are not careful about details. It is easy to convince yourself that there exists a vector field on  $X$  that has a source in each polygonal face, a saddle on each edge, a sink at each vertex, and no other zeros. As the indices of sources, saddles, and sinks are, respectively,  $+1$ ,  $-1$ ,  $+1$ , we obtain the theorem  $F - E + V = \chi(X)$  from the Poincaré-Hopf theorem. The idea of the construction is simply to place the appropriate zeros on  $X$  and then assert with conviction that it is obvious that the remainder of the field can be smoothly interpolated. The picture on one triangular face should be convincing (Figure 3-22).

The same proof outline is valid in higher dimensions. If  $X$  is  $k$  dimensional, then one uses  $k$ -dimensional generalizations of polygons and obtains the Euler characteristic of  $X$  as the alternating sum

$$\sum_{j=0}^k (-1)^j \cdot (\text{number of "faces of dimension } j").$$

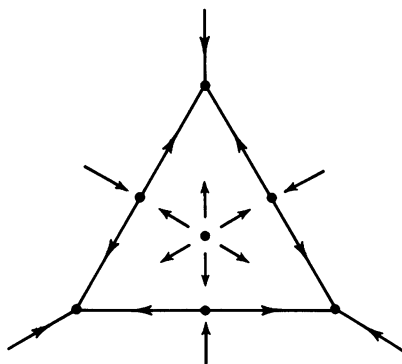


Figure 3-22

The important point to recognize is how elementary and unsophisticated this powerful invariant truly is!

### EXERCISES

1. Show that if the Euler sum  $F - E + V$  equals the Euler characteristic for triangulations that actually consist of triangles, then it does so as well for “dissections” admitting arbitrary polygons.
2. Calculate the Euler characteristics of the sphere  $S^2$  and the torus by using triangulations.
3. Think of the surface of genus  $k$  as a sphere with  $k$  tubes sewn in. Calculate its Euler characteristic by triangulating. [HINT: Remove  $2k$  non-adjacent faces from  $S^2$ ; what happens to the Euler sum? Triangulate a closed cylinder  $S^1 \times [0, 1]$  and check that the Euler sum is zero. Now sew it into two of the holes, and verify that there is no change in the Euler sum of the sliced sphere.]



## CHAPTER 4

# Integration on Manifolds

### §1 Introduction

Until now we have studied manifolds from the point of view of differential calculus. The last chapter introduces to our study the methods of integral calculus. As the new tools are developed in the next sections, the reader may be somewhat puzzled about their relevancy to the earlier material. Hopefully, the puzzlement will be resolved by the chapter's end, but a preliminary example may be helpful.

Let

$$p(z) = z^m + a_1 z^{m-1} + \cdots + a_m$$

be a complex polynomial on  $\Omega$ , a smooth compact region in the plane whose boundary contains no zero of  $p$ . In Section 3 of the previous chapter we showed that the number of zeros of  $p$  inside  $\Omega$ , counting multiplicities, equals the degree of the map

$$\frac{p}{|p|} : \partial\Omega \rightarrow S^1.$$

A famous theorem in complex variable theory, called the *argument principle*, asserts that this may also be calculated as an integral,

$$\oint_{\partial\Omega} d(\arg p).$$

You needn't understand the precise meaning of the expression in order to appreciate our point. What is important is that the number of zeros can be computed both by an intersection number and by an integral formula.

Theorems like the argument principal abound in differential topology. We shall show that their occurrence is not arbitrary or fortuitous but is a geometric consequence of a general theorem known as *Stokes theorem*. Low-dimensional versions of Stokes theorem are probably familiar to you from calculus:

1. Second Fundamental Theorem of Calculus:

$$\int_a^b f'(x) dx = f(b) - f(a).$$

2. Green's Theorem in the plane:

$$\oint_{\partial\Omega} f_1 dx + f_2 dy = \iint_{\Omega} \left( \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) dx dy.$$

3. Divergence Theorem in 3-space:

$$\iiint_{\Omega} \operatorname{div} \vec{F} dv = \iint_{\partial\Omega} \vec{F} \cdot \vec{n} dA, \quad \vec{F} = (f_1, f_2, f_3).$$

4. Classical Stokes Theorem in 3-space:

$$\begin{aligned} \oint_{\partial S} f_1 dx + f_2 dy + f_3 dz &= \iint_S \vec{n} \cdot \operatorname{curl} \vec{F} dA \\ &= \iint_S \left( \frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z} \right) dy \wedge dz \\ &\quad + \left( \frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x} \right) dz \wedge dx \\ &\quad + \left( \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) dx \wedge dy. \end{aligned}$$

The argument principal, in particular, may be easily deduced from Green's theorem provided that you know a little about complex analytic functions.

The preceding formulas share a common spirit; they all explain how to calculate an integral over a bounded region by an integral over the boundary alone. It is reasonable to expect that such relations can be generalized to arbitrary manifolds of any dimension, but it is not at all clear how they should be formulated. Part of the problem is just making sense of infinitesimal ex-

pressions like

$$f_1 dx \quad \text{or} \quad \left( \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) dx dy.$$

Elementary calculus books tend not to worry too much about these integrands, but one must be more scrupulous in higher dimensions. So we shall have to endure some formalism as the next section develops the algebra needed to define and manipulate infinitesimals. Then in Section 3 we introduce the appropriate integrands, abstract objects called differential forms.

It seems hopeless to provide at the outset a satisfactory, intuitive motivation for differential forms. Only experience with them can convince the student that these formal entities, with their special algebraic properties, provide the most natural basis for integration theory. Expressed in the language of forms, the general Stokes theorem becomes simple and elegant. And, in truth, isn't simplicity the ultimate test of the basic concepts of any theory? Why, for instance, did physics formulate the laws of mechanics in terms of momentum and energy rather than with some other arbitrary functions of mass and velocity? The reason is that those quantities enormously simplify the basic equations, thereby making the subject easier to understand. As for intuition, how forbiddingly formal were such notions as electric fields or complex numbers until their obvious convenience forced us to employ them, developing experience with time?

## §2 Exterior Algebra

In order to establish an algebraic basis for differential forms, we begin by defining certain generalizations of the dual space to a real vector space  $V$ . A  $p$ -tensor on  $V$  is a real-valued function  $T$  on the cartesian product

$$V^p = \underbrace{V \times \cdots \times V}_{p \text{ times}},$$

which is separately linear in each variable, or *multilinear*. That is, holding all but the  $j$ th variable constant, we have the usual linearity condition

$$\begin{aligned} T(v_1, \dots, v_j + av'_j, \dots, v_p) = \\ T(v_1, \dots, v_j, \dots, v_p) + aT(v_1, \dots, v'_j, \dots, v_p). \end{aligned}$$

In particular, 1-tensors are just linear functionals on  $V$ . A familiar 2-tensor is dot product on  $\mathbf{R}^k$ . Also on  $\mathbf{R}^k$  you know a  $k$ -tensor—namely, the determinant. For any  $k$  vectors  $v_1, \dots, v_k \in \mathbf{R}^k$  can be arranged into a  $k \times k$

matrix

$$\begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix},$$

and the determinant of this matrix is multilinear with respect to the row vectors; denote it by  $\det(v_1, \dots, v_k)$ .

As sums and scalar multiples of multilinear functions are still multilinear, the collection of all  $p$ -tensors is a vector space  $\mathfrak{P}(V^*)$ . Note that  $\mathfrak{P}^1(V^*) = V^*$ . Tensors may also be multiplied in a simple way; if  $T$  is a  $p$ -tensor and  $S$  a  $q$ -tensor, we define a  $p + q$  tensor  $T \otimes S$  by the formula

$$T \otimes S(v_1, \dots, v_p, v_{p+1}, \dots, v_{p+q}) = T(v_1, \dots, v_p) \cdot S(v_{p+1}, \dots, v_{p+q}).$$

$T \otimes S$  is called the *tensor product* of  $T$  with  $S$ . Note that the tensor product operation is not commutative,

$$T \otimes S \neq S \otimes T,$$

but it is easy to check that it is associative and that it distributes over addition.

Tensor product clarifies the manner in which  $\mathfrak{P}^p(V^*)$  extends  $V^*$ .

**Theorem.** Let  $\{\phi_1, \dots, \phi_k\}$  be a basis for  $V^*$ . Then the  $p$ -tensors  $\{\phi_{i_1} \otimes \dots \otimes \phi_{i_p} : 1 \leq i_1, \dots, i_p \leq k\}$  form a basis for  $\mathfrak{P}^p(V^*)$ . Consequently,  $\dim \mathfrak{P}^p(V^*) = k^p$ .

*Proof.* During the proof (only) we shall use the following notation. If  $I = (i_1, \dots, i_p)$  is a sequence of integers each between 1 and  $k$ , let

$$\phi_I = \phi_{i_1} \otimes \dots \otimes \phi_{i_p}.$$

Let  $\{v_1, \dots, v_k\}$  be the dual basis in  $V$ , and denote by  $v_I$  the sequence  $(v_{i_1}, \dots, v_{i_p})$ . By definition, if  $I$  and  $J$  are two such index sequences,  $\phi_I(v_J)$  is 1 if  $I = J$  and 0 if  $I \neq J$ . It is clear from multilinearity that two  $p$ -tensors  $T$  and  $S$  are equal if and only if  $T(v_J) = S(v_J)$  for every index sequence  $J$ . Thus if we are given  $T$ , the tensor

$$S = \sum_I T(v_I) \phi_I$$

must equal  $T$ ; hence the  $\{\phi_I\}$  span  $\mathfrak{P}^p(V^*)$ . The  $\phi_I$  are also independent, for if

$$S = \sum_I a_I \phi_I = 0,$$

then

$$0 = S(v_j) = a_j,$$

for each  $j$ . Q.E.D.

A tensor  $T$  is *alternating* if the sign of  $T$  is reversed whenever two variables are transposed:

$$T(v_1, \dots, v_i, \dots, v_j, \dots, v_p) = -T(v_1, \dots, v_j, \dots, v_i, \dots, v_p).$$

All 1-tensors are automatically alternating. The determinant is also alternating, but the dot product is not. It is useful to rephrase this condition slightly. Let  $S_p$  denote the group of permutations of the numbers 1 to  $p$ . Recall that a permutation  $\pi \in S_p$  is called *even* or *odd*, depending on whether it is expressible as a product of an even or odd number of simple transpositions. Let  $(-1)^\pi$  be  $+1$  or  $-1$ , depending on whether  $\pi$  is even or odd. For any  $p$ -tensor  $T$  and any  $\pi \in S_p$ , define another  $p$ -tensor  $T^\pi$  by

$$T^\pi(v_1, \dots, v_p) = T(v_{\pi(1)}, \dots, v_{\pi(p)}).$$

Then, clearly, the alternating  $p$ -tensors are those satisfying

$$T^\pi = (-1)^\pi T \quad \text{for all } \pi \in S_p$$

Note that  $(T^\pi)^\sigma = T^{\pi \circ \sigma}$  always holds.

There is a standard procedure for making alternating tensors out of arbitrary ones. If  $T$  is any  $p$ -tensor, define a new one  $\text{Alt}(T)$  by

$$\text{Alt}(T) = \frac{1}{p!} \sum_{\pi \in S_p} (-1)^\pi T^\pi.$$

Note that  $\text{Alt}(T)$  is in fact alternating, for it is obvious that  $(-1)^{\pi \circ \sigma} = (-1)^\pi (-1)^\sigma$ . Thus

$$[\text{Alt}(T)]^\sigma = \frac{1}{p!} \sum_{\pi \in S_p} (-1)^\pi (T^\pi)^\sigma = \frac{1}{p!} (-1)^\sigma \sum_{\pi \in S_p} (-1)^{\pi \circ \sigma} T^{\pi \circ \sigma}.$$

If we set  $\tau = \pi \circ \sigma$ , then, because  $S_p$  is a group, as  $\pi$  ranges through  $S_p$  so does  $\tau$ . Thus

$$[\text{Alt}(T)]^\sigma = (-1)^\sigma \frac{1}{p!} \sum_{\tau \in S_p} (-1)^\tau T^\tau = (-1)^\sigma \text{Alt}(T),$$

as claimed.

Also note that if  $T$  is already alternating, then  $\text{Alt}(T) = T$ , for each summand  $(-1)^\pi T^\pi$  equals  $T$ , and there are exactly  $p!$  permutations in  $S_p$ .

Since sums and scalar multiples of alternating functions continue to alternate, the alternating  $p$ -tensors form a vector subspace  $\Lambda^p(V^*)$  of  $\mathfrak{F}^p(V^*)$ . Unhappily, tensor products of alternating tensors do not alternate, but here the Alt operator can be useful. If  $T \in \Lambda^p(V^*)$  and  $S \in \Lambda^q(V^*)$ , we define their *wedge product*

$$T \wedge S \in \Lambda^{p+q}(V^*)$$

to be  $\text{Alt}(T \otimes S)$ .† The wedge product clearly distributes over addition and scalar multiplication, because Alt is a linear operation; however, proving associativity will require some work. We need a calculation.

**Lemma.** If  $\text{Alt}(T) = 0$ , then  $T \wedge S = 0 = S \wedge T$ .

*Proof.*  $S_{p+q}$  carries a natural copy of  $S_p$ —namely, the subgroup  $G$  consisting of all permutations of  $(1, \dots, p+q)$  that fix  $p+1, \dots, p+q$ . The correspondence between  $G$  and  $S_p$  assigns to each  $\pi \in G$  the permutation  $\pi'$  induced by restricting  $\pi$  to  $(1, \dots, p)$ . Note that  $(T \otimes S)^\pi = T^{\pi'} \otimes S$ , and  $(-1)^\pi = (-1)^{\pi'}$ . Thus

$$\sum_{\pi \in G} (-1)^\pi (T \otimes S)^\pi = \left[ \sum_{\pi' \in S_p} (-1)^{\pi'} T^{\pi'} \right] \otimes S = \text{Alt}(T) \otimes S = 0.$$

Now a subgroup  $G$  decomposes  $S_{p+q}$  into a disjoint union of right cosets  $G \circ \sigma = \{\pi \circ \sigma : \pi \in G\}$ . But for each such coset,

$$\sum_{\pi \in G} (-1)^{\pi \circ \sigma} (T \otimes S)^{\pi \circ \sigma} = (-1)^\sigma \left[ \sum_{\pi \in G} (-1)^\pi (T \otimes S)^\pi \right]^\sigma = 0.$$

Since  $T \wedge S = \text{Alt}(T \otimes S)$  is the sum of these partial summations over the right cosets of  $G$ , then  $T \wedge S = 0$ . Similarly,  $S \wedge T = 0$ . Q.E.D.

**Theorem.** Wedge product is associative,

$$(T \wedge S) \wedge R = T \wedge (S \wedge R),$$

justifying the notation  $T \wedge S \wedge R$ .

*Proof.* We claim that  $(T \wedge S) \wedge R$  equals  $\text{Alt}(T \otimes S \otimes R)$ . By definition,

$$(T \wedge S) \wedge R = \text{Alt}((T \wedge S) \otimes R),$$

so the linearity of Alt implies

$$(T \wedge S) \wedge R = \text{Alt}(T \otimes S \otimes R) = \text{Alt}([T \wedge S - T \otimes S] \otimes R).$$

†There is always some question about normalizing the definition of  $\wedge$ . Spivak, for example, includes some factorial multipliers. We have chosen to keep the algebra simple, but we shall be haunted by the ghost of the vanished factorials when we discuss volumes.

Since  $T \wedge S$  is alternating,

$$\begin{aligned}\text{Alt}(T \wedge S - T \otimes S) &= \text{Alt}(T \wedge S) - \text{Alt}(T \otimes S) \\ &= T \wedge S - T \wedge S = 0.\end{aligned}$$

So the lemma implies

$$\text{Alt}([T \wedge S - T \otimes S] \otimes R) = 0,$$

as needed. A similar argument shows

$$T \wedge (S \wedge R) = \text{Alt}(T \otimes S \otimes R) \quad \text{Q.E.D.}$$

The formula  $T \wedge S \wedge R = \text{Alt}(T \otimes S \otimes R)$ , derived in the proof above, obviously extends to relate the wedge and tensor products of any number of tensors. We can use it to derive a basis for  $\Lambda^p(V^*)$ . For if  $T$  is a  $p$ -tensor, then we may write

$$T = \sum t_{i_1, \dots, i_p} \phi_{i_1} \otimes \cdots \otimes \phi_{i_p},$$

where  $\{\phi_1, \dots, \phi_k\}$  is a basis for  $V^*$ , and the sum ranges over all index sequences  $(i_1, \dots, i_p)$  for which each index is between 1 and  $k$ . If  $T$  alternates, then  $T = \text{Alt}(T)$ , so

$$T = \sum t_{i_1, \dots, i_p} \text{Alt}(\phi_{i_1} \otimes \cdots \otimes \phi_{i_p}) = \sum t_{i_1, \dots, i_p} \phi_{i_1} \wedge \cdots \wedge \phi_{i_p}.$$

Henceforth we shall denote the alternating tensors  $\phi_{i_1} \wedge \cdots \wedge \phi_{i_p}$  by  $\phi_I$ , where  $I = (i_1, \dots, i_p)$ . We have shown that the  $\phi_I$  span  $\Lambda^p(V^*)$ ; however, due to a fundamental property of the wedge product, they are not independent.

Suppose that  $\phi$  and  $\psi$  are any linear functionals on  $V$ ,  $\phi, \psi \in \Lambda^1(V^*)$ . In this case, the Alt operator takes a very simple form:

$$\phi \wedge \psi = \frac{1}{2}(\phi \otimes \psi - \psi \otimes \phi).$$

Observe that

$$\phi \wedge \psi = -\psi \wedge \phi \quad \text{and} \quad \phi \wedge \phi = 0,$$

showing that  $\wedge$  is anticommutative on  $\Lambda^1(V^*)$ . As you will discover, the anticommutativity of wedge product on 1-forms is its fundamental property. In fact, the essential reason for developing the algebra of alternating tensors is to build anticommutativity into the foundation of integration theory.

Anticommutativity introduces some relations into the set of spanning tensors  $\{\phi_I\}$ . If two index sequences  $I$  and  $J$  differ only in their orderings, iterated application of anticommutativity shows that  $\phi_I = \pm \phi_J$ . And if any of the indices of  $I$  are equal,  $\phi_I = 0$ . Consequently, we can eliminate redundancy in the spanning set by allowing only those  $\phi_I$  for which the index

sequence is strictly increasing:  $1 \leq i_1 < i_2 < \cdots < i_p \leq k$ . The number of such sequences is the number of ways one can choose  $p$  items from the set  $\{1, \dots, k\}$ , namely,

$$\binom{k}{p} = \frac{k!}{p!(k-p)!}.$$

It is easy to see that the remaining tensors are linearly independent. Let  $\{v_1, \dots, v_k\}$  be the basis for  $V$  dual to  $\{\phi_1, \dots, \phi_k\}$ . For any increasing index sequence  $I = (i_1, \dots, i_p)$ , let  $v_I = (v_{i_1}, \dots, v_{i_p})$ . The definition of the Alt operator then shows that  $\phi_I(v_I) = 1/p!$ , but if  $J$  is a different increasing index sequence, then  $\phi_I(v_J) = 0$ . Thus if  $\sum a_I \phi_I = 0$  is a relation among the new spanning set, then

$$0 = \sum a_I \phi_I(v_J) = \frac{1}{p!} a_J$$

shows that each  $a_J = 0$ . We have proved

**Theorem.** If  $\{\phi_1, \dots, \phi_k\}$  is a basis for  $V^*$ , then  $\{\phi_I = \phi_{i_1} \wedge \cdots \wedge \phi_{i_p} : 1 \leq i_1, \dots, i_p \leq k\}$  is a basis for  $\Lambda^p(V^*)$ . Consequently,

$$\dim \Lambda^p(V^*) = \binom{k}{p} = \frac{k!}{p!(k-p)!}$$

Suppose that the index sequence  $I$  has length  $p$ , while  $J$  has length  $q$ . From the anticommutativity of  $\wedge$  on  $\Lambda^1(V^*)$ , you can quickly show that

$$\phi_I \wedge \phi_J = (-1)^{pq} \phi_J \wedge \phi_I.$$

The foregoing basis theorem therefore implies

**Corollary.** The wedge product satisfies the following anticommutativity relation:

$$T \wedge S = (-1)^{pq} S \wedge T,$$

whenever

$$T \in \Lambda^p(V^*) \quad \text{and} \quad S \in \Lambda^q(V^*).$$

The basis theorem also implies that  $\Lambda^k(V^*)$  is one dimensional, where  $k = \dim V$ . You have probably known this for years, although in this form the statement looks unfamiliar. We already know one nonzero alternating  $k$ -tensor on  $\mathbf{R}^k$ , the determinant tensor  $\det$ . So  $\dim \Lambda^k(\mathbf{R}^{k*}) = 1$  just means that every alternating  $k$ -multilinear function on  $\mathbf{R}^k$  is a multiple of the determinant, a fact you know as “the uniqueness of the determinant function.”



If the length of the index sequence  $I$  is greater than the dimension  $k$  of  $V$ , then  $I$  must repeat at least one integer; thus  $\phi_I = 0$ . We conclude that  $\Lambda^p(V^*) = 0$  if  $p > k$ , so the sequence of vector spaces  $\Lambda^1(V^*)$ ,  $\Lambda^2(V^*)$ ,  $\dots$  terminates at  $\Lambda^k(V^*)$ . It is useful to add one more entry to this list by defining  $\Lambda^0(V^*) = \mathbf{R}$ , which is interpreted as the constant functions on  $V$ . We extend  $\wedge$  by simply letting the wedge product of any element in  $\mathbf{R}$  with any tensor in  $\Lambda^p(V^*)$  be the usual scalar multiplication. The wedge product then makes the direct sum

$$\Lambda(V^*) = \Lambda^0(V^*) \oplus \Lambda^1(V^*) \oplus \dots \oplus \Lambda^k(V^*)$$

a noncommutative algebra, called the *exterior algebra* of  $V^*$ , whose identity element is  $1 \in \Lambda^0(V^*)$ .

