Graduate Texts in Mathematics

William Fulton

Algebraic Topology

A First Course

Springer

Graduate Texts in Mathematics 153

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces. 2nd ed.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.
61 WHITEHEAD. Elements of Homotopy

(continued after index)

William Fulton

Algebraic Topology

A First Course

With 137 Illustrations

Springer

William Fulton
Mathematics Department
University of Chicago
Chicago, IL 60637
USA

Mathematics Subject Classifications (1991): 55-01

Library of Congress Cataloging-in-Publication Data
Fulton, William, 1939–
Algebraic topology/William Fulton.
p. cm. — (Graduate texts in mathematics)
Includes bibliographical references and index.
ISBN-13: 978-0-387-94327-5
1. Algebraic topology. I. Title. II. Series.
QA612.F85 1995
514′.2—dc20 94-21786

ISBN-13: 978-0-387-94327-5 e-ISBN: 978-1-4612-4180-5
DOI: 10.1007/978-1-4612-4180-5

(EB)

9 8 7 6 5

springeronline.com

To the memory of my parents

Preface

To the Teacher. This book is designed to introduce a student to some of the important ideas of algebraic topology by emphasizing the relations of these ideas with other areas of mathematics. Rather than choosing one point of view of modern topology (homotopy theory, simplicial complexes, singular theory, axiomatic homology, differential topology, etc.), we concentrate our attention on concrete problems in low dimensions, introducing only as much algebraic machinery as necessary for the problems we meet. This makes it possible to see a wider variety of important features of the subject than is usual in a beginning text. The book is designed for students of mathematics or science who are not aiming to become practicing algebraic topologists—without, we hope, discouraging budding topologists. We also feel that this approach is in better harmony with the historical development of the subject.

What would we like a student to know after a first course in topology (assuming we reject the answer: half of what one would like the student to know after a second course in topology)? Our answers to this have guided the choice of material, which includes: understanding the relation between homology and integration, first on plane domains, later on Riemann surfaces and in higher dimensions; winding numbers and degrees of mappings, fixed-point theorems; applications such as the Jordan curve theorem, invariance of domain; indices of vector fields and Euler characteristics; fundamental groups and covering spaces; the topology of surfaces, including intersection numbers; relations with complex analysis, especially on Riemann surfaces; ideas of homology, De Rham cohomology, Čech cohomology, and the relations between them and with fundamental groups; methods of calculation such as the Mayer–Vietoris and Van Kampen theorems; and a taste of the way algebra and "functorial" ideas are used in the subject.

To achieve this variety at an elementary level, we have looked at the first nontrivial instances of most of these notions: the first homology group, the first De Rham group, the first Čech group, etc. In the case of the fundamental group and covering spaces, however, we have bowed to tradition and included the whole story; here the novelty is on the emphasis on coverings arising from group actions, since these are what one is most likely to meet elsewhere in mathematics.

We have tried to do this without assuming a graduate-level knowledge or sophistication. The notes grew from undergraduate courses taught at Brown University and the University of Chicago, where about half the material was covered in one-semester and one-quarter courses. By choosing what parts of the book to cover—and how many of the challenging problems to assign—it should be possible to fashion courses lasting from a quarter to a year, for students with many backgrounds. Although we stress relations with analysis, the analysis we require or develop is certainly not "hard analysis."

We start by studying questions on open sets in the plane that are probably familiar from calculus: When are path integrals independent of path? When are 1-forms exact? (When do vector fields have potential functions?) This leads to the notion of winding number, which we introduce first for differentiable paths, and then for continuous paths. We give a wide variety of applications of winding numbers, both for their own interest and as a sampling of what can be done with a little topology. This can be regarded as a glimpse of the general principle that algebra can be used to distinguish topological features, although the algebra (an integer!) is fairly meager.

We introduce the first De Rham cohomology group of a plane domain, which measures the failure of closed forms to be exact. We use these groups, with the ideas of earlier chapters, to prove the Jordan curve theorem. We also use winding numbers to study the singularities of vector fields. Then 1-chains are introduced as convenient objects to integrate over, and these are used to construct the first homology group. We show that for plane open sets homology, winding numbers, and integrals all measure the same thing; the proof follows ideas of Brouwer, Artin, and Ahlfors, by approximating with grids.

As a first excursion outside the plane, we apply these ideas to surfaces, seeing how the global topology of a surface relates to local behavior of vector fields. We also include applications to complex analysis. The ideas used in the proof of the Jordan curve theorem are developed more fully into the Mayer–Vietoris story, which becomes our main tool for calculations of homology and cohomology groups.

Standard facts about covering spaces and fundamental groups are discussed, with emphasis on group actions. We emphasize the construction of coverings by patching together trivial coverings, since these ideas are widely used elsewhere in mathematics (vector bundles, sheaf theory, etc.), and Čech cocycles and cohomology, which are widely used in geometry and algebra; they also allow, following Grothendieck, a very short proof of the Van Kampen theorem. We prove the relation among the fundamental group, the first homology group, the first De Rham cohomology group, and the first Čech cohomology group, and the relation between cohomology classes, differential forms, and the coverings arising from multivalued functions.

We then turn to the study of surfaces, especially compact oriented surfaces. We include the standard classification theorem, and work out the homology and cohomology, including the intersection pairing and duality theorems in this context. This is used to give a brief introduction to Riemann surfaces, emphasizing features that are accessible with little background and have a topological flavor. In particular, we use our knowledge of coverings to construct the Riemann surface of an algebraic curve; this construction is simple enough to be better known than it is. The Riemann–Roch theorem is included, since it epitomizes the way topology can influence analysis. Finally, the last part of the book contains a hint of the directions the subject can go in higher dimensions. Here we do include the construction and basic properties of general singular (cubical) homology theory, and use it for some basic applications. For those familiar with differential forms on manifolds, we include the generalization of De Rham theory and the duality theorems.

The variety of topics treated allows a similar variety of ways to use this book in a course, since many chapters or sections can be skipped without making others inaccessible. The first few chapters could be used to follow or complement a course in point set topology. A course with more algebraic topology could include the chapters on fundamental groups and covering spaces, and some of the chapters on surfaces. It is hoped that, even if a course does not get near the last third of the book, students will be tempted to look there for some idea of where the subject can lead. There is some progression in the level of difficulty, both in the text and the problems. The last few chapters may be best suited for a graduate course or a year-long undergraduate course for mathematics majors.

We should also point out some of the many topics that are omitted or slighted in this treatment: relative theory, homotopy theory, fibrations, simplicial complexes or simplicial approximation, cell complexes, homology or cohomology with coefficients, and serious homological algebra.

To the Student. Algebraic topology can be thought of as the study of the shapes of geometric objects. It is sometimes referred to in popular accounts as "rubber-sheet geometry." In practice this means we are looking for properties of spaces that are unchanged when one space is deformed into another. "Doughnuts and teacups are topologically the same." One problem of this type goes back to Euler: What relations are there among the numbers of vertices, edges, and faces in a convex polytope, such as a regular solid, in space? Another early manifestation of a topological idea came also from Euler, in the Königsberg bridge problem: When can one trace out a graph without traveling over any edge twice? Both these problems have a feature that characterizes one of the main attractions, as well as the power, of modern algebraic topology—that a global question, depending on the overall shape of a geometric object, can be answered by data that are collected locally. Since these are so appealing—and perhaps to capture your interest while we turn to other topics—they are included as problems with hints at the end of this Preface.

In fact, modern topology grew primarily out of its relation with other subjects, particularly analysis. From this point of view, we are interested in how the shape of a geometric object relates to, or controls, the answers to problems in analysis. Some typical and historically important problems here are:

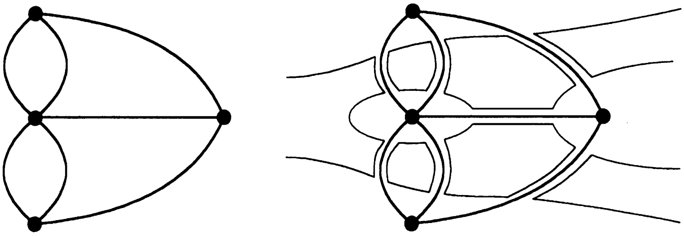
(i) whether differential forms ω on a region that are closed ($d\omega = 0$) must be exact ($\omega = d\mu$) depends on the topology of the region;
(ii) the behavior of vector fields on a surface depends on the topology of the surface; and
(iii) the behavior of integrals $\int dx/\sqrt{R(x)}$ depends on the topology of the surface $y^2 = R(x)$, here with x and y complex variables.

In this book we will begin with the first of these problems, working primarily in open sets in the plane. There is one disadvantage that must be admitted right away: this geometry is certainly flat, and lacks some of the appeal of doughnuts and teacups. Later in the book we will in fact discuss generalizations to curved spaces like these, but at the start we will stick to the plane, where the analysis is simpler. The topology of open sets in the plane is more interesting than one might think. For example, even the question of the number of connected components can be challenging. The famous Jordan curve theorem, which is one of our goals here, says that the complement of a plane set that is homeomorphic to a circle always has two components—a fact that will probably not surprise you, but whose proof is not so obvious. We will also spend some time on the second problem, which includes the popular problem of whether one can "comb the hair on a billiard ball." We will include some applications to complex analysis, later discussing some of the ideas related to the third problem.

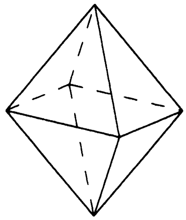
To read this book you need a basic understanding of fundamental notions of the other topology, known as point set topology or general topology. This means that you should know what is meant by words like connected, open, closed, compact, and continuous, and some of the basic facts about them. The notions we need are recalled in Appendix A; if most of this is familiar to you, you should have enough prerequisites. Because of our approach via analysis, you will also need to know some basic facts about calculus, mainly for functions of one or two variables. These calculus facts are set out in Appendix B. In algebra you will need some basic linear algebra, and basic notions about groups, especially abelian groups, which are recalled or proved in Appendix C.

There will be many sorts of *exercises*. Some exercises will be routine applications of or variations on what is done in the text. Those requiring (we estimate) a little more work or ingenuity will be called *problems*. Many will have hints at the end of the book, for you to avoid looking at. There will also be some *projects*, which are things to experiment on, speculate about, and try to develop on your own. For example, one general project can be stated right away: as we go along, try to find analogues in 3-space or n-space for what we do in the plane. (Some of this project is carried out in Part XI.)

Problem 0.1. Suppose X is a graph, which has a finite number of vertices (points) and edges (homeomorphic to a closed interval), with each edge having its endpoints at vertices, and otherwise not intersecting each other. Assume X is connected. When, and how, can you trace out X, traveling along each edge just once? Can you prove your answer?



Exercise 0.2. Let v, e, and f be the number of vertices, edges, and faces on a convex polyhedron. Compute these numbers for the five regular solids, for prisms, and some others. Find a relation among them. Experiment with other polyhedral shapes.



$v = 6$, $e = 12$, $f = 8$

(Note: This problem is "experimental." Proofs are not expected.)

Acknowledgments. I would like to thank the students who came up with good ideas that contributed to notes for the courses, especially K. Ryan, J. Silverman, J. Linhart, J. Trowbridge, and G. Gutman. Thanks also to my colleagues, for answering many of my questions and making useful suggestions, especially A. Collino, J. Harris, R. MacPherson, J.P. May, R. Narasimhan, M. Rothenberg, and S. Weinberger. I am grateful to Chandler Fulton for making the drawings. I would most like to thank those who first inspired me with some of these ideas in courses about three decades ago: H. Federer, J. Milnor, and J. Moore.

William Fulton

Preface to Corrected Edition

I am grateful to J. McClure, J. Buhler, D. Goldberg, and R.B. Burckel for pointing out errors and misprints in the first edition, and for useful suggestions.

Contents

Preface vii

PART I
CALCULUS IN THE PLANE

CHAPTER 1
Path Integrals 3
1a. Differential Forms and Path Integrals 3
1b. When Are Path Integrals Independent of Path? 7
1c. A Criterion for Exactness 10

CHAPTER 2
Angles and Deformations 17
2a. Angle Functions and Winding Numbers 17
2b. Reparametrizing and Deforming Paths 23
2c. Vector Fields and Fluid Flow 27

PART II
WINDING NUMBERS

CHAPTER 3
The Winding Number 35
3a. Definition of the Winding Number 35
3b. Homotopy and Reparametrization 38
3c. Varying the Point 42
3d. Degrees and Local Degrees 43

CHAPTER 4
Applications of Winding Numbers 48
4a. The Fundamental Theorem of Algebra 48
4b. Fixed Points and Retractions 49
4c. Antipodes 53
4d. Sandwiches 56

PART III
COHOMOLOGY AND HOMOLOGY, I

CHAPTER 5
De Rham Cohomology and the Jordan Curve Theorem 63
5a. Definitions of the De Rham Groups 63
5b. The Coboundary Map 65
5c. The Jordan Curve Theorem 68
5d. Applications and Variations 72

CHAPTER 6
Homology 78
6a. Chains, Cycles, and H_0U 78
6b. Boundaries, H_1U, and Winding Numbers 82
6c. Chains on Grids 85
6d. Maps and Homology 89
6e. The First Homology Group for General Spaces 91

PART IV
VECTOR FIELDS

CHAPTER 7
Indices of Vector Fields 97
7a. Vector Fields in the Plane 97
7b. Changing Coordinates 101
7c. Vector Fields on a Sphere 102

CHAPTER 8
Vector Fields on Surfaces 106
8a. Vector Fields on a Torus and Other Surfaces 106
8b. The Euler Characteristic 113

PART V
COHOMOLOGY AND HOMOLOGY, II

CHAPTER 9
Holes and Integrals 123
9a. Multiply Connected Regions 123

9b. Integration over Continuous Paths and Chains 127
9c. Periods of Integrals 130
9d. Complex Integration 131

CHAPTER 10
Mayer–Vietoris 137

10a. The Boundary Map 137
10b. Mayer–Vietoris for Homology 140
10c. Variations and Applications 144
10d. Mayer–Vietoris for Cohomology 147

PART VI
COVERING SPACES AND FUNDAMENTAL GROUPS, I

CHAPTER 11
Covering Spaces 153

11a. Definitions 153
11b. Lifting Paths and Homotopies 156
11c. *G*-Coverings 158
11d. Covering Transformations 163

CHAPTER 12
The Fundamental Group 165

12a. Definitions and Basic Properties 165
12b. Homotopy 170
12c. Fundamental Group and Homology 173

PART VII
COVERING SPACES AND FUNDAMENTAL GROUPS, II

CHAPTER 13
The Fundamental Group and Covering Spaces 179

13a. Fundamental Group and Coverings 179
13b. Automorphisms of Coverings 182
13c. The Universal Covering 186
13d. Coverings and Subgroups of the Fundamental Group 189

CHAPTER 14
The Van Kampen Theorem 193

14a. *G*-Coverings from the Universal Covering 193
14b. Patching Coverings Together 196
14c. The Van Kampen Theorem 197
14d. Applications: Graphs and Free Groups 201

PART VIII
COHOMOLOGY AND HOMOLOGY, III

CHAPTER 15
Cohomology 207

15a. Patching Coverings and Čech Cohomology 207
15b. Čech Cohomology and Homology 210
15c. De Rham Cohomology and Homology 213
15d. Proof of Mayer–Vietoris for De Rham Cohomology 217

CHAPTER 16
Variations 219

16a. The Orientation Covering 219
16b. Coverings from 1-Forms 220
16c. Another Cohomology Group 222
16d. *G*-Sets and Coverings 225
16e. Coverings and Group Homomorphisms 227
16f. *G*-Coverings and Cocycles 228

PART IX
TOPOLOGY OF SURFACES

CHAPTER 17
The Topology of Surfaces 233

17a. Triangulation and Polygons with Sides Identified 233
17b. Classification of Compact Oriented Surfaces 236
17c. The Fundamental Group of a Surface 242

CHAPTER 18
Cohomology on Surfaces 247

18a. 1-Forms and Homology 247
18b. Integrals of 2-Forms 251
18c. Wedges and the Intersection Pairing 252
18d. De Rham Theory on Surfaces 256

PART X
RIEMANN SURFACES

CHAPTER 19
Riemann Surfaces 263

19a. Riemann Surfaces and Analytic Mappings 263
19b. Branched Coverings 268
19c. The Riemann–Hurwitz Formula 272

CHAPTER 20
Riemann Surfaces and Algebraic Curves 277
20a. The Riemann Surface of an Algebraic Curve 277
20b. Meromorphic Functions on a Riemann Surface 281
20c. Holomorphic and Meromorphic 1-Forms 284
20d. Riemann's Bilinear Relations and the Jacobian 289
20e. Elliptic and Hyperelliptic Curves 291

CHAPTER 21
The Riemann–Roch Theorem 295
21a. Spaces of Functions and 1-Forms 295
21b. Adeles 299
21c. Riemann–Roch 303
21d. The Abel–Jacobi Theorem 306

PART XI
HIGHER DIMENSIONS

CHAPTER 22
Toward Higher Dimensions 317
22a. Holes and Forms in 3-Space 317
22b. Knots 320
22c. Higher Homotopy Groups 324
22d. Higher De Rham Cohomology 325
22e. Cohomology with Compact Supports 328

CHAPTER 23
Higher Homology 332
23a. Homology Groups 332
23b. Mayer–Vietoris for Homology 334
23c. Spheres and Degree 339
23d. Generalized Jordan Curve Theorem 343

CHAPTER 24
Duality 346
24a. Two Lemmas from Homological Algebra 346
24b. Homology and De Rham Cohomology 350
24c. Cohomology and Cohomology with Compact Supports 355
24d. Simplicial Complexes 359

APPENDICES

APPENDIX A
Point Set Topology 367
A1. Some Basic Notions in Topology 367
A2. Connected Components 369

A3. Patching 370
A4. Lebesgue Lemma 371

APPENDIX B
Analysis 373
B1. Results from Plane Calculus 373
B2. Partition of Unity 375

APPENDIX C
Algebra 378
C1. Linear Algebra 378
C2. Groups; Free Abelian Groups 380
C3. Polynomials; Gauss's Lemma 385

APPENDIX D
On Surfaces 387
D1. Vector Fields on Plane Domains 387
D2. Charts and Vector Fields 389
D3. Differential Forms on a Surface 391

APPENDIX E
Proof of Borsuk's Theorem 393

Hints and Answers 397

References 419

Index of Symbols 421

Index 425

PART I
CALCULUS IN THE PLANE

In this first part we will recall some basic facts about differentiable functions, forms, and vector fields, and integration over paths. Much of this should be familiar, although perhaps from a different point of view. At any rate, several of these notions will be needed later, so we take this opportunity to fix the ideas and notation. And of course, we will be looking particularly at the role played by the shapes (topology) of the underlying regions where these things are defined. For the facts that we use, see Appendix B either for precise statements or proofs. Most of this material is included mainly for motivation, and will be developed from a purely topological point of view later; one fact proved in the first chapter—that a closed 1-form on an open rectangle is the differential of a function—will be used later.

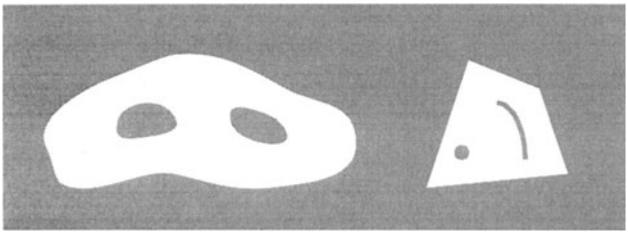
In the second chapter we will see that for any smooth path not passing through the origin, it is possible to define a smooth function that measures how the angle is changing as one moves along the path. This gives us a notion of winding number—how many times a closed path "goes around" the origin. Facts about changing variables in integrals are used to see what happens to integrals and winding numbers when paths are reparametrized and deformed. The third section includes a reinterpretation of the facts from the first chapter in vector field language, and gives a physical interpretation of these ideas to fluid flow. Although we will not use these facts in the book, we will study vector fields later, and it should be useful to have some feeling for them, if you don't already.

CHAPTER 1

Path Integrals

1a. Differential Forms and Path Integrals

In this chapter, U will denote an open set in the plane $\mathbb{R}^2$, for example, the unshaded part of



A *smooth* or $\mathscr{C}^\infty$ *function* on U is a function $f: U \to \mathbb{R}$ such that all partial derivatives of all orders[1] exist and are continuous. In particular, its partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ are $\mathscr{C}^\infty$ functions on U. Since in this chapter we will only consider $\mathscr{C}^\infty$ functions, we will sometimes just call them functions.

A function f on U is called *locally constant* if every point of U has a neighborhood on which f is constant.

[1] We will never need more than continuous second derivatives, and often much less. The few functions that we actually use, however, will be infinitely differentiable. The extra hypotheses are included so we never have to worry about differentiating any function we meet. The analytically inclined reader may enjoy supplying minimal hypotheses for each assertion.

Exercise 1.1. Prove that a function on an open set U in the plane is locally constant if and only if it is constant on each connected component of U. In other words, defining a locally constant function on U is the same as specifying a constant for each of its connected components.

If f is locally constant, then $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$ (identically, as functions on U), as follows immediately from the definitions of partial derivatives. The converse is also true and only slightly harder:

Proposition 1.2. *If f is a smooth function on U, then f is locally constant if and only if $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$.*

Proof. The point is that, in a rectangular neighborhood of a point of U, the condition $\partial f/\partial x = 0$ means that f is independent of x, i.e., that f is constant along horizontal lines. Likewise $\partial f/\partial y = 0$ means that f is constant along vertical lines, and both conditions make f constant in the rectangle. □

It may not be much, but there is a grain of topology in this:

Corollary 1.3. *The open set U is connected if and only if every smooth function f in U with $\partial f/\partial x = 0$ and $\partial f/\partial y = 0$ is constant.* □

A *differential* 1*-form*, or just a 1*-form*, on U is given by a pair of smooth functions p and q on U. We will usually denote a 1-form by ω, and we will write $\omega = p\,dx + q\,dy$. This can be regarded as just a formal notation, with the dx and dy there merely to indicate what we will do with 1-forms, namely integrate them over paths. The pair of functions (p, q) can also be identified with a vector field on U. For this interpretation, see §2c in Chapter 2.

By a *smooth path* (just called a *path* in this chapter) in U, we mean a mapping $\gamma\colon [a, b] \to U$ from a bounded interval into U that is continuous on $[a, b]$ and differentiable in the open interval (a, b); in addition, to avoid any trouble at the endpoints, we assume the two component functions of γ can be extended to $\mathscr{C}^\infty$ functions in some neighborhood of $[a, b]$. So $\gamma(t) = (x(t), y(t))$, where x and y are restrictions of smooth functions on an interval[2] $(a - \varepsilon, b + \varepsilon)$, for some

[2] In fact, there are many extensions of these functions to such neighborhoods, but we will never care about values outside the interval $[a, b]$. The assumption is useful to assure that the derivatives of these functions are continuous on the whole closed interval $[a, b]$.

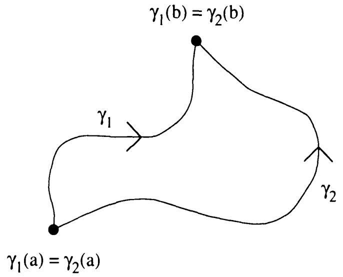
positive number ε. We call $\gamma(a)$ the *initial* point of γ, and $\gamma(b)$ the *final* point; $\gamma(a)$ and $\gamma(b)$ are called the *endpoints*, and we say that γ is a path *from* $\gamma(a)$ *to* $\gamma(b)$.

With $\omega = p\,dx + q\,dy$ as above, and γ a path given by the pair of functions $\gamma(t) = (x(t), y(t))$, the *integral* $\int_\gamma \omega$ *of* ω *along* γ is defined by the formula

$$\int_\gamma \omega = \int_a^b \left(p(x(t), y(t)) \frac{dx}{dt} + q(x(t), y(t)) \frac{dy}{dt} \right) dt .$$

Note that the integrand is continuous on $[a, b]$, so the integral exists, as a limit of Riemann sums.

The question we will be concerned with is this: given a 1-form ω on U, when does the integral $\int_\gamma \omega$ depend only on the endpoints $\gamma(a)$ and $\gamma(b)$ of γ, and not on the actual path between them?



Language is usually abused here, saying the integral is "independent of path." This happens whenever there is a "potential function":

Proposition 1.4. *If* $\omega = \partial f/\partial x\,dx + \partial f/\partial y\,dy$, *for some* $\mathscr{C}^\infty$ *function* f *on an open set containing the path* γ, *then*

$$\int_\gamma \omega = f(\gamma(b)) - f(\gamma(a)) .$$

Proof. Since, by the chain rule,

$$\frac{d}{dt}(f(\gamma(t))) = \frac{\partial f}{\partial x}(x(t), y(t)) \frac{dx}{dt} + \frac{\partial f}{\partial y}(x(t), y(t)) \frac{dy}{dt},$$

the integral is

$$\int_\gamma \omega = \int_a^b \frac{d}{dt}(f(\gamma(t)))\,dt = f(\gamma(b)) - f(\gamma(a)),$$

the last step by the fundamental theorem of calculus. $\square$

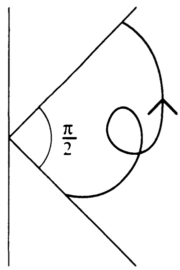
We write $df = \partial f/\partial x\, dx + \partial f/\partial y\, dy$ for this 1-form, and say that ω is the *differential* of f if $\omega = df$.

Exercise 1.5. Show that $df = dg$ on U if and only if $f - g$ is locally constant on U.

For an example, take U to be the right half plane, i.e., the set of points (x, y) with $x > 0$. Consider the function f that measures the angle in polar coordinates, measured counterclockwise from the x-axis. Analytically, $f(x, y) = \tan^{-1}(y/x)$, so

$$\begin{aligned} df &= \frac{1}{1+(y/x)^2}\left(-\frac{y}{x^2}\right)dx + \frac{1}{1+(y/x)^2}\left(\frac{1}{x}\right)dy \\ &= \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy = \frac{-y\,dx + x\,dy}{x^2+y^2}. \end{aligned}$$

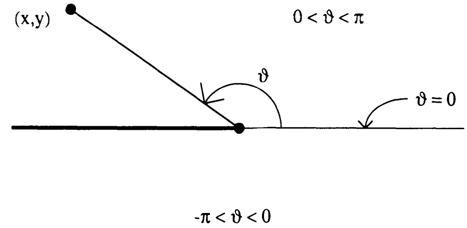
For example, if γ is any path in U from $(1, -1)$ to $(2, 2)$, then $\int_\gamma df = \pi/2$, since that is the change in angle between the two points.