One further construction is basic. Suppose that  $A: V \rightarrow W$  is a linear map. Then the transpose map  $A^*: W^* \rightarrow V^*$  extends in an obvious manner to the exterior algebras,  $A^*: \Lambda^p(W^*) \rightarrow \Lambda^p(V^*)$  for all  $p \geq 0$ . If  $T \in \Lambda^p(W^*)$ , just define  $A^*T \in \Lambda^p(V^*)$  by

$$A^*T(v_1, \dots, v_p) = T(Av_1, \dots, Av_p)$$

for all vectors  $v_1, \dots, v_p \in V$ . It is easy to check that  $A^*$  is linear and that

$$A^*(T \wedge S) = A^*T \wedge A^*S.$$

So  $A^*$  is an algebra homomorphism:  $\Lambda(W^*) \rightarrow \Lambda(V^*)$ . Note that if  $B: W \rightarrow U$  is another linear map, then  $(BA)^* = A^*B^*$ .

In particular, suppose that  $A: V \rightarrow V$  is an isomorphism, and  $\dim V = k$  as usual. Then  $A^*: \Lambda^k(V^*) \rightarrow \Lambda^k(V^*)$  is a linear map of a one-dimensional vector space, hence it must be multiplication by some constant  $\lambda \in \mathbf{R}$ , i.e.,  $A^*T = \lambda T$  for all  $T \in \Lambda^k(V^*)$ . We claim that  $\lambda$  is just the determinant of  $A$ . We know that  $\det \in \Lambda^k(\mathbf{R}^{k*})$ . So choose any isomorphism  $B: V \rightarrow \mathbf{R}^k$ , and consider  $T = B^*(\det) \in \Lambda^k(V^*)$ . Then  $A^*B^*(\det) = \lambda B^*(\det)$ , implying

$$B^{*-1}A^*B^*(\det) = \lambda(B^*)^{-1}B^*(\det) = \lambda(BB^{-1})^*(\det) = \lambda(\det)$$

or

$$(BAB^{-1})^*(\det) = \lambda(\det).$$

Now evaluate both sides of this equation on the standard ordered basis  $\{e_1, \dots, e_k\}$  for  $\mathbf{R}^k$ . A quick check of the definition of the tensor  $\det$  shows that, for any linear map  $C$ ,  $\det(Ce_1, \dots, Ce_n) = \det(C)$ . Thus

$$\lambda = \det(BAB^{-1}) = \det(A)$$

as claimed, proving

**Determinant Theorem.** If  $A: V \rightarrow V$  is a linear isomorphism, then  $A^*T = (\det A)T$  for every  $T \in \Lambda^k(V)$ , where  $k = \dim V$ . In particular, if  $\phi_1, \dots, \phi_k \in \Lambda^1(V^*)$ , then

$$A^*\phi_1 \wedge \cdots \wedge A^*\phi_k = (\det A) \phi_1 \wedge \cdots \wedge \phi_k.$$

### EXERCISES

1. Suppose that  $T \in \Lambda^p(V^*)$  and  $v_1, \dots, v_p \in V$  are linearly dependent. Prove that  $T(v_1, \dots, v_p) = 0$  for all  $T \in \Lambda^p(V^*)$ .
2. Dually, suppose that  $\phi_1, \dots, \phi_p \in V^*$  are linearly dependent, and prove that  $\phi_1 \wedge \cdots \wedge \phi_p = 0$ .
3. Suppose that  $\phi_1, \dots, \phi_k \in V^*$  and  $v_1, \dots, v_k \in V$ , where  $k = \dim V$ . Prove that

$$\phi_1 \wedge \cdots \wedge \phi_k(v_1, \dots, v_k) = \frac{1}{k!} \det [\phi_i(v_j)],$$

where  $[\phi_i(v_j)]$  is a  $k \times k$  real matrix. [HINT: If the  $\phi_i$  are dependent, then the matrix has linearly dependent rows, so Exercise 2 suffices. If not, the formula is easily checked for the dual basis in  $V$ . Now verify that the matrix does specify an alternating  $k$ -tensor on  $V$ , and use  $\dim \Lambda^k(V^*) = 1$ .]

4. More generally, show that whenever  $\phi_1, \dots, \phi_p \in V^*$  and  $v_1, \dots, v_p \in V$ , then

$$\phi_1 \wedge \cdots \wedge \phi_p(v_1, \dots, v_p) = \frac{1}{p!} \det [\phi_i(v_j)].$$

[HINT: If the  $v_i$  are dependent, use Exercise 1. If not, apply Exercise 3 to the restrictions  $\bar{\phi}_i$  of  $\phi_i$  to the  $p$ -dimensional subspace spanned by  $v_1, \dots, v_p$ .]

5. Specifically write out  $\text{Alt}(\phi_1 \otimes \phi_2 \otimes \phi_3)$  for  $\phi_1, \phi_2, \phi_3 \in V^*$ .
- \*6. (a) Let  $T$  be a nonzero element of  $\Lambda^k(V^*)$ , where  $\dim V = k$ . Prove that two ordered bases  $\{v_1, \dots, v_k\}$  and  $\{v'_1, \dots, v'_k\}$  for  $V$  are equivalently oriented if and only if  $T(v_1, \dots, v_k)$  and  $T(v'_1, \dots, v'_k)$  have the same sign. [HINT: Determinant theorem.]  
 (b) Suppose that  $V$  is oriented. Show that the one-dimensional vector space  $\Lambda^k(V^*)$  acquires a natural orientation, by defining the sign of a nonzero element  $T \in \Lambda^k(V^*)$  to be the sign of  $T(v_1, \dots, v_k)$  for any positively oriented ordered basis  $\{v_1, \dots, v_k\}$  for  $V$ .

- (c) Conversely, show that an orientation of  $\Lambda^k(V^*)$  naturally defines an orientation on  $V$  by reversing the above.
7. For a  $k \times k$  matrix  $A$ , let  $A'$  denote the transpose matrix. Using the fact that  $\det(A)$  is multilinear in both the rows and columns of  $A$ , prove that  $\det(A') = \det(A)$ . [HINT: Use  $\dim \Lambda^k(\mathbf{R}^{k*}) = 1$ ]
8. Recall that a matrix  $A$  is *orthogonal* if  $AA' = I$ . Conclude that if  $A$  is orthogonal,  $\det(A) = \pm 1$ .
9. Let  $V$  be a  $k$ -dimensional subspace of  $\mathbf{R}^N$ . Recall that a basis  $v_1, \dots, v_k$  of  $V$  is *orthonormal* if

$$v_i \cdot v_j = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Let  $A: V \rightarrow V$  be a linear map, and prove the following three conditions equivalent:

- (a)  $Av \cdot Aw = v \cdot w$  for all  $v, w \in V$ .  
 (b)  $A$  carries orthonormal bases to orthonormal bases.  
 (c) The matrix of  $A$  with respect to any orthonormal basis is orthogonal.

Such an  $A$  is called an *orthogonal transformation*. [Note, by (b), it must be an isomorphism.]

- \*10. (a) Let  $V$  be an oriented  $k$ -dimensional vector subspace of  $\mathbf{R}^N$ . Prove there is an alternating  $k$ -tensor  $T \in \Lambda^k(V^*)$  such that  $T(v_1, \dots, v_k) = 1/k!$  for all positively oriented ordered orthonormal bases. Furthermore, show that  $T$  is unique; it is called the *volume element* of  $V$ . [HINT: Use the determinant theorem, Exercises 8 and 9, plus  $\dim \Lambda^k(V^*) = 1$  for uniqueness.]  
 (b) In fact, suppose that  $\phi_1, \dots, \phi_k \in V^*$  is an ordered basis dual to some positively oriented ordered orthonormal basis for  $V$ . Show that the volume element for  $V$  is  $\phi_1 \wedge \dots \wedge \phi_k$ . [HINT: Exercise 3.]
- \*11. Let  $T$  be the volume element of  $\mathbf{R}^2$ . Prove that for any vectors  $v_1, v_2 \in \mathbf{R}^2$ ,  $T(v_1, v_2)$  is  $\pm$  one half the volume of the parallelogram spanned by  $v_1$  and  $v_2$ . Furthermore, when  $v_1$  and  $v_2$  are independent, then the sign equals the sign of the ordered basis  $\{v_1, v_2\}$  in the standard orientation of  $\mathbf{R}^2$ . Generalize to  $\mathbf{R}^3$ . Now how would you define the volume of a parallelepiped in  $\mathbf{R}^k$ ?
12. (a) Let  $V$  be a subspace of  $\mathbf{R}^N$ . For each  $v \in V$ , define a linear functional  $\phi_v \in V^*$  by  $\phi_v(w) = v \cdot w$ . Prove that the map  $v \rightarrow \phi_v$  is an isomorphism of  $V$  with  $V^*$ .

- (b) Now suppose that  $V$  is oriented and  $\dim V = 3$ . Let  $T$  be the volume element on  $V$ . Given  $u, v \in V$ , define a linear functional on  $V$  by  $w \mapsto 3!T(u, v, w)$ . By part (a), there exists a vector, which we denote  $u \times v$ , such that  $T(u, v, w) = (u \times v) \cdot w$  for all  $w \in V$ . Prove that this *cross product* satisfies  $u \times v = -v \times u$ . Furthermore, show that if  $\{v_1, v_2, v_3\}$  is a positively oriented orthonormal basis for  $V$ , then  $v_1 \times v_2 = v_3$ ,  $v_2 \times v_3 = v_1$ , and  $v_3 \times v_1 = v_2$ . (Also,  $v \times v = 0$  always.)

### §3 Differential Forms

In classical differential geometry, forms were symbolic quantities that looked like

$$\begin{aligned} & \sum_i f_i dx_i \\ & \sum_{i < j} f_{ij} dx_i \wedge dx_j \\ & \sum_{i < j < k} f_{ijk} dx_i \wedge dx_j \wedge dx_k. \end{aligned}$$

These expressions were integrated and differentiated, and because experience proved anticommutativity to be convenient, they were manipulated like alternating tensors. Modern differential forms locally reduce to the same symbolic quantities, but they possess the indispensable attribute of being globally defined on manifolds. Global definition of integrands makes possible global integration.

**Definition.** Let  $X$  be a smooth manifold with or without boundary. A  $p$ -form on  $X$  is a function  $\omega$  that assigns to each point  $x \in X$  an alternating  $p$ -tensor  $\omega(x)$  on the tangent space of  $X$  at  $x$ ;  $\omega(x) \in \Lambda^p[T_x(X)^*]$ .

Two  $p$ -forms  $\omega_1$  and  $\omega_2$  may be added point by point to create a new  $p$ -form  $\omega_1 + \omega_2$ :

$$(\omega_1 + \omega_2)(x) = \omega_1(x) + \omega_2(x).$$

Similarly, the wedge product of forms is defined point by point. If  $\omega$  is a  $p$ -form and  $\theta$  is a  $q$ -form, the  $p + q$  form  $\omega \wedge \theta$  is given by  $(\omega \wedge \theta)(x) = \omega(x) \wedge \theta(x)$ . Anticommutativity  $\omega \wedge \theta = (-1)^{pq} \theta \wedge \omega$  follows from the analogous equation at each point.

0-forms are just arbitrary real-valued functions on  $X$ .

Many examples of 1-forms can be manufactured from smooth functions. If  $\phi: X \rightarrow \mathbf{R}$  is a smooth function,  $d\phi_x: T_x(X) \rightarrow \mathbf{R}$  is a linear map at each point  $x$ . Thus the assignment  $x \mapsto d\phi_x$  defines a 1-form  $d\phi$  on  $X$ , called the

differential of  $\phi$ . In particular, the coordinate functions  $x_1, \dots, x_k$  on  $\mathbf{R}^k$  yield 1-forms  $dx_1, \dots, dx_k$  on  $\mathbf{R}^k$ . Now at each  $z \in \mathbf{R}^k$ ,  $T_z(\mathbf{R}^k) = \mathbf{R}^k$ ; check that the important differentials  $dx_1, \dots, dx_k$  have the specific action  $dx_i(z)(a_1, \dots, a_k) = a_i$ . Thus at each  $z \in \mathbf{R}^k$ , the linear functionals  $dx_1(z), \dots, dx_k(z)$  are just the standard basis for  $(\mathbf{R}^k)^*$ .

In terms of  $dx_1, \dots, dx_k$  it is easy to write down all forms on an open subset  $U$  of Euclidean space. For each strictly increasing index sequence  $I = (i_1, \dots, i_p)$ , let

$$dx_I = dx_{i_1} \wedge \cdots \wedge dx_{i_p},$$

a  $p$ -form on  $\mathbf{R}^k$ . Then from the basis theorem for the individual vector spaces

$$\Lambda^p[T_z(U)^*] = \Lambda^p(\mathbf{R}^{k*}),$$

we obtain at once

**Proposition.** Every  $p$ -form on an open set  $U \subset \mathbf{R}^k$  may be uniquely expressed as a sum  $\sum_I f_I dx_I$ , over increasing index sequences  $I = (i_1 < \cdots < i_p)$ , the  $f_I$  being functions on  $U$ .

The proposition says that forms in Euclidean space are really the symbolic quantities of classical mathematics—except that our symbols have precise meanings. Here is an easy exercise that will force you to check your understanding of the definitions: show that if  $\phi$  is a smooth function on  $\mathbf{R}^k$ , then

$$d\phi = \sum \frac{\partial \phi}{\partial x_i} dx_i.$$

(The formula itself is obvious, but what do the two sides mean?)

One of the important features of forms is that they pull back naturally under smooth mappings. If  $f: X \rightarrow Y$  is a smooth map and  $\omega$  is a  $p$ -form on  $Y$ , we define a  $p$ -form  $f^*\omega$  on  $X$  as follows. If  $f(x) = y$ , then  $f$  induces a derivative map  $df_x: T_x(X) \rightarrow T_y(Y)$ . Since  $\omega(y)$  is an alternating  $p$ -tensor on  $T_y(Y)$ , we can pull it back to  $T_x(X)$  using the transpose  $(df_x)^*$ , as described in the previous section. Define

$$f^*\omega(x) = (df_x)^*\omega[f(x)].$$

Then  $f^*\omega(x)$  is an alternating  $p$ -tensor on  $T_x(X)$ , so  $f^*\omega$  is a  $p$ -form on  $X$ , called the *pullback* of  $\omega$  by  $f$ . When  $\omega$  is a 0-form—that is, a function on  $Y$ —then  $f^*\omega = \omega \circ f$ , a function on  $X$ . Remember that  $f^*$  pulls *back* forms, it does not push them forward: when  $f: X \rightarrow Y$ ,  $f^*$  carries forms from  $Y$  to  $X$ .

Before we unravel the definition of  $f^*$ , prove the following formulas:

$$\begin{aligned} f^*(\omega_1 + \omega_2) &= f^*\omega_1 + f^*\omega_2 \\ f^*(\omega \wedge \theta) &= (f^*\omega) \wedge (f^*\theta) \\ (f \circ h)^*\omega &= h^*f^*\omega. \end{aligned}$$

Now let's see explicitly what  $f^*$  does on Euclidean space. Let  $U \subset \mathbf{R}^k$  and  $V \subset \mathbf{R}^l$  be open subsets, and let  $f: V \rightarrow U$  be smooth. Use  $x_1, \dots, x_k$  for the standard coordinate functions on  $\mathbf{R}^k$  and  $y_1, \dots, y_l$  on  $\mathbf{R}^l$ . Write  $f$  concretely as  $f = (f_1, \dots, f_k)$ , each  $f_i$  being a smooth function on  $V$ . The derivative  $df_y$  at a point  $y \in V$  is represented by the matrix

$$\left( \frac{\partial f_i}{\partial y_j}(y) \right),$$

and its transpose map  $(df_y)^*$  is represented by the transpose matrix. Consequently,

$$f^* dx_i = \sum_{j=1}^l \frac{\partial f_i}{\partial y_j} dy_j = df_i.$$

(You should convince yourself of this formula's validity before proceeding further. Can you explain the exact meaning of each term?)

Knowing the behavior of  $f^*$  on the 0-forms and on the basic 1-forms  $dx_i$  determines it completely. For an arbitrary form  $\omega$  on  $U$  may be written uniquely as

$$\omega = \sum_I a_I dx_I.$$

Now application of the abstract properties of  $f^*$  listed above gives

$$f^*(\omega) = \sum_I (f^*a_I) df_I.$$

Here the function  $f^*a_I = a_I \circ f$  is the pullback of the 0-form  $a_I$  from  $U$  to  $V$ , and we use  $f_I$  to denote  $f_{i_1} \wedge \dots \wedge f_{i_r}$ .

One example is crucially important. Suppose that  $f: V \rightarrow U$  is a diffeomorphism of two open sets in  $\mathbf{R}^k$  and  $\omega = dx_1 \wedge \dots \wedge dx_k$  (the so-called *volume form* on  $U$ ). If  $f(y) = x$ , both  $T_y(V)$  and  $T_x(U)$  are equal to  $\mathbf{R}^k$ , although in the coordinate notation we have been using, the standard basis of linear functions on  $\mathbf{R}^k$  is written as  $dy_1(y), \dots, dy_k(y)$  for  $T_y(V)$  but as  $dx_1(x), \dots, dx_k(x)$  for  $T_x(U)$ . The determinant theorem of Section 2 now gives us the formula

$$\begin{aligned} f^*\omega(y) &= (df_y)^* dx_1(x) \wedge \dots \wedge (df_y)^* dx_k(x) \\ &= \det(df_y) dy_1(y) \wedge \dots \wedge dy_k(y). \end{aligned}$$

More succinctly,

$$f^*(dx_1 \wedge \cdots \wedge dx_k) = \det(df) dy_1 \wedge \cdots \wedge dy_k,$$

where we denote by  $\det(df)$  the function  $y \rightarrow \det(df_y)$  on  $V$ .

A form  $\omega$  on an open set  $U \subset \mathbf{R}^k$  is said to be *smooth* if each coefficient function  $a_i$  in its expansion  $\sum_i a_i dx_i$  is smooth. It is clear from the preceding general calculation that when  $f: V \rightarrow U$  is a smooth map of open subsets of two Euclidean spaces, then  $f^*\omega$  is smooth if  $\omega$  is.

More generally, we define smoothness for a form  $\omega$  on  $X$  to mean that for every local parametrization  $h: U \rightarrow X$ ,  $h^*\omega$  is a smooth form on the open set  $U$  in  $\mathbf{R}^k$ . Of course, one need not really check every parametrization, only enough of them to cover  $X$ ; if  $h_\alpha: U_\alpha \rightarrow X$  is any collection of local parametrizations covering  $X$  (i.e.,  $\bigcup_\alpha h_\alpha(U_\alpha) = X$ ), then the form  $\omega$  is smooth provided that each pullback  $h_\alpha^*\omega$  is smooth. For if  $h: U \rightarrow X$  is another local parametrization, then we can write the domain  $U$  as the union of open subdomains  $h^{-1}[h_\alpha(U_\alpha)]$ . And on each  $h^{-1}[h_\alpha(U_\alpha)]$ , the identity  $h^*\omega = (h_\alpha^{-1} \circ h)^*h_\alpha^*\omega$  shows that  $h^*\omega$  is smooth.

Since we are only interested in smooth forms, from now on the word “form” will implicitly mean “smooth form.”

We close the section with one final exercise to test your grasp of these formalities. Work it out directly from the definitions, without recourse to any of the computations we made in Euclidean space.

**Exercise.** Let  $f: X \rightarrow Y$  be a smooth map of manifolds, and let  $\phi$  be a smooth function on  $Y$ . Then

$$f^*(d\phi) = d(f^*\phi).$$

## §4 Integration on Manifolds

Forms were created for integration, but what makes them the appropriate integrands? Perhaps their fundamental quality is that they automatically transform correctly when coordinates are changed. Recall the following important theorem from calculus (proofs of which can be found in many texts—for example, Spivak [2], p. 67).

**Change of Variables in  $\mathbf{R}^k$ .** Assume that  $f: V \rightarrow U$  is a diffeomorphism of open sets in  $\mathbf{R}^k$  and that  $a$  is an integrable function on  $U$ . Then

$$\int_U a dx_1 \cdots dx_k = \int_V (a \circ f) |\det(df)| dy_1 \cdots dy_k.$$

Here we follow our earlier notation by using  $\det(df)$  to denote the function

$y \rightarrow \det(df_y)$  on  $V$ . It is also easy to see that the theorem holds in the half-space  $H^k$  as well.

When we change variables by such a mapping  $f$ , functions like  $a$  are transformed into their obvious pullbacks  $a \circ f$ . Yet this transformation is not natural from the point of view of integration. Since  $f$  distorts volume as it forces  $V$  onto  $U$ , the integral of  $a \circ f$  is not the same as the integral of  $a$ . One must compensate by including the factor  $|\det(df)|$ , which measures the infinitesimal alteration of volume. Forms automatically counteract this volume change. Consider, for example, the original integrand to be not the function  $a$ , with  $dx_1 \cdots dx_k$  serving as a formal symbol of integration, but the  $k$ -form  $\omega = a dx_1 \wedge \cdots \wedge dx_k$ . Then define

$$\int_U \omega = \int_V a dx_1 \cdots dx_k.$$

As calculated in the last section,  $\omega$  pulls back to the form

$$f^*(\omega) = (a \circ f) \det(df) dy_1 \wedge \cdots \wedge dy_k$$

If  $f$  preserves orientation, then  $\det(df) > 0$ , so  $f^*\omega$  is exactly the integrand on the right in the Change of Variables Theorem. Every  $k$ -form  $\omega$  on  $U$  is  $a dx_1 \wedge \cdots \wedge dx_k$  for some function  $a$ , so if we call  $\omega$  “integrable” when  $a$  is, the theorem attains a very natural form.

**Change of Variables in  $\mathbf{R}^k$ .** Assume that  $f: V \rightarrow U$  is an orientation-preserving diffeomorphism of open sets in  $\mathbf{R}^k$  or  $H^k$ , and let  $\omega$  be an integrable  $k$ -form on  $U$ . Then

$$\int_U \omega = \int_V f^*\omega.$$

If  $f$  reverses orientation, then

$$\int_U \omega = - \int_V f^*\omega.$$

If you trace back through the previous sections, you will discover that the automatic appearance of the compensating factor  $\det(df)$  is a mechanical consequence of the anticommutative behavior of 1-forms:  $dx_i \wedge dx_j = -dx_j \wedge dx_i$ . So you see, the entire algebraic apparatus of forms exists in order to provide integrands that transform properly for integration.

This transformation property is important because it allows us to integrate forms on manifolds, where we have no recourse to standard coordinates as in Euclidean space. Let  $\omega$  be a smooth  $k$ -form on  $X$ , a  $k$ -dimensional manifold with boundary. The *support* of  $\omega$  is defined as the closure of the set of points where  $\omega(x) \neq 0$ ; we assume this closure to be compact, in which case  $\omega$  is said to be *compactly supported*. At first, assume also that the support



of  $\omega$  is contained inside a single parametrizable open subset  $W$  of  $X$ . Then if  $h: U \rightarrow W$  is an orientation-preserving diffeomorphism of  $W$  with an open subset  $U \subset H^k$ ,  $h^*\omega$  is a compactly supported, smooth  $k$ -form on  $U$ . Therefore  $h^*\omega$  is integrable, and we define  $\int_X \omega = \int_U h^*\omega$ . What if  $g: V \rightarrow W$  is another such parametrization of  $W$ ? Then  $f = h^{-1} \circ g$  is an orientation-preserving diffeomorphism  $V \rightarrow U$ , so

$$\int_U h^*\omega = \int_V f^*h^*\omega = \int_V g^*\omega.$$

Thanks to the transformation properties of forms,  $\int_X \omega$  has an intrinsic meaning, independent of the choice of parametrization. If our propaganda has not yet made you a true believer in forms, we invite you to try defining the integral of a function.

Now, in order to define the integral of an arbitrary, compactly supported, smooth  $k$ -form  $\omega$  on  $X$ , we simply use a partition of unity to break up  $\omega$  into pieces with parametrizable support. The collection of parametrizable open subsets of  $X$  forms an open cover; choose a subordinate partition of unity  $\{\rho_i\}$ . (Recall the definition, page 52.) The local finiteness property of  $\{\rho_i\}$  implies that all but finitely many of them are identically zero on the compact support of  $\omega$ . Thus only finitely many of the forms  $\rho_i\omega$  are nonzero, and each one has compact support inside a parametrizable open set. Define

$$\int_X \omega = \sum_i \int \rho_i\omega.$$

Showing that  $\int_X \omega$  does not depend on the particular partition of unity is easy. First, however, observe that if the support of  $\omega$  is actually inside some parametrizable open set, then the two definitions of  $\int_X \omega$  just given agree. Since  $\sum_i \rho_i(x) = 1$  at every  $x \in X$ ,

$$\sum_i \rho_i\omega = \omega.$$

Then the linearity of pullback and of integration on Euclidean space imply

$$\int_X \omega = \sum_i \int_X \rho_i\omega,$$

as needed. Now suppose that  $\{\rho'_j\}$  is another suitable partition of unity. Then from what we have just observed, for each  $i$ ,

$$\int_X \rho_i\omega = \sum_j \int_X \rho'_j\rho_i\omega;$$

similarly, for each  $j$ ,

$$\int_X \rho'_j \omega = \sum_i \int_X \rho_i \rho'_j \omega.$$

Then

$$\sum_i \int_X \rho_i \omega = \sum_i \sum_j \int_X \rho'_j \rho_i \omega = \sum_j \sum_i \int_X \rho_i \rho'_j \omega = \sum_j \int_X \rho'_j \omega,$$

showing that  $\int_X \omega$  is the same when calculated with either partition.

It is trivial to check that  $\int_X$  has the standard linearity properties:

$$\int_X (\omega_1 + \omega_2) = \int_X \omega_1 + \int_X \omega_2 \quad \text{and} \quad \int_X c\omega = c \int_X \omega \quad \text{if } c \in \mathbf{R}.$$

We also let you verify that our generalized integration theory continues to behave properly when domains are changed.

**Theorem.** If  $f: Y \rightarrow X$  is an orientation-preserving diffeomorphism, then

$$\int_X \omega = \int_Y f^* \omega$$

for every compactly supported, smooth  $k$ -form on  $X$  ( $k = \dim X = \dim Y$ ).

You might also convince yourself that this transformation property absolutely determines our theory. We have constructed the only linear operation on compactly supported forms that transforms naturally and that reduces to usual integration in Euclidean space.

Although we can only integrate  $k$ -forms over the  $k$ -dimensional  $X$ , we can integrate other forms over submanifolds. If  $Z$  is an oriented submanifold of  $X$  and  $\omega$  is a form on  $X$ , our abstract operations give us a natural way of “restricting”  $\omega$  to  $Z$ . Let  $i: Z \hookrightarrow X$  be the inclusion map, and define the *restriction* of  $\omega$  to  $Z$  to be the form  $i^* \omega$ . It is obvious that when  $\omega$  is a 0-form,  $i^* \omega$  is just the usual restriction of the function  $\omega$  to  $Z$ . Now if  $\dim Z = l$  and  $\omega$  is an  $l$ -form whose support intersects  $Z$  in a compact set, we define the integral of  $\omega$  over  $Z$  to be the integral of its restriction,

$$\int_Z \omega = \int_Z i^* \omega.$$

Let’s work out some specific examples. Suppose that

$$\omega = f_1 dx_1 + f_2 dx_2 + f_3 dx_3$$

is a smooth 1-form on  $\mathbf{R}^3$ , and let  $\gamma: I \rightarrow \mathbf{R}^3$  be a simple curve, a diffeomorphism of the unit interval  $I = [0, 1]$  onto  $C = \gamma(I)$  a compact one-manifold with boundary. Then

$$\int_C \omega = \int_I \gamma^* \omega.$$

If

$$\gamma(t) = (\gamma_1(t), \gamma_2(t), \gamma_3(t)),$$

then

$$\gamma^* dx_i = d\gamma_i = \frac{d\gamma_i}{dt} dt,$$

so we obtain

$$\int_C \omega = \sum_{i=1}^3 \int_0^1 f_i[\gamma(t)] \frac{d\gamma_i}{dt}(t) dt.$$

If we define  $\vec{F}$  to be the vector field  $(f_1, f_2, f_3)$  in  $\mathbf{R}^3$ , then the right side is usually called the *line integral* of  $\vec{F}$  over  $C$  and denoted  $\oint \vec{F} d\gamma$ .

Next on  $\mathbf{R}^3$ , consider a compactly supported 2-form

$$\omega = f_1 dx_2 \wedge dx_3 + f_2 dx_3 \wedge dx_1 + f_3 dx_1 \wedge dx_2.$$

We integrate  $\omega$  over a surface  $S$ , which, for simplicity, we assume to be the graph of a function.  $G: \mathbf{R}^2 \rightarrow \mathbf{R}$ ,  $x_3 = G(x_1, x_2)$ . (See Figure 4-1.) (This assumption is not really restrictive because, locally, any surface may be written as the graph of a function, although one must sometimes write  $x_1$  or  $x_2$  as a function of the other two coordinates rather than the usual  $x_3$ .)

What is  $\int_S \omega$ ? We can choose for  $S$  the parametrization  $h: \mathbf{R}^2 \rightarrow S$  defined by

$$h(x_1, x_2) = (x_1, x_2, G(x_1, x_2)).$$

Compute:

$$\begin{aligned} h^* dx_1 \wedge dx_2 &= dx_1 \wedge dx_2; \\ h^* dx_2 \wedge dx_3 &= dx_2 \wedge dG = dx_2 \wedge \left( \frac{\partial G}{\partial x_1} dx_1 + \frac{\partial G}{\partial x_2} dx_2 \right) \\ &= -\frac{\partial G}{\partial x_1} dx_1 \wedge dx_2; \end{aligned}$$

and, similarly,

$$h^* dx_3 \wedge dx_1 = -\frac{\partial G}{\partial x_2} dx_1 \wedge dx_2.$$

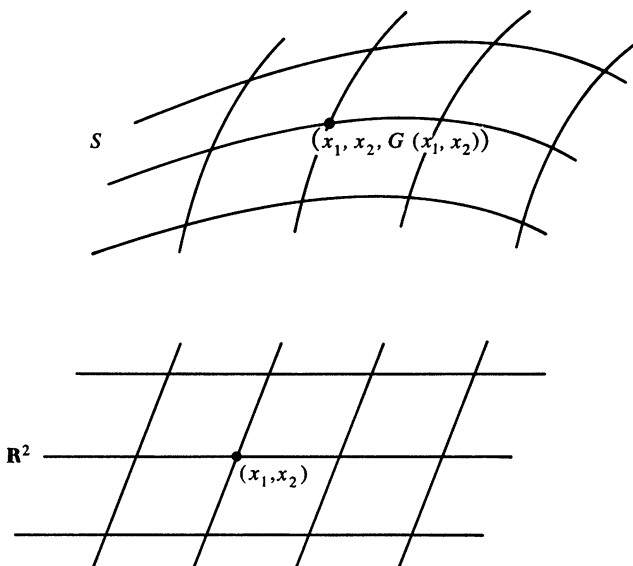


Figure 4-1

We emerge with the formula

$$\int_S \omega = \int_{\mathbb{R}^2} (n_1 f_1 + n_2 f_2 + n_3 f_3) dx_1 dx_2,$$

where

$$\vec{n} = (n_1, n_2, n_3) = \left( -\frac{\partial G}{\partial x_1}, -\frac{\partial G}{\partial x_2}, 1 \right).$$

Check that at any point  $x = (x_1, x_2, G(x_1, x_2))$  in  $S$ , the vector  $\vec{n}(x)$  is normal to the surface  $S$ ; that is,  $\vec{n}(x) \perp T_x(S)$ . We can rewrite the integral in a form you probably learned in second-year calculus. Let  $\vec{u} = \vec{n}/|\vec{n}|$  be the unit normal vector, let  $\vec{F} = (f_1, f_2, f_3)$ , and define a smooth 2-form  $dA = |\vec{n}| dx_1 \wedge dx_2$ . Then

$$\int_S \omega = \int_{\mathbb{R}^2} (\vec{F} \cdot \vec{u}) dA.$$

The form  $dA$  is usually called the *area form* of the surface  $S$ , a title you may motivate for yourself with a little exercise. First show that if  $\Delta S$  is a small rectangle in  $\mathbb{R}^3$  and  $\Delta S'$  is its projection onto the  $x_1 x_2$  plane, then

$$\text{Area}(\Delta S) = (\sec \theta) \cdot \text{area}(\Delta S'),$$

where  $\theta$  is the angle between the normal to  $\Delta S$  and the  $x_3$  axis. [First slide  $\Delta S$  on its plane until one edge is parallel to the intersection of this plane with the  $x_1x_2$  plane; then the formula is easy. Now show that for our surface

$$|\vec{n}| = \sqrt{1 + \left(\frac{\partial G}{\partial x_1}\right)^2 + \left(\frac{\partial G}{\partial x_2}\right)^2}$$

equals the secant of the angle  $\theta$  between the normal vector to  $S$  and the  $x_3$  axis. Finally, note that the area of  $\Delta S'$  is the integral of  $dx_1 \wedge dx_2$  over  $\Delta S'$ .] (See Figure 4-2.)

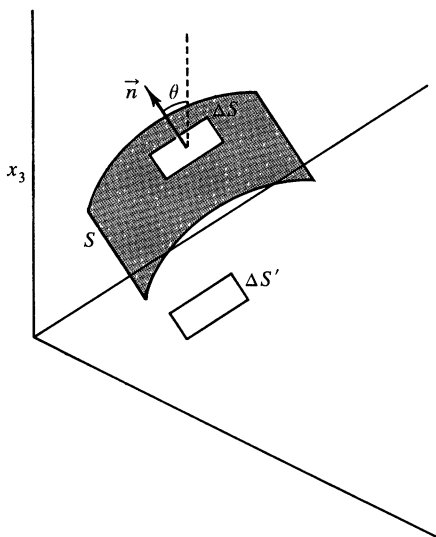


Figure 4-2

## EXERCISES

1. Let  $Z$  be a finite set of points in  $X$ , considered as a 0-manifold. Fix an orientation of  $Z$ , an assignment of orientation numbers  $\sigma(z) = \pm 1$  to each  $z \in Z$ . Let  $f$  be any function on  $X$ , considered as a 0-form, and check that

$$\int_Z f = \sum_{z \in Z} \sigma(z) \cdot f(z).$$

2. Let  $X$  be an oriented  $k$ -dimensional manifold with boundary, and  $\omega$  a compactly supported  $k$ -form on  $X$ . Recall that  $-X$  designates the oriented manifold obtained simply by reversing the orientation on  $X$ .

Check that

$$\int_{-x} \omega = - \int_x \omega.$$

3. Let  $c: [a, b] \rightarrow X$  be a smooth curve, and let  $c(a) = p$ ,  $c(b) = q$ . Show that if  $\omega$  is the differential of a function on  $X$ ,  $\omega = df$ , then

$$\int_a^b c^* \omega = f(q) - f(p).$$

4. Let  $c: [a, b] \rightarrow X$  be a smooth curve, and let  $f: [a_1, b_1] \rightarrow [a, b]$  be a smooth map with  $f(a_1) = a$  and  $f(b_1) = b$ . Show that the integrals

$$\int_a^b c^* \omega \quad \text{and} \quad \int_{a_1}^{b_1} (c \circ f)^* \omega$$

are the same (i.e.,  $\int_c \omega$  is independent of orientation-preserving reparametrization of  $c$ ).

- \*5. A *closed curve* on a manifold  $X$  is a smooth map  $\gamma: S^1 \rightarrow X$ . If  $\omega$  is a 1-form on  $X$ , define the *line integral* of  $\omega$  around  $\gamma$  by

$$\oint_{\gamma} \omega = \int_{S^1} \gamma^*(\omega).$$

For the case  $X = \mathbf{R}^k$ , write  $\oint_{\gamma} \omega$  explicitly in terms of the coordinate expressions of  $\gamma$  and  $\omega$ .

6. Let  $h: \mathbf{R}^1 \rightarrow S^1$  be  $h(t) = (\cos t, \sin t)$ . Show that if  $\omega$  is any 1-form on  $S^1$ , then

$$\int_{S^1} \omega = \int_0^{2\pi} h^* \omega.$$

- \*7. Suppose that the 1-form  $\omega$  on  $X$  is the differential of a function,  $\omega = df$ . Prove that  $\oint_{\gamma} \omega = 0$  for all closed curves  $\gamma$  on  $X$ . [HINT: Exercises 3 and 5.]

- \*8. Define a 1-form  $\omega$  on the punctured plane  $\mathbf{R}^2 - \{0\}$  by

$$\omega(x, y) = \left( \frac{-y}{x^2 + y^2} \right) dx + \left( \frac{x}{x^2 + y^2} \right) dy.$$

- (a) Calculate  $\int_C \omega$  for any circle  $C$  of radius  $r$  around the origin.

- (b) Prove that in the half-plane  $\{x > 0\}$ ,  $\omega$  is the differential of a function. [HINT: Try  $\arctan(y/x)$  as a random possibility.]  
 (c) Why isn't  $\omega$  the differential of a function globally on  $\mathbf{R}^2 - \{0\}$ ?

- \*9. Prove that a 1-form  $\omega$  on  $S^1$  is the differential of a function if and only if  $\int_{S^1} \omega = 0$ . [HINT: "Only if" follows from Exercise 6. Now let  $h$  be as in Exercise 5, and define a function  $g$  on  $\mathbf{R}$  by

$$g(t) = \int_0^t h^* \omega.$$

Show that if  $\int_{S^1} \omega = 0$ , then  $g(t + 2\pi) = g(t)$ . Therefore  $g = f \circ h$  for some function  $f$  on  $S^1$ . Check  $df = \omega$ .]

- \*10. Let  $\nu$  be any 1-form on  $S^1$  with nonzero integral. Prove that if  $\omega$  is any other 1-form, then there exists a constant  $c$  such that  $\omega - c\nu = df$  for some function  $f$  on  $S^1$ .

11. Suppose that  $\omega$  is a 1-form on the connected manifold  $X$ , with the property that  $\oint_\gamma \omega = 0$  for all closed curves  $\gamma$ . Then if  $p, q \in X$ , define  $\int_p^q \omega$  to be  $\int_0^1 c^* \omega$  for a curve  $c: [0, 1] \rightarrow X$  with  $c(0) = p$ ,  $c(1) = q$ . Show that this is well defined (i.e., independent of the choice of  $c$ ). [HINT: You can paste any two such curves together to form a closed curve, using a trick first to make the curves constant in neighborhoods of zero and one. For this last bit, use Exercise 4.]

- \*12. Prove that any 1-form  $\omega$  on  $X$  with the property  $\oint_\gamma \omega = 0$  for all closed curves  $\gamma$  is the differential of a function,  $\omega = df$ . [HINT: Show that the connected case suffices. Now pick  $p \in X$  and define  $f(x) = \int_p^x \omega$ . Check that  $df = \omega$  by calculating  $f$  in a coordinate system on a neighborhood  $U$  of  $x$ . Note that you can work entirely in  $U$  by picking some  $p' \in U$ , for  $f(x) = f(p') + \int_{p'}^x \omega$ .]

13. Let  $S$  be an oriented two-manifold in  $\mathbf{R}^3$ , and let  $\vec{n}(x) = (n_1(x), n_2(x), n_3(x))$  be the outward unit normal to  $S$  at  $x$ . (See Exercise 19 of Chapter 3, Section 2 for definition.) Define a 2-form  $dA$  on  $S$  by

$$dA = n_1 dx_2 \wedge dx_3 + n_2 dx_3 \wedge dx_1 + n_3 dx_1 \wedge dx_2.$$

(Here each  $dx_i$  is restricted to  $S$ .) Show that when  $S$  is the graph of a

function  $F: \mathbf{R}^2 \rightarrow \mathbf{R}$ , with orientation induced from  $\mathbf{R}^2$ , then this  $dA$  is the same as that defined in the text.

14. Let

$$\omega = f_1 dx_2 \wedge dx_3 + f_2 dx_3 \wedge dx_1 + f_3 dx_1 \wedge dx_2$$

be an arbitrary 2-form in  $\mathbf{R}^3$ . Check that the restriction of  $\omega$  to  $S$  is the form  $(\vec{F} \cdot \vec{n}) dA$ , where

$$\vec{F}(x) = (f_1(x), f_2(x), f_3(x)).$$

[HINT: Check directly that if  $u, v \in T_x(S) \subset \mathbf{R}^3$ , then  $\omega(x)(u, v)$  equals one half the determinant of the matrix

$$\begin{pmatrix} \vec{F}(x) \\ u \\ v \end{pmatrix}.$$

If  $\vec{F}(x) \in T_x(S)$ , this determinant is zero, so only the normal component of  $\vec{F}(x)$  contributes.]

## §5 Exterior Derivative

Forms cannot only be integrated, they can also be differentiated. We have already seen how to do this in general for 0-forms, obtaining from a smooth function  $f$  the 1-form  $df$ . In Euclidean space, it is obvious how to continue. If  $\omega = \sum a_I dx_I$  is a smooth  $p$ -form on an open subset of  $\mathbf{R}^k$ , we simply differentiate its coefficient functions. Define the *exterior derivative* of  $\omega$  to be the  $(p+1)$  form  $d\omega = \sum da_I \wedge dx_I$ . The following theorem lists the most important properties of this definition.

**Theorem.** The exterior differentiation operator  $d$ , defined on smooth forms on the open  $U \subset \mathbf{R}^k$  (or  $H^k$ ), possesses the following three properties:

1. Linearity:

$$d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2.$$

2. The Multiplication Law:

$$d(\omega \wedge \theta) = (d\omega) \wedge \theta + (-1)^p \omega \wedge d\theta$$

if  $\omega$  is a  $p$ -form.



## 3. The Cocycle Condition:

$$d(d\omega) = 0.$$

Furthermore, this is the only operator that exhibits these properties and agrees with the previous definition of  $df$  for smooth functions  $f$ .

*Proof.* Part (a) is obvious; parts (b) and (c) are computational. We will write out (c), leaving (b) as an exercise. If  $\omega = \sum_I a_I dx_I$ , then

$$d\omega = \sum_I da_I \wedge dx_I = \sum_I \left( \sum_i \frac{\partial a_I}{\partial x_i} dx_i \right) \wedge dx_I.$$

Then

$$d(d\omega) = \sum_I \sum_i \left( \sum_j \frac{\partial^2 a_I}{\partial x_i \partial x_j} dx_j \right) \wedge dx_i \wedge dx_I.$$

Using the fact that

$$\frac{\partial^2 a_I}{\partial x_i \partial x_j} = \frac{\partial^2 a_I}{\partial x_j \partial x_i},$$

but

$$dx_j \wedge dx_i = -dx_i \wedge dx_j,$$

we cancel terms in the sum two by two, showing  $d(d\omega) = 0$ .