Although the function $f(x, y) = \tan^{-1}(y/x)$, at least as it stands, is not defined where $x = 0$, the expression we found for df makes sense everywhere except at the origin, and is a smooth 1-form on the open set $\mathbb{R}^2 \setminus \{(0, 0)\}$. Let us denote this 1-form by ω_ϑ:

$$\omega_\vartheta = \frac{-y\,dx + x\,dy}{x^2+y^2} \quad \text{on} \quad \mathbb{R}^2 \setminus \{(0,0)\}.$$

In fact, although y/x cannot be extended across the y-axis, the function $\tan^{-1}(y/x)$ can, at least away from the origin. This is clear if we think of it geometrically as the angle in polar coordinates, which can be extended, for example, to the complement of the negative x-axis:



However, there is trouble in trying to extend this angle function to be well defined everywhere on $\mathbb{R}^2 \setminus \{(0,0)\}$. In fact, we can use our last proposition to show that there is no smooth function g on $\mathbb{R}^2 \setminus \{(0,0)\}$ with $dg = \omega_\vartheta$. For example, if $\gamma(t) = (\cos(t), \sin(t))$, $0 \le t \le 2\pi$, is the counterclockwise path around the unit circle, we calculate using the definition of the path integral:

$$\begin{aligned}\int_\gamma \omega_\vartheta &= \int_0^{2\pi} (-\sin(t)\cdot(-\sin(t)) + \cos(t)\cdot\cos(t))\,dt \\ &= \int_0^{2\pi} 1\,dt \;=\; 2\pi.\end{aligned}$$

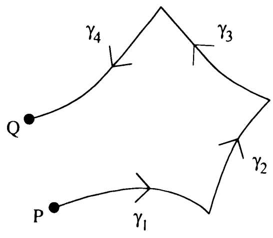
Since $\gamma(0) = \gamma(2\pi)$, it follows from Proposition 1.4 that ω_ϑ cannot be the differential of any function.

Exercise 1.6. On which of the following open sets U is there a smooth function g with $dg = \omega_\vartheta$ on U? Prove your answers. (i) The upper half plane $\{(x,y)\colon y > 0\}$. (ii) The union of the upper half plane and the right half plane. (iii) The left half plane. (iv) The lower half plane. (v) The complement of the negative x-axis. (vi) The annulus $\{(x,y)\colon 1 < x^2 + y^2 < 2\}$. (vii) *Challenge*. The points of the form $(re^t\cos(t), re^t\sin(t))$, $0 < t < 4\pi$, $1/2 < r < 2$.

Exercise 1.7. Is $\omega = (x\,dx + y\,dy)/(x^2+y^2)^2$ the differential of a function on $\mathbb{R}^2 \setminus \{(0,0)\}$?

1b. When Are Path Integrals Independent of Path?

It will be useful to generalize the notion of smooth path in order to allow integration over a sequence of such paths. Let us define a *segmented path* γ to be a sequence of paths $\gamma_1, \gamma_2, \ldots, \gamma_n$, where each γ_i is a smooth path, and the final point of each γ_i is the initial point of the next γ_{i+1}, for $i = 1, 2, \ldots, n-1$.



We sometimes write $\gamma = \gamma_1 + \ldots + \gamma_n$ for this segmented path. The *initial* point of γ is defined to be the initial point of γ_1, and the *final* point of γ is defined to be the final point of γ_n. The segmented path is *closed* if the final point of γ_n is the initial point of γ_1. If ω is a 1-form on an open set containing (the images of) these paths, we define

$$\int_\gamma \omega = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega + \ldots + \int_{\gamma_n} \omega .$$

If γ is a segmented path in U from P to Q, and $\omega = df$ in U, then it follows from Proposition 1.4 (the "interior endpoints" canceling) that

$$\int_\gamma \omega = f(Q) - f(P).$$

We'll show now that the converse of this is also true:

Proposition 1.8. *Let ω be a 1-form on U. The following are equivalent*: (i) $\int_\gamma \omega = \int_\delta \omega$ *for all segmented paths γ and δ in U with the same initial and final points*; (ii) $\int_\tau \omega = 0$ *for all segmented paths τ in U that are closed*; *and* (iii) $\omega = df$ *for some smooth function f on U.*

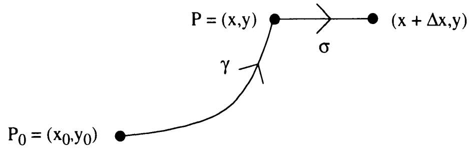
Proof. The preceding remark shows that (iii) implies (i). To show that (ii) implies (i), we use the notion of the *inverse* of a path $\sigma\colon [a,b] \to U$, which is the path $\sigma^{-1}\colon [a,b] \to U$ defined by $\sigma^{-1}(t) = \sigma(b + a - t)$; note that the integral of any ω along σ^{-1} is the negative of the integral of ω along σ (cf. Exercise 2.12). Given γ and δ as in (ii), form the closed segmented path τ which is first the sequence of paths making up γ, and then the inverses of the paths that make up δ, but taken in the reverse order. Then $\int_\tau \omega = \int_\gamma \omega - \int_\delta \omega$, from which the fact that (ii) implies (i) follows. That (i) implies (ii) is obvious, by comparing a closed path with a constant path. To show that (i) implies (iii), it is enough to find such a function on each connected component of U, so we can assume U is connected, and hence path-connected (see Appendix A2). Choose and fix an arbitrary point P_0 in U, and define a function f on U by the formula

$$f(P) = \int_\gamma \omega,$$

where γ is any segmented path from P_0 to P in U. (See Exercise 1.9 below.) By assumption, this is a well-defined function on U. We claim that $\partial f/\partial x = p$ and $\partial f/\partial y = q$, where $\omega = p\,dx + q\,dy$. For the first, we must look at the limit of

$$\frac{f(x+\Delta x, y) - f(x,y)}{\Delta x}$$

as Δx approaches zero, and $P = (x,y)$ is any point in U. To estimate this, let σ be the path from (x,y) to $(x+\Delta x, y)$ given by the formula $\sigma(t) = (x+t, y)$ $0 \le t \le \Delta x$, assuming for the moment that Δx is positive. Let γ be any segmented path from P_0 to P.



Since $\gamma + \sigma$ is a segmented path from P_0 to $(x+\Delta x, y)$,

$$\begin{aligned}\frac{f(x+\Delta x, y) - f(x,y)}{\Delta x} &= \frac{1}{\Delta x}\left(\int_{\gamma+\sigma}\omega - \int_\gamma \omega\right) = \frac{1}{\Delta x}\left(\int_\sigma \omega\right) \\ &= \frac{1}{\Delta x}\int_0^{\Delta x} p(x+t, y)\,dt.\end{aligned}$$

By the mean value theorem, this last expression is equal to $p(x^*, y)$ for some x^* between x and $x + \Delta x$. Letting $\Delta x \to 0$, we have, since p is continuous,

$$\lim_{\Delta x \to 0} \frac{1}{\Delta x}\int_0^{\Delta x} p(x+t, y)\,dt = p(x,y),$$

as required. If Δx is negative, use instead the path $\sigma(t) = (x-t, y)$, $0 \le t \le |\Delta x|$, and the argument is the same, with the modification

$$\begin{aligned}\frac{1}{\Delta x}\left(\int_\sigma \omega\right) &= \frac{-1}{|\Delta x|}\int_0^{|\Delta x|} p(x-t, y)(-1)\,dt \\ &= \frac{1}{|\Delta x|}\int_0^{|\Delta x|} p(x-t, y)\,dt = p(x^*, y),\end{aligned}$$

with $x + \Delta x \leq x^* \leq x$. The proof that $df/\partial y = q$ is similar, interchanging the roles of x and y, and is left as an exercise. □

Exercise 1.9. Show that an open set U in the plane is connected if and only if there is a segmented path between any two points of U. *Challenge*. Can you show that any two points in a connected open set can be connected by an *arc*, i.e., a path that is one-to-one, and whose tangent vector never vanishes?

1c. A Criterion for Exactness

We want a practical criterion to tell if a given 1-form ω is the differential of some function without going to the work of constructing such a function. We shall need another fact from calculus, the equality of mixed partial derivatives: $\partial/\partial x(\partial f/\partial y) = \partial/\partial y(\partial f/\partial x)$. This translates to a simple

Criterion 1.10. *$\omega = p\,dx + q\,dy$ cannot be the differential of a function unless $\partial q/\partial x = \partial p/\partial y$.*

This necessary condition, however, is not always sufficient. For example, if ω_ϑ is the 1-form on $U = \mathbb{R}^2 \setminus \{(0,0)\}$ that we looked at earlier, you can verify easily that ω_ϑ satisfies this condition—either by direct calculation, or by the fact that any point in U has a neighborhood on which the restriction of ω_ϑ is the differential of a function—but we have seen that ω_ϑ cannot be the differential of any function on U. We also saw that the restrictions of ω_ϑ to some simpler open sets, like the right half plane, are the differentials of functions. We will see that it is the topology of U that is controlling this situation.

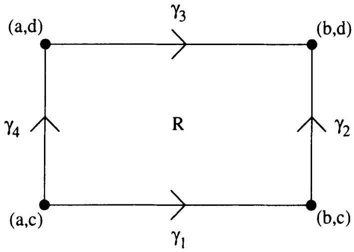
A *2-form* on U is an expression $h\,dx\,dy$, where h is a $\mathscr{C}^\infty$ function on U. Logically, as before, a 2-form can be identified with the function h that defines it, with the "$dx\,dy$" playing only a formal role. The notation indicates that we will use 2-forms for integrating over two-dimensional regions. All we will need is double integrals $\iint_R h\,dx\,dy$ over rectangles $R = [a,b] \times [c,d]$, defined as limits of Riemann sums.

If $\omega = p\,dx + q\,dy$ is a 1-form on U, define $d\omega$ to be the 2-form

$$d\omega = \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right) dx\,dy.$$

So our criterion can be stated: if $\omega = df$, then $d\omega = 0$; or simply that $d(df) = 0$ for all functions f.

Let $R = [a, b] \times [c, d]$ be a closed (bounded) rectangle, and consider the four boundary segments:



In formulas,

$$\gamma_1(t) = (t, c), \quad a \le t \le b; \qquad \gamma_2(t) = (b, t), \quad c \le t \le d;$$
$$\gamma_3(t) = (t, d), \quad a \le t \le b; \qquad \gamma_4(t) = (a, t), \quad c \le t \le d.$$

We will need Green's theorem for a rectangle (see Appendix B). This says that if ω is a 1-form on an open set containing the rectangle R, then

$$\int_{\partial R} \omega = \iint_R d\omega,$$

where

$$\int_{\partial R} \omega = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega - \int_{\gamma_3} \omega - \int_{\gamma_4} \omega.$$

We will need only the following consequence:

Lemma 1.11. *If $d\omega = 0$, then $\int_{\partial R} \omega = 0$, i.e.,*

$$\int_{\gamma_1} \omega + \int_{\gamma_2} \omega = \int_{\gamma_3} \omega + \int_{\gamma_4} \omega.$$

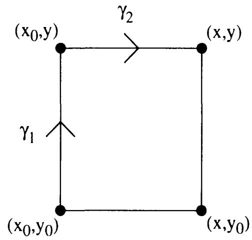
We can apply this to show that, on the plane, or a half plane, or any rectangle, the necessary condition $d\omega = 0$ is actually sufficient for integrals to be path-independent:

Proposition 1.12. *Let U be a product of two open finite or infinite intervals, i.e.,*

$$U = \{(x, y) : a < x < b \quad \text{and} \quad c < y < d\},$$

with $-\infty \le a < b \le \infty$ *and* $-\infty \le c < d \le \infty$. *If* ω *is any* 1*-form on* U *such that* $d\omega = 0$, *then there is a function* f *on* U *with* $\omega = df$.

Proof. Fix a point $P_0 = (x_0, y_0)$ in U. For $P = (x, y)$ in U, let $f(P) = \int_\gamma \omega$, where $\gamma = \gamma_1 + \gamma_2$ is the path shown:

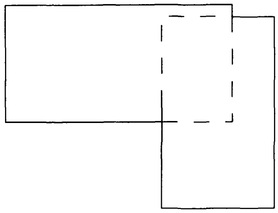


Assume for the moment that $x \ge x_0$ and $y \ge y_0$. The formulas are $\gamma_1(t) = (x_0, y_0 + t)$, $0 \le t \le y - y_0$, and $\gamma_2(t) = (x_0 + t, y)$, $0 \le t \le x - x_0$. The last calculation in the proof of Proposition 1.8 shows that $\partial f/\partial x = p$, where $\omega = p\,dx + q\,dy$. If $y < y_0$, the same is true, replacing $\gamma_1(t)$ by $(x_0, y_0 - t)$, $0 \le t \le y_0 - y$; and similarly if $x < x_0$, replace $\gamma_2(t)$ by $(x_0 - t, y)$, $0 \le t \le x_0 - x$.

Similarly, define a function g by $g(P) = \int_{\gamma^*} \omega$, where γ^* is the path that first goes horizontally from (x_0, y_0) to (x, y_0) and then goes vertically from (x, y_0) to (x, y). The same argument, again left as an exercise, shows that $\partial g/\partial y = q$. Our assumptions on U imply that the closed rectangle with opposite corners at P_0 and P is contained in U, so that Green's theorem can be applied, and it follows from Lemma 1.11 that $f(P)$ is equal to $g(P)$. It follows that $\partial f/\partial x = p$ and $\partial f/\partial y = q$, which means that $df = \omega$. □

Exercise 1.13. Show that the proposition is also true when U is the inside of a disk, i.e., $U = \{(x, y): (x - a)^2 + (y - b)^2 < r^2\}$. Can you prove it when U is any convex region, or any starshaped region? (Convex means that the straight line between any two points in the region is contained in the region, and starshaped means that there is a point P_0 in the region such that for any point P in the region, the straight line from P_0 to P is contained in the region.)

A 1-form ω is called *closed* if $d\omega = 0$, and it is *exact* if $\omega = df$ for some function f. So all exact forms are closed, and the last proposition says that, when U is a rectangle, all closed forms are exact. There are many other regions U for which this is true, besides rectangles and those in the preceding exercise. For example, if U is the union of any two rectangles, each as in the proposition, then any closed 1-form ω in U is exact.

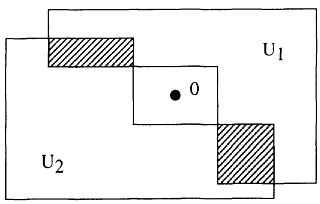


To see this, let U_1 and U_2 be the two open rectangles whose union is U. By the proposition, there are functions f_1 in U_1 and f_2 in U_2 such that $df_1 = \omega$ on U_1 and $df_2 = \omega$ on U_2. Since $U_1 \cap U_2$ is connected, and since $d(f_2 - f_1) = \omega - \omega = 0$ on $U_1 \cap U_2$, it follows that $f_2 - f_1$ is a constant function on $U_1 \cap U_2$. If we replace f_2 by $f_2 - c$, where c is this constant, we can assume f_1 and f_2 agree on $U_1 \cap U_2$. This means that there is a function f on $U = U_1 \cup U_2$ whose restriction to U_1 is f_1 and whose restriction to U_2 is f_2. Moreover, $df = \omega$ on all of U, since this condition is a local condition, to be verified at each point of U; and every point is either in U_1 or U_2, where we know $df_1 = \omega$ and $df_2 = \omega$.

In fact, this argument proves:

Lemma 1.14. *Suppose U_1 and U_2 are open sets, and $U_1 \cap U_2$ is connected. Let $U = U_1 \cup U_2$, and let ω be a 1-form on U. If the restrictions of ω to U_1 and U_2 are both exact, then ω is exact on U.* □

The connectedness of $U_1 \cap U_2$ is crucial for this. For example, suppose U_1 and U_2 are the regions indicated, with the shaded overlap, and the origin outside in the middle.



Let ω be the restriction of ω_ϑ to U, then ω_ϑ is the differential of a function on each of the regions, but not on their union, as can be seen by integrating ω along a path around the origin in U.

Exercise 1.15. Show that if U is a union of open sets $U_1, \ldots, U_n$, and ω is a 1-form on U such that the restriction of ω to each U_i is exact, and $(U_1 \cup U_2 \cup \ldots \cup U_i) \cap U_{i+1}$ is connected for $1 \leq i \leq n-1$, then ω is exact on U.

If ω is a closed 1-form in an open set U, although ω may not be exact on all of U, it follows from Proposition 1.12 that it is always locally exact. That is, any point in U has a neighborhood (say rectangular), so that the restriction of ω to this neighborhood is the differential of a function. This can be used to calculate path integrals, by cutting the path into pieces, on each of which Proposition 1.4 can be applied:

Proposition 1.16. *If ω is a closed 1-form on U, and γ: $[a,b] \to U$ is a smooth path, then there is a subdivision $a = t_0 < t_1 < \ldots < t_n = b$ and a collection of open subsets $U_1, \ldots, U_n$ of U so that γ maps $[t_{i-1}, t_i]$ into U_i, and the restriction of ω to U_i is the differential of a function f_i. Let $P_i = \gamma(t_i)$. Then, for any such choices,*

$$\int_\gamma \omega = (f_1(P_1) - f_1(P_0)) + (f_2(P_2) - (f_2(P_1)) + \ldots + (f_n(P_n) - f_n(P_{n-1})).$$

Proof. For each point P in $\gamma([a,b])$, choose a neighborhood U_P of P on which the restriction of ω is exact. The open sets $\gamma^{-1}(U_P)$ form an open covering of the compact interval $[a,b]$, so a finite number of them cover the interval. From this it is not hard to construct the subdivision. One quick way to do it is to use the Lebesgue covering lemma (see §A4 of Appendix A), which guarantees that, if the subdivision is small enough, each subinterval will be mapped into one of the neighborhoods U_P. Having fixed such a subdivision, choose one of these open sets containing the image of $[t_{i-1}, t_i]$, call it U_i, and choose a function f_i on U_i with $df_i = \omega$ on U_i. Let γ_i: $[t_{i-1}, t_i] \to U_i$ be the restriction of γ to $[t_{i-1}, t_i]$. Then

$$\int_\gamma \omega = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega + \ldots + \int_{\gamma_n} \omega$$
$$= \sum_{i=1}^{n} (f_i(\gamma_i(t_i)) - f_i(\gamma_i(t_{i-1}))) = \sum_{i=1}^{n} (f_i(P_i) - f_i(P_{i-1})). \quad \square$$

The following two lemmas will be used in Part III to prove the Jordan curve theorem. They are special cases of general theorems to be proved in Chapter 9, but we can prove them directly with the methods of this section. For any positive number r, let $\gamma_{P,r}$ be the counterclockwise circle of radius r about P:

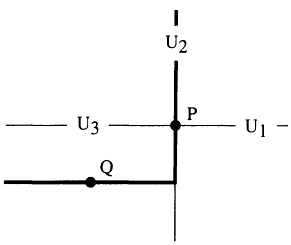
$$\gamma_{P,r}(t) = P + r(\cos(2\pi t), \sin(2\pi t)), \qquad 0 \leq t \leq 1.$$

Lemma 1.17. *Suppose $U = \mathbb{R}^2 \setminus \{P\}$, and let $r > 0$. If ω is a closed 1-form on U such that $\int_{\gamma_{P,r}} \omega = 0$, then ω is exact.*

Proof. Let U_1, U_2, U_3, and U_4 be the half planes to the right of P, above P, to the left of P, and below P. By Proposition 1.12, there are functions f_i on U_i, unique up to the addition of constants, with $df_i = \omega$ on U_i. By adjusting the constants, we may assume $f_2 = f_1$ on $U_1 \cap U_2$, and $f_3 = f_2$ on $U_2 \cap U_3$, and $f_4 = f_3$ on $U_3 \cap U_4$. Then $f_4 = f_1 + c$ on $U_4 \cap U_1$ for some constant c. By Proposition 1.16, $\int_{\gamma_{P,r}} \omega = c$. The hypothesis implies that $c = 0$, which means exactly that the four functions f_i agree on overlaps, so define a function f on U such that $df = \omega$. □

Lemma 1.18. *Suppose $U = \mathbb{R}^2 \setminus \{P, Q\}$, and let $0 < r < |P - Q|$. If ω is a closed 1-form on U such that $\int_{\gamma_{P,r}} \omega = 0$ and $\int_{\gamma_{Q,r}} \omega = 0$, then ω is exact.*

Proof. The proof is similar. Suppose for definiteness that Q is located to the southwest of P. Let U_1 be the half plane to the right of P, and choose f_1 on U_1 so that $\omega = df_1$ on U_1. Let U_2 be the half plane above P, and choose f_2 on U_2 so that $\omega = df_2$ on U_2, and $f_2 = f_1$ on $U_1 \cap U_2$. Let U_3 be the quarter plane to the left of P and above Q, and choose f_3 on U_3 so that $\omega = df_3$ on U_3, and $f_3 = f_2$ on $U_2 \cap U_3$.



Let U_4 be the quarter plane to the right of Q and below P, and choose f_4 on U_4 so that $\omega = df_4$ on U_4, and $f_4 = f_3$ on $U_3 \cap U_4$. Then f_4 and f_1 differ by a constant on $U_4 \cap U_1$, and the hypothesis that $\int_{\gamma_{P,r}} \omega = 0$ implies as in the preceding lemma that this constant is 0. Let U_5 be the half plane to the left of Q, and choose f_5 on U_5 such that $\omega = df_5$ on U_5 and $f_5 = f_3$ on $U_3 \cap U_5$. Let U_6 be the half plane below Q, and choose f_6 on U_6 such that $\omega = df_6$ on U_6 and $f_6 = f_5$ on $U_5 \cap U_6$. The hypothesis that $\int_{\gamma_{Q,r}} \omega = 0$ implies that $f_6 = f_4$ on $U_4 \cap U_6$. The functions f_i on U_i agree on overlaps, defining a function f on U such that $\omega = df$, which completes the proof in this case.

The proof when Q is located southeast of P, or the special cases when P and Q are on a horizontal or vertical line, are similar and left to the reader. □

Problem 1.19. Generalize the preceding lemmas from one or two to n points.

For any point $P = (x_0, y_0)$, define the 1-form $\omega_{P,\vartheta}$ on $\mathbb{R}^2 \setminus \{P\}$ by the formula

$$\omega_{P,\vartheta} = \frac{-(y - y_0)\,dx + (x - x_0)dy}{(x - x_0)^2 + (y - y_0)^2}.$$

Problem 1.20. For any two points P and Q, show that the 1-form $\omega = \omega_{P,\vartheta} - \omega_{Q,\vartheta}$ is exact on $\mathbb{R}^2 \setminus L$, where L is the line segment from P to Q. *Challenge*. Find a function whose differential is ω.

CHAPTER 2

Angles and Deformations

2a. Angle Functions and Winding Numbers

Any point in the plane can be expressed in polar coordinates, i.e., it can be written in the form $(r\cos(\vartheta), r\sin(\vartheta))$ for some $r \geq 0$ and some real number ϑ. The radius r is unique, being the distance from the origin, or the square root of the sums of the squares of the Cartesian coordinates. At the origin, $r = 0$, and ϑ can be any number. We often denote the origin simply by 0 instead of $(0, 0)$. Except for the origin, the angle ϑ is determined only up to adding integral multiples of 2π. We call any of these numbers *an angle* for the point.

Suppose $\gamma\colon [a, b] \to \mathbb{R}^2 \setminus \{0\}$ is a $\mathscr{C}^\infty$ path to the complement of the origin, given in Cartesian coordinates by $\gamma(t) = (x(t), y(t))$. We want to describe this in polar coordinates, that is, to find $\mathscr{C}^\infty$ functions $r(t)$ and $\vartheta(t)$ so that

$$(2.1) \quad \gamma(t) = r(t)(\cos(\vartheta(t)), \sin(\vartheta(t))) = (r(t)\cos(\vartheta(t)), r(t)\sin(\vartheta(t)))$$

for all $a \leq t \leq b$. There is no problem with the function r: it is the distance from the origin, defined by

$$r(t) = \|\gamma(t)\| = \sqrt{x(t)^2 + y(t)^2}.$$

The angle function is not so simple. At any time t, there are many possible angles to choose, all differing by multiples of 2π. If we choose them say all lying in the interval $(-\pi, \pi]$ they would be unique, but then would not vary continuously if the point crosses the negative x-axis.

The initial angle can be chosen arbitrarily. In other words, choose any number ϑ_a so that

$$\gamma(a) = (x(a), y(a)) = (r(a)\cos(\vartheta_a), r(a)\sin(\vartheta_a)).$$

We will require that our function $\vartheta(t)$ satisfy the initial condition $\vartheta(a) = \vartheta_a$. Motivated by the discussion in §1b, we have a candidate for the derivative $\vartheta'(t)$, which is the rate of change of angle: it should be

$$\vartheta'(t) = \frac{-y(t)x'(t) + x(t)y'(t)}{x(t)^2 + y(t)^2} = \frac{-y(t)x'(t) + x(t)y'(t)}{r(t)^2}.$$

So we can define $\vartheta(t)$ to be the unique function with this initial condition and this derivative. That is, we define $\vartheta(t)$ by

$$\vartheta(t) = \vartheta_a + \int_a^t \frac{-y(\tau)x'(\tau) + x(\tau)y'(\tau)}{r(\tau)^2}\,d\tau.$$

Proposition 2.2. *With these definitions, the functions $r(t)$ and $\vartheta(t)$ are $\mathscr{C}^\infty$ functions, and equation* (2.1) *is satisfied for all t in* $[a, b]$.