Uniqueness follows easily. Suppose that  $D$  were another operator satisfying (a), (b), (c), and such that  $Df = df$  for functions. Then  $D(dx_i) = 0$ . For by (b),

$$D(dx_{i_1} \wedge \cdots \wedge dx_{i_p}) = \sum_j \pm dx_{i_1} \wedge \cdots \wedge D dx_{i_j} \wedge \cdots \wedge dx_{i_p},$$

but

$$D(dx_{i_j}) = D(Dx_{i_j}) = 0.$$

Now let  $\omega = \sum_I a_I dx_I$  be any  $p$ -form. Then by (a) and (b),

$$D\omega = \sum_I [D(a_I) \wedge dx_I + a_I D(dx_I)].$$

Since  $D(dx_I) = 0$  and  $D(a_I) = da_I$ ,  $D\omega = d\omega$ . Q.E.D.

**Corollary.** Suppose that  $g: V \rightarrow U$  is a diffeomorphism of open sets of  $\mathbf{R}^k$  (or  $H^k$ ). Then for every form  $\omega$  on  $U$ ,  $d(g^*\omega) = g^*(d\omega)$ .

*Proof.* Just check that the operator  $D = (g^{-1})^* \circ d \circ g^*$  satisfies (a), (b), and (c). You proved the corollary earlier for functions, so  $D$  and  $d$  agree for functions on  $U$ . Consequently,  $D = d$ , or  $d \circ g^* = g^* \circ d$ . Q.E.D.

Just as the natural transformation law

$$\int_V g^* \omega = \int_U \omega$$

allowed us to define integration on manifolds, so the relation  $d \circ g^* = g^* \circ d$  permits us to differentiate forms on manifolds. Suppose that  $\omega$  is a  $p$ -form on  $X$ , a manifold with boundary. We define its exterior derivative  $d\omega$  locally. If  $\phi: U \rightarrow X$  is any local parametrization, define  $d\omega$  on the image set  $\phi(U)$  to be  $(\phi^{-1})^* d(\phi^* \omega)$ . If  $\psi: V \rightarrow X$  is another parametrization with overlapping image, then on the overlap we know

$$(\phi^{-1})^* d(\phi^* \omega) = (\psi^{-1})^* d(\psi^* \omega).$$

For set  $g = \phi^{-1} \circ \psi$ . By the corollary,

$$g^* d(\phi^* \omega) = d(g^* \phi^* \omega) = d(\psi^* \omega),$$

so

$$(\psi^{-1})^* d(\psi^* \omega) = (\psi^{-1})^* g^* d(\phi^* \omega) = (\phi^{-1})^* d(\phi^* \omega).$$

As every point of  $X$  lies inside the image of some parametrization,  $d\omega$  is a well defined  $(p+1)$  form globally defined on  $X$ . We leave you to translate the properties of the exterior derivative on Euclidean space into the general setting, proving the next theorem.

**Theorem.** The exterior differentiation operator defined for forms on arbitrary manifolds with boundary has the following properties:

1.  $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$ .
2.  $d(\omega \wedge \theta) = (d\omega) \wedge \theta + (-1)^p \omega \wedge d\theta$ , where  $\omega$  is a  $p$ -form.
3.  $d(d\omega) = 0$ .
4. If  $f$  is a function,  $df$  agrees with the earlier definition.
5. If  $g: Y \rightarrow X$  is a diffeomorphism, then  $d \circ g^* = g^* \circ d$ .

The two Euclidean operations,  $\int$  and  $d$ , extend to manifolds for analogous reasons: both transform naturally under coordinate change. However, there is actually a great difference in the depth of the two transformation properties. The Change of Variables Theorem for integration (which we invoked in the last section but did not prove) is rather subtle, requiring a precise analysis of the way Euclidean volume is distorted by diffeomorphisms. In contrast, the fact that  $d$  commutes with pullback,  $d \circ g^* = g^* \circ d$ , is a simple consequence

of the definition. Moreover,  $g$  need not be a diffeomorphism; the formula is valid for arbitrary maps.

**Theorem.** Let  $g: Y \rightarrow X$  be any smooth map of manifolds with boundary. Then for every form  $\omega$  on  $X$ ,  $d(g^*\omega) = g^*(d\omega)$ .

*Proof.* When  $\omega$  is a 0-form, you proved the formula at the end of Section 3. It follows also when  $\omega = df$  is the differential of a 0-form, for  $d\omega = 0$  implies  $g^*(d\omega) = 0$ , and

$$d(g^*\omega) = d(g^*df) = d(dg^*f) = 0.$$

Furthermore, part (2) of the previous theorem shows that if this theorem holds for some  $\omega$  and  $\theta$ , then it is also valid for  $\omega \wedge \theta$ . But locally, every form on  $X$  is expressible as a wedge product of a 0-form and a number of differentials of 0-forms, since in Euclidean space every form is  $\sum a_I dx_I$ . Because the theorem is local (the two forms  $d(g^*\omega)$  and  $g^*(d\omega)$  are equal if they are equal in a neighborhood of every point), we are done. Q.E.D.

Before closing this section, let us calculate completely the operator  $d$  in  $\mathbf{R}^3$ . In effect, it is probably quite well known to you, although expressed in terms of vector fields rather than forms.

**0-forms.** If  $f$  is a function on  $\mathbf{R}^3$ , then

$$df = g_1 dx_1 + g_2 dx_2 + g_3 dx_3,$$

where

$$(g_1, g_2, g_3) = \left( \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3} \right) = \overrightarrow{\text{grad}}(f),$$

the *gradient* vector field of  $f$ .

**1-forms.** Let

$$\omega = f_1 dx_1 + f_2 dx_2 + f_3 dx_3.$$

Then

$$\begin{aligned} d\omega &= df_1 \wedge dx_1 + df_2 \wedge dx_2 + df_3 \wedge dx_3 \\ &= g_1 dx_2 \wedge dx_3 + g_2 dx_3 \wedge dx_1 + g_3 dx_1 \wedge dx_2, \end{aligned}$$

where

$$g_1 = \frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}, \quad g_2 = \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}, \quad g_3 = \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}.$$

If we define  $\vec{F}$  and  $\vec{G}$  to be the vector fields  $(f_1, f_2, f_3)$  and  $(g_1, g_2, g_3)$ , then  $\vec{G} = \text{curl } \vec{F}$ .

**2-forms.** For

$$\begin{aligned}\omega &= f_1 dx_2 \wedge dx_3 + f_2 dx_3 \wedge dx_1 + f_3 dx_1 \wedge dx_2, \\ d\omega &= \left( \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3} \right) dx_1 \wedge dx_2 \wedge dx_3 \\ &= (\text{div } \vec{F}) dx_1 \wedge dx_2 \wedge dx_3.\end{aligned}$$

**3-forms.**  $d(\text{any 3-form}) = 0$ . (Why?)

So the classical operators of vector calculus in 3-space—the gradient, curl, and divergence—are really the  $d$  operator in vector field form. Show that the cocycle condition  $d^2 = 0$  on  $R^3$  is equivalent to the two famous formulas  $\text{curl}(\text{grad } f) = 0$  and  $\text{div}(\text{curl } \vec{F}) = 0$ .

## EXERCISES

1. Calculate the exterior derivatives of the following forms in  $R^3$ :

- (a)  $z^2 dx \wedge dy + (z^2 + 2y) dx \wedge dz$ .
- (b)  $13x dx + y^2 dy + xyz dz$ .
- (c)  $f dg$ , where  $f$  and  $g$  are functions.
- (d)  $(x + 2y^3)(dz \wedge dx + \frac{1}{2}dy \wedge dx)$ .

2. Show that the vector field

$$\left( \frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right)$$

has curl zero, but that it cannot be written as the gradient of any function.

## §6 Cohomology with Forms†

A  $p$ -form  $\omega$  on  $X$  is *closed* if  $d\omega = 0$  and *exact* if  $\omega = d\theta$  for some  $(p-1)$  form  $\theta$  on  $X$ . Exact forms are all closed, since  $d^2 = 0$ , but it may not be true that closed forms are all exact. In fact, whether or not closed forms on  $X$  are actually exact turns out to be a purely topological matter. You are probably familiar with this from calculus, although perhaps expressed

†This section is not required reading for subsequent sections, but some later exercises do refer to it.

in vector field language. All gradient vector fields have curl zero, but the converse depends on the domain of definition, as the preceding Exercise 2 shows. In the language of forms, Exercise 2 says that the 1-form

$$\omega = \left( \frac{-y}{x^2 + y^2} \right) dx + \left( \frac{x}{x^2 + y^2} \right) dy$$

is closed but not exact.

In this section we shall make a rudimentary study of the closed versus exact distinction. The subject is quite pretty, so, as in the winding number section, we will leave most of the verifications for you. Hopefully, this will compensate for the dearth of interesting exercises in the last few sections.

In order to measure the failure of the implication closed  $\Rightarrow$  exact, we define an equivalence relation on the vector space of closed  $p$ -forms on  $X$ . Two closed  $p$ -forms  $\omega$  and  $\omega'$  are called *cohomologous*, abbreviated  $\omega \sim \omega'$ , if their difference is exact:  $\omega - \omega' = d\theta$ . (Check that this is indeed an equivalence relation.) The set of equivalence classes is denoted  $H^p(X)$ , the  $p$ th *DeRham cohomology group* of  $X$ .  $H^p(X)$  is more than a set; it has a natural real vector space structure. For if  $\omega_1 \sim \omega'_1$  and  $\omega_2 \sim \omega'_2$ , then  $\omega_1 + \omega_2 \sim \omega'_1 + \omega'_2$ ; also, if  $c \in \mathbf{R}$ , then  $c\omega_1 \sim c\omega'_1$ . Thus the vector space operations on closed  $p$ -forms naturally define addition and scalar multiplication of cohomology classes. The 0 cohomology class in the vector space  $H^p(X)$  is just the collection of exact forms, since  $\omega + d\theta \sim \omega$  always.

Suppose that  $f: X \rightarrow Y$  is a smooth map, so that  $f^*$  pulls  $p$ -forms on  $Y$  back to  $p$ -forms on  $X$ . Use the relation  $f^* \circ d = d \circ f^*$  to check that  $f^*$  pulls back closed forms to closed forms and exact forms to exact forms. In fact, if  $\omega \sim \omega'$ , then  $f^*\omega \sim f^*\omega'$ . Thus  $f^*$  pulls cohomology classes on  $Y$  back to cohomology classes on  $X$ ; that is,  $f^*$  defines a mapping  $f^\#: H^p(Y) \rightarrow H^p(X)$ . Since  $f^*$  is linear, you can easily check that  $f^\#$  is linear. (Remember that  $f^\#$  pulls back: i.e., when  $f: X \rightarrow Y$ , then  $f^\#: H^p(Y) \rightarrow H^p(X)$ .)

The numbered statements in the following discussion are for you to prove.

1. If  $X \xrightarrow{f} Y \xrightarrow{g} Z$ , then  $(g \circ f)^\# = f^\# \circ g^\#$ .

In some simple cases, we can easily compute  $H^p(X)$ . For example,  $H^p(X) = 0$  for all  $p > \dim X$ . The next easiest case is

2. The dimension of  $H^0(X)$  equals the number of connected components in  $X$ . [HINT: There are no exact zero forms. Show that a zero form—that is, a function—is closed if and only if it is constant on each component of  $X$ .]

In obtaining information about other cohomology groups, we shall define an operator  $P$  on forms. Just like the operators  $d$  and  $\int$ ,  $P$  is first defined

in Euclidean space and then extended to manifolds by local parametrizations. And as with the earlier operators, the reason  $P$  may be so extended is that it transforms properly under diffeomorphisms.

Suppose that  $U$  is an open set in  $\mathbf{R}^k$  and  $\omega$  is a  $p$ -form on  $\mathbf{R} \times U$ . Then  $\omega$  may be uniquely expressed as a sum

$$\omega = \sum_I f_I(t, x) dt \wedge dx_I + \sum_J g_J(t, x) dx_J. \quad (1)$$

Here  $t$  is the standard coordinate function on  $\mathbf{R}$ ,  $x_1, \dots, x_k$  are the standard coordinate functions on  $\mathbf{R}^k$ ,  $I$  and  $J$  are increasing index sequences, respectively of length  $p - 1$  and  $p$ . The operator  $P$  transforms  $\omega$  into a  $p - 1$  form  $P\omega$  on  $\mathbf{R} \times U$ , defined by

$$P\omega(t, x) = \sum_I \left[ \int_0^t f_I(s, x) ds \right] dx_I.$$

Notice that  $P\omega$  does not involve a  $dt$  term.

Now let  $\phi: V \rightarrow U$  be a diffeomorphism of open subsets of  $\mathbf{R}^k$ , and let  $\Phi: \mathbf{R} \times V \rightarrow \mathbf{R} \times U$  be the diffeomorphism  $\Phi = \text{identity} \times \phi$ . Prove the essential transformation property

3.  $\Phi^*P\omega = P\Phi^*\omega$ . [HINT: This is not difficult if you avoid writing everything in coordinates. Just note that  $\Phi^*dt = dt$  and that  $\Phi^*$  converts each of the two sums in the expression (1) for  $\omega$  into sums of the same type.]

Now copy the arguments used to put  $d$  and  $\int$  on manifolds to prove

4. There exists a unique operator  $P$ , defined for all manifolds  $X$ , that transforms  $p$ -forms on  $\mathbf{R} \times X$  into  $p - 1$  forms on  $\mathbf{R} \times X$  and that satisfies the following two requirements:
- (1) If  $\phi: X \rightarrow Y$  is a diffeomorphism, and  $\Phi = \text{identity} \times \phi$ , then  $\Phi^* \circ P = P \circ \Phi^*$ .
  - (2) If  $X$  is an open subset of  $\mathbf{R}^k$ ,  $P$  is as defined above.

The main attraction of this operator is the following marvelous formula. (No doubt it appears anything but marvelous at first, but wait!)

5. Let  $\pi: \mathbf{R} \times X \rightarrow X$  be the usual projection operator and  $i_a: X \rightarrow \mathbf{R} \times X$  be any embedding  $x \rightarrow (a, x)$ . Then

$$dP\omega + Pd\omega = \omega - \pi^*i_a^*\omega.$$

[HINT: Essentially this is a question of unraveling notation. For example, if  $\omega$  is expressed by the sum (1), then  $\pi^*i_a^*\omega = \sum_J g_J(x, a) dx_J$ .]

The first important consequence of this formula is

**6. (Poincaré Lemma)** The maps

$$i_a^\# : H^p(\mathbf{R} \times X) \rightarrow H^p(X) \quad \text{and} \quad \pi^\# : H^p(X) \rightarrow H^p(\mathbf{R} \times X)$$

are inverses of each other. In particular,  $H^p(\mathbf{R} \times X)$  is isomorphic to  $H^p(X)$ .

[HINT:  $\pi \circ i_a = \text{identity}$ , so Exercise 1 implies  $i_a^\# \circ \pi^\# = \text{identity}$ . For  $\pi^\# \circ i_a^\#$ , interpret Exercise 5 for closed forms  $\omega$ .]

Take  $X$  to be a single point, so  $H^p(X) = 0$  if  $p > 0$ . Now the Poincaré lemma implies by induction:

**Corollary.**  $H^p(\mathbf{R}^k) = 0$  if  $k > 0$ ; that is, every closed  $p$ -form on  $\mathbf{R}^k$  is exact if  $p > 0$ .

A slightly more subtle consequence is

7. If  $f, g : X \rightarrow Y$  are homotopic maps, then  $f^\# = g^\#$ . [HINT: Let  $H : \mathbf{R} \times X \rightarrow Y$  be a smooth map such that  $H(a, x) = f(x)$  and  $H(b, x) = g(x)$ . Then

$$f^\# = i_a^\# \circ H^\# \quad \text{and} \quad g^\# = i_b^\# \circ H^\#.$$

But it is clear from Exercise 6 that  $i_a^\# = i_b^\#$ .]

Now strengthen the corollary to Exercise 6 by proving

8. If  $X$  is contractible, then  $H^p(X) = 0$  for all  $p > 0$ .

We conclude this section with one deeper result.

**Theorem.**  $H^p(S^k)$  is one dimensional for  $p = 0$  and  $p = k$ . For all other  $p$ ,  $H^p(S^k) = 0$  ( $k > 0$ ).

Here is an inductive approach to the theorem. Assume the theorem for  $S^{k-1}$ , and we shall prove it for  $S^k$ . Let  $U_1$  be  $S^k$  minus the south pole, and let  $U_2$  be  $S^k$  minus the north pole. By stereographic projection, both  $U_1$  and  $U_2$  are diffeomorphic to  $\mathbf{R}^{k-1}$ . Prove

9.  $U_1 \cap U_2$  is diffeomorphic to  $\mathbf{R} \times S^{k-1}$ . [HINT: Stereographic projection shows that  $U_1 \cap U_2$  is diffeomorphic to  $\mathbf{R}^{k-1} - \{0\}$ .]

We shall now apply a classical technique of algebraic topology, the “Mayer-Vietoris argument,” to prove the following key fact:

**Proposition.** For  $p > 1$ , the vector spaces  $H^p(U_1 \cup U_2)$  and  $H^{p-1}(U_1 \cap U_2)$  are isomorphic.

Begin with a closed  $p$ -form  $\omega$  on  $U_1 \cup U_2 = S^k$ . Since  $U_1$  is contractible, Exercise 8 implies that the restriction of  $\omega$  to  $U_1$  is exact; thus  $\omega = d\phi_1$  on  $U_1$ . Similarly,  $\omega = d\phi_2$  on  $U_2$ . Now, consider the  $p-1$  form  $\nu = \phi_1 - \phi_2$  on  $U_1 \cap U_2$ . Since  $d\phi_1 = d\phi_2$  on  $U_1 \cap U_2$ ,  $\nu$  is closed. Thus we have a procedure for manufacturing closed  $p-1$  forms on  $U_1 \cap U_2$  out of closed  $p$ -forms on  $U_1 \cup U_2$ .

This procedure is easily reversed. Find functions  $\rho_1$  and  $\rho_2$  on  $S^k$  such that  $\rho_1$  vanishes in a neighborhood of the north pole and  $\rho_2$  vanishes in a neighborhood of the south pole, but  $\rho_1 + \rho_2 = 1$  everywhere. Now given a closed  $p-1$  form  $\nu$  on  $U_1 \cap U_2$ , define the form  $\phi_1$  on  $U_1$  to be  $\rho_1\nu$ . Although  $\nu$  itself may blow up at the north pole,  $\rho_1$  kills it off, so that  $\phi_1$  is smoothly defined on all of  $U_1$ . Similarly, set  $\phi_2 = -\rho_2\nu$  on  $U_2$ . Note that  $\phi_1 - \phi_2 = \nu$  on  $U_1 \cap U_2$ . Define a  $p$ -form  $\omega$  on  $U_1 \cup U_2$  by setting  $\omega = d\phi_1$  on  $U_1$  and  $\omega = d\phi_2$  on  $U_2$ . Since  $d\phi_1 - d\phi_2 = d\nu = 0$  on  $U_1 \cap U_2$ ,  $\omega$  is a well-defined smooth form on all of  $U_1 \cup U_2$ , and it is certainly closed.

10. Show that these two procedures prove the proposition. (What happens for  $p = 1$ ?)

Now we are almost done.

$$H^p(U_1 \cup U_2) = H^p(S^k),$$

and by Exercise 9 and the Poincaré lemma,

$$H^{p-1}(U_1 \cap U_2) \simeq H^{p-1}(S^{k-1}).$$

Thus

$$H^p(S^k) \simeq H^{p-1}(S^{k-1}) \quad \text{if } p > 1.$$

Two items are missing. We must start the induction by proving  $\dim H^1(S^1)$  is 1, and we must supply the missing datum  $H^1(S^k) = 0$  for  $k > 1$ . The latter is an easy consequence of Stokes theorem, so we delay its proof until Exercise 10, Section 7. And you will discover that you have already proved  $H^1(S^1)$  to be one dimensional if you simply collect together Exercises 9 and 10 of Section 4 and Exercise 2 of Section 5.

## §7 Stokes Theorem

There is a remarkable relationship among the operators  $\int$  and  $d$  on forms and the operation  $\partial$  which to each manifold with boundary associates its boundary (remarkable because  $\partial$  is a purely geometric operation



and  $d$  and  $\int$  are purely analytic). In one dimension, the relationship is announced by the fundamental theorems of calculus, and in two and three dimensions it is the subject of the classical theorems of Green, Gauss, and Stokes. In general, suppose that  $X$  is any compact-oriented  $k$ -dimensional manifold with boundary, so  $\partial X$  is a  $k - 1$  dimensional manifold with the boundary orientation.

**The Generalized Stokes Theorem.** If  $\omega$  is any smooth  $(k - 1)$  form on  $X$ , then

$$\int_{\partial X} \omega = \int_X d\omega.$$

*Proof.* Both sides of the equation are linear in  $\omega$ , so we may assume  $\omega$  to have compact support contained in the image of a local parametrization  $h: U \rightarrow X$ ,  $U$  being an open subset of  $\mathbf{R}^k$  or  $H^k$ .

First, assume  $U$  is open in  $\mathbf{R}^k$ ; i.e.,  $h(U)$  does not intersect the boundary. Then

$$\int_{\partial X} \omega = 0 \quad \text{and} \quad \int_X d\omega = \int_U h^*(d\omega) = \int_U dv,$$

where  $v = h^*\omega$ . Since  $v$  is a  $(k - 1)$  form in  $k$ -space, we may write it as

$$v = \sum_{i=1}^k (-1)^{i-1} f_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_k.$$

Here the  $\widehat{dx_i}$  means that the term  $dx_i$  is omitted from the product. Then

$$dv = \left( \sum_i \frac{\partial f_i}{\partial x_i} \right) dx_1 \wedge \cdots \wedge dx_k$$

and

$$\int_{\mathbf{R}^k} dv = \sum_i \int_{\mathbf{R}^k} \frac{\partial f_i}{\partial x_i} dx_1 \cdots dx_k. \quad (2)$$

The integral over  $\mathbf{R}^k$  is computed as usual by an iterated sequence of integrals over  $\mathbf{R}^1$ , which may be taken in any order (Fubini's theorem). Integrate the  $i$ th term first with respect to  $x_i$ :

$$\int_{\mathbf{R}^{k-1}} \left( \int_{-\infty}^{\infty} \frac{\partial f_i}{\partial x_i} dx_i \right) dx_1 \cdots \widehat{dx_i} \cdots dx_k.$$

Of course,

$$\int_{-\infty}^{\infty} \frac{\partial f_i}{\partial x_i} dx_i$$

is the function of  $x_1, \dots, \hat{x}_i, \dots, x_k$  that assign to any  $(k-1)$  tuple  $(b_1, \dots, \hat{b}_i, \dots, b_k)$  the number  $\int_{-\infty}^{\infty} g'(t) dt$ , where  $g(t) = f_i(b_1, \dots, t, \dots, b_k)$ . Since  $\nu$  has compact support,  $g$  vanishes outside any sufficiently large interval  $(-a, a)$  in  $\mathbf{R}^1$ . Therefore the Fundamental Theorem of Calculus gives

$$\int_{-\infty}^{\infty} g'(t) dt = \int_{-a}^a g'(t) dt = g(a) - g(-a) = 0 - 0 = 0.$$

Thus  $\int_X d\omega = 0$ , as desired.

When  $U \subset H^k$ , the preceding analysis works for every term of (2) except the last. Since the boundary of  $H^k$  is the set where  $x_k = 0$ , the last integral is

$$\int_{\mathbf{R}^{k-1}} \left( \int_0^{\infty} \frac{\partial f_k}{\partial x_k} dx_k \right) dx_1 \cdots dx_{k-1}.$$

Now compact support implies that  $f_k$  vanishes if  $x_k$  is outside some large interval  $(0, a)$ , but although  $f_k(x_1, \dots, x_{k-1}, a) = 0$ ,  $f_k(x_1, \dots, x_{k-1}, 0) \neq 0$ . Thus applying the Fundamental Theorem of Calculus as above, we obtain

$$\int_X d\omega = \int_{\mathbf{R}^{k-1}} -f_k(x_1, \dots, x_{k-1}, 0) dx_1 \cdots dx_{k-1}.$$

On the other hand,

$$\int_{\partial X} \omega = \int_{\partial H^k} \nu.$$

Since  $x_k = 0$  on  $\partial H^k$ ,  $dx_k = 0$  on  $\partial H^k$  as well. Consequently, if  $i < k$ , the form  $(-1)^{i-1} f_i dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx^k$  restricts to 0 on  $\partial H^k$ . So the restriction of  $\nu$  to  $\partial H^k$  is  $(-1)^{k-1} f(x_1, \dots, x_{k-1}, 0) dx_1 \wedge \cdots \wedge dx_{k-1}$ , whose integral over  $\partial H^k$  is therefore  $\int_{\partial X} \omega$ .

Now  $\partial H^k$  is naturally diffeomorphic to  $\mathbf{R}^{k-1}$  under the obvious map  $(x_1, \dots, x_{k-1}) \rightarrow (x_1, \dots, x_{k-1}, 0)$ , but this diffeomorphism does not always carry the usual orientation of  $\mathbf{R}^{k-1}$  to the boundary orientation of  $\partial H^k$ . Let  $e_1, \dots, e_k$  be the standard ordered basis for  $\mathbf{R}^k$ , so  $e_1, \dots, e_{k-1}$  is the standard ordered basis for  $\mathbf{R}^{k-1}$ . Since  $H^k$  is the upper half-space, the outward unit normal to  $\partial H^k$  is  $-e_k = (0, \dots, 0, -1)$ . Thus in the boundary orientation of  $\partial H^k$ , the sign of the ordered basis  $\{e_1, \dots, e_{k-1}\}$  is defined to be the sign of the ordered basis  $\{-e_k, e_1, \dots, e_{k-1}\}$  in the standard orientation of  $H^k$ . The latter sign is easily seen to be  $(-1)^k$ , so the usual diffeomorphism  $\mathbf{R}^k \rightarrow \partial H^k$  alters orientation by this factor  $(-1)^k$ .

The result is the formula

$$\begin{aligned}\int_{\partial X} \omega &= \int_{\partial H^k} (-1)^{k-1} f_k(x_1, \dots, x_{k-1}, 0) dx_1 \cdots dx_{k-1} \\ &= (-1)^k \int_{\mathbf{R}^{k-1}} (-1)^{k-1} f_k(x_1, \dots, x_{k-1}, 0) dx_1 \cdots dx_{k-1}.\end{aligned}$$

Since  $(-1)^k(-1)^{k-1} = -1$ , it is exactly the formula we derived for  $\int_x d\omega$ .  
Q.E.D.

The classical versions of Stokes theorem are included in the exercises, along with some typical applications. But, for us, the essential value of Stokes theorem is that it provides a fundamental link between analysis and topology. The final sections of the book are devoted to exploring this link.

### EXERCISES

1. Show that Stokes theorem for a closed interval  $[a, b]$  in  $\mathbf{R}^1$  is just the Fundamental Theorem of Calculus. (See Exercise 1, Section 4.)
2. Prove the classical Green's formula in the plane: let  $W$  be a compact domain in  $\mathbf{R}^2$  with smooth boundary  $\partial W = \gamma$ . Prove

$$\int_{\gamma} f dx + g dy = \int_W \left( \frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx dy.$$

3. Prove the Divergence Theorem: let  $W$  be a compact domain in  $\mathbf{R}^3$  with smooth boundary, and let  $\vec{F} = (f_1, f_2, f_3)$  be a smooth vector field on  $W$ . Then

$$\int_W (\operatorname{div} \vec{F}) dx dy dz = \int_{\partial W} (\vec{n} \cdot \vec{F}) dA.$$

(Here  $\vec{n}$  is the outward normal to  $\partial W$ . See Exercises 13 and 14 of Section 4 for  $dA$ , and page 178 for  $\operatorname{div} \vec{F}$ .)

4. Prove the classical Stokes theorem: let  $S$  be a compact oriented two-manifold in  $\mathbf{R}^3$  with boundary, and let  $\vec{F} = (f_1, f_2, f_3)$  be a smooth vector field in a neighborhood of  $S$ . Prove

$$\int_S (\operatorname{curl} \vec{F} \cdot \vec{n}) dA = \int_{\partial S} f_1 dx_1 + f_2 dx_2 + f_3 dx_3.$$

(Here  $\vec{n}$  is the outward normal to  $S$ . For  $dA$ , see Exercises 13 and 14 of Section 4, and for  $\operatorname{curl} \vec{F}$ , see page 178.)

5. The Divergence Theorem has an interesting interpretation in fluid dynamics. Let  $D$  be a compact domain in  $\mathbf{R}^3$  with a smooth boundary  $S = \partial D$ . We assume that  $D$  is filled with an incompressible fluid whose density at  $x$  is  $\rho(x)$  and whose velocity is  $\vec{v}(x)$ . The quantity

$$\int_S \rho(\vec{v} \cdot \vec{n}) dA$$

measures the amount of fluid flowing out of  $S$  at any fixed time. If  $x \in D$  and  $B_\epsilon$  is the ball of radius  $\epsilon$  about  $x$ , the “infinitesimal amount” of fluid being added to  $D$  at  $x$  at any fixed time is

$$(*) \quad \lim_{\epsilon \rightarrow 0} \int_{\partial B_\epsilon} \frac{\rho(\vec{v} \cdot \vec{n}) dA}{\text{vol}(B_\epsilon)}.$$

Show that  $(*) = \text{div } \rho \vec{v}$ , and deduce from the Divergence Theorem that the amount of fluid flowing out of  $D$  at any fixed time equals the amount being added.

6. The Divergence Theorem is also useful in electrostatics. Let  $D$  be a compact region in  $\mathbf{R}^3$  with a smooth boundary  $S$ . Assume  $0 \in \text{Int}(D)$ . If an electric charge of magnitude  $q$  is placed at 0, the resulting force field is  $q\vec{r}/r^3$ , where  $\vec{r}(x)$  is the vector to a point  $x$  from 0 and  $r(x)$  is its magnitude. Show that the amount of charge  $q$  can be determined from the force on the boundary by proving Gauss's law:

$$\int_S \vec{F} \cdot \vec{n} dA = 4\pi q.$$

[HINT: Apply the Divergence Theorem to a region consisting of  $D$  minus a small ball around the origin.]

- \*7. Let  $X$  be compact and boundaryless, and let  $\omega$  be an exact  $k$ -form on  $X$ , where  $k = \dim X$ . Prove that  $\int_X \omega = 0$ . [HINT: Apply Stokes theorem. Remember that  $X$  is a manifold with boundary, even though  $\partial X$  is empty.]
- \*8. Suppose that  $X = \partial W$ ,  $W$  is compact, and  $f: X \rightarrow Y$  is a smooth map. Let  $\omega$  be a closed  $k$ -form on  $Y$ , where  $k = \dim X$ . Prove that if  $f$  extends to all of  $W$ , then  $\int_X f^* \omega = 0$ .
- \*9. Suppose that  $f_0, f_1: X \rightarrow Y$  are homotopic maps and that the compact, boundaryless manifold  $X$  has dimension  $k$ . Prove that for all closed  $k$ -

forms  $\omega$  on  $Y$ ,

$$\int_X f_0^* \omega = \int_X f_1^* \omega.$$

[HINT: Previous exercise.]

10. Show that if  $X$  is a simply connected manifold, then  $\oint_\gamma \omega = 0$  for all closed 1-forms  $\omega$  on  $X$  and all closed curves  $\gamma$  in  $X$ . [HINT: Previous exercise.]
11. Prove that if  $X$  is simply connected, then all closed 1-forms  $\omega$  on  $X$  are exact. (See Exercise 11, Section 4.)
12. Conclude from Exercise 11 that  $H^1(S^k) = 0$  if  $k > 1$ .
13. Suppose that  $Z_0$  and  $Z_1$  are compact, cobordant,  $p$ -dimensional submanifolds of  $X$ . Prove that

$$\int_{Z_1} \omega = \int_{Z_2} \omega$$

for every closed  $p$ -form  $\omega$  on  $X$ .

14. (a) Suppose that  $\omega_1$  and  $\omega_2$  are cohomologous  $p$ -forms on  $X$ , and  $Z$  is a compact oriented  $p$ -dimensional submanifold. Prove that

$$\int_Z \omega_1 = \int_Z \omega_2.$$

- (b) Conclude that integration over  $Z$  defines a map of the cohomology group  $H^p(X)$  into  $\mathbf{R}$ , which we denote by

$$\int_Z : H^p(X) \rightarrow \mathbf{R}.$$

Check that  $\int_Z$  is a linear functional on  $H^p(X)$ .

- (c) Suppose that  $Z$  bounds; specifically, assume that  $Z$  is the boundary of some compact oriented  $p+1$  dimensional submanifold-with-boundary in  $X$ . Show that  $\int_Z$  is zero on  $H^p(X)$ .
- (d) Show that if the two compact oriented manifolds  $Z_1$  and  $Z_2$  in  $X$  are cobordant, then the two linear functionals  $\int_{Z_1}$  and  $\int_{Z_2}$  are equal.

## §8 Integration and Mappings

Our primary application of Stokes theorem is the following theorem, which relates the analytic operation of integration to the topological behavior of mappings.

**Degree Formula.** Let  $f: X \rightarrow Y$  be an arbitrary smooth map of two compact, oriented manifolds of dimension  $k$ , and let  $\omega$  be a  $k$ -form on  $Y$ . Then

$$\int_X f^* \omega = \deg(f) \int_Y \omega.$$

Thus the mapping  $f$  alters the integral of every form by an integer multiple, that integer being the purely topological invariant  $\deg(f)$ . The theorem has numerous applications, one simply being the argument principle described in the chapter introduction. Let

$$p(z) = z^m + a_1 z^{m-1} + \cdots + a_m$$

be a complex polynomial, and  $\Omega$  a region in the plane whose smooth boundary contains no zeros of  $p$ . We wish to prove that the number of zeros of  $p$  inside  $\Omega$ , counting multiplicities, is given by the integral

$$\frac{1}{2\pi} \int_{\partial\Omega} d \arg p(z).$$

We had better interpret the meaning of the integral. Recall that each non-zero complex number  $w$  may be written as  $re^{i\theta}$ , where  $r$  is the norm of  $w$  and  $\theta$  is, by definition, its *argument*. But  $\theta = \arg(w)$  is not really a well-defined function; since  $e^{i\theta} = e^{i(\theta+2\pi)}$ , the values  $\theta + 2\pi n$  all qualify to be  $\arg(w)$ , where  $n$  is any integer. Happily, the ambiguity can be dispelled by passing to the exterior derivative. For in a suitable neighborhood of any point, we can always choose values for  $\arg(w)$  at every point so as to obtain a smooth function of  $w$ ; let's call it  $\arg_0(w)$ . Then every other smooth function  $\phi(w)$  in that neighborhood which satisfies the necessary formula  $w = |w|e^{i\phi(w)}$  equals  $\arg_0(w) + 2\pi n$  for some integer  $n$ . Since  $\phi$  and  $\arg_0$  differ by a constant,  $d\phi = d\arg_0$ . This smooth 1-form, defined on  $\mathbb{C} - \{0\}$ , is what we call  $d \arg$ . Although the notation suggests that it is the differential of a function, we see that this assumption is misleading; only locally do suitable functions exist.

In the argument principal, the integrand is the 1-form

$$z \rightarrow d \arg p(z) = p^*(d \arg).$$

This integrand is defined and smooth on the complex plane minus the zeros of  $p$ . Since we have already proven that the zeros of  $p$  in  $\Omega$  are counted by the degree of the map  $f: \partial\Omega \rightarrow S^1$ , where

$$f(z) = \frac{p(z)}{|p(z)|} = e^{i \arg p(z)},$$

we must identify the integral of  $d \arg p(z)$  with  $2\pi \deg(f)$ . Apply the degree formula to the restriction of the 1-form  $d \arg$  to  $S^1$ . If

$$w = f(z) = e^{i \arg p(z)},$$

then  $\arg(w) = \arg p(z)$ . Since  $\arg p(z)$  is actually a smooth function, at least locally, we get

$$d \arg p(z) = d[f^* \arg(w)] = f^* d \arg(w).$$

Thus

$$\int_{\partial\Omega} d \arg p(z) = \deg(f) \int_{S^1} d \arg(w).$$

Calculating the integral over  $S^1$  is absolutely trivial. Obviously, removal of a single point, say  $w = 1$ , from  $S^1$  will not change the integral. But we may parametrize  $S^1 - \{1\}$  by  $\theta \rightarrow e^{i\theta}$ ,  $\theta \in (0, 2\pi)$ .  $\arg(w)$  is a smooth function on  $S^1 - \{1\}$  that pulls back to the identity  $\theta \rightarrow \theta$  on  $(0, 2\pi)$ . Therefore

$$\int_{S^1} d \arg(w) = \int_0^{2\pi} d\theta = 2\pi,$$

and we are done.

At the heart of the degree formula lies the following theorem, which should remind you strongly of a fundamental property of degree.

**Theorem.** If  $X = \partial W$  and  $f: X \rightarrow Y$  extends smoothly to all of  $W$ , then  $\int_X f^* \omega = 0$  for every  $k$ -form  $\omega$  on  $Y$ . (Here  $X, Y, W$  are compact and oriented and  $k = \dim X = \dim Y$ .)

*Proof.* Let  $F: W \rightarrow Y$  be an extension of  $f$ . Since  $F = f$  on  $X$ ,

$$\int_X f^* \omega = \int_{\partial W} F^* \omega = \int_W F^* d\omega.$$

But  $\omega$  is a  $k$ -form on a  $k$ -dimensional manifold, so  $d\omega = 0$ . (All  $k + 1$  forms on  $k$ -dimensional manifolds are automatically 0.) Q.E.D.

**Corollary.** If  $f_0, f_1: X \rightarrow Y$  are homotopic maps of compact oriented  $k$ -dimensional manifolds, then for every  $k$ -form  $\omega$  on  $Y$

$$\int_X f_0^* \omega = \int_X f_1^* \omega.$$

*Proof.* Let  $F: I \times X \rightarrow Y$  be a homotopy. Now

$$\partial(I \times X) = X_1 - X_0,$$

so

$$0 = \int_{\partial(I \times X)} (\partial F)^* \omega = \int_{X_1} (\partial F)^* \omega - \int_{X_0} (\partial F)^* \omega$$

(0 according to the theorem). But when we identify  $X_0$  and  $X_1$  with  $X$ ,  $\partial F$  becomes  $f_0$  on  $X_0$  and  $f_1$  on  $X_1$ . Q.E.D.

A local version of the degree formula around regular values is very easily established, and its proof shows most concretely the reason why the factor  $\deg(f)$  appears.

**Lemma.** Let  $y$  be a regular value of the map  $f: X \rightarrow Y$  between oriented  $k$ -dimensional manifolds. Then there exists a neighborhood  $U$  of  $y$  such that the degree formula

$$\int_S f^* \omega = \deg(f) \int_Y \omega$$

is valid for every  $k$ -form  $\omega$  with support in  $U$ .

*Proof.* Because  $f$  is a local diffeomorphism at each point in the preimage  $f^{-1}(y)$ ,  $y$  has a neighborhood  $U$  such that  $f^{-1}(U)$  consists of disjoint open sets  $V_1, \dots, V_N$ , and  $f: V_i \rightarrow U$  is a diffeomorphism for each  $i = 1, \dots, N$  (Exercise 7, Chapter 1, Section 4). If  $\omega$  has support in  $U$ , then  $f^* \omega$  has support in  $f^{-1}(U)$ ; thus

$$\int_X f^* \omega = \sum_{i=1}^N \int_{V_i} f^* \omega.$$

But since  $f: V_i \rightarrow U$  is a diffeomorphism, we know that

$$\int_{V_i} f^* \omega = \sigma_i \int_U \omega,$$

the sign  $\sigma_i$  being  $\pm 1$ , depending on whether  $f: V_i \rightarrow U$  preserves or reverses orientation. Now, by definition,  $\deg(f) = \sum \sigma_i$ , so we are done. Q.E.D.



Finally, we prove the degree formula in general. Choose a regular value  $y$  for  $f: X \rightarrow Y$  and a neighborhood  $U$  of  $y$  as in the lemma. By the Isotopy Lemma of Chapter 3, Section 6, for every point  $z \in Y$  we can find a diffeomorphism  $h: Y \rightarrow Y$  that is isotopic to the identity and that carries  $y$  to  $z$ . Thus the collection of all open sets  $h(U)$ , where  $h: Y \rightarrow Y$  is a diffeomorphism isotopic to the identity, covers  $Y$ . By compactness, we can find finitely many maps  $h_1, \dots, h_n$  such that  $Y = h_1(U) \cup \dots \cup h_n(U)$ . Using a partition of unity, we can write any form  $\omega$  as a sum of forms, each having support in one of the sets  $h_i(U)$ ; therefore, since both sides of the degree formula

$$\int_X f^* \omega = \deg(f) \int_Y \omega$$

are linear in  $\omega$ , it suffices to prove the formula for forms supported in some  $h(U)$ .

So assume that  $\omega$  is a form supported in  $h(U)$ . Since  $h \sim \text{identity}$ , then  $h \circ f \sim f$ . Thus the corollary above implies

$$\int_X f^* \omega = \int_X (h \circ f)^* \omega = \int_X f^* h^* \omega.$$

As  $h^* \omega$  is supported in  $U$ , the lemma implies

$$\int_X f^* (h^* \omega) = \deg(f) \int_Y h^* \omega.$$

Finally, the diffeomorphism  $h$  is orientation preserving; for  $h \sim \text{identity}$  implies  $\deg(h) = +1$ . Thus the change of variables property gives

$$\int_Y h^* \omega = \int_Y \omega,$$

and

$$\int_X f^* \omega = \deg(f) \int_Y \omega,$$

as claimed.

## EXERCISES

1. Check that the 1-form  $d \arg$  in  $\mathbb{R}^2 - \{0\}$  is just the form

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

discussed in earlier exercises. [HINT:  $\theta = \arctan (y/x)$ .] (This form is also often denoted  $d\theta$ .) In particular, you have already shown that  $d \arg$  is closed but not exact.

2. Let  $\gamma$  be a smooth closed curve in  $\mathbb{R}^2 - \{0\}$  and  $\omega$  any closed 1-form on  $\mathbb{R}^2 - \{0\}$ . Prove that

$$\oint_{\gamma} \omega = W(\gamma, 0) \int_{S^1} \omega,$$

where  $W(\gamma, 0)$  is the *winding number* of  $\gamma$  with respect to the origin.  $W(\gamma, 0)$  is defined just like  $W_2(\gamma, 0)$ , but using degree rather than degree mod 2; that is,  $W(\gamma, 0) = \deg(\gamma/|\gamma|)$ . In particular, conclude that

$$W(\gamma, 0) = \frac{1}{2\pi} \oint_{\gamma} d \arg.$$

3. We can easily define *complex valued forms* on  $X$ . The forms we have used heretofore are *real forms*. Create imaginary  $p$ -forms by multiplying any real form by  $i = \sqrt{-1}$ . Then complex forms are sums  $\omega_1 + i\omega_2$ , where  $\omega_1$  and  $\omega_2$  are real. Wedge product extends in the obvious way, and  $d$  and  $\int$  are defined to commute with multiplication by  $i$ :

$$d\omega = d\omega_1 + i d\omega_2, \quad \int_X \omega = \int_X \omega_1 + i \int_X \omega_2.$$

Stokes theorem is valid for complex forms, for it is valid for their real and imaginary parts. We can now use our apparatus to prove a fundamental theorem in complex variable theory: the Cauchy Integral Formula.