Proof. The function $r(t)$ is $\mathscr{C}^\infty$ since it is a composite of $\mathscr{C}^\infty$ functions, and $\vartheta(t)$ is $\mathscr{C}^\infty$ since it is the integral of a $\mathscr{C}^\infty$ function. Let

$$u(t) = (\cos(\vartheta(t)), \sin(\vartheta(t))) \quad \text{and} \quad v(t) = (-\sin(\vartheta(t)), \cos(\vartheta(t))).$$

These are perpendicular unit vectors for all t. We want to show that

$$\frac{1}{r(t)}\gamma(t) = u(t)$$

for all t. We are assuming that they are equal for $t = a$. It suffices to show that both sides of this equation have the same dot product with vectors $u(t)$ and $v(t)$, since the difference $(1/r(t))\gamma(t) - u(t)$ would then be perpendicular to two independent vectors, and so would be zero. For the right side of the equation $u(t)$, these dot products are identically 1 and 0, respectively, so it suffices to prove that the same is true for the left side. Since we know the equality of the vectors for $t = a$, it suffices to prove that the derivatives of these dot products vanish. We are therefore reduced to showing that

$$\frac{d}{dt}\left(\frac{1}{r(t)}\gamma(t)\cdot u(t)\right) \equiv 0 \quad \text{and} \quad \frac{d}{dt}\left(\frac{1}{r(t)}\gamma(t)\cdot v(t)\right) \equiv 0.$$

These are simple verifications, using the usual rules of calculus. Omitting the variable t, and writing $\dot{r}$ in place of $r'(t)$, etc., the first of these derivatives is

$$\frac{r\dot{x}-x\dot{r}}{r^2}\cos(\vartheta)+\frac{r\dot{y}-y\dot{r}}{r^2}\sin(\vartheta)+\frac{x}{r}\dot{\vartheta}(-\sin(\vartheta))+\frac{y}{r}\dot{\vartheta}\,(\cos(\vartheta))$$

$$= \frac{1}{r^3}\left(\cos(\vartheta)[r^2\dot{x}-xr\dot{r}+yr^2\dot{\vartheta}]+\sin(\vartheta)[r^2\dot{y}-yr\dot{r}-xr^2\dot{\vartheta}]\right).$$

Using the identities $r\dot{r}=x\dot{x}+y\dot{y}$ and $r^2\dot{\vartheta}=-y\dot{x}+x\dot{y}$, one sees that the terms in brackets vanish. The proof that the other derivative vanishes is similar and left as an exercise. □

Exercise 2.3. (a) Show that this function $\vartheta(t)$ is unique, and that in fact it is the only continuous function with $\vartheta(a)=\vartheta_a$ such that (2.1) holds. (b) Show that if ϑ_a is replaced by $\vartheta_a+2\pi n$ for some n, then the corresponding angle function is $\vartheta(t)+2\pi n$.

Using this angle function, we can define the (total signed) *change in angle* of the path γ to be $\vartheta(b)-\vartheta(a)$. Note by the preceding exercise that this is independent of the choice of initial angle. Equivalently, this change in angle is

$$\int_a^b \frac{-y(t)x'(t)+x(t)y'(t)}{r(t)^2}\,dt = \int_\gamma \frac{-y\,dx+x\,dy}{x^2+y^2} = \int_\gamma \omega_\vartheta.$$

If a segmented path does not pass through the origin, i.e., it is a path in $\mathbb{R}^2\setminus\{0\}$, we want to define its *winding number*, which should be the "net" number of times γ goes around the origin, counting the counterclockwise motion as positive, and the clockwise motion as negative. In other words, it is the total signed change in angle, divided by 2π. We will denote it by $W(\gamma,0)$, the "0" indicating that we are going around the origin. What we have just done shows what the definition should be. We define the *winding number of* γ *around* *the origin* by

$$W(\gamma, 0) = \frac{1}{2\pi}\int_\gamma \omega_\vartheta = \frac{1}{2\pi}\int_\gamma \frac{-y\,dx + x\,dy}{x^2+y^2}.$$

Proposition 2.4. *For any closed segmented path that does not pass through the origin, the winding number of the path around the origin is an integer.*

Proof. The formula given before the proposition defines a number $W(\gamma, 0)$ for any segmented path γ that does not pass through the origin. Suppose γ starts at P and goes to Q, and we choose an angle ϑ_P for P. We claim that $\vartheta_P + 2\pi W(\gamma, 0)$ is an angle for Q. This will prove the result, for when $P = Q$ it says that ϑ_P and $\vartheta_P + 2\pi W(\gamma, 0)$ define the same angle, so they must differ by an integer multiple of 2π.

It suffices to prove the claim for a smooth path, since if γ is a sum of smooth paths, the assertion for each of them implies it for the sum. But when $\gamma\colon [a, b] \to \mathbb{R}^2 \setminus \{0\}$ is smooth, $\vartheta_P + 2\pi W(\gamma, 0)$ is just the value of our angle function $\vartheta(t)$ at $t = b$, starting with $\vartheta(a) = \vartheta_P$, so the result follows from Proposition 2.2. □

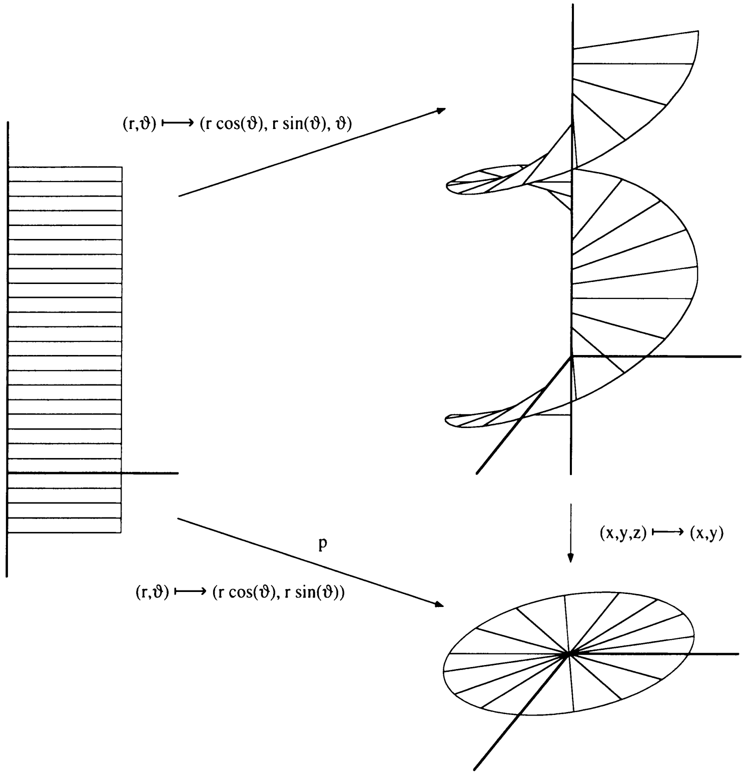
Exercise 2.5. Use Proposition 1.16 to give another proof of this proposition.

Exercise 2.6. Let $\gamma(t) = (k\cos(nt), k\sin(nt))$, $0 \le t \le 2\pi$, where k is a positive number and n is an integer. Show that $W(\gamma, 0) = n$.

In earlier centuries, before modern rigor required functions to be single-valued, the 1-form ω_ϑ would have been written as the differential $d\vartheta$ of the multivalued function ϑ. (In fact, this is still a common and useful notation for this 1-form, as long as one realizes that ϑ is not a function, so that Proposition 1.4 is not contradicted!) The graph of this multivalued function can be visualized in 3-space, as the locus of points (x, y, z) of the form $(r\cos(\vartheta), r\sin(\vartheta), \vartheta)$ for some $r > 0$ and real number ϑ. This is closely related to the *polar coordinate* mapping

$$p\colon \{(r, \vartheta)\colon r > 0\} \to \mathbb{R}^2 \setminus \{0\}, \qquad (r, \vartheta) \mapsto (r\cos(\vartheta), r\sin(\vartheta)).$$

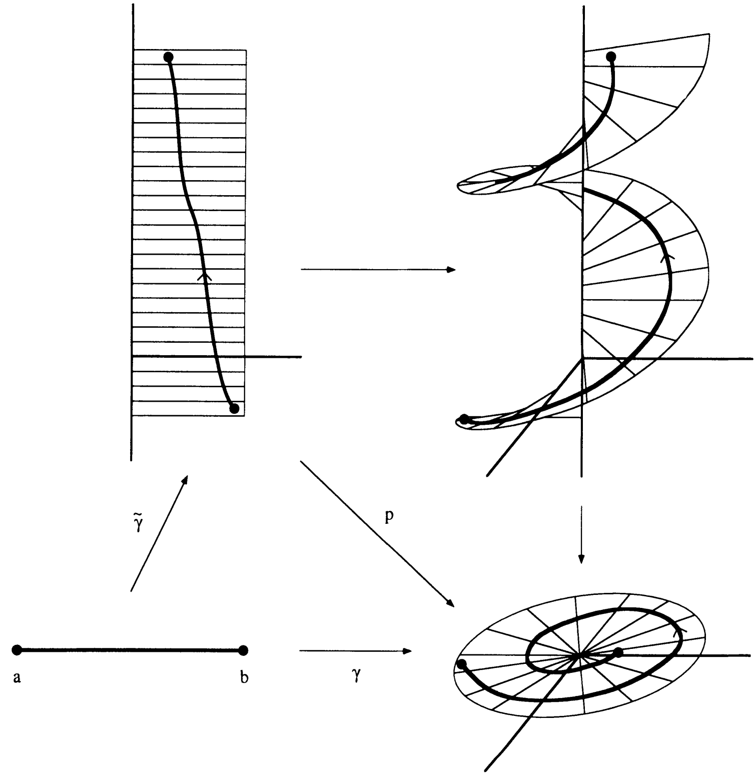
Picturing these together:



This illustrates the fact that, although the distance from the origin is a continuous function on $\mathbb{R}^2 \setminus \{0\}$, the counterclockwise angle from the x-axis cannot be defined continuously. What we have proved, however, is that one can define such an angle continuously along a curve. Proposition 2.2 (with Exercise 2.3) can be expressed geometrically as follows:

Corollary 2.7. *Let* $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{0\}$ *be a smooth path with starting point* $P_a = (r_a \cos(\vartheta_a), r_a \sin(\vartheta_a))$. *Then there is a unique smooth path* $\tilde{\gamma}\colon [a,b] \to R$, *where* $R = \{(r, \vartheta)\colon r > 0\}$ *is the right half plane, with starting point* (r_a, ϑ_a), *and with* $p \circ \tilde{\gamma} = \gamma$. □

We say that the path $\tilde{\gamma}$ is the *lifting* of the path γ with starting point (r_a, ϑ_a). On the picture,



Problem 2.8. Show that the locus of points $(r\cos(\vartheta), r\sin(\vartheta), \vartheta)$ is one connected component of the surface in the complement of the z-axis in $\mathbb{R}^3$ defined by the equation $y \cdot \cos(z) = x \cdot \sin(z)$.

Exercise 2.9. Show that every point in $\mathbb{R}^2 \setminus \{0\}$ has a neighborhood V such that $p^{-1}(V)$ is a disjoint union of open sets V_i, each of which is mapped homeomorphically (in fact, diffeomorphically) by p onto V.

This exercise verifies that p is a "covering map," a class of maps we will study in Chapter 11. Corollary 2.7 is a special case of a general theorem about covering maps.

Winding numbers can be described around any point $P = (x_0, y_0)$ in place of the origin. One can do this either by translating everything, or directly by using the form

$$\omega_{P,\vartheta} = \frac{-(y - y_0)\,dx + (x - x_0)dy}{(x - x_0)^2 + (y - y_0)^2}.$$

Exercise 2.10. For any closed segmented path $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$, define the winding number $W(\gamma, P)$ of γ around P by the formula

$$W(\gamma, P) = \frac{1}{2\pi}\int_\gamma \omega_{P,\vartheta}.$$

Generalize the assertions of this section to these winding numbers.

2b. Reparametrizing and Deforming Paths

Path integrals do not depend on the parametrization of the path, in the following sense. Suppose ω is a 1-form on an open set U in the plane, and $\gamma\colon [a,b] \to U$ is a smooth path, and suppose $\varphi\colon [a',b'] \to [a,b]$ is a $\mathscr{C}^\infty$ function (as usual, extending to a neighborhood of $[a',b']$) that maps a' to a and b' to b. The path $\gamma \circ \varphi$ is called a *reparametrization* of γ.

Lemma 2.11. *With these assumptions,*

$$\int_{\gamma\circ\varphi} \omega = \int_\gamma \omega.$$

Proof. Write $\omega = p\,dx + q\,dy$, and $\gamma(t) = (x(t), y(t))$, and calculate, using the chain rule and change of variables formulas:

$$\begin{aligned}\int_{\gamma\circ\varphi} \omega &= \int_{a'}^{b'} [p(x(\varphi(s)), y(\varphi(s)))\cdot (x\circ\varphi)'(s) \\ &\qquad\qquad + q(x(\varphi(s)), y(\varphi(s)))\cdot(y\circ\varphi)'(s)]\,ds \\ &= \int_{a'}^{b'} [p(x(\varphi(s)), y(\varphi(s)))\cdot x'(\varphi(s)) \\ &\qquad\qquad + q(x(\varphi(s)), y(\varphi(s)))\cdot y'(\varphi(s))]\cdot(\varphi)'(s)\,ds \\ &= \int_a^b [p(x(t), y(t))\cdot x'(t) + q(x(t), y(t))\cdot y'(t)]dt = \int_\gamma \omega. \quad \square\end{aligned}$$

Exercise 2.12. If $\gamma\colon [a,b] \to U$ is a path in U, let $\gamma^{-1}\colon [a,b] \to U$ be the same path traveled backward: $\gamma^{-1}(t) = \gamma(a+b-t)$. Show that

$$\int_{\gamma^{-1}} \omega = -\int_\gamma \omega.$$

Show more generally that if $\varphi\colon [a',b'] \to [a,b]$ is any $\mathscr{C}^\infty$ function that maps a' to b and b' to a, then $\int_{\gamma\circ\varphi}\omega = -\int_\gamma \omega$.

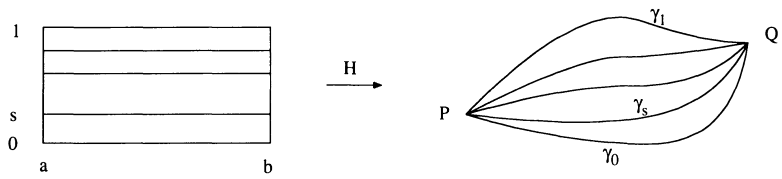
The following problem shows how one could avoid dealing with segmented paths: by turning around the corners *very* slowly!

Problem 2.13. (a) Construct a $\mathscr{C}^\infty$ mapping φ from $[0, 1]$ to $[a, b]$ taking 0 to a and 1 to b, such that all derivatives of φ vanish at 0 and at 1. (b) Use this to show that any smooth path can be reparametrized to a path with the same endpoints but such that all derivatives vanish at the endpoints. (c) If γ_1: $[a, b] \to U$ and γ_2: $[b, c] \to U$ are $\mathscr{C}^\infty$ paths with $\gamma_1(b) = \gamma_2(b)$, and all derivatives of γ_1 and γ_2 vanish at b, verify that the path γ: $[a, c] \to U$ that agrees with γ_1 on $[a, b]$ and γ_2 on $[b, c]$ is a $\mathscr{C}^\infty$ path. (d) If $\gamma = \gamma_1 + \ldots + \gamma_n$ is a segmented path, show that there is a $\mathscr{C}^\infty$ path γ^*: $[0, n] \to U$, so that the restriction of γ^* to $[k-1, k]$ is a reparametrization of γ_k for each k.

Problem 2.14. Given a 1-form ω on an open set U, show that the following are equivalent: (i) there is a function f on U with $df = \omega$; (ii) $\int_\gamma \omega = \int_\delta \omega$ whenever γ and δ are $\mathscr{C}^\infty$ paths in U with the same initial and final points; and (iii) $\int_\gamma \omega = 0$ whenever γ is a $\mathscr{C}^\infty$ closed path in U (i.e., with the initial point equal to the final point).

Problem 2.15. Given a 1-form ω on an open set U, show that the following are equivalent: (i) $d\omega = 0$; (ii) $\int_{\partial R} \omega = 0$ for all closed rectangles R contained in U; and (iii) every point in U has a neighborhood such that $\int_{\partial R} \omega = 0$ for all closed rectangles R contained in the neighborhood. Is the same true if rectangles are replaced by disks?

Next we turn to the question of what happens when the path is moved, or deformed. We consider first the case of a deformation of paths with fixed endpoints. This will be given by a family of paths γ_s, for simplicity all defined on the same interval $[a, b]$, with the parameter s varying in another interval which we take to be the unit interval $[0, 1]$. We will assume this is a smooth family, in the sense that the coordinates of the point $\gamma_s(t)$ are smooth functions of both s and t. This means that we are given a mapping H from $[a, b] \times [0, 1]$ to U, which we assume is $\mathscr{C}^\infty$ (this, as usual, means that the two coordinate functions can be extended to be infinitely differentiable functions on some open neighborhood of the rectangle). We assume that $H(a, s) = P$ and $H(b, s) = Q$ for all $0 \le s \le 1$. Set $\gamma_s(t) = H(t, s)$, so each γ_s is a path in U from P to Q.



We call H a *smooth homotopy* from the path γ_0 to the path γ_1. We say that two paths from an interval $[a,b]$ to U, with the same initial and final points, are *smoothly homotopic* in U if there is such a homotopy from one to the other.

Proposition 2.16. *If γ and δ are smoothly homotopic paths from P to Q in an open set U, and ω is a closed 1-form in U, then*

$$\int_\gamma \omega = \int_\delta \omega .$$

Proof. First we sketch a proof using only ideas from calculus. Let V be a neighborhood of the rectangle $R = [a,b] \times [0,1]$ mapped into U by an extension of H, and let $x(t,s)$ and $y(t,s)$ be the coordinate functions of this mapping from V to U. Define a "pull-back" form ω^* on V by the formula

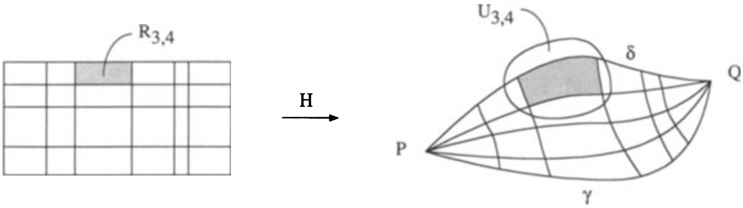
$$\begin{aligned}\omega^* = &\left(p(x(t,s),y(t,s))\frac{\partial x}{\partial t} + q(x(t,s),y(t,s))\frac{\partial y}{\partial t}\right) dt \\ &+ \left(p(x(t,s),y(t,s))\frac{\partial x}{\partial s} + q(x(t,s),y(t,s))\frac{\partial y}{\partial s}\right) ds .\end{aligned}$$

A little calculation, left to you, shows that $d\omega^* = 0$ on V. By Green's theorem, we therefore know that the integral of ω^* around the boundary of R must be zero. Simpler calculations show that the integral of ω^* along the bottom of the rectangle is $\int_\gamma \omega$, that along the top is $\int_\delta \omega$, and that the integrals along the two sides are zero. Putting this all together gives the required equality.

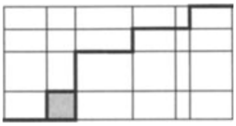
Here is another proof, more topological in flavor. Each point in the image $H(R)$ of the rectangle has a neighborhood on which ω is exact. Applying the Lebesgue lemma, it follows that, if we subdivide the rectangle small enough, by choosing

$$a = t_0 < t_1 < \ldots < t_n = b \quad \text{and} \quad 0 = s_0 < s_1 < \ldots < s_m = 1 ,$$

then each subrectangle $R_{i,j} = [t_{i-1}, t_i] \times [s_{j-1}, s_j]$ is mapped by H into an open set $U_{i,j}$ on which ω is the differential of some function $f_{i,j}$.



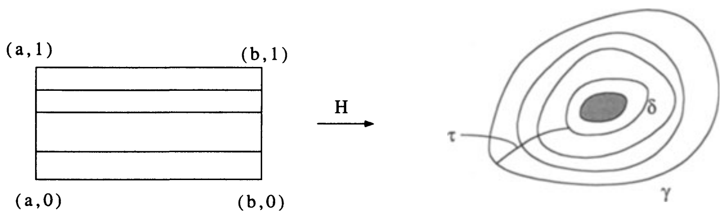
Since ω is the differential of a function on $U_{i,j}$, the integral of $\omega \circ H$ around the boundary of $R_{i,j}$ is zero, i.e., the integral along the bottom and right side of $R_{i,j}$ is the same as the integral along the left side and top. Now $\int_\gamma \omega$ is the integral of ω along the bottom and right side of the original rectangle, and one can successively replace integrals over the bottom and right sides by integrals over the left and top sides, of each of the small rectangles, until one has the integral over the left and top sides of the whole rectangle, which is $\int_\delta \omega$.



More succinctly, integrate ω over all the boundaries of the small rectangles. Each inside edge is integrated over twice, once in each direction, which leaves the integrals over the outside, giving

$$0 = \sum_{i,j} \int_{\partial R_{i,j}} \omega = \int_{\partial R} \omega = \int_\gamma \omega - \int_\delta \omega . \qquad \square$$

There is another kind of deformation or homotopy that is also important. This is a deformation of closed paths, through a family of closed paths, but allowing the endpoints to vary. It is given by a $\mathscr{C}^\infty$ mapping H from a rectangle $[a, b] \times [0, 1]$ into U, with the property that $H(a, s) = H(b, s)$ for every s in $[0, 1]$. Each of the paths γ_s given by $\gamma_s(t) = H(t, s)$ is a closed path, starting and ending at the point $\tau(s) = H(a, s) = H(b, s)$. These endpoints are allowed to vary along the path τ.



We call H a *smooth homotopy* from the path $\gamma = \gamma_0$ to the path $\delta = \gamma_1$, and we say that two closed paths γ and δ are *smoothly homotopic* if there is such a smooth homotopy between them.

Proposition 2.17. *If γ and δ are smoothly homotopic closed paths in an open set U, and ω is a closed 1-form in U, then*

$$\int_\gamma \omega = \int_\delta \omega .$$

Proof. Either of the proofs of the last proposition works for this one. The only point to notice is that the integrals of ω over the two sides of the rectangle may not be zero, but since the two paths given by H on these two sides are the same (namely τ), these integrals $\int_\tau \omega$ cancel. The details are left as an exercise. □

When U is the complement of a point P, and $\omega = (1/2\pi)\omega_{P,\vartheta}$, this specializes to the important:

Corollary 2.18. *If γ and δ are smoothly homotopic closed paths in $\mathbb{R}^2 \setminus \{P\}$, then $W(\gamma, P) = W(\delta, P)$.*

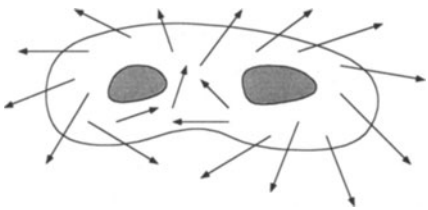
Of course, it is crucial for this that the homotopy stays in the complement of the point P!

Problem 2.19. Prove that being smoothly homotopic is an equivalence relation.

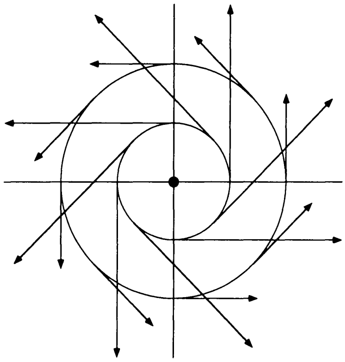
2c. Vector Fields and Fluid Flow

We defined a 1-form on an open set U in the plane to be a pair of smooth functions p and q on U. Such a pair of functions can also be identified with a *vector field* on U, which assigns to each point (x, y) in U the vector

$$V(x, y) = p(x, y)\mathbf{i} + q(x, y)\mathbf{j} .$$



The vector field corresponding to the 1-form ω_ϑ is perpendicular to the position vector, pointing in a counterclockwise direction, with length that is the inverse of the distance to the origin:



The path integral

$$\int_\gamma p\,dx + q\,dy = \int_a^b \left(p(x(t), y(t)) \frac{dx}{dt} + q(x(t), y(t)) \frac{dy}{dt} \right) dt$$

can be written, using the dot product, as

$$\int_a^b V(\gamma(t)) \cdot \gamma'(t)\,dt = \int_a^b \left(V(\gamma(t)) \cdot \frac{\gamma'(t)}{\|\gamma'(t)\|} \right) \|\gamma'(t)\|\,dt,$$

which is the integral of the projection of the vector field along the tangent to the curve. (For this last formula, we must assume the tangent vector is not zero.) If the vector field represents a force, it is the work done by the force along the curve.

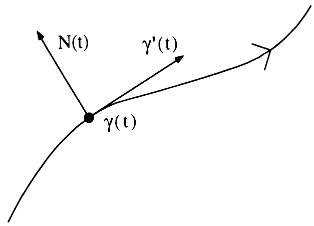
The translation of the equation $\omega = df$, or $p = \partial f/\partial x$ and $q = \partial f/\partial y$, into vector field language says that the corresponding vector field is the *gradient* of f:

$$\text{grad}(f) = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}.$$

The function f (or sometimes $-f$) is called a *potential* function.

We'll finish this chapter with a quick sketch of the interpretation of these ideas for case where the vector field gives the velocity of a fluid flowing in an open set in the plane. So $V(x, y)$ is the velocity vector at (x, y). We assume the flow is in a steady state, which means that, as written, this velocity vector depends only on the point (and not on time). Let p and q be the components of V, as above.

The integral $\int_\gamma p\,dx + q\,dy$ represents the *circulation of the fluid along the path* γ, per unit of time. This can be seen from the interpretation as the integral of the projection of the velocity vector in the tangent direction along the curve. There is another important integral, $\int_\gamma q\,dx - p\,dy$, that represents the *flux of the fluid across the path* γ (per unit of time). To see this, let $N(t) = (-dy/dt, dx/dt)$ be the normal vector, which is perpendicular to and 90° counterclockwise from the tangent vector.



Then

$$\begin{aligned}\int_\gamma q\,dx - p\,dy &= \int_a^b \left(q(x(t), y(t))\frac{dx}{dt} - p(x(t), y(t))\frac{dy}{dt}\right) dt \\ &= \int_a^b V(\gamma(t)) \cdot N(t)\,dt = \int_a^b \left(V(\gamma(t)) \cdot \frac{N(t)}{\|N(t)\|}\right) \|\gamma'(t)\|\,dt.\end{aligned}$$

This is the integral of the projection of the velocity vector in the normal direction, which measures the flow across γ, from right to left in the direction along γ, per unit of time.