- (a) Let  $z$  be the standard complex coordinate function on  $\mathbb{C} = \mathbb{R}^2$ . Check that  $dz = dx + i dy$ .
- (b) Let  $f(z)$  be a complex valued function on an open subset  $U$  of  $\mathbb{C}$ . Prove that the 1-form  $f(z) dz$  is closed if and only if  $f(z) = f(x, y)$  satisfies the Cauchy-Riemann equation

$$\frac{\partial f}{\partial \bar{y}} = i \frac{\partial f}{\partial x}.$$

Express  $f$  in terms of its real and imaginary parts  $f = f_1 + if_2$ , and show that the Cauchy-Riemann equation is actually two real equations. If  $f(z) dz$  is closed, the function  $f$  is called *holomorphic* in  $U$ .

- (c) Show that the product of two holomorphic functions is holomorphic.
- (d) Check that the complex coordinate function  $z$  is holomorphic. Conclude that every complex polynomial is holomorphic.

- (e) Suppose that  $f$  is holomorphic in  $U$  and  $\gamma_1, \gamma_2$  are two homotopic closed curves in  $U$ . Prove that

$$\oint_{\gamma_1} f(z) dz = \oint_{\gamma_2} f(z) dz.$$

[HINT: Use Exercise 9, Section 7.]

- (f) If  $f$  is holomorphic in a simply connected region  $U$ , show that  $\oint_{\gamma} f(z) dz = 0$  for every closed curve  $\gamma$  in  $U$ . [HINT: Part (e).]  
 (g) Prove that the function  $f(z) = 1/z$  is holomorphic in the punctured plane  $\mathbf{C} - \{0\}$ . Similarly,  $1/(z - a)$  is holomorphic in  $\mathbf{C} - \{a\}$ .  
 (h) Let  $C_r$  be a circle of radius  $r$  around the point  $a \in \mathbf{C}$ . Prove that

$$\int_{C_r} \frac{1}{z - a} dz = 2\pi i,$$

by direct calculation.

- (i) Suppose that  $f(z)$  is a holomorphic function in  $U$  and  $C_r$  is the circle of radius  $r$  around  $a \in U$ . Prove that

$$\int_{C_r} \frac{f(z)}{z - a} dz = 2\pi i \cdot f(a).$$

[HINT: By part (e), this does not depend on  $r$ . Note that  $|f(z) - f(a)| < \epsilon_r$  on  $C_r$ , where  $\epsilon_r \rightarrow 0$  as  $r \rightarrow 0$ . So let  $r \rightarrow 0$  and use a simple continuity argument.]

- (j) Prove the *Cauchy Integral Formula*: If  $f$  is holomorphic in  $U$  and  $\gamma$  is a closed curve in  $U$  not passing through  $a \in U$ , then

$$\frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{z - a} dz = W(\gamma, a) \cdot f(a).$$

[HINT: Use part (i) and Exercise 2.]

4. Construct a  $k$ -form on  $S^k$  with nonzero integral. [HINT: Construct a compactly supported  $k$ -form in  $\mathbf{R}^k$  with nonzero integral, and project stereographically.]
5. (a) Prove that a closed  $k$ -form  $\omega$  on  $S^k$  is exact if and only if  $\int_{S^k} \omega = 0$ .  
 [HINT:  $\dim H^k(S^k) = 1$ . Now use previous exercise.]  
 (b) Conclude that the linear map  $\int_{S^k} : H^k(S^k) \rightarrow \mathbf{R}$  is an isomorphism.

6. Prove that a compactly supported  $k$ -form  $\omega$  on  $\mathbf{R}^k$  is the exterior derivative of a compactly supported  $k - 1$  form if and only if  $\int_{\mathbf{R}^k} \omega = 0$ . [HINT: Use stereographic projection to carry  $\omega$  to a form  $\omega'$  on  $S^k$ . By Exercise 5,  $\omega' = dv$ . Now  $dv$  is zero in a contractible neighborhood  $U$  of the north pole  $N$ . Use this to find a  $k - 2$  form  $\mu$  on  $S^k$  such that  $v = d\mu$  near  $N$ . Then  $v - d\mu$  is zero near  $N$ , so it pulls back to a compactly supported form on  $\mathbf{R}^k$ .]
7. Show that on any compact oriented  $k$ -dimensional manifold  $X$ , the linear map  $\int_X : H^k(X) \rightarrow \mathbf{R}$  is an isomorphism. In particular, show  $\dim H^k(X) = 1$ . [HINT: Let  $U$  be an open set diffeomorphic to  $\mathbf{R}^k$ , and let  $\omega$  be a  $k$ -form compactly supported in  $U$  with  $\int_X \omega = 1$ . Use Exercise 6 to show that every compactly supported form in  $U$  is cohomologous to a scalar multiple of  $\omega$ . Now choose open sets  $U_1, \dots, U_N$  covering  $X$ , each of which is deformable into  $U$  by a smooth isotopy. Use Exercise 7 of Section 6 and a partition of unity to show that any  $k$ -form  $\theta$  on  $X$  is cohomologous to  $c\omega$  for some  $c \in \mathbf{R}$ . Identify  $c$ .]
8. Let  $f: X \rightarrow Y$  be a smooth map of compact oriented  $k$ -manifolds. Consider the induced map on the top cohomology groups,  $f^*: H^k(Y) \rightarrow H^k(X)$ . Integration provides canonical isomorphisms of both  $H^k(Y)$  and  $H^k(X)$  with  $\mathbf{R}$ , so under these isomorphisms the linear map  $f^*$  must correspond to multiplication by some scalar. Prove that this scalar is the degree of  $f$ . In other words, the following square commutes:

$$\begin{array}{ccc}
 H^k(Y) & \xrightarrow{f^*} & H^k(X) \\
 \int_Y \downarrow & & \downarrow \int_X \\
 \mathbf{R} & \xrightarrow[\deg(f)]{\text{multiplication by}} & \mathbf{R}
 \end{array}$$

## §9 The Gauss-Bonnet Theorem

We begin this section with a discussion of volume. Suppose that  $X$  is a compact oriented  $k$ -dimensional manifold in  $\mathbf{R}^N$ . For each point  $x \in X$ , let  $v_x(x)$  be the volume element on  $T_x(X)$ , the alternating  $k$ -tensor that has value  $1/k!$  on each positively oriented orthonormal basis for  $T_x(X)$ . (See Exercise 10, Section 2.) It is not hard to show that the  $k$ -form  $v_x$  on  $X$  is smooth; it is called the *volume form* of  $X$ . For example, the volume form on

$\mathbf{R}^k$  is just  $dx_1 \wedge \cdots \wedge dx_k$ . The ungraceful  $1/k!$  is necessitated by the way we normalized the definition of wedge product. (Compare with the footnote on page 156). The integral  $\int_X \nu_X$  is defined to be the *volume* of  $X$ .

The volume form is valuable because it provides a means of integrating functions. If  $g$  is a function on  $X$ , then  $g\nu_X$  is a  $k$ -form on  $X$ , thus we can define  $\int_X g$  to be  $\int_X g\nu_X$ . (When  $X$  is  $\mathbf{R}^k$ ,  $\nu_X = dx_1 \wedge \cdots \wedge dx_k$ , so that  $\int_{\mathbf{R}^k} g$  is the usual integral.)

Of course, you realize that the volume form is geometrical, not topological; it depends strongly on the precise manner in which the manifold sits inside Euclidean space. Consequently, integration of functions is not a natural topological operation; it does not transform properly under diffeomorphism.

(For most diffeomorphisms  $h: Y \rightarrow X$ , that is,  $\int_Y h^*g \neq \int_X g$ .)

The reason that integration of functions transforms improperly, as we explained in Section 4, is that diffeomorphisms distort volume. We can quantitatively measure the distortion with the aid of volume forms. In fact, if  $f: X \rightarrow Y$  is any smooth map of two  $k$ -dimensional manifolds with boundary, the pullback  $f^*\nu_Y$  is a  $k$ -form on  $X$ . At every point  $x \in X$ ,  $\nu_X(x)$  is a basis for  $\Lambda^k[T_x(X)^*]$ , so  $(f^*\nu_Y)(x)$  must be a scalar multiple of  $\nu_X(x)$ . This scalar is called the *Jacobian* of  $f$  at  $x$  and is denoted  $J_f(x)$ . For motivation, note that the tensor  $\nu_X(x)$  assigns to the  $k$ -tuple  $(v_1, \dots, v_k)$  plus or minus the volume of the parallelepiped it spans in the vector space  $T_x(X)$ , multiplied by that awkward factor  $1/k!$  (Compare with Exercise 11, Section 2.).  $(f^*\nu_Y)(x)$  assigns  $\pm$  the volume of the parallelepiped spanned by  $df_x(v_1), \dots, df_x(v_k)$  in  $T_{f(x)}(Y)$ , multiplied by the same factor. Thus the magnitude of  $J_f(x)$  is the factor by which  $df_x$  expands or contracts volume; its sign reflects whether  $df_x$  preserves or reverses orientation. In this sense,  $J_f$  measures everywhere the infinitesimal change of volume and orientation effected by  $f$ .

We apply these generalities to study the geometry of *hypersurfaces*,  $k$ -dimensional submanifolds of  $\mathbf{R}^{k+1}$ . Now, the hypersurface  $X$  is oriented if and only if we can smoothly choose between the two unit normal vectors to  $X$  at every point (Exercise 18, Section 2, Chapter 3). If, in particular,  $X$  is a compact hypersurface, we know from the Jordan-Brouwer Separation Theorem that  $X$  is orientable as the boundary of its “inside”; so we can just choose  $\vec{n}(x)$  as the outward pointing normal.

The map  $g: X \rightarrow S^k$ , defined by  $g(x) = \vec{n}(x)$ , is called the *Gauss map* of the oriented hypersurface  $X$ , and its Jacobian  $J_g(x) = \kappa(x)$  is called the *curvature* of  $X$  at  $x$ . For example, if  $X$  is a  $k$ -sphere of radius  $r$ , then  $\kappa(x) = 1/r^k$  everywhere (Exercise 6.) As the radius increases, the curvature decreases, for large spheres are flatter than small ones. Of course, when  $X = \mathbf{R}^k$ , then  $\kappa = 0$ , since  $g$  is constant.

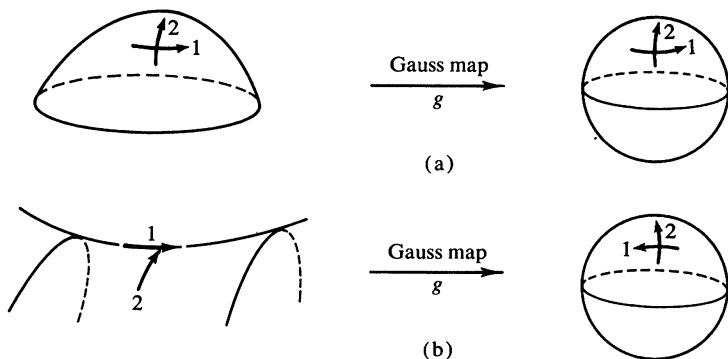


Figure 4-3

Thus the magnitude of  $\kappa(x)$ , in some sense, measures how curved  $X$  is at  $x$ ; the more curved the space, the faster the normal vector turns. For surfaces, the sign of  $\kappa(x)$ , indicating whether or not the Gauss map is disorienting, serves to distinguish between local convex appearance and local saddlelike appearance: in part (a) of Figure 4-3, the Gauss map preserves orientation, while in part (b) it reverses orientation.

The curvature is a strictly geometric characteristic of spaces; it obviously is not preserved by topological transformations. But one of the loveliest theorems in mathematics implies that the global integral of the curvature on compact, even-dimensional hypersurfaces is a topological invariant. Thus no matter how we kick, twist, or stretch the space, all the local changes in curvature must exactly cancel. Moreover, since  $\int_X \kappa$  is a global topological invariant of  $X$ , you will probably not be shocked to learn that it is expressible in terms of (what else?) the Euler characteristic.

**The Gauss-Bonnet Theorem.** If  $X$  is a compact, even-dimensional hypersurface in  $\mathbf{R}^{k+1}$ , then

$$\int_X \kappa = \frac{1}{2} \gamma_k \chi(X)$$

where  $\chi(X)$  is the Euler characteristic of  $X$  and the constant  $\gamma_k$  is the volume of the unit  $k$ -sphere  $S^k$ .

Of course, when  $X$  is odd dimensional, the formula is false, since the Euler characteristic is automatically zero.

The first part of the proof is an application of the degree formula to convert the integral into a topological expression.

$$\int_X \kappa = \int_X J_g \nu_X = \int_X g^* \nu_{S^k} = \deg(g) \int_{S^k} \nu_{S^k} = \deg(g) \cdot \gamma_k.$$

So, in order to prove Gauss-Bonnet, we must show that the degree of the Gauss map equals one-half the Euler characteristic of  $X$ :  $\deg(g) = \frac{1}{2}\chi(X)$ . To do so, we shall use the Poincaré-Hopf theorem.

Choose a unit vector  $a \in S^k$  such that both  $a$  and  $-a$  are regular values of  $g$ . (Why can we do so?) Let  $\vec{v}$  be the vector field on  $X$  whose value at a point  $x$  is the projection of the vector  $-a$  onto  $T_x(X)$ :

$$\vec{v}(x) = (-a) - [-a \cdot \vec{n}(x)]\vec{n}(x) = [a \cdot g(x)]g(x) - a.$$

(See Figure 4-4.) A point  $z \in X$  is a zero of  $\vec{v}$  if and only if  $g(z)$  is a multiple of  $a$ ; that is,  $g(z) = \pm a$ . Since  $a$  and  $-a$  are regular values of  $g$ , and  $X$  is

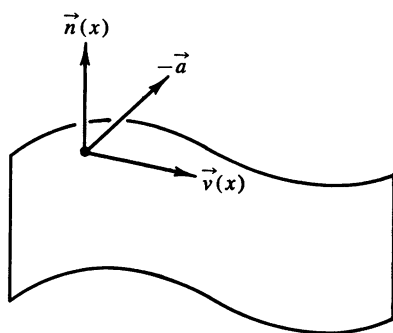


Figure 4-4

compact,  $\vec{v}$  has only finitely many zeros. Let us denote the translation map  $y \mapsto y - a$  in  $\mathbb{R}^{k+1}$  by  $T$ , so we can write the mapping  $\vec{v} : X \rightarrow \mathbb{R}^{k+1}$  as  $\vec{v} = T \circ [a \cdot g]$ .

**Lemma.** If  $g(z) = a$ , then  $d\vec{v}_z = dT_a \circ dg_z$ ; and if  $g(z) = -a$ , then  $d\vec{v}_z = -dT_a \circ dg_z$ .

*Proof.* Calculate the derivative at  $z$  of the map  $f : X \rightarrow \mathbb{R}^{k+1}$  defined by  $f(x) = [a \cdot g(x)]g(x)$ : if  $w \in T_z(X)$ , then the vector  $df_z(w) \in \mathbb{R}^{k+1}$  is

$$df_z(w) = [a \cdot g(z)] dg_z(w) + [a \cdot dg_z(w)]g(z) \quad (3)$$

[To check this, write  $w$  as the tangent to a curve  $c(t)$ , so  $df_z(w)$  is the tangent to the curve  $f(c(t))$ . Now apply the usual product rule to each coordinate of  $f(c(t))$ .] Since  $a \in S^k$ , the product  $a \cdot a$  equals 1; thus the first term in Eq. (3) is  $dg_z(w)$  if  $g(z) = +a$  and  $-dg_z(w)$  if  $g(z) = -a$ . For the second term, differentiate the constant function  $g(x) \cdot g(x) = 1$  to show that  $g(z) \cdot dg_z(w) = 0$ . (This result simply expresses the fact that the image of  $dg_z$  is tangent to

$S^k$  at  $g(z)$ , thus perpendicular to the vector  $g(z)$ .) Consequently, if  $g(z) = \pm a$ ,  $a \cdot dg_z(w) = 0$ , and the second term vanishes. Q.E.D.

**Corollary.** The index of the vector field  $\vec{v}$  at its zero  $z$  is  $+1$  if  $g: X \rightarrow S^k$  preserves orientation at  $z$ , and  $-1$  if  $g$  reverses orientation at  $z$ .

*Proof.* According to Exercises 3 and 5 of Chapter 3, Section 5, the derivative  $d\vec{v}_z: T_z(X) \rightarrow \mathbf{R}^{k+1}$  actually carries  $T_z(X)$  into itself; furthermore, if  $d\vec{v}_z$  is an isomorphism of  $T_z(X)$ , then  $\text{ind}_z(\vec{v})$  equals the sign of the determinant of  $d\vec{v}_z: T_z(X) \rightarrow T_z(X)$ . Now the linear map  $dT_a: \mathbf{R}^{k+1} \rightarrow \mathbf{R}^{k+1}$  is just the identity, so, considered as linear maps of  $T_z(X)$  into  $\mathbf{R}^{k+1}$ ,

$$d\vec{v}_z = \pm dg_z.$$

First, suppose that  $g(z) = +a$ , so  $d\vec{v}_z = +dg_z$ . Since

$$d\vec{v}_z: T_z(X) \longrightarrow T_z(X) \quad \text{and} \quad dg_z: T_z(X) \longrightarrow T_a(S^k),$$

the two subspaces  $T_z(X)$  and  $T_a(S^k)$  must be identical. Moreover, they have the same orientation, for the outward unit normal to  $X$  at  $z$  is  $\vec{n}(z) = g(z) = a$ , while the outward unit normal to  $S^k$  at  $a$  is  $a$  itself. Since  $a$  is a regular value of  $g$ , the linear map  $dg_z = d\vec{v}_z$  is an isomorphism. Since  $\det(d\vec{v}_z) = \det(dg_z)$ , the exercise implies that  $\text{ind}_z(\vec{v}) = +1$  if  $dg_z$  preserves orientation and  $-1$  if  $dg_z$  reverses orientation.

If  $g(z) = -a$ , then  $d\vec{v}_z = -dg_z$ . Again, we conclude that the two subspaces  $T_z(X)$  and  $T_{-a}(S^k)$  in  $\mathbf{R}^{k+1}$  are identical, including their orientations. The regularity of  $-a$  implies that  $d\vec{v}_z$  is an isomorphism, so the exercise applies. Here

$$\det(d\vec{v}_z) = \det(-dg_z) = (-1)^k \det(dg_z),$$

which still equals  $\det(dg_z)$ , because  $k$  is even. So again,  $\text{ind}_z(\vec{v}) = +1$  if  $dg_z$  preserves orientation and  $-1$  if  $dg_z$  reverses orientation. Q.E.D.

The Gauss-Bonnet theorem follows directly. For by the Poincaré-Hopf theorem, the sum of the indices of  $\vec{v}$  equals  $\chi(X)$ . By the corollary, if we first add the indices at zeros where  $g(z) = +a$ , we get  $I(g, \{a\}) = \deg(g)$ ; similarly, adding the indices at zeros where  $g(z) = -a$  gives  $I(g, \{-a\}) = \deg(g)$ . Therefore  $\chi(X) = 2 \deg(g)$ , completing the proof. (*Question:* Why does the proof fail for odd-dimensional hypersurfaces?). Since this computation is a famous result in its own right, let us record it as a theorem.

**Theorem.** For even-dimensional manifolds, the Euler characteristic equals twice the degree of the Gauss map.



A final remark: One can give an alternative definition of curvature simply by using metric properties of the manifold  $X$  itself, not properties of its embedding in  $\mathbf{R}^N$ . For two dimensions, this fact was realized by Gauss and is called the *theorem egregium*. Chern has given an “intrinsic” proof of the Gauss-Bonnet theorem, that is, one that completely avoids embedding (see [16]). For an accessible two-dimensional version of this proof, see [17]. (A “piecewise linear” form of the intrinsic Gauss-Bonnet theorem is derived below, in Exercise 11.)

## EXERCISES

1. Let  $f: \mathbf{R}^2 \rightarrow \mathbf{R}$  be a smooth function, and let  $S \subset \mathbf{R}^3$  be its graph. Prove that the volume form on  $S$  is just the form  $dA$  described in Section 4.†
2. If  $S$  is an oriented surface in  $\mathbf{R}^3$  and  $(n_1, n_2, n_3)$  is its unit normal vector, prove that the volume form is

$$n_1 dx_2 \wedge dx_3 + n_2 dx_3 \wedge dx_1 + n_3 dx_2 \wedge dx_1.$$

In particular, show that the volume form on the unit sphere is

$$x_1 dx_2 \wedge dx_3 + x_2 dx_3 \wedge dx_1 + x_3 dx_1 \wedge dx_2.$$

[HINT: Show that the 2-form just described is the 2-form defined by:

$$(v, w) \rightarrow \frac{1}{2} \det \begin{pmatrix} v \\ w \\ n \end{pmatrix} \text{ for pairs of vectors } v \text{ and } w \text{ in } \mathbf{R}^3.]$$

3. Let  $A: \mathbf{R}^n \rightarrow \mathbf{R}^n$  be a rotation (i.e., an orthogonal linear mapping). Show that the map of  $S^{n-1}$  onto  $S^{n-1}$  induced by  $A$  preserves the volume form (that is,  $A^*$  applied to the volume form gives the volume form back again).
4. Let  $C: [a, b] \rightarrow \mathbf{R}^3$  be a parametrized curve in  $\mathbf{R}^3$ . Show that its “volume” (i.e., the integral over  $C$  of the volume form) is just the arc length:

$$\int_a^b \left| \frac{dC}{dt} \right| dt.$$

5. Let  $f$  be a smooth map of the interval  $[a, b]$  into the positive real numbers. Let  $S$  be the surface obtained by rotating the graph of  $f$  around the

†Volume form is unfortunate terminology in two dimensions. “Area form” would be better.

$x$  axis in  $\mathbf{R}^3$ . There is a classical formula which says that the surface area of  $S$  is equal to

$$\int_a^b 2\pi f \sqrt{1 + (f')^2} dt.$$

Derive this formula by integrating the volume form over  $S$ . [HINT: Use Exercise 2.]