The flow is called *irrotational* if the circulation around all small closed loops is zero. By Problem 2.15, this is true precisely when $d\omega = 0$, i.e., $\partial q/\partial x - \partial p/\partial y = 0$. The function $\partial q/\partial x - \partial p/\partial y$ is called the *curl* of the vector field.

The flow is called *incompressible* if the net flow across small closed loops is zero. Applying Problem 2.15 to the 1-form $q\,dx - p\,dy$, this is equivalent to the condition that $\partial p/\partial x + \partial q/\partial y = 0$. The function $\partial p/\partial x + \partial q/\partial y$ is called the *divergence* of the vector field.

Exercise 2.20. If R is a rectangle (or disk) in U, show that the integral of the curl (identified with the 2-form $(\partial q/\partial x - \partial p/\partial y)\,dx\,dy$) over R is the circulation around ∂R, and the integral of the divergence is the flux across ∂R.

If the open set U is a simple one, such as an open rectangle or disk, and if the flow is irrotational, then we know that there is a potential function f, i.e., $V = \mathrm{grad}(f)$. If the fluid is also incompressible, the equation we just found says that

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0,$$

which is the condition for f to be a *harmonic* function.

Although we have no topological excuses for them, here are a handful of typical applications of and variations on these notions.

Exercise 2.21. (a) Show that the following functions are harmonic: (i) $a + bx + cy$ for any a, b, c; (ii) $x^2 - y^2$; (iii) $\log(r)$, $r = \sqrt{x^2 + y^2}$ on the complement of the origin. (b) Find all harmonic polynomial functions of x and y of degree at most three. (c) Find all harmonic functions that have the form $h(r)$.

Problem 2.22. (a) State and prove an analogue of Green's theorem when R is the region between two rectangles, one contained in the other. (b) State and prove an analogue of Green's theorem when R is a disk, or the region between two concentric disks.

Problem 2.23. If R is a region, with boundary ∂R, for which Green's theorem is known, prove the following two formulas of Green, for functions f and g on a region containing the closure of R:

(i) $$\int_{\partial R} f \cdot \left(-\frac{\partial g}{\partial y}dx + \frac{\partial g}{\partial x}dy\right)$$
$$= \iint_R \left(f \cdot \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}\right) + \mathrm{grad}(f) \cdot \mathrm{grad}(g)\right) dx\, dy;$$

(ii) $$\int_{\partial R} f \cdot \left(-\frac{\partial g}{\partial y}dx + \frac{\partial g}{\partial x}dy\right) - g \cdot \left(-\frac{\partial f}{\partial y}dx + \frac{\partial f}{\partial x}dy\right)$$
$$= \iint_R \left(f \cdot \left(\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}\right) - g \cdot \left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right)\right) dx\, dy.$$

Problem 2.24. (a) If f is harmonic in R and vanishes on ∂R, show that f must be identically zero. (b) If two harmonic functions on R have the same restriction to ∂R, show that they must agree everywhere on R.

Problem 2.25. If f is harmonic on a disk, show that the value of f at the center of the disk is the average value of f on the boundary of the disk.

Exercise 2.26. If $V(x,y) = -(1/2\pi)(x,y)/\|(x,y)\|^2$ and γ is a closed path not passing through the origin, show that the flux across γ is the winding number $W(\gamma, 0)$, i.e., $W(\gamma, 0) = \int_a^b V(\gamma(t)) \cdot N(t)\,dt$.

These physical notions suggest a way to generalize the notion of a winding number to higher dimensions. If f is a $\mathscr{C}^\infty$ map from a rectangle or disk R to $\mathbb{R}^3 \setminus (0)$, let $N_f(P)$ be the cross product of the columns of the Jacobian matrix of f at P, and let

$$V(x,y,z) = (1/4\pi)(x,y,z)/\|(x,y,z)\|^3.$$

It is a good project to develop a notion of an "engulfing number," setting $W(f, 0) = \iint_R V(f(P)) \cdot N_f(P)$.

PART II

WINDING NUMBERS

The notion of winding numbers is generalized to arbitrary continuous paths, and the facts we proved in the smooth case using calculus are proved here by purely topological arguments. We also look at what happens when the point being wound around is varied. Finally, we use the idea of winding numbers to define the degree of a mapping from one circle to another, and to define the local degree of a mapping from one open plane set to another.

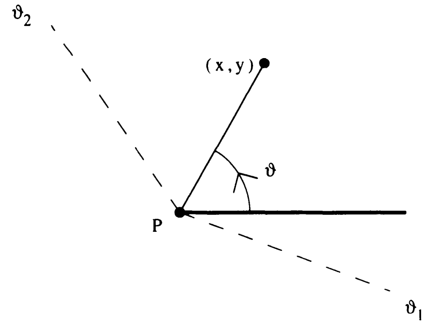
In Chapter 4 there are several applications of winding numbers, some written out, and many left as exercises and problems. Many of these results generalize to higher dimensions; the names attached to some of them refer to these generalizations.

CHAPTER 3

The Winding Number

3a. Definition of the Winding Number

For any point P in the plane, and any sector with vertex at P, we can define a continuous (even $\mathscr{C}^\infty$) angle function on the sector, although the choice is unique only up to adding integral multiples of 2π. Here angles are measured with reference to P, counterclockwise from the horizontal line to the east of P:



If $p = p_P$ is the corresponding polar coordinate map:

$$p(r, \vartheta) \;=\; P + (r\cos(\vartheta), r\sin(\vartheta)),$$

a sector is the image of a strip $\{(r, \vartheta)\colon r > 0 \text{ and } \vartheta_1 < \vartheta < \vartheta_2\}$, where ϑ_1 and ϑ_2 are any real numbers with $0 < \vartheta_2 - \vartheta_1 \leq 2\pi$. The fact that this ϑ is a continuous function on the sector can be seen directly, using formulas like the arc tangent, or topologically as follows: p is one-to-one and continuous from the strip onto the sector, and it maps small open rectangles $\{(r, \vartheta)\colon a<r<b, \alpha<\vartheta<\beta\}$ onto small open sectors (bounded by two straight lines and arcs of two circles), so p is open as well as continuous; it is therefore a homeomorphism (in fact a diffeomorphism). This means that r and ϑ are continuous functions of x and y.

For any continuous path $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$, we define its *winding number* $W(\gamma, P)$ as follows:

Step 1. Subdivide the interval into $a = t_0 \leq t_1 \leq \ldots \leq t_n = b$, so that each subinterval $[t_{i-1}, t_i]$ is mapped into some sector with vertex at P. Such a subdivision exists by the Lebesgue lemma, since each point in the image of γ is contained in some such sector.

Step 2. Choose such a sector U_i containing $\gamma([t_{i-1}, t_i])$ and a corresponding angle function ϑ_i on U_i, for $1 \leq i \leq n$. Let $P_i = \gamma(t_i), 0 \leq i \leq n$. Define

$$
\begin{aligned}
W(\gamma, P) &= \frac{1}{2\pi}[(\vartheta_1(P_1) - \vartheta_1(P_0)) + (\vartheta_2(P_2) - \vartheta_2(P_1)) \\
&\qquad + \ldots + (\vartheta_n(P_n) - \vartheta_n(P_{n-1}))] \\
&= \frac{1}{2\pi}\sum_{i=1}^{n} (\vartheta_i(P_i) - \vartheta_i(P_{i-1})) .
\end{aligned}
$$

Each term represents the net change in angle along that part of the path.

Proposition 3.1. (*a*) *The definition of a winding number is independent of the choices made in Steps* 1 *and* 2. (*b*) *If γ is a closed path, i.e., $\gamma(a) = \gamma(b)$, then $W(\gamma, P)$ is an integer.*

Proof. (a) To see that it is independent of the choices of the sectors U_i and the angle functions ϑ_i, suppose U_i' and ϑ_i' were other choices. Then ϑ_i and ϑ_i' would differ by a constant (in fact, a multiple of 2π) on the component of the intersection of U_i and U_i' that contains $\gamma([t_{i-1}, t_i])$. So the difference in the values of ϑ_i at the two points P_{i-1} and P_i is the same as the difference in values of ϑ_i' at these two points, and adding over i shows that the winding number doesn't change.

So it is enough to show that the definition is independent of the choice of subdivision. Suppose we add one point to a given subdivision that satisfies the condition in Step 1, say by inserting a point t^* between some t_{i-1} and t_i. If U_i and ϑ_i are chosen for $[t_{i-1}, t_i]$, these same U_i and ϑ_i can also be chosen for the two subintervals $[t_{i-1}, t^*]$ and $[t^*, t_i]$. And if $P^* = \gamma(t^*)$, then

$$(\vartheta_i(P_i) - \vartheta_i(P^*)) + (\vartheta_i(P^*) - \vartheta_i(P_{i-1})) \;=\; \vartheta_i(P_i) - \vartheta_i(P_{i-1}),$$

so again the sum is unchanged. It follows that if we insert any finite number of points into a given subdivision, the definition of the winding number will not change. But then, if two subdivisions both satisfy the condition in Step 1, the common refinement of both of them, obtained by including all division points for each, will define the same winding number as each of them.

(b) In general, the claim is that, even when γ is not closed, if ϑ_a is an angle for the initial point $\gamma(a)$, then $\vartheta_a + 2\pi \cdot W(\gamma, P)$ is an angle for the endpoint $\gamma(b)$. When the path is closed, this implies that $W(\gamma, P)$ is an integer. This claim is evident for the restriction of γ to each subinterval $[t_{i-1}, t_i]$ on which there is a continuous angle function, and the general case follows by adding up the results (or inducting on the number of subintervals). □

Exercise 3.2. Show that if γ is smooth, this definition agrees with that in Chapter 2.

Exercise 3.3. Show that if $\gamma\colon [a,b] \to U$ is a closed path, and U is an open set in $\mathbb{R}^2 \setminus \{P\}$ on which there is a continuous angle function (for example, a sector with vertex at P), then $W(\gamma, P) = 0$.

Exercise 3.4. Show that winding numbers are invariant by translation, in the following sense. Let $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$, and let v be any vector in the plane. Let $\gamma + v$ be the path defined by $(\gamma + v)(t) = \gamma(t) + v$. Show that

$$W(\gamma + v, P + v) \;=\; W(\gamma, P).$$

Problem 3.5. Show that for any continuous path $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$, there are continuous functions $r\colon [a,b] \to \mathbb{R}^+$ (the positive real numbers) and $\vartheta\colon [a,b] \to \mathbb{R}$, so that

$$\gamma(t) \;=\; P + (r(t)\cos(\vartheta(t)), r(t)\sin(\vartheta(t))), \qquad a \le t \le b.$$

Show that r is uniquely determined, and ϑ is uniquely determined up to adding a constant integral multiple of 2π. Show in fact that $r(t) = \|\gamma(t) - P\|$, and if γ' denotes the restriction of γ to the interval

$[a,t]$ (so $\gamma^t(u)=\gamma(u)$ for $a \leq u \leq t$), and ϑ_a is an angle for $\gamma(a)$, then one may take

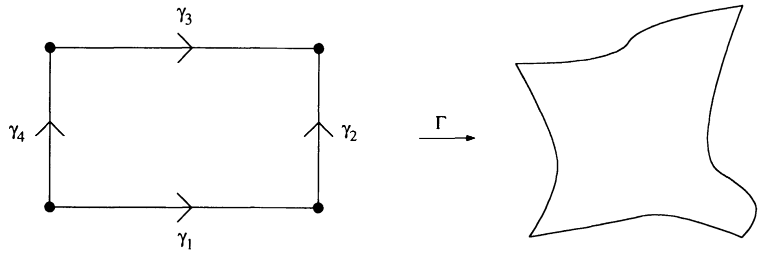
$$\vartheta(t) = \vartheta_a + 2\pi \cdot W(\gamma^t, P).$$

Equivalently, for any choice of ϑ_a, there is a unique continuous mapping $\tilde{\gamma}$: $[a,b] \to \{(r,\vartheta): r>0\}$ such that $p_P \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}(a) = (r(a), \vartheta_a)$, where p_P is the polar coordinate mapping. Such a $\tilde{\gamma}$ is called a *lifting* of γ with starting point $(r(a), \vartheta_a)$.

3b. Homotopy and Reparametrization

Suppose $R=[a,b] \times [c,d]$ is a closed rectangle, and $\Gamma: R \to \mathbb{R}^2$ is a continuous mapping. The restrictions of Γ to the four sides of the rectangle define four paths γ_1, γ_2, γ_3, and γ_4:

$$\begin{array}{lll} \gamma_1(t) = \Gamma(t,c), & \gamma_3(t) = \Gamma(t,d), & a \leq t \leq b; \\ \gamma_2(s) = \Gamma(b,s), & \gamma_4(s) = \Gamma(a,s), & c \leq s \leq d. \end{array}$$



Theorem 3.6. *For any point P not in $\Gamma(R)$,*

$$W(\gamma_1, P) + W(\gamma_2, P) = W(\gamma_3, P) + W(\gamma_4, P).$$

Proof. In fact, the second proofs we gave for Proposition 2.16 and 2.17 work equally well in this case. As there, use the Lebesgue lemma to subdivide the rectangle into subrectangles $R_{i,j}$, such that Γ maps $R_{i,j}$ into a sector $U_{i,j}$ (with vertex at P), on which there is a continuous angle function $\vartheta_{i,j}$. Then, by the canceling of inside edges as before,

$$W(\Gamma|_{\partial R}, P) = \sum_{i,j} W(\Gamma|_{\partial R_{i,j}}, P),$$

where $W(\Gamma|_{\partial R}, P)$ is defined by the equation

$$W(\Gamma|_{\partial R}, P) = W(\gamma_1, P) + W(\gamma_2, P) - W(\gamma_3, P) - W(\gamma_4, P),$$

and $W(\Gamma|_{\partial R_{i,j}}, P)$ is defined similarly as the signed sum of the winding numbers around the small rectangles. But these winding numbers around the small rectangles are all zero, since each $R_{i,j}$ is mapped into a region where there is an angle function $\vartheta_{i,j}$. So

$$W(\gamma_1, P) + W(\gamma_2, P) - W(\gamma_3, P) - W(\gamma_4, P) \; = \; 0 . \qquad \square$$

This theorem implies that the winding numbers have the same invariance under homotopies as in the smooth case considered in §2b. Again, there are two kinds of homotopies we want to consider, first for paths with fixed endpoints, and second for closed paths. If γ: $[a,b] \to U$ and δ: $[a,b] \to U$ are paths with the same initial and final points, a *homotopy from* γ *to* δ *with fixed endpoints* is a continuous mapping H: $[a,b] \times [0,1] \to U$ such that

$$\begin{array}{llll} H(t,0) = \gamma(t) & \text{and} & H(t,1) = \delta(t) & \text{for all } \; a \leq t \leq b; \\ H(a,s) = \gamma(a) = \delta(a) & \text{and} & H(b,s) = \gamma(b) = \delta(b) & \text{for all } \; 0 \leq s \leq 1 . \end{array}$$

The paths γ_s defined by $\gamma_s(t) = H(t,s)$ give a continuous family of smooth paths from $\gamma_0 = \gamma$ to $\gamma_1 = \delta$. The paths γ and δ are called *homotopic with fixed endpoints* if there is such a homotopy H.

On the other hand, if γ and δ are closed paths in U, again defined on the same interval $[a,b]$, a *homotopy from* γ *to* δ *through closed paths* is a continuous H: $[a,b] \times [0,1] \to U$, such that

$$\begin{array}{llll} H(t,0) = \gamma(t) & \text{and} & H(t,1) = \delta(t) & \text{for all } \; a \leq t \leq b; \\ H(a,s) = H(b,s) & & & \text{for all } \; 0 \leq s \leq 1 . \end{array}$$

The paths γ and δ are called *homotopic closed paths* if there is such a homotopy H.

Exercise 3.7. Prove that the relation of being homotopic with fixed endpoints, or as closed paths, is an equivalence relation.

Corollary 3.8. *If two paths γ and δ in $\mathbb{R}^2 \setminus \{P\}$ are homotopic, either as paths with the same endpoints, or as closed paths, then*

$$W(\gamma, P) \; = \; W(\delta, P) .$$

Proof. This is an immediate consequence of the theorem, applied to the homotopy $H = \Gamma$. In the first case, the winding numbers of the constant paths from the sides of the rectangle are both zero, and in the second case they are the same, so their winding numbers cancel in the result. $\square$

We next consider what happens to the winding number by a change of parameter, generalizing what we saw in the last chapter in the smooth case.

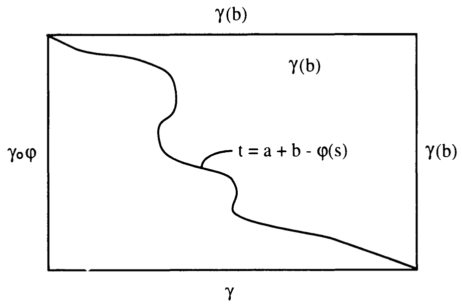
Corollary 3.9. *Let γ: $[a,b] \to \mathbb{R}^2 \setminus \{P\}$ be a continuous path, and φ: $[c,d] \to [a,b]$ a continuous function.*

(a) If $\varphi(c) = a$ and $\varphi(d) = b$, then $W(\gamma \circ \varphi, P) = W(\gamma, P)$.
(b) If $\varphi(c) = b$ and $\varphi(d) = a$, then $W(\gamma \circ \varphi, P) = -W(\gamma, P)$. In particular, if $\gamma^{-1}(t) = \gamma(a + b - t)$, $a \leq t \leq b$, then

$$W(\gamma^{-1}, P) \;=\; -W(\gamma, P).$$

Proof. For (a), define Γ: $[a,b] \times [\mathrm{c}, \mathrm{d}] \to U$ by the formula

$$\Gamma(t,s) = \gamma(\min(t + \varphi(s) - a, b)).$$



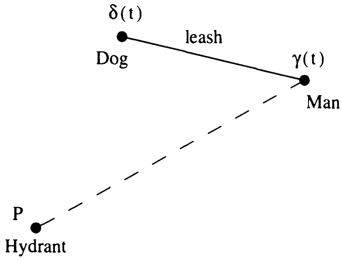
Then Γ is continuous, since the minimum or composite of two continuous functions is continuous. The paths from the sides of Γ are $\gamma_1 = \gamma$, $\gamma_4 = \gamma \circ \varphi$, and γ_2 and γ_3 are constant paths at the point $\gamma(\mathrm{b})$. So Theorem 3.6 applies and gives (a). Similarly for (b), use

$$\Gamma(s,t) \;=\; \gamma(\max(t + \varphi(s) - b, a)).$$

In this case $\gamma_1 = \gamma$, $\gamma_2 = \gamma \circ \varphi$, and γ_3 and γ_4 are constant. □

Exercise 3.10. Give a direct proof of Corollary 3.9 from the definition of the winding number, in case the change of coordinates φ is a monotone increasing or monotone decreasing function. (Monotone increasing means that $\varphi(t) < \varphi(s)$ if $t < s$.)

As another application, we have the "dog-on-a-leash" theorem of Poincaré and Bohl. This says if the leash is kept shorter than the distance from the walker to the fire hydrant, then the walker and the dog both wind around the hydrant the same number of times:



Theorem 3.11 (Dog-on-a-Leash). *Suppose* $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$ *and* $\delta\colon [a,b] \to \mathbb{R}^2 \setminus \{P\}$ *are closed paths, and the line segment between* $\gamma(t)$ and $\delta(t)$ *never hits the point P. Then*

$$W(\gamma, P) = W(\delta, P).$$

Proof. Define $H\colon [a,b] \times [0,1] \to \mathbb{R}^2$ by the formula

$$H(t,s) = (1-s)\gamma(t) + s\,\delta(t), \qquad a \le t \le b, \quad 0 \le s \le 1.$$

This is a continuous homotopy from γ to δ through closed paths. The hypotheses imply that Γ maps the rectangle into $\mathbb{R}^2 \setminus \{P\}$. The result therefore follows from Corollary 3.8. □

Corollary 3.12. *If* $\gamma\colon [a,b] \to \mathbb{R}^2$ *and* $\delta\colon [a,b] \to \mathbb{R}^2$ *are closed paths such that* $\|\gamma(t) - \delta(t)\| < \|\gamma(t) - P\|$ *for all* t *in* $[a,b]$, *then* $W(\gamma, P) = W(\delta, P)$.

Proof. The hypothesis implies that neither path hits P, and that the line segment between $\gamma(t)$ and $\delta(t)$ doesn't hit P. □

Exercise 3.13. Show that if U is an open rectangle (bounded or unbounded), then any two paths from $[a,b]$ to U with the same endpoints are homotopic, and any two closed paths in U are homotopic. Show that the same is true for any convex open set U. Can you prove it when U is just starshaped?

Problem 3.14. Let γ and δ be paths from an interval to $\mathbb{R}^2 \setminus \{P\}$ with the same endpoints. Show that the following are equivalent:

(i) γ and δ are homotopic in $\mathbb{R}^2 \setminus \{P\}$;
(ii) $W(\gamma, P) = W(\delta, P)$; and
(iii) if $\tilde{\gamma}$ and $\tilde{\delta}$ are liftings of γ and δ with the same initial point, as in Problem 3.5, then $\tilde{\gamma}$ and $\tilde{\delta}$ have the same final point.

Problem 3.15. Let γ and δ be closed paths from a closed interval to

$\mathbb{R}^2 \setminus \{P\}$. Show that γ and δ are homotopic through closed paths in $\mathbb{R}^2 \setminus \{P\}$ if and only if $W(\gamma, P) = W(\delta, P)$.

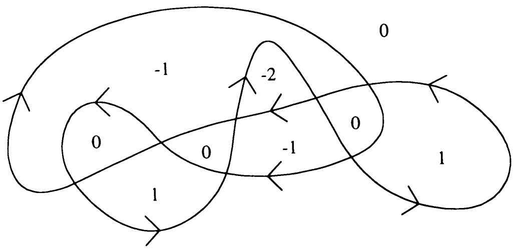
3c. Varying the Point

We want to study what happens to the winding number if a closed path γ: $[a,b] \to \mathbb{R}^2$ is fixed, and the point P is allowed to vary, but always so that the path does not pass through P. Let us denote the image of the path γ by Supp(γ), and call it the *support* of γ, i.e.,

$$\mathrm{Supp}(\gamma) \;=\; \gamma([a,b]).$$

Since the interval is compact, the support is a compact, and hence closed and bounded, subset of the plane. The complement of the support is an open set, which may have many—even infinitely many—connected components, each of which is open. Since the support is bounded, however, there is one connected component of $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$ that contains all points outside some large disk; this component is called the *unbounded* component.

Proposition 3.16. *As a function of P, the function $W(\gamma, P)$ is constant on each connected component of $\mathbb{R}^2 \setminus Supp(\gamma)$. It vanishes on the unbounded component.*



Proof. For the first statement, we must show that $W(\gamma, P)$ is a locally constant function of P in $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$. Given P, choose a disk around P contained in $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$. We must show that $W(\gamma, P') = W(\gamma, P)$ for all P' in the disk. Let $v = P' - P$ be the vector from P to P'. By Exercise 3.4,

$$W(\gamma, P) \;=\; W(\gamma + v, P + v) \;=\; W(\gamma + v, P').$$

The homotopy $H(t,s) = \gamma(t) + sv$, $a \le t \le b$, $0 \le s, \le 1$, is a homotopy through closed paths from γ to $\gamma + v$ that never hits the point P', so

$$W(\gamma + v, P') = W(\gamma, P').$$

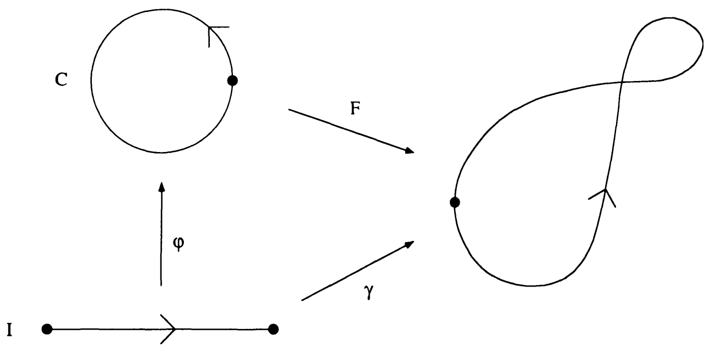
To show that the winding number vanishes on the unbounded component, it suffices to show it vanishes on one such point. For example, we can take P far out on the negative x-axis, so that the support of γ is contained in a half plane to the right of P. But then there is an angle function on this half plane, so the winding number is zero by Exercise 3.3. □

The following exercise gives alternative proofs of the proposition:

Exercise 3.17. (a) Prove directly from the definition that for any path γ, whether closed or not, the function $P \mapsto W(\gamma, P)$ is continuous on the complement of $\mathrm{Supp}(\gamma)$, and approaches zero when $\|P\|$ goes to infinity. (b) Use (a) and the fact that, when γ is closed, $W(\gamma, P)$ takes values in the integers to give another proof of the proposition. (c) For a closed path γ, show directly that $W(\gamma, P)$ is constant on path components of $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$ by applying Theorem 3.6 to the mapping $\Gamma(t, s) = \gamma(t) - \sigma(s)$ and the point $P = 0$, where σ is any path in $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$.

3d. Degrees and Local Degrees

If I is any closed interval, and C is any circle, a continuous closed path γ from I to an open set U is essentially the same thing as a continuous mapping from C into U:



This can be realized explicitly as follows. Let us assume that $I = [0, 1]$, since this is the most common convention. Suppose C has center (x_0, y_0) and radius r. Let $\varphi\colon I \to C$ be the function that wraps I around C:

$$\varphi(t) = (x_0, y_0) + (r\cos(2\pi t), r\sin(2\pi t)), \qquad 0 \le t \le 1.$$

So φ is a one-to-one mapping of I onto C, except that $\varphi(0) = \varphi(1)$. It follows that if F is a mapping from C into an open set U, then $\gamma = F \circ \varphi$ is a mapping from I to U with $\gamma(0) = \gamma(1)$, and that any such γ can be realized in this way for a unique F.

Lemma 3.18. *The mapping γ is continuous if and only if F is continuous.*

Proof. In fact, φ realizes C as the quotient space of I with its endpoints 0 and 1 identified. Concretely, this means that a subset X of C is open in C if and only if $\varphi^{-1}(X)$ is open in I. This is easy to verify directly, or one can argue that the quotient space is a compact Hausdorff space, and the induced mapping from it to C, being continuous and one-to-one, must be a homeomorphism. If follows that for an open subset V of U, $F^{-1}(V)$ is open if and only if $\gamma^{-1}(V) = \varphi^{-1}(F^{-1}(V))$ is open, which proves the lemma. □

For any continuous $F\colon C \to \mathbb{R}^2 \setminus \{P\}$, we can define the *winding number of F around P*, denoted $W(F, P)$, to be the winding number $W(\gamma, P)$ of the path $\gamma = F \circ \varphi$.

Exercise 3.19. Identify $\mathbb{R}^2$ with the complex numbers $\mathbb{C}$, and let $f\colon \mathbb{C} \to \mathbb{C}$ be the mapping that takes a complex number z to its nth power z^n, where n is an integer. Let C be any circle centered at the origin, and let F be the restriction of f to C. Show that $W(F, 0) = n$. If $f(z) = -z$, show that $W(F, 0) = 1$.

Proposition 3.20. *Suppose C is the boundary of the closed disk D, and $F\colon C \to \mathbb{R}^2 \setminus \{P\}$ extends to a continuous function from D to $\mathbb{R}^2 \setminus \{P\}$. Then $W(F, P) = 0$.*

Proof. If D is the disk of radius r about the point (x_0, y_0) as above, and $\gamma\colon [0,1] \to \mathbb{R}^2 \setminus \{P\}$ is the path corresponding to F, and $\tilde{F}\colon D \to \mathbb{R}^2 \setminus \{P\}$ is such an extension of F, then

$$H(t, s) = \tilde{F}((x_0, y_0) + s(r\cos(2\pi t), r\sin(2\pi t))), \qquad 0 \le t \le 1, \quad 0 \le s \le 1,$$

gives a homotopy from γ to the constant path at the point $\tilde{F}((x_0, y_0))$. This homotopy stays inside $\mathbb{R}^2 \setminus \{P\}$, and since the winding number of a constant path is zero, the claim follows from Corollary 3.8. □

Problem 3.21. If F_0 and F_1 are mappings from a circle C to U, corresponding to paths γ_0 and γ_1 from $[0, 1]$ to U, show that γ_0 and γ_1 are homotopic through closed paths if and only if F_0 and F_1 are *homotopic* mappings, i.e., there is a continuous mapping

$$H: C \times [0,1] \to U$$

with $H(P \times 0) = F_0(P)$ and $H(P \times 1) = F_1(P)$ for all P in C.

Problem 3.22. Show that the converse of Proposition 3.20 is true: if $W(\gamma, P) = 0$, then γ has a continuous extension to a map from D to $\mathbb{R}^2 \setminus \{P\}$.

Problem 3.23. Let C be a circle centered at the origin, and let $F: C \to \mathbb{R}^2$ be a continuous mapping such that the vector $F(P)$ is never tangent to the curve C at P, i.e., the dot product $F(P) \cdot P$ is not zero for all P in C. Show that $W(F, 0) = 1$.

The same idea lets us define the *degree* of any continuous mapping F from one circle C to another circle C', which measures how many times the first circle is wound around the second by F. Let P' be the center of the circle C', and define the degree of F by the formula

$$\deg(F) = W(F, P').$$

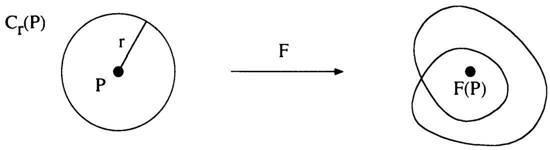
Exercise 3.24. Show that one could take any point inside C' in place of P'.

Exercise 3.25. (a) Show that if F is not surjective, then its degree is zero. (b) Give an example of a surjective mapping from C to C' that has degree zero. (c) Show that if F and G are homotopic mappings from C to C', i.e., if there is a continuous mapping H from $C \times [0, 1]$ to C', with $F(Q) = H(Q \times 0)$ and $G(Q) = H(Q \times 1)$ for all Q in C, then F and G have the same degree. (d) Show that if C is the boundary of the disk D, then F extends to a continuous mapping from D to C' if and only if F is homotopic to a constant mapping from C to C', and then the degree of F is zero.

Problem 3.26. (a) Prove that two mappings from C to C' have the same degree if and only if they are homotopic. (b) Deduce that, if C is the boundary of the disk D, and the degree of a mapping is zero, then the mapping extends to a continuous mapping from D to C'. (c) Deduce also that if S^1 is the unit circle centered at the origin, and $F: S^1 \to S^1$ is a continuous mapping with degree n, then F is homotopic to the mapping that takes $(\cos(\vartheta), \sin(\vartheta))$ to $(\cos(n\vartheta), \sin(n\vartheta))$, i.e., in the terminology of complex numbers, the restriction of the nth power mapping $z \mapsto z^n$ to the unit circle.

Problem 3.27. If $F: C \to C'$ and $G: C' \to C''$ are continuous mappings of circles, what can you say about the relation among the degrees of F, G, and the composite $G \circ F$? Can you prove your answer?

These ideas can also be used to define an important notion of a *local degree*. Suppose U and V are open sets in the plane, and $F: U \to V$ is a continuous mapping, and let P be a point in U. Assume that P has some small neighborhood such that $F(Q) \neq F(P)$ for all Q in that neighborhood with $Q \neq P$. Choose a positive number r so that no point within a distance r of P has the same image as P, and let $C_r(P)$ be the circle of radius r about P. Then F restricts to a continuous mapping from $C_r(P)$ to $\mathbb{R}^2 \setminus \{F(P)\}$.



Define the *local degree of F at P*, denoted $\deg_P(F)$, to be the winding number of this mapping of the circle around the point $F(P)$. In other words, $\deg_P(F) = W(\gamma_r, F(P))$, where

$$\gamma_r(t) \;=\; F(P + r(\cos(2\pi t), \sin(2\pi t))), \qquad 0 \leq t \leq 1 .$$

To know that this is well defined, we need

Lemma 3.28. *This winding number is independent of choice of r.*

Proof. If r' is another, $H(t, s) = F(P + ((1-s)r + sr')(\cos(2\pi t), \sin(2\pi t)))$ gives a homotopy from γ_r to $\gamma_{r'}$. □

Equivalently, the local degree of F at P is the winding number of the mapping from the unit circle S^1 to itself given by

$$Q \;\mapsto\; \frac{F(P + rQ) - F(P)}{\|F(P + rQ) - F(P)\|} .$$

Problem 3.29. Show that if $F: \mathbb{R}^2 \to \mathbb{R}^2$ is a linear mapping, given by a (2×2)-matrix, and the determinant is not zero, then the local degree of F at the origin is $+1$ if this determinant is positive, and -1 if this determinant is negative.

Problem 3.30. Show that if $F: U \to V$ is a $\mathscr{C}^\infty$ mapping, and the Jacobian determinant of F is not zero at P, then the local degree is defined, and is $+1$ or -1, depending on the sign of this determinant.

Problem 3.31. Suppose $F: \mathbb{C} \to \mathbb{C}$ is given by a complex polynomial. Show that the local degree of F at a complex number z is the multiplicity of z as a root of $F(T) - F(z)$. (This multiplicity is the highest power of $T - z$ that divides $F(T) - F(z)$.)

Problem 3.32. If $F: C \to C'$ is a map between circles in the plane, and P is a point in C such that $F(Q) \neq F(P)$ for all Q in a neighborhood of P in C, give a precise definition of a local degree of F at P; this should be $+1$ if F is increasing at P, -1 if F is decreasing, and 0 if F has a local maximum or minimum at P (all expressed in terms of counterclockwise angles). Show that if P' is any point of C' such that $F^{-1}(P')$ is finite, then

$$\deg(F) = \sum_{P \in F^{-1}(P')} \deg_P F,$$

where $\deg_P F$ is the local degree of F at P. This implies in particular that the right side is independent of choice of P'.

CHAPTER 4

Applications of Winding Numbers

4a. The Fundamental Theorem of Algebra

The set $\mathbb{C}$ of complex numbers is identified as usual with the real plane $\mathbb{R}^2$, the number $z = x + iy$ being identified with the point (x, y). We will use the fact that for any complex polynomial

$$g(T) = a_0T^n + a_1T^{n-1} + \ . \ . \ . \ + a_{n-1}T + a_n\,,$$

with coefficients a_i complex numbers, the mapping $z \mapsto g(z)$ is a continuous mapping from $\mathbb{C}$ to $\mathbb{C}$. This follows from the fact that addition and multiplication of complex numbers are continuous. The goal of this section is to show that, if $n > 0$ and $a_0 \neq 0$, then the polynomial has a root, i.e., $g(z) = 0$ for some z. We may divide by a_0, so we can assume $g(T)$ has leading coefficient $a_0 = 1$. If $g(T)$ has no root, g is a mapping of $\mathbb{C}$ into the complement of the origin.

Restrict g to a circle C_r of radius r centered at the origin. This gives a mapping from C_r to $\mathbb{C} \setminus \{0\}$, which we denote by g_r. Since g_r extends to a continuous mapping of the disk D_r of radius r into $\mathbb{C} \setminus \{0\}$, it follows from Proposition 3.20 that the winding number $W(g_r, 0)$ must be zero. The idea is to compare g_r with the mapping f_r given similarly by the polynomial $f(T) = T^n$. The restriction of this to the circle is $f_r(z) = z^n$, and a simple calculation shows that $W(f_r, 0) = n$ (see Exercise 3.19). We will apply the Dog-on-a-Leash Theorem 3.11 to show that, for r sufficiently large, f_r and g_r must have the same winding number, which will be the desired contradiction. For this, it suffices to show that for r sufficiently large,

$$|f_r(z) - g_r(z)| < |f_r(z) - 0| \qquad \text{for} \quad z \in C_r.$$

Now $|f_r(z) - 0| = |z^n| = r^n$, and

$$\begin{aligned} |f_r(z) - g_r(z)| &= |a_1 z^{n-1} + \ldots + a_{n-1} z + a_n| \\ &\leq |a_1| r^{n-1} + \ldots + |a_{n-1}| r + |a_n|, \end{aligned}$$

which is less than r^n if r is large, e.g., if $|a_i| < r^i/n$ for all i. This completes the proof of

Proposition 4.1 (Fundamental Theorem of Algebra). *Any complex polynomial of degree greater than zero has a root.*

If z_1 is a root of $g(T)$, then $g(T) = (T - z_1) \cdot h(T)$, where $h(T)$ is a polynomial of degree $n - 1$. By induction, we see that $g(T)$ factors into linear factors: $g(T) = a_0 \cdot \Pi_{i=1}^{n} (T - z_i)$.

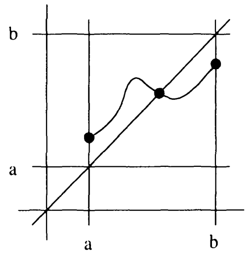
Exercise 4.2. (a) Suppose f: $\mathbb{C} \to \mathbb{C}$ is a continuous function such that for some $R > 0$, $|f(z)| < |z|^n$ for $|z| = R$. Show that $z^n + f(z) = 0$ has a solution z with $|z| < R$. (b) Suppose g: $\mathbb{C} \to \mathbb{C}$ is a continuous function such that $g(z)/z^n$ approaches a nonzero constant as $|z| \to \infty$. Show that g is surjective.

There is another topological proof of the Fundamental Theorem of Algebra that is in some ways even simpler, but requires a little more about the local structure of a mapping given by a polynomial. One shows that g extends to a continuous mapping of the Riemann sphere to itself, taking the point at infinity to itself (which is essentially what the above calculation showed), and which is an open mapping (see Problem 3.31, and §19a for details). The image is compact, so closed, and since both open and closed, it is the whole sphere.

4b. Fixed Points and Retractions

One of the most important applications of topological ideas in other areas of mathematics, and in science in general, is to give criteria to guarantee that a continuous mapping from a space to itself must have a point that is mapped to itself.

We start with the case of a closed interval $[a, b]$. We claim that any continuous function f: $[a, b] \to [a, b]$ must have a fixed point. This can be "seen" by looking at the graph of the mapping:



To prove it rigorously, consider the function $g(x) = f(x) - x$. This is a continuous function on the interval $[a, b]$, with $g(a) \geq 0$ and $g(b) \leq 0$. Since the image $g([a, b])$ of the interval by g must be connected, it must contain the interval $[g(b), g(a)]$, so it must contain 0. This means that $g(x) = 0$ for some x, which says that $f(x) = x$.

This is closely related to another property of an interval: that there is no continuous mapping from an interval $[a, b]$ onto its endpoints $[a, b]$ that maps a to a and b to b. This fact is obvious since the image of a connected set must be connected. In general if Y is a subspace of a topological space X, a *retraction* from X to Y is a continuous mapping $r: X \to Y$ such that $r(P) = P$ for all P in Y. In this case Y is called a *retract* of X. So there is no continuous retraction of an interval onto its boundary. This proves again that any continuous function f from the interval $[-1, 1]$ to itself must have fixed points, for otherwise the mapping $x \mapsto (x - f(x))/|x - f(x)|$ would be a continuous retraction of $[-1, 1]$ onto $\{-1, 1\}$.

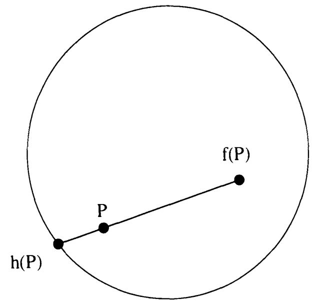
Next we turn to the case of a closed disk D. We first show there is no retraction onto its boundary circle $C = \partial D$, and then use this as above to show that any map from the disk to itself must have a fixed point.

Proposition 4.3. *There is no retraction from a closed disk onto its boundary circle.*

Proof. If C is the boundary of the disk D, the identity mapping from C to itself has degree 1. A retraction would give an extension of the identity mapping to a mapping from D to C, which would imply by Proposition 3.20 that the degree is 0. □

Proposition 4.4 (Brouwer). *Any continuous mapping from a closed disk to itself must have a fixed point.*

Proof. We may assume the disk D is the unit disk centered at the origin, with boundary C the unit circle (see Exercise 4.7 below). Suppose $f\colon D \to D$ is a continuous mapping with no fixed point. The idea is to define a mapping $h\colon D \to C$ that takes a point P to the point in C hit by the ray from $f(P)$ to P:



This mapping h will be the identity on C, so will be a retraction of D onto C. To finish the proof, we must verify that h is continuous. To do this, we find an explicit formula for h. We know that

$$h(P) \;=\; P + t\cdot U \qquad \text{with} \quad U = \frac{P - f(P)}{\|P - f(P)\|},$$

and t is the positive number determined by the property that $\|h(P)\| = 1$. A little calculation shows that

$$t \;=\; -P\cdot U + \sqrt{1 - P\cdot P + (P\cdot U)^2}$$

is such a positive number. With this t, the above formula for h is then continuous, and is easily checked to be the identity on C. □

The following exercise gives a proof with less calculation:

Exercise 4.5. If $f\colon D^2 \to D^2$ has no fixed point, define $g\colon D^2 \to \mathbb{R}^2 \setminus \{0\}$ by setting $g(P) = P - f(P)$. Show that $g(P)\cdot P > 0$ for all P in S^1, so the restriction of g to S^1 is homotopic to the identity mapping of S^1 (by the homotopy $H(P \times s) = (1-s)P + s\,g(P)$). Apply Proposition 3.20.

Exercise 4.6. Deduce Proposition 4.3 from Proposition 4.4.

One says that a topological space X *has the fixed point property* if every continuous mapping $f\colon X \to X$ has some fixed point, i.e., there is a point P in X with $f(P) = P$. So any closed interval and any closed disk have the fixed point property.

Exercise 4.7. Show that a space that is homeomorphic to a space with the fixed point property has the fixed point property. Show that if Y is a retract of a space X that has the fixed point property, then Y has the fixed point property.

Exercise 4.8. Which of the following spaces have the fixed point property? (i) a closed rectangle; (ii) the plane; (iii) an open interval; (iv) an open disk; (v) a circle; (vi) a sphere S^2; (vii) a torus $S^1 \times S^1$; and (viii) a solid torus $S^1 \times D^2$.

Exercise 4.9. Let D be a disk with boundary circle C, and let $f\colon D \to \mathbb{R}^2$ be a continuous mapping. Suppose P is a point in $\mathbb{R}^2$ that is not in the image C, and the winding number of the restriction $f|_C$ of f to C around P is not zero. Show that there is some point Q in D such that $f(Q) = P$.

Exercise 4.10. Suppose D and D' are disks with boundary circles C and C'. Suppose $f\colon D \to \mathbb{R}^2$ is a continuous mapping that maps C into C', such that the degree of this map from C to C' is not zero. Show that $f(D)$ must contain D'.

Exercise 4.11. Show that if $f\colon D^2 \to \mathbb{R}^2 \setminus \{0\}$ is a continuous mapping, there is some P in $S^1 = \partial D^2$ and some $\lambda > 0$ such that $f(P) = \lambda \cdot P$, and there is some Q in S^1 and some $\mu < 0$ such that $f(Q) = \mu \cdot Q$.

Exercise 4.12. Suppose F is a continuous mapping from the positive octant $\{(x, y, z)\colon x \geq 0, y \geq 0, z \geq 0\}$ to itself. Show that there is a unit vector P in this octant, and a nonnegative number λ, such that $F(P) = \lambda \cdot P$.

Exercise 4.13. If all the entries of a (2×2)-matrix are nonnegative, show by direct calculation that at least one of its eigenvalues must be nonnegative. Prove the same for a (3×3)-matrix A.

Exercise 4.14. If $f\colon D^2 \to \mathbb{R}^2$ is a continuous mapping, show that either: (i) there is either some point $Q \in D^2$ such that $f(Q) = Q$; or (ii) there is some P_1 in S^1 and some $\lambda_1 > 1$ such that $f(P_1) = \lambda_1 \cdot P_1$, and there is some point P_2 in S^1 and some $\lambda_2 < 1$ such that $f(P_2) = \lambda_2 \cdot P_2$.

Exercise 4.15. If $f\colon C \to C$ is a continuous mapping with no fixed point, show that degree of f must be 1. In particular, if f has no fixed point, show that f must be surjective.

Exercise 4.16. If $f\colon C \to \mathbb{R}^2$ is continuous and $W(f, P) \neq 0$, show that every ray emanating from P meets $f(C)$.

Exercise 4.17. If $f\colon D^2 \to \mathbb{R}^2$ is continuous and $f(P) \cdot P \neq 0$ for all P in S^1, show that there is some Q in D^2 with $f(Q) = 0$.

Problem 4.18. Let $D^\infty = \{(a_0, a_1, a_2, \ldots): \Sigma_{n=0}^{\infty} a_n^{\,2} \leq 1\}$, the unit ball in the metric space of infinite sequences such that $\Sigma_{n=0}^{\infty} a_n^{\,2} < \infty$. (a) Find a continuous mapping $f\colon D^\infty \to D^\infty$ that has no fixed point. (b) Find a continuous retraction of D^∞ onto $S^\infty = \{(a_0, a_1, \ldots): \Sigma_{n=0}^{\infty} a_n^{\,2} = 1\}$.

4c. Antipodes

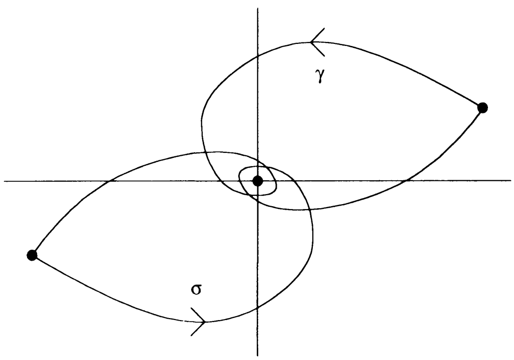
The *antipode* of a point in a circle C or sphere S is the opposite point, i.e., the point hit by the ray from the point through the center. We denote the antipode of P by P^*. With the center at the origin, $P^* = -P$. The *antipodal map* is the mapping that takes each point to its antipode.

Exercise 4.19. Show that the degree of the antipodal map from a circle to itself is 1.

Lemma 4.20 (Borsuk). *If C and C' are circles, and $f\colon C \to C'$ is a continuous map such that $f(P^*) = f(P)^*$ for all P, then the degree of f is odd.*

Proof. There is no loss of generality in assuming $C = C' = S^1$. Let $\gamma(t) = f(\cos(t), \sin(t)), 0 \leq t \leq \pi$. Since $\gamma(\pi) = -\gamma(0)$, the change in angle along γ must be $2\pi n + \pi$ for some integer n. So $W(\gamma, 0) = n + {}^1\!/_2$. Now let

$$\sigma(t) \;=\; f(\cos(t+\pi), \sin(t+\pi)) \;=\; -\gamma(t), \qquad 0 \leq t \leq \pi.$$



It follows easily from the definition of the winding number that $W(\sigma, 0) = W(\gamma, 0)$, and so $\deg(f) = W(\gamma, 0) + W(\sigma, 0) = 2n + 1$. □

Lemma 4.21. *There is no continuous mapping f from a sphere S to a circle C such that $f(P^*) = f(P)^*$ for all P in S.*

Proof. Again we can take $S = S^2$. Consider the mapping $g\colon D^2 \to C$ given by the formula

$$g(x, y) = f((x, y, \sqrt{1 - x^2 - y^2})).$$

This is continuous, since the projection from the upper hemisphere of the sphere to the disk (being one-to-one and continuous) is a homeomorphism, and g is the composite of f with the inverse. Now $g(P^*) = f(P^*) = f(P)^* = g(P)^*$ for P in S^1, so by Lemma 4.20 the degree of the restriction of g to S^1 must be odd. But since this map extends over the disk, its degree must be zero, a contradiction. □

Proposition 4.22 (Borsuk–Ulam). *For any continuous mapping $f\colon S \to \mathbb{R}^2$ from a sphere S to the plane, there is a point P in S such that $f(P) = f(P^*)$.*

So there are always two antipodal points on the earth with the same temperature and humidity—or any other two real values, provided they vary continuously over the earth.

Proof. We may take $S = S^2$. If there is no such P, consider the function $g\colon S^2 \to S^1$ given by

$$g(P) = \frac{f(P) - f(-P)}{\|f(P) - f(-P)\|}.$$

Then $g(-P) = -g(P)$, contradicting Lemma 4.21. □

A fact which was unquestionably obvious until people started looking for a proof is the fact that open sets of different dimensions cannot be homeomorphic. (The fact that they cannot be *diffeomorphic* can be reduced, using Jacobian matrices, to the fact that vector spaces of different dimensions cannot be isomorphic.) The fact that there are continuous maps from intervals onto squares makes the assertion less obvious than might have been thought. The first case is easy: an open interval cannot be homeomorphic to an open set in the plane or any $\mathbb{R}^n$, for the reason that removing a point disconnects an interval, but does not disconnect an open set in $\mathbb{R}^n$, $n \geq 2$. The next case is less obvious:

Corollary 4.23 (Invariance of Dimension). *An open set in* $\mathbb{R}^2$ *cannot be homeomorphic to an open set in* $\mathbb{R}^n$ *for* $n \geq 3$.

Proof. In fact, no subset of the plane can be homeomorphic to a set which contains a solid ball in $\mathbb{R}^n$, $n \geq 3$, for a homeomorphism from a ball D^n (of some small radius) in the plane would embed a two-sphere $S^2 \subset D^3 \subset D^n$ in the plane, contradicting the proposition. □

As you must expect, these results are also true in higher dimensions, but more machinery is needed to extend the proofs. We'll come back to this in the last part of the book.

Exercise 4.24. Show that if $f\colon C \to C'$ is a map between circles such that $f(P^*) = f(P)$ for all P, then the degree of f is even.

Exercise 4.25. If $f\colon C \to \mathbb{R}^2 \setminus \{Q\}$ is a continuous mapping such that Q lies on the line segment between $f(P)$ and $f(P^*)$ for all P in C, show that the winding number of f around Q is odd.

Exercise 4.26. Suppose D is a disk with boundary C, and $f\colon D \to \mathbb{R}^2$ is a continuous mapping such that $f(P^*) = -f(P)$ for all P in C. Show that there is some point Q in D with $f(Q) = 0$.

Exercise 4.27. If f and g are continuous real-valued functions on a sphere S such that $f(P^*) = -f(P)$ and $g(P^*) = -g(P)$ for all P, show that f and g must have a common zero on the sphere.

Exercise 4.28. State and prove the analogue of the Borsuk–Ulam theorem for mappings from a circle to $\mathbb{R}$.

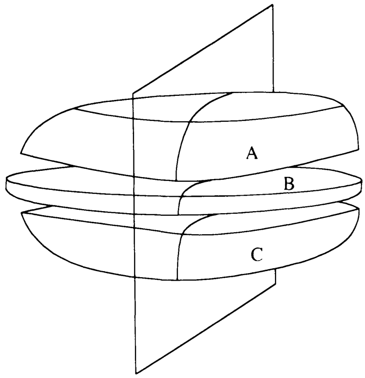
Exercise 4.29. If $f\colon C \to C'$ is a continuous mapping between circles such that $f(P^*) \neq f(P)$ for all P, show that $\deg(f) \neq 0$, so f must be surjective.

Problem 4.30. Let $f\colon S \to S'$ be a continuous mapping between spheres. Show that if $f(P) \neq f(P^*)$ for all P, then f must be surjective.

Exercise 4.31. Let $f\colon C \to C$ be continuous. If $\deg(f) \neq 1$, show that there is a P in C with $f(P) = P$ and there is a Q in C with $f(Q) = Q^*$.

4d. Sandwiches

A pleasant application of the Borsuk–Ulam theorem is the so-called ham sandwich problem: to cut both slices of bread and the ham between in two equal parts with one slice of a knife.