6. Prove that for the  $n - 1$  sphere of radius  $r$  in  $\mathbf{R}^n$ , the Gaussian curvature is everywhere  $1/r^{n-1}$ . [HINT: Show only that the derivative of the Gauss map is everywhere just  $1/r$  times the identity.]
7. Compute the curvature of the hyperboloid  $x^2 + y^2 - z^2 = 1$  at the point  $(1, 0, 0)$ . [HINT: The answer is  $-1$ .]
8. Let  $f = f(x, y)$  be a smooth function on  $\mathbf{R}^2$ , and let  $S \subset \mathbf{R}^3$  be its graph. Suppose that

$$f(0) = \frac{\partial f}{\partial x}(0) = \frac{\partial f}{\partial y}(0) = 0.$$

Let  $\kappa_1$  and  $\kappa_2$  be the eigenvalues of the Hessian

$$\frac{1}{2} \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \quad \text{at } 0.$$

Show that the curvature of  $S$  at  $(0, 0, 0)$  is just the product  $\kappa_1 \kappa_2$ . [HINT: It's enough simply to consider the case  $f(x, y) = \kappa_1 x^2 + \kappa_2 y^2$ . Why?]

9. A surface  $S \subset \mathbf{R}^3$  is called “ruled” if through every point of  $S$  there passes a straight line contained in  $S$ . Show that a ruled surface has Gaussian curvature less than or equal to zero everywhere.
10. Let  $T = T_{a,b}$  be the standard torus consisting of points in  $\mathbf{R}^3$  that are a distance  $a$  from the circle of radius  $b$  in the  $xy$  plane ( $a < b$ ). At what points is the curvature positive, at what points is it negative, and at what points is it zero?
11. Let  $T_1, T_2, \dots, T_k$  be a collection of closed triangles in  $\mathbf{R}^3$ . The set  $S = T_1 \cup \dots \cup T_k$  is called a polyhedral surface if the following statements are true.† [See part (a) of Figure 4-5.]

†We learned about this “intrinsic” form of the Gauss-Bonnet theorem for polyhedral surfaces from Dennis Sullivan.

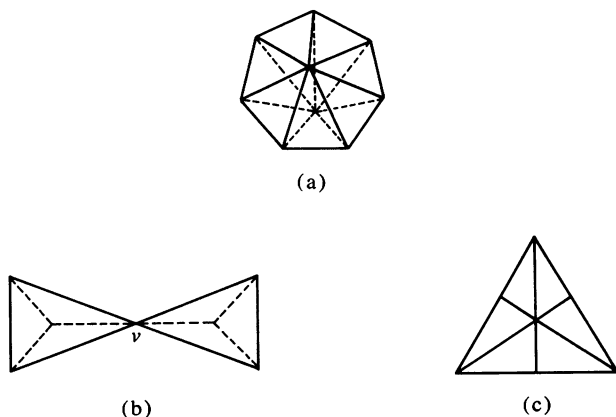


Figure 4-5

- (a) Each side of  $T_i$  is also the side of exactly one other triangle  $T_j$ .  
 (b) No two triangles have more than one side in common.  
 (c) If  $T_{i_1}, \dots, T_{i_s}$  are the triangles in the collection having  $v$  as a vertex, and  $s_{i_k}$  is the side opposite  $v$  in  $T_{i_k}$ , then  $\cup s_{i_k}$  is connected. [This disqualifies example (b) in Figure 4-5.]

For each vertex  $v$  we define  $\kappa(v)$  to be  $2\pi$  minus the sum of the angles at the vertex. Prove:

- (i) If each triangle  $T_i$  is subdivided into smaller triangles (for example, as indicated in part (c) of Figure 4-5), the total sum  $\sum \kappa(v)$  over all vertices is unchanged.  
 (ii)

$$\sum_v \kappa(v) = 2\pi \cdot \chi(S),$$

where the Euler characteristic  $\chi(S)$  is defined to be number of vertices minus the number of sides plus the number of triangles. (Compare with Chapter 3, Section 7.) [HINT: Every triangle has three sides, and every side is contained in two triangles.]

12. Let  $X$  be an oriented  $n - 1$  dimensional manifold and  $f: X \rightarrow \mathbf{R}^n$  an immersion. Show that the Gauss map  $g: X \rightarrow S^{n-1}$  is still defined even though  $X$  is not, properly speaking, a submanifold of  $\mathbf{R}^n$ . Prove that when  $X$  is even dimensional, the degree of  $g$  is one-half the Euler characteristic of  $X$ . Show this is not the case for odd-dimensional manifolds by showing that the Gauss map of the immersion

$$S^1 \longrightarrow \mathbf{R}^2, \quad t \longrightarrow [\cos(nt), \sin(nt)]$$

has degree  $n$ .

# Measure Zero and Sard's Theorem

If  $a = (a_1, \dots, a_n)$  and  $b = (b_1, \dots, b_n)$  are two  $n$ -tuples with  $a_1 < b_1, \dots, a_n < b_n$ , we denote by  $S(a, b)$  the *rectangular solid* consisting of the points  $(x_1, \dots, x_n) \in \mathbf{R}^n$  satisfying  $a_i < x_i < b_i, i = 1, \dots, n$ . The product

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

is the *volume* of  $S = S(a, b)$ , denoted  $\text{vol}(S)$ . (If  $b_1 - a_1 = \dots = b_n - a_n$ ,  $S$  is a *cube*.)

As defined in Chapter 1, Section 7, a subset  $A \subset \mathbf{R}^n$  has *measure zero* if for every  $\epsilon > 0$  there exists a countable covering of  $A$  by solids  $S_1, S_2, \dots$  such that  $\sum \text{vol}(S_i) < \epsilon$ . (*Problem:* Show that the  $S_i$  may be taken to be cubes.)

It is obvious that subsets of sets with measure zero also have measure zero. And we showed in Chapter 1, Section 7, that countable unions of sets of measure zero are still of measure zero. From this fact, we can exhibit many sets of measure zero.

**Exercise.** Show that the natural copy of  $\mathbf{R}^{n-1}$  inside  $\mathbf{R}^n$ —namely,  $\{(x_1, \dots, x_{n-1}, 0)\}$ —has measure zero. [HINT: Show that every compact subset of  $\mathbf{R}^{n-1}$  sits inside a single rectangular solid in  $\mathbf{R}^n$  with volume  $< \epsilon$ .]

What is surprisingly hard to demonstrate is that there exist sets not of measure zero! We deduce this fact with a clever proof due to Von Neumann.

**Proposition.** Let  $S$  be a rectangular solid and  $S_1, S_2, \dots$  a covering of its closure of  $\bar{S}$  by other solids. Then  $\sum \text{vol}(S_j) \geq \text{vol}(S)$ .

*Proof.* Call  $(z_1, \dots, z_n)$  an *integer point* of  $\mathbf{R}^n$  if each  $z_i$  is an integer. Let  $S = S(a, b)$ . The number of integers in the interval  $a_i < z_i < b_i$  is less than  $b_i - a_i + 1$  and at least as large as  $b_i - a_i - 1$ . Assume, for the moment, that the length  $b_i - a_i$  of each side of  $S$  is greater than 1. Then the number of integer points in  $S$  is less than

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

and at least

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

If  $S_1, S_2, \dots$  is a cover of  $\bar{S}$ , then, by compactness, some finite number  $S_1, \dots, S_N$  also is a cover. Now the number of integer points in  $S$  is at most equal to the sum of the numbers of integer points in  $S_1, \dots, S_N$ . Thus if  $S_j = S(a^j, b^j)$ , we get

$$\prod_{i=1}^n (b_i - a_i - 1) \leq \sum_{j=1}^N \prod_{i=1}^n (b_i^j - a_i^j + 1).$$

For each positive number  $\lambda$ , define  $\lambda S(a, b) = S(\lambda a, \lambda b)$ . Since  $\overline{\lambda S}$  is covered by  $\lambda S_1, \dots, \lambda S_N$  the preceding calculation gives

$$\prod_{i=1}^n (\lambda b_i - \lambda a_i - 1) \leq \sum_{j=1}^N \prod_{i=1}^n (\lambda b_i^j - \lambda a_i^j + 1).$$

Now, no matter what the size of  $S$ , if  $\lambda$  is large enough, then  $\lambda S$  will have sides of length greater than 1. So the last inequality is valid for large  $\lambda$  without any restriction on  $S$ .

You have now only to divide both sides of the inequality by  $\lambda^n$  and then let  $\lambda \rightarrow \infty$  to complete the proof. Q.E.D.

For the proof of Sard's theorem, we will need the measure zero form of Fubini's theorem. Suppose that  $n = k + l$ , and write  $\mathbf{R}^n = \mathbf{R}^k \times \mathbf{R}^l$ . For each  $c \in \mathbf{R}^k$ , let  $V_c$  be the "vertical slice"  $\{c\} \times \mathbf{R}^l$ . We shall say that a subset of  $V_c$  has measure zero in  $V_c$  if, when we identify  $V_c$  with  $\mathbf{R}^l$  in the obvious way, the subset has measure zero in  $\mathbf{R}^l$ .

**Fubini Theorem** (for measure zero). Let  $A$  be a closed subset of  $\mathbf{R}^n$  such that  $A \cap V_c$  has measure zero in  $V_c$  for all  $c \in \mathbf{R}^k$ . Then  $A$  has measure zero in  $\mathbf{R}^n$ .

*Proof.* Because  $A$  may be written as a countable union of compacts, we may, in fact, assume  $A$  compact. Also, by induction on  $k$ , it suffices to prove the theorem for  $k = 1$  and  $l = n - 1$ . We will divide the proof into several lemmas.

**Lemma 1.** Let  $S_1, \dots, S_N$  be a covering of the closed interval  $[a, b]$  in  $\mathbf{R}^1$ . Then there exists another cover  $S'_1, \dots, S'_M$  such that each  $S'_j$  is contained in some  $S_i$ , and

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

*Proof.* Left to the reader.

Given a set  $I \subset \mathbf{R}$ , let  $V_I = I \times \mathbf{R}^{n-1}$  in  $\mathbf{R}^n$ .

**Lemma 2.** Let  $A$  be a compact subset of  $\mathbf{R}^n$ . Suppose that  $A \cap V_c$  is contained in an open set  $U$  of  $V_c$ . Then for any suitably small interval  $I$  about  $c$  in  $\mathbf{R}$ ,  $A \cap V_I$  is contained in  $I \times U$ .

*Proof.* If not, there would exist a sequence of points  $(x_j, c_j)$  in  $A$  such that  $c_j \rightarrow c$  and  $x_j \notin U$ . Replace this sequence by a convergent one to get a contradiction. Q.E.D.

*Proof of Fubini.* Since  $A$  is compact, we may choose an interval  $I = [a, b]$  such that  $A \subset V_I$ . For each  $c \in I$ , choose a covering of  $A \cap V_c$  by  $(n - 1)$  dimensional rectangular solids  $S_1(c), \dots, S_{N_c}(c)$  having a total volume less than  $\epsilon$ . Choose an interval  $J(c)$  in  $\mathbf{R}$  so that the rectangular solids  $J(c) \times S_i(c)$  cover  $A \cap V_I$  (Lemma 2). The  $J(c)$ 's cover the line interval  $[a, b]$ , so we can use Lemma 1 to replace them with a finite collection of subintervals  $J'_j$  with total length less than  $2(b - a)$ . Each  $J'_j$  is contained in some interval  $J(c_j)$ , so the solids  $J'_j \times S_i(c_j)$  cover  $A$ ; moreover, they have total volume less than  $2\epsilon(b - a)$ . Q.E.D.

We need one more fact from measure theory—namely, that measure zero makes sense on manifolds. To show this, we prove

**Theorem.** Let  $U$  be an open set of  $\mathbf{R}^n$ , and let  $f: U \rightarrow \mathbf{R}^n$  be a smooth map. If  $A \subset U$  is of measure zero, then  $f(A)$  is of measure zero.

*Proof.* We may assume that  $\bar{A}$  is compact and contained in  $U$ , since  $A$  may be written as a countable union of subsets with this property. Let  $W$  be an open neighborhood of  $A$  such that  $\bar{W}$  is compact and contained in  $U$ .

Since  $\bar{W}$  is compact, there exists a constant  $M$  such that  $|f(x) - f(y)| < M|x - y|$  for all pairs  $x, y \in \bar{W}$ . It follows that there exists another constant  $M'$  such that if  $S$  is any cube in  $W$ , then  $f(S)$  is contained inside a cube  $S'$  of volume less than  $M' \text{vol}(S)$ .

Now show that the set  $A$  can be covered by a sequence of cubes  $S_1, S_2$ , etc., each contained in  $W$ , with  $\sum \text{vol}(S_i)$  less than any prescribed  $\epsilon$ . Thus we get a covering  $S'_1, S'_2, \dots$  of  $f(A)$  by cubes with  $\sum \text{vol}(S'_i) < M'\epsilon$ . As  $\epsilon$  is arbitrary,  $f(A)$  has measure zero. Q.E.D.

**Exercise.** Combine this theorem with the first exercise to prove

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

(Incidentally, this weak form of Sard's theorem is all we used to prove the Whitney embedding theorem.)

Now we may define a subset  $A$  of a  $k$ -dimensional manifold  $X$  to be of measure zero provided that, for every local parametrization  $f: U \rightarrow X$ , the preimage  $f^{-1}(A)$  has measure zero in  $U \subset \mathbf{R}^k$ . Because of the last theorem, it suffices to check that around any point in  $X$  there exists one local parametrization  $f: U \rightarrow X$  for which  $f^{-1}(A)$  has measure zero.

We are now ready for the task of proving

**Sard's Theorem.** Let  $f: X \rightarrow Y$  be a smooth map of manifolds, and let  $C$  be the set of critical points of  $f$  in  $X$ . Then  $f(C)$  has measure zero in  $Y$ .

Our proof is taken virtually verbatim from Milnor [1] pp. 16–19. By the Second Axiom of Countability, we can find a countable collection of open sets  $(U_i, V_i)$ ,  $U_i$  open in  $X$  and  $V_i$  open in  $Y$ , such that the  $U_i$ 's cover  $X$ ,  $f(U_i) \subset V_i$ , and the  $U_i$ 's and  $V_i$ 's are diffeomorphic to open sets in  $\mathbf{R}^n$ . Therefore it suffices to prove

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

The theorem is certainly true for  $n = 0$ , so we will assume that it is true for  $n - 1$  and prove it for  $n$ . We begin by partitioning  $C$  into a sequence of nested subsets  $C \supset C_1 \supset C_2 \supset \dots$ , where  $C_1$  is the set of all  $x \in U$  such that  $(df)_x = 0$ , and  $C_i$  for  $i \geq 1$  is the set of all  $x$  such that the partial

derivatives of  $f$  of order  $\leq i$  vanish at  $x$ . (Note that the  $C_i$ 's are closed subsets of  $C$ .) We will first prove

**Lemma 1.** The image  $f(C - C_1)$  has measure zero.

*Proof.* Around each  $x \in C - C_1$ , we will find an open set  $V$  such that  $f(V \cap C)$  has measure zero. Since  $C - C_1$  is covered by countably many of these neighborhoods (by the Second Axiom of Countability), this will prove that  $f(C - C_1)$  has measure zero.

Since  $x \notin C_1$ , there is some partial derivative, say  $\partial f / \partial x_1$ , that is not zero at  $x$ . Consider the map  $h: U \rightarrow \mathbf{R}^n$  defined by

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

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

We have constructed  $g$  so that it has the following property: it maps points of the form  $(t, x_2, \dots, x_n)$  in  $V'$  into points of the form  $(t, y_2, \dots, y_p)$  in  $\mathbf{R}^p$  (i.e., first coordinates the same). Therefore for each  $t$ ,  $g$  induces a map  $g'$  of  $(t \times \mathbf{R}^{n-1}) \cap V'$  into  $t \times \mathbf{R}^{p-1}$ . Since the derivative of  $g$  has the form

$$\left( \frac{\partial g_i}{\partial x_j} \right) = \left( \begin{array}{c|c} 1 & 0 \\ \hline * & \left( \frac{\partial g'_i}{\partial x_j} \right) \end{array} \right),$$

a point of  $t \times \mathbf{R}^{n-1}$  is critical for  $g'$  if and only if it is critical for  $g$  [because the determinant of the matrix on the right is just  $\det(\partial g'_i / \partial x_j)$ ]. By induction, Sard's theorem is true for  $n - 1$ , so the set of critical values of  $g'$  has measure zero. Consequently, by Fubini's theorem, the set of critical values of  $g$  is of measure zero.† Q.E.D.

Next we will prove

**Lemma 2.**  $f(C_k - C_{k+1})$  is of measure zero for  $k \geq 1$ .

*Proof.* This is a similar argument, but easier. For each  $x \in C_k - C_{k+1}$  there is some  $(k + 1)$ st derivative of  $f$  that is not zero. Thus we can find a  $k$ th derivative of  $f$ , say  $\rho$ , that (by definition) vanishes on  $C_k$  but has a first derivative, say  $\partial \rho / \partial x_1$ , that does not vanish. Then the map  $h: U \rightarrow \mathbf{R}^n$ , de-

†The critical values of  $g$  need not be a closed set; however, the critical points of  $g$  are closed (as a subset of the domain of  $g$ ). Since the domain of  $g$  is a countable union of compact sets, the same is true of the critical points, hence also of the critical values. This means that we can apply Fubini.



defined by  $h(x) = (\rho(x), x_2, \dots, x_n)$ , maps a neighborhood  $V$  of  $x$  diffeomorphically onto an open set  $V'$ . By construction,  $h$  carries  $C_k \cap V$  into the hyperplane  $0 \times \mathbf{R}^{n-1}$ . Therefore the map  $g = f \circ h^{-1}$  has all its critical points of type  $C^k$  in the hyperplane  $0 \times \mathbf{R}^{n-1}$ . Let  $\bar{g}: (0 \times \mathbf{R}^{n-1}) \cap V' \rightarrow \mathbf{R}^p$  be the restriction of  $g$ . By induction, the set of critical values of  $\bar{g}$  is of measure zero. Moreover, the critical points of  $g$  of type  $C_k$  are obviously critical points of  $\bar{g}$ . This proves that the image of these critical points is of measure zero, and, therefore,  $f(C_k \cap V)$  is of measure zero. Since  $C_k - C_{k-1}$  is covered by countably many such sets  $V$ , we are done. Q.E.D.

Finally, we will prove

**Lemma 3.** For  $k > n/p - 1$ ,  $f(C_k)$  is of measure zero.

Combined with the two preceding lemmas, this will prove the theorem.

*Proof of Lemma 3.* Let  $S \subset U$  be a cube whose sides are of length  $\delta$ . If  $k$  is sufficiently large ( $k > n/p - 1$ , to be precise), we will prove that  $f(C_k \cap S)$  has measure zero. Since  $C_k$  can be covered by countably many such cubes, this will prove that  $f(C_k)$  has measure zero.

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

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

where

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

for  $x \in C_k \cap S$ ,  $x + h \in S$ . Here  $a$  is a constant that depends only on  $f$  and  $S$ . Now subdivide  $S$  into  $r^n$  cubes whose sides are of length  $\delta/r$ . Let  $S_1$  be a cube of the subdivision that contains a point  $x$  of  $C_k$ . Then any point of  $S_1$  can be written as  $x + h$  with

$$(2) \quad |h| < \sqrt[n]{n} \left( \frac{\delta}{r} \right).$$

From (1), it follows that  $f(S_1)$  lies in a cube with sides of length  $b/r^{k+1}$  centered about  $\delta(x)$ , where  $b = 2a(\sqrt[n]{n} \delta)^{k+1}$  is constant. Hence  $f(C^k \cap S)$  is contained in the union of at most  $r^n$  cubes having total volume

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

If  $k + 1 > n/p$ , then evidently  $v$  tends to 0 as  $r \rightarrow \infty$ , so  $f(C_k \cap S)$  must have measure zero. This completes the proof of Sard's theorem. Q.E.D.

# Classification of Compact One-Manifolds

In this appendix we prove the following classification theorem, using Morse functions. For a proof based on arc length instead, see the appendix of Milnor [1].

**Theorem.** Every compact one-dimensional manifold with boundary is diffeomorphic either to a circle or a closed interval.

The proof requires a straightforward technical lemma, which we prove at the start.

**Smoothing Lemma.** Let  $g$  be a function on  $[a, b]$  that is smooth and has positive derivative everywhere except at one interior point  $c$ . Then there exists a globally smooth function  $\tilde{g}$  that agrees with  $g$  near the endpoints and has positive derivative everywhere.

*Proof.* Let  $\rho$  be a smooth nonnegative function that vanishes outside a compact subset of  $(a, b)$ , which equals 1 near  $c$ , and which satisfies  $\int_a^b \rho = 1$ . Define

$$\tilde{g}(x) = g(a) + \int_a^x [k\rho(s) + g'(s)(1 - \rho(s))] ds,$$

where the constant

$$k = g(b) - g(a) - \int_a^b g'(s)(1 - \rho(s)) ds.$$

(Note  $k > 0$ .) We leave you to verify that  $\tilde{g}$  has the desired properties. Q.E.D.

To prove the theorem, choose a Morse function  $f$  on  $X$ . Let  $S$  be the union of the critical points of  $f$  and the boundary points of  $X$ . As  $S$  is finite,  $X - S$  consists of a finite number of connected one-manifolds  $L_1, \dots, L_N$ .

**Proposition.**  $f$  maps each  $L_i$  diffeomorphically onto an open interval in  $\mathbf{R}^1$ .

*Proof.* Let  $L$  be any of the  $L_i$ . Because  $f$  is a local diffeomorphism and  $L$  is connected,  $f(L)$  is open and connected in  $\mathbf{R}^1$ . Moreover,  $f(L)$  is contained in the compact set  $f(X)$ , so  $f(L) = (a, b)$ . It suffices to show that  $f$  is one to one on  $L$ , for then  $f^{-1} : (a, b) \rightarrow L$  is defined, and it is (locally) smooth because  $f$  is a local diffeomorphism.