Suppose we have three bounded objects A, B, C in space. The problem is to show that there is one plane that cuts each of them in half (by volume). Here is what we need to know about the volume of each object $X = A$, B, or C:

(i) For any fixed line l, there is a unique point in l such that the plane perpendicular to l through the point cuts X in half. Call this point $P_{l,X}$.
(ii) The point in (i) varies continuously as the line varies continuously. More precisely, take a big sphere S with X inside, and consider for each $Q \in S$ the line $l(Q)$ going from Q to its antipodal point. Then the map from Q to $P_{l(Q),X}$ is continuous.

These properties are intuitively evident, and follow from elementary properties of volume. For example, the first follows from the fact that the volume on one side increases continuously as the point moves along the line. We will take them as axioms, or consider only objects for which we know them. (For a detailed discussion, see Chinn and Steenrod (1966).)

Given a body X inside a sphere S as above, define a continuous real-valued function $f_X \colon S \to \mathbb{R}$ by defining $f_X(Q)$ to be the distance from Q to the point $P_{l(Q),X}$. Note that $f_X(Q^*) = d - f_X(Q)$, where d is the diameter of the sphere, and Q^* is the antipodal point to Q.

Now for three bodies, take S containing all three, and consider the mapping $g: S \rightarrow \mathbb{R}^2$ given by

$$g(Q) = (f_A(Q) - f_C(Q), f_B(Q) - f_C(Q)).$$

Then $g(Q^*) = -g(Q)$ for all Q, so by Proposition 4.22 some Q must be mapped to the origin, which means that $P_{l(Q),\mathrm{A}} = P_{l(Q),\mathrm{B}} = P_{l(Q),\mathrm{C}}$, as required. This proves

Proposition 4.32 (Stone–Tukey). *Given three bounded measurable objects A, B, and C in space, there is a plane that divides each in half by volume.*

The same idea is used in the following proposition, which looks quite different.

Proposition 4.33 (Lusternik–Schnirelman–Borsuk). *It is impossible to cover a sphere with three closed sets, none of which contains a pair of antipodal points.*

Proof. Suppose the sphere S is covered by three such closed sets K_1, K_2, and K_3. For each i, define a real-valued continuous function f_i on S, whose value at P is the minimum distance from P to K_i (see the following exercise). Consider the mapping $g: S \rightarrow \mathbb{R}^2$ given by

$$g(P) = (f_1(P) - f_3(P), f_2(P) - f_3(P)).$$

By Proposition 4.22, there is a point P with $g(P^*) = g(P)$. For such P, $f_i(P) - f_j(P) = f_i(P^*) - f_j(P^*)$ for all i and j. But P must be in one of the sets K_i, and P^* in another K_j. Since $P \notin K_j$ and $P^* \notin K_i$,

$$0 > -f_j(P) = f_i(P) - f_j(P) = f_i(P^*) - f_j(P^*) = f_i(P^*) > 0,$$

a contradiction. □

Exercise 4.34. Show that for any compact set K in space, the function ρ on $\mathbb{R}^3$ that maps a point P to its distance from K is continuous. Note by compactness that if $\rho(P) = r$, then there is a point Q in K of distance r from P.

Exercise 4.35. Show that the unit ball D^3 is not the union of three closed sets, each with diameter less than 2.

Exercise 4.36. Show that the "three" in Proposition 4.33 cannot be replaced by "four."

Exercise 4.37. State and prove the analogue of Proposition 4.33 for a circle.

Exercise 4.38. Show how Proposition 4.33 implies Lemma 4.21, and hence Proposition 4.22.

Exercise 4.39. For a subset X of a sphere, let $X^* = \{P^*: P \in X\}$. Suppose A, B, and C are disjoint closed subsets in a sphere, none containing a pair of antipodal points. Show that the six sets A, B, C, A^*, B^*, and C^* cannot cover the sphere.

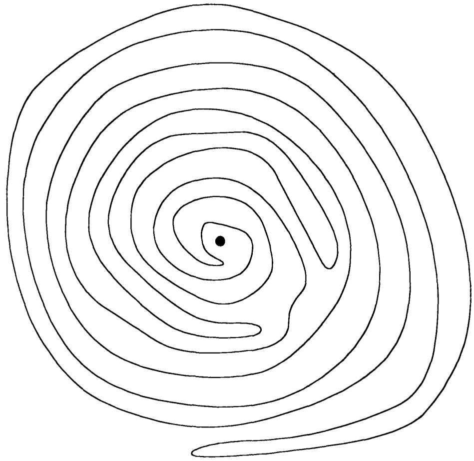
Problem 4.40. True/False. If a sphere is a union of two closed sets A and B, then either A or B must contain a closed connected set X such that $X^* = X$.

It is an excellent project to try to generalize the results of this chapter, including the exercises, to higher dimensions, assuming, for example, that there is no continuous retraction from D^n onto S^{n-1}, and that there is no continuous mapping f from the sphere S^n to the sphere S^{n-1} such that $f(P^*) = f(P)^*$ for all P. This will be carried out in Chapter 23.

PART III
COHOMOLOGY AND HOMOLOGY, I

We have seen that the topology of an open set U in the plane is related to the question of whether all closed 1-forms on U are exact. This is formalized by introducing the vector space H^1U of closed forms modulo exact forms. What we learned in the first few chapters amounts to some calculations of these first De Rham cohomology groups. There is also a 0th De Rham group H^0U, which measures how many connected components U has.

A central theme will be how the topology of a union $U \cup V$ of two open sets compares with the topology of U and V and the intersection $U \cap V$. We construct a linear map from $H^0(U \cap V)$ to $H^1(U \cup V)$, and describe the kernel and cokernel of this map. We use these groups and this map to prove the famous Jordan curve theorem, which says that any subset of the plane homeomorphic to a circle separates the plane into exactly two connected pieces, an "inside" and an "outside."



This is one of those facts that seem intuitively evident, although a complicated enough maze may raise a few doubts. In fact, it went unquestioned until mathematicians realized that continuous mappings can be pretty horrible—for example, that there can be a continuous mapping from an interval onto a square. When X is a closed polygon, it is not too hard to give an elementary proof, and you might enjoy seeing if you can.

A basic question, which motivates Chapter 6 and the continuation in Chapter 9, stems from three different ways one can compare two closed paths γ and δ in an open set U in the plane, supposing for simplicity that they are differentiable:

(1) Are the winding numbers of γ and δ around all points not in U the same?
(2) Are all integrals of closed 1-forms ω on U along γ and δ the same?
(3) Are the paths homotopic, or related by deformations of some kind?

In studying questions like this, we will find it useful to generalize and formalize some of the ideas of previous chapters. In Chapter 6, paths and segmented paths are generalized to the notion of 1-chains, which are arbitrary sums and differences of (nonconstant) paths. We define what it means for two such chains to be homologous: the difference should be a linear combination of boundaries of maps of a rectangle into the region. For an open set in the plane, we show that this notion is equivalent to saying that the two chains have the same winding number around all points outside the open set. The main technique is to approximate general paths by rectangular paths along sides of a grid.[3] We introduce the first homology group H_1U, which is the group of closed 1-chains up to homology, and a 0th group H_0U, which is a free abelian group on the connected components of U. In Part V we will develop general tools for computing these homology and cohomology groups.

[3] The use of grids in these chapters follows ideas of L.E.J. Brouwer, E. Artin, and L. Ahlfors, see Ahlfors (1979).

CHAPTER 5

De Rham Cohomology and the Jordan Curve Theorem

5a. Definitions of the De Rham Groups

Define, for an open set U in the plane: the *zeroth De Rham group*,

$$H^0U = \{\text{locally constant functions on } U\}.$$

This is a vector space, by the ordinary addition of functions, and multiplication of functions by real scalars. As we have seen, to give a locally constant function on U is the same as giving a constant on each connected component of U. If U has n connected components, then H^0U is an n-dimensional vector space; if $U_1, \ldots, U_n$ are the connected components of U, and e_i is the function that is 1 on U_i and 0 on U_j for all $j \neq i$, then $e_i, \ldots, e_n$ is a basis for H^0U. Similarly, if U has infinitely many components, H^0U is an infinite-dimensional vector space.

The closed 1-forms on U also form a vector space, since the sum of closed forms, and a constant times a closed 1-form, are also closed. The exact 1-forms on U are a subspace of this vector space, since $d(f_1 + f_2) = df_1 + df_2$ and $d(cf) = cdf$. Whenever we have a subspace W of a vector space V, we can form the quotient space V/W of equivalence classes: two vectors in V are equivalent if their difference is in W (see Appendix C). Define the *first De Rham cohomology group of* U, denoted H^1U, by

$$H^1U = \{\text{closed 1-forms on } U\}/\{\text{exact 1-forms on } U\}.$$

If ω is a 1-form on U, we may write $[\omega]$ for the equivalence class in H^1U containing the 1-form ω.

Later we will have general methods for calculating H^1U. In this section we use the ideas of Part I to compute a few simple examples. For any point $P = (x_0, y_0)$, let ω_P denote the 1-form

$$\omega_P = \frac{1}{2\pi}\omega_{P,\vartheta} = \frac{1}{2\pi}\frac{-(y-y_0)\,dx + (x-x_0)\,dy}{(x-x_0)^2 + (y-y_0)^2}.$$

We have seen that ω_P is a closed 1-form on any open set not containing P.

Proposition 5.1. (*a*) *If U is an open rectangle, then $H^1U = 0$.* (*b*) *If $U = \mathbb{R}^2 \setminus \{P\}$, then $[\omega_P]$ is a basis for H^1U.* (*c*) *If $U = \mathbb{R}^2 \setminus \{P, Q\}$, with $P \neq Q$, then $[\omega_P]$ and $[\omega_Q]$ form a basis for H^1U.*

Proof. Assertion (a) is a translation of Proposition 1.12. To prove (b), fix a positive number r, and let $\gamma_{P,r}$ denote the counterclockwise circle of radius r about P:

$$\gamma_{P,r}(t) = P + r(\cos(2\pi t), \sin(2\pi t)), \qquad 0 \leq t \leq 1.$$

Note first that $[\omega_P]$ is not 0 in H^1U, for if ω_P were exact, its integral around $\gamma_{P,r}$ would be zero by Proposition 1.4, and we know that this integral is 1. Let ω be any closed 1-form on U, and let $c = \int_{\gamma_{P,r}} \omega$. To show that $[\omega_P]$ spans H^1U, it suffices to show that $\omega - c \cdot \omega_P$ is exact, for then $[\omega] = c \cdot [\omega_P]$ in H^1U. Now $\omega - c \cdot \omega_P$ is a closed 1-form on U whose integral along γ is 0, and Lemma 1.17 implies that such a form is exact.

For (c), fix a positive number r less than the distance between P and Q. To see that $[\omega_P]$ and $[\omega_Q]$ are linearly independent, we must show that $a \cdot \omega_P + b \cdot \omega_Q$ is not exact unless a and b are both zero. The integral of this form along $\gamma_{P,r}$ is a, and the integral along $\gamma_{Q,r}$ is b, and by Proposition 1.4, these integrals must vanish if the form is exact. To show that $[\omega_P]$ and $[\omega_Q]$ span H^1U, let ω be any closed 1-form on U. Let $a = \int_{\gamma_{P,r}} \omega$ and let $b = \int_{\gamma_{Q,r}} \omega$. To complete the proof of (c), we must show that $\omega - a \cdot \omega_P - b \cdot \omega_Q$ is exact on U, for then $[\omega] = a \cdot [\omega_P] + b \cdot [\omega_Q]$ in H^1U. Now $\omega - a \cdot \omega_P - b \cdot \omega_Q$ is a closed 1-form on U whose integrals along $\gamma_{P,r}$ and $\gamma_{Q,r}$ both vanish, and an appeal to Lemma 1.18 completes the proof. □

Problem 5.2. Generalize the proposition from one and two to n points.

There is one other result we will need later in this chapter:

Proposition 5.3. *If A is a connected closed subset of $\mathbb{R}^2$, and P and Q are points in A, then $[\omega_P] = [\omega_Q]$ in $H^1(\mathbb{R}^2 \setminus A)$.*

Proof. In order to show that $\omega = \omega_P - \omega_Q$ is exact on $\mathbb{R}^2 \setminus A$ it suffices by Proposition 1.8 to show that $\int_\gamma \omega = 0$ whenever γ is a segmented closed path in $\mathbb{R}^2 \setminus A$. But

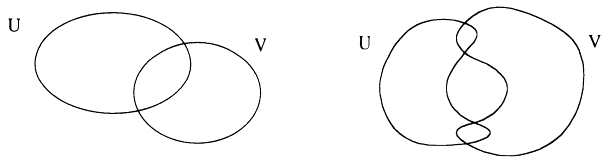
$$\int_\gamma \omega = \int_\gamma \omega_P - \int_\gamma \omega_Q = W(\gamma, P) - W(\gamma, Q).$$

Now Proposition 3.16 implies that the winding numbers $W(\gamma, P)$ and $W(\gamma, Q)$ are equal, since P and Q belong to the same connected component of $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$. □

Exercise 5.4. With A as in the proposition, and P in A, show that $[\omega_P] = 0$ in $H^1(\mathbb{R}^2 \setminus A)$ if and only if A is unbounded.

5b. The Coboundary Map

Our basic tool for studying H^0U and H^1U, and hence the topology of U, will be a method to study the relations among these groups for open sets U, V and their union $U \cup V$ and intersection $U \cap V$.



That there might be some relation can be seen already in Lemma 1.14. This showed, for example, that if $U \cap V$ is connected, i.e., $H^0(U \cap V)$ has dimension at most one, and if $H^1U = 0$, and $H^1V = 0$, then $H^1(U \cup V)$ must vanish as well. We want to generalize this to more complicated open sets.

For any two open sets U and V in the plane, we will define a linear map

$$\delta\colon H^0(U \cap V) \to H^1(U \cup V).$$

To do this, we will need the following basic fact:

Lemma 5.5. *Given open sets U and V in $\mathbb{R}^2$, any $\mathscr{C}^\infty$ function on*

$U \cap V$ can be written as the difference of two $\mathscr{C}^\infty$ functions, one that extends to a $\mathscr{C}^\infty$ function on U, the other to a $\mathscr{C}^\infty$ function on V.

Proof. This is a consequence of the existence of a partition of unity, for the simple (but not so trivial) case of the covering of $U \cup V$ by the open sets U and V. This says that there are $\mathscr{C}^\infty$ functions φ and ψ on $U \cup V$ such that $\varphi + \psi \equiv 1$ on $U \cup V$, such that the closure (in $U \cup V$) of the support of φ is contained in U, and the closure (in $U \cup V$) of the support of ψ is contained in V. (The support of a function is the set where it is not zero.) For the proof, see Appendix B2.

To use it to prove the lemma, given a $\mathscr{C}^\infty$ function f on $U \cap V$, define f_1 on U by the rule

$$f_1 = \begin{cases} \psi \cdot f & \text{on } U \cap V, \\ 0 & \text{on } U \setminus U \cap V. \end{cases}$$

The assumption on the support of ψ insures that any point in the set $U \setminus U \cap V$ has a neighborhood disjoint from the support of ψ, from which it follows that f_1 is $\mathscr{C}^\infty$ on all of U. Similarly, define a $\mathscr{C}^\infty$ function f_2 on V by

$$f_2 = \begin{cases} -\varphi \cdot f & \text{on } U \cap V, \\ 0 & \text{on } V \setminus U \cap V. \end{cases}$$

Then $f_1 - f_2 = (\psi + \varphi)f = f$ on $U \cap V$, as required. □

Construction of δ: $H^0(U \cap V) \to H^1(U \cup V)$. Given a locally constant function f on $U \cap V$, use the lemma to find $\mathscr{C}^\infty$ functions f_1 and f_2 on U and V respectively so that $f = f_1 - f_2$ on $U \cap V$. Since f is locally constant,

$$df_1 - df_2 = d(f_1 - f_2) = df = 0 \quad \text{on } U \cap V.$$

This means df_1 and df_2 agree on $U \cap V$, so there is a unique 1-form ω on $U \cup V$ that agrees with df_1 on U and with df_2 on V. This 1-form ω is closed, since it is even exact on each of U and V.

We define $\delta(f)$ to be the equivalence class in $H^1(U \cup V)$ determined by this closed form ω, i.e., $\delta(f) = [\omega]$. For this to be well defined, we must see how ω depends on the way we write f as the difference $f_1 - f_2$. We claim that a different choice would lead to a closed 1-form that differs from ω by an exact 1-form. To see this, suppose f_1' and f_2' were another choice of functions on U and V with $f_1' - f_2' = f$ on $U \cap V$. (These primes have nothing to do with derivatives!) Let ω' be the 1-form on $U \cap V$ that is df_1' on U and is df_2'

on V. Now since $f_1' - f_2' = f_1 - f_2$ on $U \cap V$,

$$f_1' - f_1 = f_2' - f_2$$

on $U \cap V$, so there is a $\mathscr{C}^\infty$ function g on $U \cup V$ that is $f_1' - f_1$ on U and is $f_2' - f_2$ on V. Then $dg = \omega' - \omega$, as required.

Lemma 5.6. *The mapping δ is a linear mapping of vector spaces, i.e., $\delta(f+g) = \delta(f) + \delta(g)$ and $\delta(c \cdot f) = c \cdot \delta(f)$, for f, g locally constant functions, and c a constant.*

Proof. This is just a matter of making the choices "linearly," and is better (and probably easier) to check for yourself than to read. Write $f = f_1 - f_2$ and $g = g_1 - g_2$ as in the construction of $\delta(f)$ and $\delta(g)$. Then $f + g = (f_1 + g_1) - (f_2 + g_2)$. If ω_f is the 1-form that is df_1 on U and df_2 on V, and ω_g is the 1-form that is dg_1 on U and dg_2 on V, then $\omega_f + \omega_g$ is the 1-form that is $d(f_1 + g_1)$ on U and $d(f_2 + g_2)$ on V. Therefore $\delta(f+g)$ is represented by the 1-form $\omega_f + \omega_g$, which by definition represents the sum $\delta(f) + \delta(g)$. This proves that δ preserves sums. The proof that it preserves multiplication by a scalar is similar, and left as an exercise. □

This map δ is called the *coboundary* map. In order to use it to compare $U \cap V$ with $U \cup V$, we need a description of its kernel and its image. The following two propositions do this.

Proposition 5.7. *A locally constant function f on $U \cap V$ is in the kernel of δ if and only if there are locally constant functions f_1 on U and f_2 on V so that $f = f_1 - f_2$ on $U \cap V$. In particular, if U and V are connected, then the kernel of δ consists of the constant functions on $U \cap V$.*

Proof. If $f = f_1 - f_2$ with f_1 and f_2 locally constant on U and V, these can be chosen for the construction of $\delta(f)$, and the corresponding form ω is zero. Conversely, if $\delta(f)$ is zero, the form ω of the construction from an equation $f = f_1 - f_2$ must be exact. Write $\omega = dg$. Then $df_1 = dg$ on U, and $df_2 = dg$ on V. This means that $f_1 - g$ is locally constant on U, and $f_2 - g$ is locally constant on V. And

$$f = f_1 - f_2 = (f_1 - g) - (f_2 - g)$$

is the difference of two such functions.

If U and V are connected, locally constant functions on them must be constant, and since f is the difference of two such functions, f is also constant. □

Exercise 5.8. Show that if U and V are connected, and $H^1(U \cup V) = 0$, then $U \cap V$ is also connected.

Proposition 5.9. *The class $[\omega]$ of a closed 1-form ω on $U \cup V$ is in the image of δ if and only if the restrictions of ω to U and to V are exact. In particular, if $H^1U = 0$ and $H^1V = 0$, then δ is surjective.*

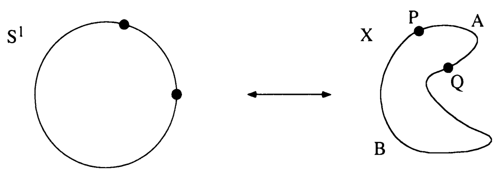
Proof. By construction, if ω is in the image of δ, there are functions f_1 on U and f_2 on V with $\omega = df_1$ on U and $\omega = df_2$ on V. Conversely, if there are such functions, the function $f = f_1 - f_2$ on $U \cap V$ is locally constant, since $df = df_1 - df_2 = \omega - \omega = 0$ on $U \cap V$, and $[\omega] = \delta(f)$ by construction. □

5c. The Jordan Curve Theorem

If X is a closed subset of the plane, there is a close relation between the topology of X and the topology of its complement. We prove some important cases of this fact here.

Theorem 5.10 (Jordan Curve Theorem). *If $X \subset \mathbb{R}^2$ is homeomorphic to a circle, then its complement $\mathbb{R}^2 \setminus X$ has two connected components, one bounded, the other unbounded. Any neighborhood of any point on X meets both of these components.*

Proof. Let P and Q be any two points in X. By considering a homeomorphism of X with the circle, we can write X as a union of two subsets A and B, each homeomorphic to a closed interval, with $A \cap B = \{P, Q\}$.



We will apply the results of the preceding section to the open sets $U = \mathbb{R}^2 \setminus A$ and $V = \mathbb{R}^2 \setminus B$. Note that

$$U \cup V = \mathbb{R}^2 \setminus \{P, Q\} \quad \text{and} \quad U \cap V = \mathbb{R}^2 \setminus X.$$

To show that $\mathbb{R}^2 \setminus X$ has two components, we want to show that the dimension of $H^0(U \cap V) = H^0(\mathbb{R}^2 \setminus X)$ is 2.

We know from Proposition 5.1(c) that $H^1(U \cup V) = H^1(\mathbb{R}^2 \setminus \{P, Q\})$ is a vector space of dimension 2, with a basis the classes of ω_P and ω_Q. Each of U and V is the complement of a subset homeomorphic to an interval. We will need an analogue of the Jordan curve theorem, but where the circle is replaced by an interval:

Theorem 5.11. *If $Y \subset \mathbb{R}^2$ is homeomorphic to a closed interval, then $\mathbb{R}^2 \setminus Y$ is connected.*

We postpone a discussion of this theorem to the end of this section, and show now how to use it to prove the Jordan curve theorem. Consider the boundary map

$$\delta\colon\ H^0(U \cap V) = H^0(\mathbb{R}^2 \setminus X) \quad \to \quad H^1(U \cup V) = H^1(\mathbb{R}^2 \setminus \{P, Q\}).$$

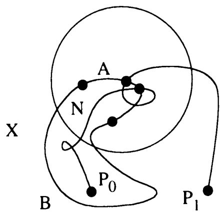
Our goal is to show that the image and kernel of δ are both one dimensional, which will imply by the rank-nullity theorem (see Appendix C) that $H^0(\mathbb{R}^2 \setminus X)$ is two dimensional, which means that $\mathbb{R}^2 \setminus X$ has two connected components.

Since, by Theorem 5.11, U and V are connected, we can apply Proposition 5.7, so the kernel of δ consists of the constant functions on $\mathbb{R}^2 \setminus X$, which is therefore one dimensional. We claim that the image of δ consists of those linear combinations $a \cdot [\omega_P] + b \cdot [\omega_Q]$ of the basis elements that have $a + b = 0$. This means that $[\omega_P] - [\omega_Q]$ forms a basis for the image of δ, and completes the proof of the claim that the image is one dimensional.

By Proposition 5.9, the image of δ consists of those linear combinations $a \cdot [\omega_P] + b \cdot [\omega_Q]$ such that the restrictions of $a \cdot \omega_P + b \cdot \omega_Q$ to U and to V are exact. Since P and Q are in the same connected component of $\mathbb{R}^2 \setminus U$ and of $\mathbb{R}^2 \setminus V$, it follows from Proposition 5.3 that $\omega_P - \omega_Q$ is exact on U and on V. If $a + b = 0$, it follows that $a \cdot \omega_P + b \cdot \omega_Q = a \cdot (\omega_P - \omega_Q)$ is exact on U and on V, so $a \cdot [\omega_P] + b \cdot [\omega_Q]$ is in the image of δ if $a + b = 0$. Conversely, suppose the restriction of $\omega = a \cdot \omega_P + b \cdot \omega_Q$ to U (and to V) is exact. Let $\gamma = \gamma_{0,r}$ be a circle about the origin, with r so large that X is contained inside this circle. Since ω is exact on U, Proposition 1.4 guarantees that $\int_\gamma \omega = 0$. Since P and Q are inside the circle, $\int_\gamma \omega_P = 1$ and $\int_\gamma \omega_Q = 1$, so

$$0 \;=\; \int_\gamma \omega \;=\; a \cdot \int_\gamma \omega_P + b \cdot \int_\gamma \omega_Q \;=\; a + b.$$

This completes the proof that $\mathbb{R}^2 \setminus X$ has two connected components. Note that since X is bounded, one of these components must contain everything outside some large disk; this is the unbounded component, while the other must be bounded. To verify the last assertion of the theorem, suppose N is a neighborhood of some point in X. We may divide X into two pieces, as in the preceding discussion, so that one of them, say A, lies entirely in N. Take two points, say P_0 and P_1, one in each of the two components of $\mathbb{R}^2 \setminus X$.

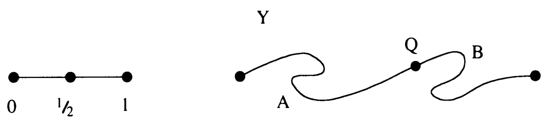


By Theorem 5.11, $\mathbb{R}^2 \setminus B$ is connected, so there is a path $\gamma(t)$, $0 \leq t \leq 1$, from P_0 to P_1 in $\mathbb{R}^2 \setminus B$. This path must hit A, for if it didn't, it would connect the two points in $\mathbb{R}^2 \setminus X$. By looking at the first and last time the path hits A, i.e., the minimal and maximal t such that $\gamma(t)$ is in A (which is a closed set in $[0, 1]$), it follows that N must contain points of both components: for t near to but less than the minimum, $\gamma(t)$ will be in N and the component of P_0, and for t near to but greater than the maximum, $\gamma(t)$ will be in N and the other component. □

Remark 5.12. (1) It follows in particular that X is nowhere dense: no point has a neighborhood contained in X. Note that an arbitrary continuous image of a circle or interval need not have this property.

(2) The fact that points on X are "accessible" from both sides is strengthened in Problem 10.24.

***Proof of Theorem* 5.11.** The proof of Theorem 5.11 uses the same ideas, but with one new wrinkle. Choose a homeomorphism from the interval $[0, 1]$ to Y. Let A be the subset of Y corresponding to the left half of the interval $[0, {}^1\!/_2]$, and B the subset corresponding to the right half $[{}^1\!/_2, 1]$, and $Q = A \cap B$ the point corresponding to ${}^1\!/_2$.



We apply the basic construction to $U = \mathbb{R}^2 \setminus A$ and $V = \mathbb{R}^2 \setminus B$, with $U \cap V = \mathbb{R}^2 \setminus Y$ and $U \cup V = \mathbb{R}^2 \setminus \{Q\}$. It might seem that little will be gained by this, since each of U, V, and $U \cap V$ is of the same form: the complement of a subset homeomorphic to an interval. But we will see that some progress has been made.

Suppose $\mathbb{R}^2 \setminus Y$ is not connected, and let P_0 and P_1 be points in two different connected components. The claim we will prove is: *P_0 and P_1 must be in different connected components of $\mathbb{R}^2 \setminus A$ or in different connected components of $\mathbb{R}^2 \setminus B$* (or both). We use the map

$$\delta\colon H^0(U \cap V) = H^0(\mathbb{R}^2 \setminus Y) \to H^1(U \cup V) = H^1(\mathbb{R}^2 \setminus \{Q\}).$$

Now $H^1(\mathbb{R}^2 \setminus \{Q\})$ is generated by the class $[\omega_Q]$ of ω_Q. By Proposition 5.9 the image of δ consists of classes $a \cdot [\omega_Q]$ such that $a \cdot \omega_Q$ is exact on U and V. If $a \cdot \omega_Q$ is exact on U, the integral of $a \cdot \omega_Q$ around a large circle is zero, which implies that $a = 0$. So the image of δ is zero. In other words, every locally constant function on $\mathbb{R}^2 \setminus Y$ is in the kernel of δ.

By Proposition 5.7, a function f in the kernel of δ must have the form $f_1 - f_2$, where f_1 and f_2 are locally constant functions on U and on V, respectively. Take f to be any locally constant function on $U \cap V$ that takes on different values at the two points P_0 and P_1, which is possible since they are in different components. If the claim were false, and P_0 and P_1 were in the same component of U and of V, then each of f_1 and f_2 would have to take on the same values at P_0 and P_1. But then their difference f would also take on the same values, which is a contradiction.

The progress made by proving this claim comes from the fact that the two sets A and B are *smaller* than Y. To take advantage of this, we can argue as follows: Let Y_1 be one of the halves of Y such that P_0 and P_1 are in different components of $\mathbb{R}^2 \setminus Y_1$. Repeat the argument, cutting Y_1 into two pieces (corresponding to cutting the half interval $[0, 1/2]$ or $[1/2, 1]$ into equal pieces). For one of the two pieces, say Y_2, by the same argument, the two points P_0 and P_1 are still in different components of $\mathbb{R}^2 \setminus Y_2$. Continuing in this way, we get a nested sequence of subsets

$$Y \supset Y_1 \supset Y_2 \supset Y_3 \supset \ldots \supset Y_n \supset \ldots$$

with the property that P_0 and P_1 are in different components of $\mathbb{R}^2 \setminus Y_n$ for all n, with the intersection of all these subsets Y_n being a single point P in Y.

Since the complement of a point is connected, there is a path from P_0 to P_1 in $\mathbb{R}^2$ that doesn't pass through P. Some neighborhood N of

P is disjoint from this path, and, for large n, Y_n is contained in N. But this forces P_0 and P_1 to be in the same component of $\mathbb{R}^2 \setminus Y_n$, a contradiction. This finishes the proof of Theorem 5.11, and hence of the full Jordan curve theorem. □

5d. Applications and Variations

The same ideas can be used to calculate the number of connected components of the complements of many other subsets in the plane. For many of these variations, we need the following generalization of Proposition 5.1, which will be proved in Chapter 9 (see §9c and Lemma 9.1).

> ($*$) Let K be a compact, nonempty subset of the plane. (a) If K is connected, then $H^1(\mathbb{R}^2 \setminus K)$ is one-dimensional, generated by $[\omega_P]$ for any $P \in K$. (b) If K is not connected, and P and Q are in different components of K, then $[\omega_P]$ and $[\omega_Q]$ are linearly independent in $H^1(\mathbb{R}^2 \setminus K)$; if K has exactly two connected components, then $[\omega_P]$ and $[\omega_Q]$ form a basis for $H^1(\mathbb{R}^2 \setminus K)$.

Exercise 5.13. Show that, if A and B are compact connected subsets in the plane such that $A \cap B$ is not connected (and not empty), then $\mathbb{R}^2 \setminus (A \cup B)$ is not connected.

Exercise 5.14. Show that if Y is a subset of the plane homeomorphic to a closed rectangle or a closed disk, then the complement is connected.

Exercise 5.15. Show that if X is a subset of the plane homeomorphic to a closed annulus, then the complement has two connected components.

Exercise 5.16. Show that if X is a subset of the plane homeomorphic to a figure 8, or a "theta" Θ, then the complement has three connected components.

Here is a simple application of the Jordan curve theorem. Let D be a closed disk, D° its interior, and C its boundary circle.

Proposition 5.17. *Let $f\colon D \to \mathbb{R}^2$ be a continuous, one-to-one mapping. Then $\mathbb{R}^2 \setminus f(C)$ has two connected components, which are*

$$f(D^\circ) \quad \text{and} \quad \mathbb{R}^2 \setminus f(D).$$

In particular, $f(D^\circ)$ is an open subset of the plane.

Proof. Recall that a continuous, one-to-one mapping on a compact set such as D or C must be a homeomorphism onto its image. So the Jordan curve theorem applies to the image of C, and its complement has two connected components. The first displayed set is connected since it is the continuous image of a connected space, and the second is connected by Exercise 5.14. They are disjoint, and their union is $\mathbb{R}^2 \setminus f(C)$. It follows immediately that they must be the two components of $\mathbb{R}^2 \setminus f(C)$. In particular, since the components of an open subset of the plane are open, it follows that $f(D^\circ)$ is open. □

The following is another intuitively "obvious" result that is not so easy to prove by hand (try it!):

Corollary 5.18 (Invariance of Domain). *If U is an open set in the plane, and $F\colon U \to \mathbb{R}^2$ is a continuous, one-to-one mapping, then $F(U)$ is an open subset of $\mathbb{R}^2$, and F is a homeomorphism of U onto $F(U)$.*

Proof. Take any P in U, and a closed disk D containing P and contained in U. By the proposition, the image of the interior of the disk must be open. This gives an open neighborhood of the image point in $F(U)$, which implies that $F(U)$ is open, and that F is a homeomorphism of U with $F(U)$. □

In particular, if two subsets of the plane are homeomorphic, and one is an open subset, the other must also be open. So if one is a domain (a connected open subset), the other must be as well.

Another application is another proof of a result from Chapter 4.

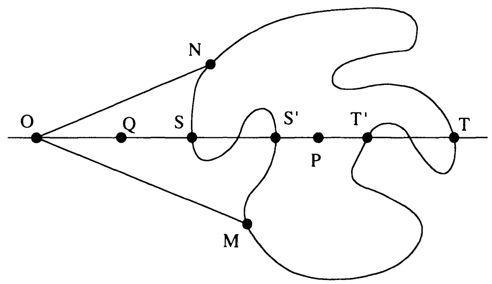
Corollary 5.19. *There is no subset of the plane that is homeomorphic to a two-sphere.*

Proof. Suppose $f\colon S \to \mathbb{R}^2$ maps the sphere homeomorphically onto a subset X of $\mathbb{R}^2$. Let A and B be the upper and lower closed hemispheres, with intersection C, and let $A^\circ = A \setminus C$ and $B^\circ = B \setminus C$. Note that A is homeomorphic to a disk, with C corresponding to the boundary. The complement of the image $f(C)$ of the circle in the plane has two connected components, and by the proposition (applied to the map from A to the plane), $f(A^\circ)$ must be one of them. Since $f(A^\circ)$ is contained in $f(A)$, it must be the bounded component. The same reasoning applies to the map from B, so $f(B^\circ)$ must also be the bounded component of the complement of $f(C)$. But then $f(A^\circ) = f(B^\circ)$, which contradicts the fact that f is one-to-one. □

This section concludes with a few related results, mostly in the form of exercises and problems, which can be sampled according to your interest or perseverance. The next proposition verifies that a Jordan curve winds once around each point inside. (This result will not be used elsewhere.)

Proposition 5.20. *Let C be a circle, $F: C \to \mathbb{R}^2$ a one-to-one, continuous mapping. Then $W(F,P) = \pm 1$ for P in the bounded component of $\mathbb{R}^2 \setminus F(C)$.*

Proof. In the course of the proof, we will make several assertions about winding numbers, leaving their proofs as exercises in using the properties proved in Chapter 3. The mapping F is a homeomorphism onto its image $X = F(C)$. By the Jordan curve theorem, the complement has two components, and we know the winding number is 0 for points in the unbounded component, and it is constant in the bounded component. So it suffices to find one point P with $W(F,P) = \pm 1$. Take a horizontal line L that has points of X on both sides of it. Take a point O on L so that X lies in the half plane to the right of O. Let S and T be the nearest and farthest points from O in $X \cap L$, and let M and N be points on X below and above L such that no other points of X lie on the line segments from O to M and from O to N.



The points M and N correspond by F to two points m and n of the circle C. Let A be the image by F of the counterclockwise arc from m to n, and let B be the image of the other clockwise arc from m to n. Let γ_A be a closed path starting at O, going along the segment to M, then traveling along A to N, and back to O along the segment;

define γ_B similarly using B in place of A. We choose these paths along A and B by using the mapping F, so that

$$W(\gamma_A, P) - W(\gamma_B, P) = W(F, P)$$

for all P not in X or the line segments from O to M or N.

The answer will depend on whether the point T is in A or in B. Suppose first that T is in A. Then $W(\gamma_B, T) = 0$, so $W(\gamma_B, T') = 0$ for all T' in $A \cap L$. For a point Q on the line between O and S, $W(\gamma_B, Q) = 1$. It follows that S is not in A, so it must be in B. Since $W(\gamma_A, Q) = 1$, we must have $W(\gamma_A, S) = 1$, and therefore $W(\gamma_A, S') = 1$ for all S' in $B \cap L$. Let S' be the point in $B \cap L$ that is farthest from O, and let T' be the point in $A \cap L$ that is closest to O, and let P be any point on L between S' and T'. Then

$$W(F, P) = W(\gamma_A, P) - W(\gamma_B, P) = W(\gamma_A, S') - W(\gamma_B, T') = 1.$$

If T is in B, a similar argument shows that $W(F, P) = -1$. □

The proof gives a criterion to tell whether the winding number is $+1$ or -1: it is $+1$ when four points M, N, S, and T chosen as in the proof have the same relative position as the four corresponding points of the circle. (This proof, from Ahlfors (1979), also shows directly, without the Jordan curve theorem, that the complement of X has at least two components.)

It is a fact that if X is a subset of $\mathbb{R}^2$ homeomorphic to a circle, then the bounded component of the complement is homeomorphic to an open disk, and the unbounded component is homeomorphic to a disk minus a point. In fact, the Riemann mapping theorem of complex analysis implies that any connected plane domain with $H^1U = 0$ is analytically isomorphic to (in particular, diffeomorphic to) an open disk. Even more is true: any homeomorphism of S^1 with X can be extended to a homeomorphism from $\mathbb{R}^2$ onto $\mathbb{R}^2$, mapping the inside and outside of the circle to the two components of $\mathbb{R}^2 \setminus X$. This last is called the Schoenflies theorem. For elementary (but not so simple) proofs, see Newman (1939). This is a case where the analogue in higher dimensions is more complicated: a homeomorphism of S^2 in $\mathbb{R}^3$, or of S^{n-1} in $\mathbb{R}^n$ for $n \geq 3$, will separate the space into two connected components, as we will see in Chapter 23, but the inside need not be homeomorphic to an open disk; for a picture of a counterexample, "Alexander's horned sphere," see Hocking and Young (1988). However, if the embedding extends to an embedding of a product of S^{n-1} with an interval into $\mathbb{R}^n$, then this wild behavior cannot occur (see Bredon (1993), §19).

Problem 5.21. Prove the following strong form of Euler's theorem. Let X be a subset of the plane that is a union of $v \geq 1$ points and $e \geq 0$ edges. The edges are assumed to be images of continuous maps from $[0, 1]$ to $\mathbb{R}^2$, each of which maps 0 and 1 into the set of vertices, and maps the open interval $(0, 1)$ one-to-one into the complement of the set of vertices. In addition, these open edges are assumed to be disjoint. Suppose X has k connected components. Show that $\mathbb{R}^2 \setminus X$ has $f = e - v + k + 1$ connected components, i.e.,

$$v - e + f = 2 + (k - 1).$$

You may enjoy comparing your argument in the last problem with that given in Rademacher and Toeplitz (1957), §12, as well as the applications given there to the problem of coloring maps. Can you spot where they make implicit assumptions amounting to what we proved in this chapter?

Problem 5.22. Show that the following graphs cannot be embedded in the plane: (i) the graph with vertices $P_1, P_2, P_3, Q_1, Q_2, Q_3$, with an edge between each P_i and each Q_j; and (ii) the graph with five vertices, and an edge between each pair of distinct vertices. (It is a theorem of Kuratowski that any finite graph not containing a subgraph homeomorphic to one of these two examples *can* be embedded in the plane.)

Problem 5.23. Let U be any connected open set in the plane. (a) Show that if $X \subset U$ is homeomorphic to a closed interval, then $U \setminus X$ is connected. (b) Show that if $X \subset U$ is homeomorphic to a circle, then $U \setminus X$ has two connected components. (c) If $X \subset U$ is a graph as in Problem 5.1, show that $U \setminus X$ has $e - v + k + 1$ connected components.

Exercise 5.24. Show that Theorem 5.11, the Jordan curve theorem (without mention of bounded or unbounded components), and the result of Problem 5.21 remain valid when $\mathbb{R}^2$ is replaced by a sphere.

Problem 5.25. Find two graphs in a sphere that are homeomorphic, but such that there is no homeomorphism of the sphere taking one onto the other.

Problem 5.26. Show that there is no one-to-one continuous map from a Moebius band into the plane.

Problem 5.27. Suppose X is a subset of the plane homeomorphic to a circle, and P_1 and P_2 are points in the complement that are joined by a path that crosses X n times. Show that P_1 and P_2 are in the same component of the complement if n is even, and the opposite component if n is odd. (A complete answer should include a precise definition of what it means for a path to cross X at a point!)

It is again an excellent project to speculate on the higher-dimensional generalizations of the results of this chapter. For example, if you assume the fact that if X is any subset of $\mathbb{R}^n$ homeomorphic to a sphere S^{n-1} (resp. a ball D^n), then $\mathbb{R}^n \setminus X$ has two (resp. one) connected components, can you state and prove the invariance of domain for open sets in $\mathbb{R}^n$?

Problem 5.28. A topological *surface with boundary* is defined to be a Hausdorff space such that every point P has a neighborhood homeomorphic either to the open disk $D^\circ = \{(x, y) \colon x^2 + y^2 < 1\}$ or the half disk $\{(x, y) \in D^\circ \colon y \geq 0\}$, with P corresponding to the origin; P is an *interior* or *boundary* point according to which case occurs. (a) Show that this notion is well defined: a point cannot be both an interior point and a boundary point. (b) Show that homeomorphic surfaces have homeomorphic boundaries (so the Moebius band is not homeomorphic to a cylinder).

CHAPTER 6

Homology

6a. Chains, Cycles, and H_0U

As we have seen, it frequently happens that one wants to compute winding numbers or integrals along a succession of paths, counting some positively and some negatively. For example, the integral around the boundary of a rectangle is the sum of integrals over two of its sides, minus the sum of the integrals over the other two sides. In this section we formalize these ideas, by introducing the notion of a 1-*chain*. A 1-chain γ in U is an expression of the form

$$\gamma = n_1\gamma_1 + n_2\gamma_2 + \ . \ . \ . + n_r\gamma_r,$$

where each γ_i is a continuous path in U, and each n_i is an integer. For simplicity, so we will not have to mention the interval each path might be defined on, we will take all paths from now on to be defined on the unit interval $[0, 1]$. The paths that are *constant*, that is, that map $[0, 1]$ to one point of U, can be ignored in the present story. For example, their winding numbers are zero, all integrals over them are zero. We will agree that if we meet a sum $\Sigma n_i\gamma_i$ where some of the γ_i are constant paths, we simply throw away any constant paths that occur. Another way to say this is that we identify two expressions $\Sigma n_i\gamma_i$ and $\Sigma n_i'\gamma_i'$ if their difference $\Sigma n_i\gamma_i - \Sigma n_i'\gamma_i'$ is a linear combination of constant paths.

We will make this notion more precise in a moment, but for now we note that, whatever it means, it is clear how we should define the winding number of a 1-chain γ with respect to P:

$$W(\gamma, P) = n_1 W(\gamma_1, P) + \ldots + n_r W(\gamma_r, P);$$

this will be defined provided P is not in the support of any of the paths γ_i. In Chapter 9 we will also define the integral of a closed 1-form along a path, and it will extend additively in the same way to integrals over 1-chains. Two 1-chains should be regarded as the same when each path occurs with the same multiplicity in both 1-chains. A 1-chain will have a unique expression as shown, provided the paths γ_i are all taken to be distinct and nonconstant, and all the coefficients are taken to be nonzero. (There is also the *zero* 1-chain, written $\gamma = 0$, which has no paths at all.)

To make this precise, define a *1-chain* in U to be a function that assigns to every nonconstant path in U some integer, with the property that the function is zero for all but a finite number of paths. If $\gamma_1, \ldots, \gamma_r$ are the paths for which the value is not zero, and the value of this function on γ_i is n_i, we write the 1-chain as $n_1\gamma_1 + \ldots + n_r\gamma_r$. From this definition it is clear how to add and subtract 1-chains: one just adds or subtracts the corresponding values on each path, or the coefficients in such expressions. In this way the 1-chains form an abelian group,[4] with the operation in the group written additively. Any path γ is identified with the 1-chain $1 \cdot \gamma$, the corresponding function taking the value 1 on γ and 0 on all other paths.

In practice, we will not use this "functional" terminology, but just write 1-chains as formal linear combinations of paths. Either way, specifying a 1-chain is the same as specifying a finite number of paths, and assigning an integer to each. It should be emphasized that in this definition two paths are the same only if defined by exactly the same mapping. However, the following problem indicates some common variations that are possible.

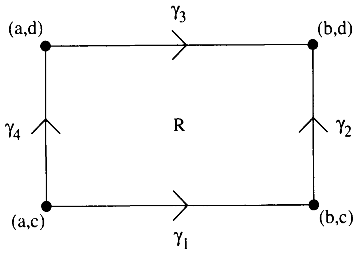
Problem 6.1. Call two paths equivalent if they differ by a monotone increasing reparametrization. Show that this is an equivalence relation. Show how to define an abelian group of 1-chains by using equivalence classes of paths instead of paths. Do the same if paths are called equivalent when they are homotopic with the same endpoints.

[4] If you are familiar with the language from algebra, the chains are elements of the free abelian group on the set of nonconstant paths. See Appendix C for more about free abelian groups.

Some particular paths and 1-chains will be important. For any two points P and Q, the *straight path* from P to Q will be the path

$$\gamma(t) \;=\; P + t(Q-P) = (1-t)P + tQ, \qquad 0 \le t \le 1.$$

If R is a bounded rectangle with sides parallel to the axes, the *boundary* ∂R of R is the 1-chain $\gamma_1 + \gamma_2 - \gamma_3 - \gamma_4$, where each γ_i is the straight path shown:



If $R = [a,b] \times [c,d]$, then γ_1 is the straight path from (a,c) to (b,c), γ_2 from (b,c) to (b,d), γ_3 from (a,d) to (b,d), and γ_4 from (a,c) to (a,d).

If D is a disk of radius r around a point P, the *boundary* $\gamma = \partial D$ is the counterclockwise path around the circle

$$\gamma(t) \;=\; P + (r\cos(2\pi t), r\sin(2\pi t)), \qquad 0 \le t \le 1.$$

We want to define the notion of a *closed* 1-*chain*, also called a 1-*cycle*. This should mean that each point occurs as many times as an initial point as it does as a final point, counting multiplicities correctly. For example, any closed path, such as the preceding ∂D, is a closed 1-chain, and the boundary ∂R of a rectangle is closed. The precise definition is as follows. For $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$, if γ_i is a path from P_i to Q_i, we define γ to be a *closed* 1-*chain* if for each point T occurring as a starting or ending point of any γ_i,

$$\sum_{P_i=T} n_i \;=\; \sum_{Q_j=T} n_j .$$

Exercise 6.2. Verify that the sum or difference of closed 1-chains is closed.

Define a 0-*chain* in U to be a formal finite linear combination $m_1P_1 + \ldots + m_sP_s$ of points in U, with integer coefficients. Precisely, it is a function from U to $\mathbb{Z}$ that is zero outside a finite set (or an element of the free abelian group on the set of points of U). In practice it means to pick out a finite set of points and assign a positive or negative multiplicity to each. For any 1-chain $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$, where $\gamma_i\colon [0, 1] \to U$ is a path, define the *boundary* $\partial\gamma$ of γ to be the 0-chain

$$\partial\gamma = n_1(\gamma_1(1) - \gamma_1(0)) + \ldots + n_r(\gamma_r(1) - \gamma_r(0)).$$

The preceding definition of a closed 1-chain can be said simply in this language: a 1-chain is closed exactly when its boundary is zero.

Let Z_0U be the group of 0-chains. A 0-chain ζ is called a *0-boundary* if there is a 1-chain γ such that $\zeta = \partial\gamma$. These 0-boundaries form a subgroup of Z_0U which is denoted by B_0U. For example, if P and Q are in the same component of U, then the 0-chain $Q - P$ is in B_0U, since it is the boundary of any path from P to Q. The quotient group Z_0U/B_0U is called the *0th homology group* of U, and is denoted H_0U:

$$H_0U = Z_0U/B_0U.$$

We will see that although the groups Z_0U and B_0U are large (even uncountable), the quotient group is small: it simply measures how many connected components U has:

Proposition 6.3. *The group H_0U is canonically isomorphic to the free abelian group on the set of path-connected components of U.*

Proof. Let F be the free abelian group on the set of path-connected components of U. The map that takes a point to the path component containing it determines a surjective homomorphism from Z_0U to F. We claim that the kernel of this homomorphism is exactly the group of boundaries B_0U. This will conclude the proof, since such a homomorphism determines a canonical isomorphism of Z_0U/B_0U with F (see Appendix C). Any boundary is in the kernel, since the endpoints of a path must be in the same component. Conversely, if a 0-cycle is in the kernel, the total of the coefficients appearing in front of points in any given component must be zero. Such a 0-cycle can be written (not necessarily uniquely) in the form $\Sigma(Q_i - P_i)$, where, in each term, P_i and Q_i are in the same component. As we saw before the proof, such a 0-cycle is a boundary. □

We will use this proposition to determine the number of connected components of U, by finding other ways of calculating H_0U. If we show that H_0U has rank n, we will know that U has exactly n connected components. This depends on the algebraic fact that a free abelian group has a well-defined rank; this is proved in Appendix C.

(Alternatively, one could replace all the integer coefficients in all our 0-chains and 1-chains by real numbers. Then we would find that H_0U is a real vector space of dimension n, where n is the number of components, and appeal to the fact that a vector space has a well-defined dimension.)

There is a homomorphism from H_0U to the integers $\mathbb{Z}$, defined by the map that takes each connected component of U to 1. In other words, it takes the class of a 0-cycle $\zeta = \Sigma n_i P_i$ to the sum Σn_i of the coefficients. This is called the *degree* homomorphism. It is an isomorphism exactly when U is connected.

6b. Boundaries, H_1U, and Winding Numbers

The group of 1-chains on U is denoted C_1U. The subgroup of closed 1-chains, or 1-cycles, is denoted Z_1U. There are some closed 1-chains in U, called 1-*boundaries*, that play a particularly simple role. They will turn out to be exactly those 1-chains for which winding numbers around points not in U vanish, and for which all integrals of closed 1-forms in U also vanish. These come from boundaries of continuous mappings Γ from a square $R = [0, 1] \times [0, 1]$ into U. For such a mapping, define the 1-chain $\partial\Gamma$ by the formula

$$\partial\Gamma = \gamma_1 + \gamma_2 - \gamma_3 - \gamma_4,$$

where γ_1, γ_2, γ_3, and γ_4 are the paths obtained by restricting Γ to the four sides of the square, as in §3b. (We should note here that one or more of these four paths γ_i could be constant paths, in which case we omit them from the formula for $\partial\Gamma$.) We call a 1-chain γ a *boundary*, or a *boundary* 1-*chain*, or 1-*boundary*, in U if it can be written as a finite linear combination (with integer coefficients) of boundaries of such maps on rectangles. Two closed 1-chains in U are *homologous* if the difference between them is a boundary in U.

We will need to know that some other 1-chains are boundaries. The following lemma considers what happens when one reparametrizes, subdivides, or deforms a path.

Lemma 6.4. (*a*) *Let* $\gamma\colon [0, 1] \to U$ *be a path. Let* $\varphi\colon [0, 1] \to [0, 1]$ *be a continuous function. If* $\varphi(0) = 0$ *and* $\varphi(1) = 1$, *then* $\gamma - \gamma\circ\varphi$ *is a boundary in* U; *if* $\varphi(0) = 1$ *and* $\varphi(1) = 0$, *then* $\gamma + \gamma\circ\varphi$ *is a boundary.*

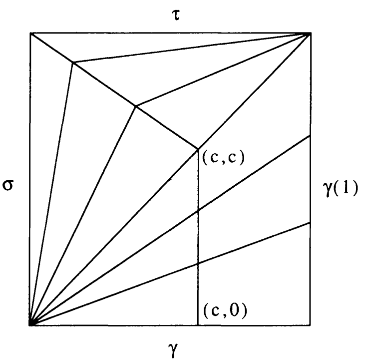
(*b*) *Let* $\gamma\colon [0, 1] \to U$ *be a path, let* $0 \le c \le 1$, *and let* σ *and* τ *be* *the restriction of* γ *to* $[0,c]$ *and* $[c,1]$, *but scaled to be defined on the unit interval, i.e.*,

$$\sigma(t) = \gamma(c \cdot t), \quad 0 \le t \le 1; \qquad \tau(t) = \gamma(c + (1-c)\cdot t), \quad 0 \le t \le 1.$$

Then $\gamma - \sigma - \tau$ *is a boundary in* U.