Let  $p$  be any point of  $L$  and set  $c = f(p)$ . We show that every other point  $q \in L$  can be joined to  $p$  by a curve  $\gamma : [c, d] \rightarrow L$  (or  $\gamma : [d, c] \rightarrow L$ ), such that  $f \circ \gamma = \text{identity}$  and  $\gamma(d) = q$ . Since  $f(q) = d \neq c = f(p)$ , this result shows that  $f$  is one to one (and, in fact, virtually constructs  $f^{-1}$ ). So let  $Q$  be the set of points  $q$  that can be so joined. We leave to you the trivial verification that  $Q$  is both open and closed, thanks to the fact that  $f$  is a local diffeomorphism. Therefore  $Q = L$ . Q.E.D.

Now we invoke a lemma that you are asked to prove in Exercise 8, Chapter 2, Section 2.

**Lemma.** Let  $L$  be a subset of  $X$  diffeomorphic to an open interval of  $\mathbf{R}^1$ , where  $\dim X = 1$ . Then its closure  $\bar{L}$  contains at most two points not in  $L$ .

Here the diffeomorphism  $f$  of  $L_i$  to an open interval of  $\mathbf{R}$  extends to the closure  $\bar{L}_i$ , so Figure A-1 cannot occur. Thus each  $\bar{L}_i$  has exactly two boundary points. Also, since  $X$  is a manifold, each point  $p \in S$  is either on the boundary of one or two  $\bar{L}_i$ ; and in the former case,  $p \in \partial X$ . (Why cannot Figure A-2 occur?)

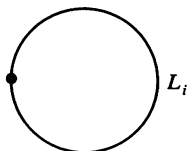


Figure A-1

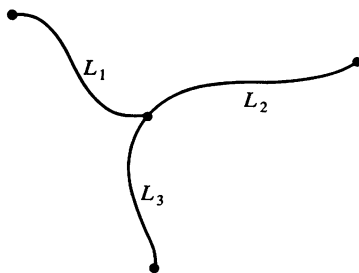


Figure A-2

We will call a sequence  $L_1, \dots, L_k$  a *chain*, provided that each consecutive pair  $\bar{L}_j$  and  $\bar{L}_{j+1}$  have a common boundary point  $p_j$ ,  $j = 1, \dots, k-1$ . Let  $p_0$  denote the other boundary point of  $L_1$ , and  $p_k$  the other boundary point for  $L_k$ . Since there are only finitely many  $L_i$  altogether, there clearly exists a maximal chain, a chain that cannot be extended by appending another  $L_i$ . We finish our proof by establishing the

**Claim.** If  $L_1, \dots, L_k$  is a maximal chain, then it contains every  $L_i$ . If  $\bar{L}_k$  and  $\bar{L}_0$  have a common boundary point, then  $X$  is diffeomorphic to the circle; if not,  $X$  is diffeomorphic to a closed interval.

*Proof.* Suppose that  $L$  is not included in the chain. Then  $\bar{L}$  cannot share the boundary point  $p_0$  or  $p_k$ ; otherwise the chain could be lengthened. It also cannot share any other  $p_j$ , since Figure A-2 is prohibited. Thus  $\bigcup_{j=1}^k \bar{L}_j$  does not intersect any of the  $\bar{L}$  excluded from the chain. Consequently, this union is both open and closed in  $X$ , and therefore

$$X = \bigcup_{j=1}^k \bar{L}_j$$

by connectivity.

Now the map  $f$  behaves nicely on each  $L_j$ , but it may reverse directions as it crosses some boundary points (Figure A-3). We will simply straighten it out in the least delicate manner. Set  $a_j = f(p_j)$ , so that  $f$  maps  $L_j$  diffeomorphically onto  $(a_{j-1}, a_j)$  or  $(a_j, a_{j-1})$ , whichever interval makes sense. For each  $j = 1, \dots, k$ , pick an affine map  $\tau_j: \mathbf{R}^1 \rightarrow \mathbf{R}^1$  carrying  $a_{j-1}$  to  $j-1$  and  $a_j$  to  $j$ . (An affine map is simply a linear one plus a constant,  $t \rightarrow \alpha t + \beta$ .) Define  $f_j: \bar{L}_j \rightarrow [j-1, j]$  by  $f_j = \tau_j \circ f$ .

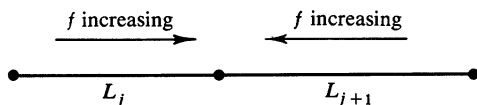


Figure A-3

If  $a_0 \neq a_k$ , then the  $f_j$  agree on points of common definition. Therefore they fit together to define a map  $F: X \rightarrow [0, k]$ , by  $F = f_j$  on  $\bar{L}_j$ .  $F$  is continuous and a diffeomorphism except at the points  $p_1, \dots, p_{k-1}$ . Using the Smoothing Lemma, simply modify  $F$  to be a diffeomorphism globally.

If  $a_0 = a_k$ , set  $g_j = \exp [i(2\pi/k)f_j]$ . Then we can define  $G: X \rightarrow S^1$  by setting  $G = g_j$  on  $\bar{L}_j$ .  $G$  is continuous and a diffeomorphism except at  $p_1, \dots, p_{k-1}$ . Again use the Smoothing Lemma (applied in local coordinates on  $S^1$ ) to make  $G$  a global diffeomorphism. Q.E.D.

# Bibliography

As collateral reading, we highly recommend the following two books:

- [1] J. MILNOR, *Topology from a Differential Viewpoint*. University of Virginia Press, 1965.
- [2] M. SPIVAK, *Calculus on Manifolds*. New York: Benjamin, 1965.

We also recommend that you look at some of the following books, which elaborate on matters we have touched on here. We don't propose that you read these books from cover to cover. The assumption is that you will have a few hours free now and then to find out a little more about ideas that have piqued your interest, not that you will spend a whole semester pursuing any topic. Therefore most of these references are not course books; or, if they are, the relevant sections are explicitly cited. You may find the following rating system helpful.

- G = The book is elementary. Prerequisites: only a first course in analysis and linear algebra.
- PG = The book is a bit less elementary; a little mathematical sophistication is helpful (e.g., some modern algebra, some algebraic topology).
- R = Graduate level mathematics, but a perceptive undergraduate can gain some insights from it.
- X = The book is hard going for a graduate student, meant to be read more for inspiration than for comprehension.

Here are some references for the material in Chapter 1.

- [3] J. MILNOR, *Morse Theory*. Princeton, N.J.: Princeton University Press, No. 51, 1963.  
We showed in Chapter 1, Section 7 that every manifold possesses Morse functions. For the implications of this fact, look at the first part of this book (pp. 1–42). A little algebraic topology is helpful but not indispensable [PG].
- [4] M. MORSE, *Pits, Peaks, and Passes*. Produced by the Committee on Educational Media, Mathematical Association of America. Released by Martin Learning Aids, 1966.  
In this engaging film, Marston Morse shows how geography tells us about topology. In two dimensions, the critical points of a Morse function correspond to the mountains, valleys, and passes on the graph surface [G].
- [5] A. WALLACE, *Differential Topology, First Steps*. New York: Benjamin, 1968.  
The author discusses some formidable subjects (Morse theory and surgery) from an elementary point of view and also gives a Morse theoretical classification of two-manifolds [PG].
- [6] A. GRAMAIN, *Cours d'initiation à la topologie algébrique*, Orsay, Faculté des Sciences, 1970.  
Here is another place where you can find two-manifolds classified using Morse theory. For beginners, this book is a little more pedestrian and easier to read than Wallace. (However, it is written in French.) [G]
- [7] L. AHLFORS and L. SARIO, *Riemann Surfaces*. Princeton, N.J.: Princeton University Press, 1960.  
Here you can find a classical treatment of the classification of two-manifolds.
- [8] R. ABRAHAM, *Transversal Mappings and Flows*. New York: Benjamin, 1967.  
The transversality theorem has important applications in the theory of dynamical systems. Before glancing through Abraham's book for this subject, however, you should read the Smale paper cited below [R].
- [9] M. GOLUBITSKY and V. GUILLEMIN, *Stable Mappings and Their Singularities*. New York: Springer, 1973. Another interesting application of transversality is presented here: the study of singularities of mappings [R].

Concerning the material in Chapters 2 and 3, we strongly recommend that you read Milnor's book [1]. In particular, Section 7 of Milnor contains an introduction to framed cobordism. If the material interests you, we recommend you pursue the subject further through some of the following sources.

- [10] L. PONTRYAGIN, "Smooth Manifolds and Their Applications in Homotopy Theory," *Amer. Math. Society Translations*, Series 2, **II**(1959), 1-114.

This is fairly easy to read; don't be put off by the unbending lemma-theorem-corollary format or by the Slavisms that have managed to creep into the translation. It is helpful to know a little homotopy theory [R].

- [11] P. ALEXANDROFF and H. HOPF, *Topologie*. New York: Chelsea, 1965 (reprint of the original edition published in Berlin in 1935).

One of the classical treatises on topology. It is not easy reading—it is written in German, which will deter some of you—but you will find some delightful applications of the ideas we have discussed [X].

The usual formulation of the Lefschetz Fixed-Point Theorem includes a prescription for computing the global Lefschetz number in terms of homology theory. If you are interested in seeing how this goes, and you've had some algebraic topology, you might look at

- [12] M. GREENBERG, *Lectures on Algebraic Topology*. New York: Benjamin, 1967, Section 30 [PG].

If you want to go a little deeper into the topological aspects of flows and vector fields than we did, a good place to start is

- [13] W. HUREWICZ, *Lectures on Ordinary Differential Equations*. Cambridge, Mass.: The MIT Press, 1958, Chapter 5, pp. 102-115 [G].

We also recommend a very readable survey article:

- [14] S. SMALE, "Differentiable Dynamical Systems," *Bulletin of the A.M.S.*, **73** (1967), 747-817 [PG].

For the material in Chapter 4, our main reference is Spivak [2]. It probably wouldn't do you any harm, however, to go back to a calculus text like Apostol and remind yourself what the usual Green's theorem, Divergence Theorem, and Stokes theorem are.

For the argument principle, look at

- [15] AHLFORS, *Complex Analysis*. New York: McGraw-Hill, 1953, p. 123 [PG].

The form of the Gauss-Bonnet theorem we proved in Chapter 4, Section 9 is the so-called "extrinsic" Gauss-Bonnet theorem (for hypersurfaces in  $\mathbb{R}^n$ ). There is a much more subtle "intrinsic" Gauss-Bonnet theorem:

- [16] S. CHERN, "A Simple Intrinsic Proof of the Gauss-Bonnet Formula for Closed Riemannian Manifolds," *Annals of Math*, **45** (1944), 747-752 [X].

You'll probably find the Chern paper hard to read unless you have a fairly strong background in differential geometry. A very readable discussion



of the two-dimensional intrinsic Gauss-Bonnet theorem can be found, however, in Chapter 7 of

- [17] I. M. SINGER and H. A. THORPE, *Lecture Notes on Elementary Geometry and Topology*. Glenview, Ill.: Scott, Foresman, 1967 [PG].

(The reason that the two-dimensional case is simpler than the  $n$ -dimensional case is that the group of  $2 \times 2$  orthogonal matrices is abelian!)

For the relationship between the intrinsic and extrinsic forms of the Gauss-Bonnet theorem, look at Chapter 8 of Singer-Thorpe.

For an introduction to cohomology through forms, as in Section 6, we recommend

- [18] M. SPIVAK, *A Comprehensive Introduction to Differential Geometry*. Vol. 1, Boston, Mass.: Publish or Perish, Inc.

Apropos, we highly recommend Spivak as a general textbook. It is the best introduction we know of to graduate level differential geometry. Unfortunately, it is rather hard to obtain. Bookstores don't seem to carry it. You can, however, get it by writing directly to Publish or Perish, Inc., 6 Beacon St., Boston, Mass. 02108.



# Index

- Almost every, 40
- Alt operator, 155
- Alternating tensor, 155
- Antipodal map, 38
- Appropriate for intersection theory, 107
- Arcwise connected, 38
- Area form, 170
- Argument, 188
- Argument principle, 151
  
- Ball in  $\mathbf{R}^k$ , 5
- Borsuk-Ulam theorem, 91
- Boundary:
  - definition, 57
  - of homotopy space  $X \times I$ , 99
  - orientation of, 97
  - tangent space of, 60
- Boundary Theorem, 80
- Brouwer fixed-point theorem, 65
  
- Canonical immersion, 14
- Canonical submersion, 20
- Cartesian product, 4
- Cauchy integral formula, 192
- Cauchy-Riemann equation, 192
- Chain rule, 8, 11
- Change of variables, 165
  - on manifolds, 168
  - in  $\mathbf{R}^k$ , 166
- Classification of one-manifolds, 64, 208
- Classification of two-manifolds, 124
- Closed curve, 172
- Closed form, 178
- Cobordant, 84, 187
- Cocycle condition, 174
- Codimension, 24
- Cohomologous forms, 179
- Cohomology:
  - class, 179
  - group, 179

- Commutative, 9
- Compactly supported form, 166
- Compactly supported isotopy, 142
- Complementary dimensions, 77
- Complex valued form, 192
- Contractible, 38
- Coordinate function, 4
- Coordinate system, 3
  - linear, 6
- Critical point, 40
  - nondegenerate, 41
- Critical value, 21
- Cross section, 55
- Curl, 178
- Curvature:
  - of hypersurfaces in  $\mathbf{R}^k$ , 195
  - of one-manifolds in  $\mathbf{R}^2$ , 75
- Curve, 12
  - closed, 172
  - Jordan curve theorem, 85
- De Rham cohomology group, 179
- Deformation, 38
- Degree:
  - Hopf degree theorem, 146
  - and integration (*see* Degree formula)
  - mod 2, 80
  - oriented, 105
- Degree formula, 188
- Derivative:
  - of forms (*see* Exterior derivative)
  - of maps of Euclidean spaces, 8
  - of maps of manifolds, 10
- Derivative map of tangent bundles, 50
- Determinant tensor, 153
- Determinant theorem, 160
- Diagonal, 7
  - tangent space of, 12, 76
- Diffeomorphism, 3
  - local, 13, 14
- Differential:
  - of forms (*see* Exterior derivative)
  - of a function, 163
- Dimension, 3
  - complementary, 77
- Direct sum orientation, 100
- Divergence, 178
- Divergence theorem, 152
- Embedding, 17
  - Whitney embedding theorem, 48, 53
- Epsilon Neighborhood Theorem, 69
- Equivalent maps, 14
- Equivalently oriented bases, 95
- Euler characteristic, 116, 148
- Even permutation, 155
- Exact form, 178
- Exact sequence, 103
- Extension theorem:
  - for maps into spheres, 145
  - for transversal maps, 72
- Exterior algebra, 159
- Exterior derivative:
  - on manifolds, 176
  - in  $\mathbf{R}^k$ , 174
- Fixed point, 33
  - Brouwer theorem, 65
  - Lefschetz, 33, 120
- Flow, 134
- Form, 162
  - area form, 170
  - closed, 178
  - complex valued, 192
  - pull-back, 163
  - restriction, 168
  - smooth, 165
  - support, 166
  - volume form (*see* Volume form)

- Fubini theorem, 204
- Fundamental Theorem of Algebra, 82, 110
  
- Gauss-Bonnet theorem, 196
- Gauss map, 195
- General position, 74
- Genus, 124
- Global property, 3
- Gradient field, 140, 177
- Graph, 7
  - tangent space of, 12
- Green's theorem, 152
  
  
- Half space, 57, 63
- Hessian matrix, 41
- Holomorphic function, 192
- Homotopy, 33
  - homotopy class, 34
  - Transversality Homotopy Theorem, 70
- Hopf Degree Theorem, 146
- Hypersurface, 86, 195
  
  
- Immersion, 14
  - canonical, 14
  - Local Immersion Theorem, 15
  - Whitney Immersion Theorem, 55
- Independent functions, 23
- Index, 133
  - Poincaré-Hopf Index Theorem, 134
- Integral:
  - Cauchy integral formula, 192
  - of forms on manifolds, 167
  - of forms on  $\mathbf{R}^k$ , 166
  - of forms on submanifolds, 168
  - of functions, 195
  - line integral, 169
- Interior, 57
- Intersection, transversal, 30
  - tangent space of, 32
- Intersection number, mod 2:
  - of map and submanifold, 78
  - self-intersection, 79
  - of two maps, 83
  - of two submanifolds, 79
- Intersection number, oriented:
  - of map and submanifold, 107
  - self-intersection, 115
  - of two maps, 113
  - of two submanifolds, 112
- Inverse Function Theorem, 13, 19, 56
- Inward unit normal vector, 63
- Isotopy, 142
- Isotopy lemma, 142
  
  
- Jacobian, 195
- Jacobian matrix, 8
- Jordan-Brouwer Separation Theorem, 89
- Jordan Curve Theorem, 85
  
  
- Klein bottle, 48, 49
  
  
- Lefschetz map, 33, 120
- Lefschetz number:
  - global, 19
  - local, 121, 127
  - local computation, 130
- Lie group, 23
- Line integral, 169, 172
- Linear coordinates, 6
- Local Immersion Theorem, 15
- Local property, 2
- Local Submersion Theorem, 20
- Locally finite cover, 56

- Manifold, 3
  - one-dimensional, 64, 208
  - two-dimensional, 124
- Manifold with boundary, 57
  - tangent space of, 59
- Mayer-Vietoris argument, 181
- Measure zero, 39, 202
- Möbius band, 63
- Mod 2 intersection number (*see*
  - Intersection number, mod 2)
- Morse function, 43
- Morse lemma, 42
- Multilinear, 153
- Multiplicity, 110
  
- Negatively oriented basis, 95
- Neighborhood:
  - epsilon, 69
  - tubular, 76
- Nondegenerate critical point, 41
- Nondegenerate zero, 139
- Nonretraction theorem, 65
- Normal Bundle, 71
  - of diagonal, 76
  - of submanifold (relative), 76
  - trivial, 76
- Normal space, 71
- Normal vector:
  - field, 106
  - inward, 63, 98
  - outward, 64, 98
  
- Odd permutation, 155
- One-manifolds, classification, 64, 208
- Open ball, 5
- Orientable manifold, 96
- Orientation:
  - boundary, 97
  - direct sum, 100
  - for manifolds, 96
  - opposite, 97
- Orientation (*cont.*):
  - preimage, 100
  - product, 97
  - for vector spaces, 95
- Orientation number (for
  - zero-dimensional spaces), 96
- Orientation preserving map:
  - of oriented manifolds, 96
  - of vector spaces, linear, 96
- Oriented bases:
  - equivalently, 95
  - negatively, 95
  - positively, 95
- Oriented intersection number (*see*
  - Intersection number, oriented)
- Oriented manifold, 97
- Orthogonal group, 22
- Orthonormal basis, 161
  
- Parametrization, 3, 57
- Partition of unity, 52
- Permutation group, 155
- Poincaré-Hopf Index Theorem, 134
- Poincaré lemma, 181
- Positively oriented basis, 95
- Preimage, 21
  - orientation, 100
  - tangent space, 33
- Preimage Theorem, 21
- Product, 58
  - manifold, 4, 58
  - map, 7
  - orientation, 97
  - tensor, 154
  - wedge, 156, 162
- Projection map, 55, 71
- Proper map, 17
- Pullback:
  - cohomology classes, 179
  - forms, 163
  - vector fields, 134

- Rectangular solid, 202
- Regular point, 40
- Regular value, 21
- Relative normal bundle, 76
- Relatively open, 2
- Restriction, of forms, 168
- Retraction, 65
  
- Saddle point, 123
- Sard's theorem, 39, 62, 205
- Second Fundamental Theorem of Calculus, 152
- Self-intersection number:
  - mod 2, 79
  - oriented, 115
- Short exact sequence, 103
- Sign, of a basis, 95
- Simply connected, 38
- Sink, 123
- Smooth family of maps, 39
- Smooth form, 165
- Smooth map, 1, 2
- Source, 122
- Sphere, 4
- Sphere bundle, 55
- Splitting proposition, 126
- Stability Theorem, 35
- Stable class, 34
- Stable property, 34
- Stack of Records Theorem, 26
- Stereographic Projection, 6
- Stokes Theorem:
  - classical, 152
  - generalized, 183
- Submanifold, 5
- Submersion, 20
  - canonical, 20
  - local submersion theorem, 20
- Support, Compact (*see* Compactly supported form)
- Support, of a form, 166
  
- Tangency, to a vector field, 134
- Tangent bundle, 50
- Tangent space, 9
  - of boundary, 60
  - of diagonal, 12, 76
  - of graph, 12
  - of intersection, 32
  - of manifold with boundary, 59
  - of preimage, 33
  - of product, 12
- Tangent vector, 9
- Tensor, 153
  - alternating, 155
  - determinant, 153
- Tensor product, 154
- Torus, 6
- Transpose map, 159
- Transversal intersection, 30
- Transversal map, 28, 72
- Transversality Theorem, 68
- Triangulation, 148
- Trivial normal bundle, 76
- Tubular Neighborhood Theorem, 76
  
- Unit Normal (*see* Normal vector)
- Upper half space, 57, 63
- Urysohn Theorem, 56
  
- Vector, tangent, 9
- Vector field, 55, 132
  - gradient, 177
  - tangent to, 134
  - zero of (*see* Zero of vector field)
- Velocity vector, 12
- Volume, 194
- Volume element, 161
- Volume form:
  - on Euclidean space, 164
  - on a manifold, 194

- Wedge product:
  - of forms, 162
  - of tensors, 156
- Whitney Embedding Theorem, 48, 53
- Whitney Immersion Theorem, 55
- Winding number:
  - mod 2, 86, 88
  - oriented, 144, 192
- Zero of vector field, 50, 132
  - nondegenerate, 139