(*c*) *If* γ *and* δ *are paths in* U *that are homotopic, either as paths with the same endpoints, or as closed paths, then* $\gamma - \delta$ *is a boundary in* U.

Proof. Part (a) was proved in the course of proving Corollary 3.9. Part (b) is trivial if $c = 0$ or $c = 1$. Otherwise, we construct a mapping Γ: $[0,1] \times [0,1] \to U$ as indicated in the following diagram:



A little scratch-work in analytic plane geometry produces the formula for Γ:

$$\Gamma(t,s) = \begin{cases} \gamma(t) & \text{if } s \le t, \\ \gamma(c \cdot s + (1-c)\cdot t) & \text{if } s \ge t. \end{cases}$$

Note that the two expressions for Γ agree where $s = t$, so they define a continuous mapping. Moreover, $\Gamma(t,0) = \gamma(t)$, $\Gamma(1,s) = \gamma(1)$, $\Gamma(t,1) = \tau(t)$, and $\Gamma(0,s) = \sigma(s)$, so $\partial\Gamma = \gamma - \tau - \sigma$, which proves (b). Part (c) follows directly from the definition. □

Exercise 6.5. Show that if U is starshaped, then every closed 1-chain on U is a 1-boundary.

The 1-boundaries form a subgroup, denoted B_1U, of the group Z_1U of 1-cycles. The quotient group Z_1U/B_1U is called the *first homology group* of U, and is denoted H_1U. One of our aims in this book will be to study and use this group. We will see that, as the 0th group

H_0U measures the simplest topological fact about U—how many connected components it has—the 1st group H_1U measures how many "holes" there are in U. Two closed 1-chains are homologous exactly when they have the same image in H_1U, in which case we say that they define the same *homology class*.

The support $\mathrm{Supp}(\gamma)$ of a 1-chain γ in the plane is the union of the supports of the paths that appear in γ with nonzero coefficients. If P is a point not in the support, we define the *winding number* of γ around P to be

$$W(\gamma, P) \;=\; n_1 W(\gamma_1, P) + \ldots + n_r W(\gamma_r, P),$$

where $\gamma = n_1\gamma_1 + n_2\gamma_2 + \ldots + n_r\gamma_r$.

Proposition 6.6. *If γ is a closed 1-chain, then, for any P not in the support of γ, $W(\gamma, P)$ is an integer.*

Proof. Let $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$ be any 1-chain, with γ_i a path. For each point T that occurs as an endpoint of any of the paths γ_i, choose an angle ϑ_T for T with respect to P. (Such an angle is measured counterclockwise from a horizontal line to the right from P; it is determined only up to adding multiples of 2π.) If γ_i is a path from P_i to Q_i, then $W(\gamma_i, P) = (1/2\pi)(\vartheta_{Q_i} - \vartheta_{P_i}) + N_i$ for some integer N_i. Therefore,

$$\begin{aligned} W(\gamma, P) \;&=\; \sum_{i=1}^{r} n_i\left(\frac{1}{2\pi}(\vartheta_{Q_i} - \vartheta_{P_i}) + N_i\right) \\ &=\; \frac{1}{2\pi}\sum_{i=1}^{r} n_i(\vartheta_{Q_i} - \vartheta_{P_i}) + \sum_{i=1}^{r} n_i N_i . \end{aligned}$$

Suppose $\partial\gamma = \Sigma_{i=1}^{r} n_i(Q_i - P_i) = m_1T_1 + \ldots + m_sT_s$. Then we have

$$W(\gamma, P) \;=\; \frac{1}{2\pi}(m_1\vartheta_{T_1} + \ldots + m_s\vartheta_{T_s}) + \sum_{i=1}^{r} n_i N_i .$$

In particular, if γ is closed, i.e., $\partial\gamma = 0$, then the first sum vanishes, so $W(\gamma, P)$ is an integer. □

Lemma 6.7. *If γ is a 1-boundary in $\mathbb{R}^2 \setminus \{P\}$, then $W(\gamma, P) = 0$. If two 1-chains differ by a 1-boundary in $\mathbb{R}^2 \setminus \{P\}$, then they have the same winding number around P.*

Proof. If $\gamma = \partial\Gamma$ is the boundary of a map from $[0,1] \times [0,1]$ into $\mathbb{R}^2 \setminus \{P\}$, Theorem 3.6 implies that $W(\gamma, P) = 0$. For a general boundary $\gamma = \Sigma n_i(\partial\Gamma_i)$, $W(\gamma, P) = \Sigma n_i W(\partial\Gamma_i, P) = 0$. □

Proposition 6.8. *If γ is a closed 1-chain on $\mathbb{R}^2$, then the function $P \mapsto W(\gamma, P)$ is constant on connected components of $\mathbb{R}^2 \setminus Supp(\gamma)$, and vanishes on the unbounded component.*

Proof. To show that the function is locally constant, it suffices to show that it is constant on a disk D about a point P that does not meet the support of γ. We want to show that $W(\gamma, P) = W(\gamma, Q)$, with Q a point of D. Let v be the vector from P to Q. Let $\gamma = \Sigma_{i=1}^{r} n_i\gamma_i$, with each γ_i a path. We know from Exercise 3.4 that $W(\gamma, P) = W(\gamma + v, P + v) = W(\gamma + v, Q)$, where $\gamma + v$ is the 1-chain $\Sigma_{i=1}^{r} n_i(\gamma_i + v)$, so it is enough to show that

$$W(\gamma, Q) \;=\; W(\gamma + v, Q).$$

By the lemma, this follows if we verify that the difference of the 1-chains γ and $\gamma + v$ is a 1-boundary in $\mathbb{R}^2 \setminus \{Q\}$. Define mappings Γ_i from $[0, 1] \times [0, 1]$ to $\mathbb{R}^2 \setminus \{Q\}$ by the formula

$$\Gamma_i(t, s) \;=\; \gamma_i(t) + s \cdot v.$$

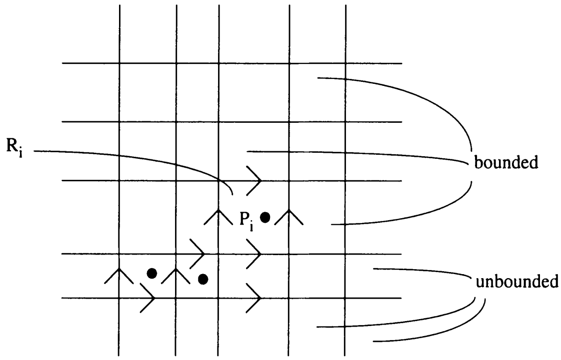
The boundary of Γ_i has the paths γ_i and $\gamma_i + v$ on the bottom and top, and straight line paths from endpoints of γ_i to their translations by v, along the sides. The fact that γ is closed means that these straight line paths from the sides cancel in the sum $\Sigma_{i=1}^{r} n_i(\partial\Gamma_i)$. Therefore,

$$\sum_{i=1}^{r} n_i(\partial\Gamma_i) \;=\; \sum_{i=1}^{r} n_i\gamma_i - \sum_{i=1}^{r} n_i(\gamma_i + v) \;=\; \gamma - (\gamma + v).$$

This shows that $\gamma - (\gamma + v)$ is a boundary in $\mathbb{R}^2 \setminus \{Q\}$, and concludes the proof that the function is locally constant. On the unbounded component, we may take the point P far to the left of the support of γ, so that there is one angle function ϑ defined on all the paths occurring in γ. With $\gamma = \Sigma_{i=1}^{r} n_i\gamma_i$, and γ_i a path from P_i to Q_i, we have $W(\gamma, P) = (1/2\pi)\Sigma_{i=1}^{r} n_i(\vartheta(Q_i) - \vartheta(P_i))$, which is zero since the boundary of γ is zero. □

6c. Chains on Grids

A *grid G* will be a finite union of lines in the plane, each parallel to x-axis or the y-axis, with at least two horizontal and two vertical lines. These lines cut the plane into a finite number of rectangular regions, some bounded and some unbounded. By a *rectangular* 1-chain for a given grid we shall mean a 1-chain μ of the form $\mu = n_1\sigma_1 + \ldots + n_r\sigma_r$, where each σ_j is a straight path along one of the sides of the bounded rectangles, from left to right or from bottom to top.



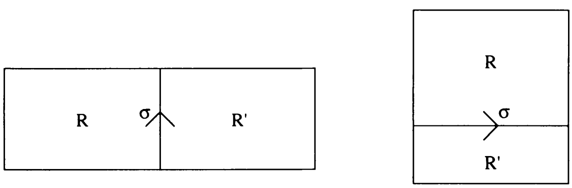
Let $R_1, \ldots, R_r$ be the bounded rectangles for the grid G, numbered in any order, and choose a point P_i in the interior of R_i for each i. The next lemma proves an elementary but important fact: a closed rectangular 1-chain is completely determined by its winding numbers about these points.

Lemma 6.9. *If μ is a closed rectangular 1-chain for G, then*

$$\mu = n_1\partial R_1 + \ldots + n_r\partial R_r,$$

where $n_i = W(\mu, P_i)$.

Proof. Since the winding number of ∂R_i around P_j is 1 if $i = j$ and 0 otherwise, the winding number of each side of the displayed equation around each P_j is the same. Let $\tau = \mu - \Sigma n_i\partial R_i$ be the difference, which has winding number zero around each P_j. We must show that τ is zero as a 1-chain. Suppose an edge σ occurs in τ with nonzero coefficient m. Suppose that σ is a vertical or horizontal line between two rectangles R and R' of the grid, with R to the left of or above R'. Assume first that R is bounded:



The trick is to consider the closed 1-chain $\tau' = \tau - m \cdot \partial R$. Let P and

P' be interior points in R and R', respectively. Since $W(\partial R, P) = 1$ and $W(\partial R, P') = 0$, we have

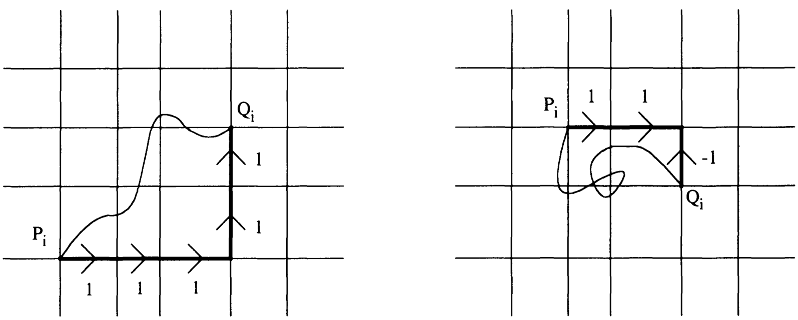
$$W(\tau', P) = W(\tau, P) - m \cdot W(\partial R, P) = -m,$$
$$W(\tau', P') = W(\tau, P') - m \cdot W(\partial R, P') = 0.$$

But the edge σ does not appear in τ', so P and P' belong to the same connected component of the complement of the support of τ', which implies by Proposition 6.8 that $W(\tau', P) = W(\tau', P')$, a contradiction. If R is unbounded, and R' is bounded, the argument is similar, using $\tau' = \tau + m \cdot \partial R'$. □

Next we need an approximation lemma that will assure that, for the purposes of winding numbers and integration, all paths and 1-chains can be replaced by rectangular 1-chains.

Lemma 6.10. *Let γ be any 1-chain in an open set U. Then there is a grid G, and a rectangular 1-chain μ for G, with the support of μ contained in U, such that $\gamma - \mu$ is a 1-boundary in U. If γ is closed, then μ is also closed.*

Proof. By Lemma 6.4(b) we can subdivide any of the paths that occur in γ, and the difference between γ and the 1-chain with subdivided paths will be a boundary. Using the Lebesgue lemma as usual to subdivide, we may therefore assume that each path γ_i occurring in γ maps $[0, 1]$ into some open rectangle U_i contained in U. Let P_i and Q_i be the starting and ending points of γ_i. Take any grid G that has a vertical and horizontal line passing through each P_i and each Q_i. Let μ_i be a rectangular 1-chain for G that goes from P_i to Q_i, involving only edges on the closed rectangle with corners at P_i and Q_i; in particular, the boundary of μ_i is $Q_i - P_i$. (Note that μ_i is in U_i, so μ_i is a chain in U.)



It suffices to verify that $\gamma_i - \mu_i$ is a boundary in U_i, for if $\gamma = \Sigma n_i \gamma_i$, then $\mu = \Sigma n_i \mu_i$ will be the required rectangular 1-chain, with $\gamma - \mu = \Sigma n_i(\gamma_i - \mu_i)$ a boundary. But since $\gamma_i - \mu_i$ is a closed 1-chain on a starshaped open set U_i, this follows from Exercise 6.5. □

We are now ready to prove the main goal of this chapter: that the geometric condition for 1-cycles to be homologous in an open set is equivalent to the numerical condition of having the same winding number around all points outside the open set.

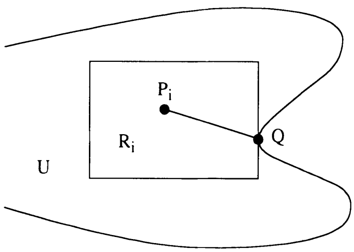
Theorem 6.11. *Suppose γ and δ are closed 1-chains on an open set U in the plane. Then the following are equivalent:*

(1) *γ and δ are homologous, i.e., $\gamma - \delta$ is a boundary in U; and*

(2) *$W(\gamma, P) = W(\delta, P)$ for all points P not in U.*

Proof. We saw in Lemma 6.7 that (1) implies (2). For the converse, by looking at $\tau = \gamma - \delta$, it suffices to show that if τ is a closed 1-chain such that $W(\tau, P) = 0$ for all $P \notin U$, then τ is a boundary. By Lemma 6.10, there is a closed rectangular 1-chain μ for some grid G so that $\tau - \mu$ is a boundary in U. By Lemma 6.9, $\mu = \Sigma n_i \partial R_i$, where $n_i = W(\mu, P_i)$.

We claim that $n_i = 0$ unless the entire closed rectangle R_i is contained in U. For if Q is a point of R_i that is not in U, then $W(\tau, Q) = 0$ by assumption, so $W(\mu, Q) = 0$ by Lemma 6.7. Since the straight line from P_i to Q lies in the complement of the support of μ, it follows from Proposition 6.8 that $W(\mu, P_i) = W(\mu, Q)$, so $n_i = W(\mu, P_i) = 0$.



It follows that $\mu = \Sigma n_i \partial R_i$, with each R_i contained in U, and such a 1-chain is visibly a boundary in U. So $\tau = (\tau - \mu) + \mu$ is also a boundary. □

For example, if $U = \mathbb{R}^2 \setminus \{P\}$ is the complement of a point, two 1-cycles are homologous in U exactly when their winding numbers around P coincide. In other words, the winding number gives an isomorphism

$$H_1(\mathbb{R}^2 \setminus \{P\}) \xrightarrow{\cong} \mathbb{Z}, \qquad [\gamma] \mapsto W(\gamma, P),$$

where $[\gamma]$ denotes the class of a 1-cycle γ in the homology group.

Exercise 6.12. Suppose $U = \mathbb{R}^2 \setminus \{P_1, \ldots, P_n\}$ is the complement of n points in the plane. Show that the mapping that takes a closed 1-chain γ to $(W(\gamma, P_1), \ldots, W(\gamma, P_n))$ determines an isomorphism of H_1U with the free abelian group $\mathbb{Z}^n$.

Exercise 6.13. State and prove the analogue of Theorem 6.11 when γ and δ are arbitrary 1-chains in U with the same boundary.

6d. Maps and Homology

If $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$ is a 1-chain in an open set U, with γ_i paths, and $F \colon U \to U'$ is a continuous mapping from U to another open set U', define $F_*\gamma$ to be the 1-chain in U' defined by

$$F_*\gamma = n_1(F \circ \gamma_1) + \ldots + n_r(F \circ \gamma_r).$$

F also maps 0-chains in U to 0-chains in U': $F_*(\Sigma m_i P_i) = \Sigma m_i F(P_i)$.

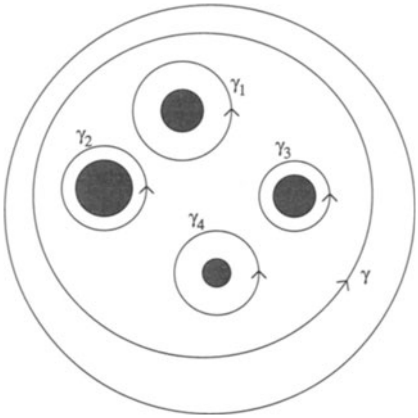
Exercise 6.14. Show that $\gamma \mapsto F_*\gamma$ is a homomorphism from the group of 1-chains in U to the group of 1-chains in U'. If γ is a closed 1-chain, show that $F_*\gamma$ is also closed. Show in fact that $F_*(\partial\gamma) = \partial(F_*\gamma)$ for any 1-chain. Show that if γ is a boundary in U, then $F_*\gamma$ is a boundary in U'.

From this exercise and Theorem 6.11, we deduce the following fact, which is not so obvious from the definition of the winding number:

Proposition 6.15. *If γ and δ are closed 1-chains in U with the same winding number around all points not in U, then $F_*\gamma$ and $F_*\delta$ are closed 1-chains in U' with the same winding number around all points not in U'.* □

For example, take U to be the region inside one disk D and outside a disjoint union of closed disks $A_1, \ldots, A_n$ contained in D. Let γ

be a circular path in U containing the disks A_i, and let γ_i be a circular path around A_i.



Since γ and $\Sigma\gamma_i$ have the same winding numbers around each point not in U, we conclude that for any $F\colon U \to U'$, and any $Q \notin U'$,

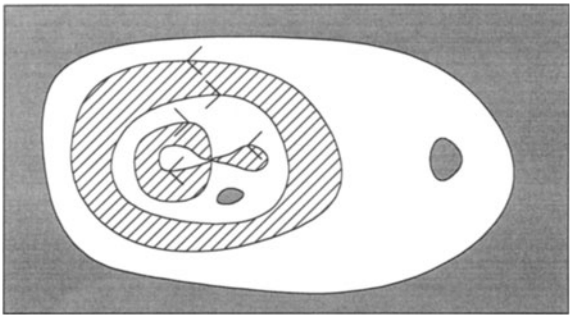
$$W(F_*\gamma, Q) = W(F\circ\gamma_1, Q) + \ldots + W(F\circ\gamma_n, Q).$$

Here is an application, which is a substantial generalization of Exercise 4.9.

Corollary 6.16. *Let $F\colon U \to \mathbb{R}^2$ be a continuous mapping, and suppose γ is a closed 1-chain in U such that $W(\gamma, P) = 0$ for all points P not in U. Let Q be a point in $\mathbb{R}^2$, not in $F(Supp(\gamma))$, such that $W(F_*\gamma, Q) \neq 0$. Then there is a point P in U such that $W(\gamma, P) \neq 0$ and $F(P) = Q$.*

Proof. Let $Z = F^{-1}(Q)$, a closed (conceivably empty) subset of U disjoint from Supp(γ). Apply the proposition to the open set $U^\circ = U \setminus Z$, the restriction of F to U°, and the 1-chains γ and $\delta = 0$. If the assertion of the corollary is false, then $W(\gamma, P) = 0$ for all P in Z, so $W(\gamma, P) = 0$ for all P not in U°, so γ is homologous to 0 in U°. But since Q is not in $F(U^\circ)$, the fact that $W(F_*\gamma, Q) \neq 0$ shows that $F_*\gamma$ is not homologous to 0 in $F(U^\circ)$, contradicting Proposition 6.15. $\square$

For example, with U the complement of the gray area, the 1-chain γ that is the sum of the paths shown has a nonzero winding number only around points in the striped area, so any point with $W(F_*\gamma, Q) \neq 0$ would be the image of a point from one of the striped regions.



Problem 6.17. Under the conditions of the corollary, assume that, at each point P of $F^{-1}(Q)$, the local degree of F at P is defined. Show that

$$W(F_*\gamma, Q) = \sum_{P \in F^{-1}(Q)} \deg_P(F) \cdot W(\gamma, P).$$

Problem 6.18. Let E be the closed set obtained from a closed disk D by removing the interiors of k disjoint disks $D_1, \ldots, D_k$ contained in D, for some $k \geq 0$. Let C be the boundary of D, and C_i the boundary of D_i. Suppose $F\colon E \to \mathbb{R}^2$ is a continuous mapping such that for each point P in any of the boundary circles, the vector from P to $P + F(P)$ is not tangent to that circle. Show that, if $k \neq 1$, there must be a point Q in E with $F(Q) = 0$. What about the case $k = 1$?

6e. The First Homology Group for General Spaces

The general definitions of this chapter make sense without change for any topological space, although, of course, one does not have winding numbers in arbitrary spaces. Any topological space X has abelian groups Z_0X of 0-chains, B_0X of 0-boundaries, with 0th homology group $H_0X = Z_0X/B_0X$, which is canonically isomorphic to the free abelian group on the set of path-connected components of X. Similarly, one has the abelian group C_1X of 1-chains on X, with the subgroups Z_1X of 1-cycles and 1-boundaries B_1X, and the 1st homology group $H_1X = Z_1X/B_1X$. There are no changes in the definitions, other than replacing a "U" by an "X."

Exercise 6.19. Show that a continuous mapping $F: X \to Y$ determines a homomorphism from Z_1X to Z_1Y taking B_1X to B_1Y.

This determines a homomorphisms of the quotient groups, denoted

$$F_*: H_1X \to H_1Y.$$

Exercise 6.20. Show that if $F: X \to Y$ and $G: Y \to Z$ are continuous, then $(G \circ F)_* = G_* \circ F_*$ as homomorphisms from H_1X to H_1Z. If F is the identity map on X, show that F_* is the identity map on H_1X.

The result of this exercise is expressed by saying that the diagram

$$\begin{tikzcd} H_1X \arrow[rr, "F_*"] \arrow[dr, "(G\circ F)_*"'] & & H_1Y \arrow[dl, "G_*"] \\ & H_1Z & \end{tikzcd}$$

commutes: starting with an element in the upper left group H_1X, mapping it to H_1Z by either route gives the same answer.

For example, if $Z = X$, and F and G are homeomorphisms that are inverses to each other, so that $G \circ F$ is the identity map of X, then $(G \circ F)_*$ is the identity map of H_1X, so the composite

$$H_1X \xrightarrow{F_*} H_1Y \xrightarrow{G_*} H_1X$$

is the identity map on H_1X. Similarly $F \circ G$ is the identity map on Y, so $G_* \circ F_*$ is the identity map on H_1Y. It follows that F_* and G_* are inverse isomorphisms between H_1X and H_1Y. In particular,

Proposition 6.21. *If X and Y are homeomorphic*, then H_1X *and* H_1Y *are isomorphic abelian groups.*

Similarly, if Y is contained in X, and $r: X \to Y$ is a continuous retract, the identity map on H_1Y must factor into a composite of homomorphisms $H_1Y \to H_1X \to H_1Y$. For example, if $H_1X = 0$, and $H_1Y \neq 0$, this shows that there can be no such retract.

Exercise 6.22. Compute H_1X for X a circle and for X a disk, and show again that a circle is not a retract of the disk it bounds.

Two continuous maps F and G from X to Y are *homotopic* if there is a continuous mapping $H: X \times [0,1] \to Y$ such that $F(x) = H(x, 0)$ and $G(x) = H(x, 1)$ for all x in X.

Proposition 6.23. *If F and G are homotopic maps from X to Y, then $F_* = G_*$.*

Proof. If $\gamma = \Sigma n_i \gamma_i$ is a 1-cycle on X, set $\Gamma_i(t, s) = H(\gamma_i(t), s)$. Then $F_*\gamma - G_*\gamma = \Sigma n_i \partial \Gamma_i$, the other terms canceling each other since γ is a cycle. □

A subspace Y of a space X is called a *deformation retract* if there is a continuous retract $r: X \to Y$ such that the identity map from X to X is homotopic to the map $i \circ r$, where i is the inclusion of Y in X. A space X is called *contractible* if it contains a point that is a deformation retract of X.

Exercise 6.24. (a) Show that the circle $S^1 \subset \mathbb{R}^2 \setminus \{0\}$ is a deformation retract. (b) Give an example of a retract that is not a deformation retract. (c) Show that any two maps from any space to a contractible space are homotopic. In particular, every point in a contractible space is a deformation retract of the space.

Exercise 6.25. Show that if Y is a deformation retract of X, then the map from H_1Y to H_1X determined by the inclusion of Y in X is an isomorphism. Show that $H_1(X) = 0$ if X is contractible.

Exercise 6.26. Show that $F: X \to Y$ determines a homomorphism from Z_0X to Z_0Y taking B_0X to B_0Y, and so a homomorphism, also denoted F_*, from H_0X to H_0Y. Verify the analogues of the assertions in Exercises 6.19 and 6.20, and Propositions 6.21 and 6.23.

If X is a subspace of the plane that is not open, one does have a notion of the winding number around points not in X, but the situation is more complicated, as the following problems indicate.

Problem 6.27 (For those who know the Tietze extension theorem). Suppose X and X' are closed subsets of the plane, and $F: X \to X'$ is a continuous mapping. Suppose γ and δ are two closed 1-chains on X such that $W(\gamma, P) = W(\delta, P)$ for all P not in X. Show that $W(F_*\gamma, P') = W(F_*\delta, P')$ for all P' not in X'.

Problem 6.28. Let X be a closed set in the plane. Show that if γ and δ are homologous 1-cycles on X, then $W(\gamma, P) = W(\delta, P)$ for all P not in X. Is the converse true?

PART IV
VECTOR FIELDS

A mapping from an open set in the plane to the plane can be regarded as a vector field, and winding numbers can be used to define the index of a vector field at a singularity. The ideas of Chapter 6 can be used to relate sums of indices to winding numbers around regions. This is applied to show that vector fields on a sphere must have singularities: one cannot comb the hair on a billiard ball. The same ideas are used in the next chapter to study more interesting surfaces. These chapters are inserted here to indicate some other interesting things one can do with winding numbers; a reader in a hurry to move on can skip to Part V or VI.

Chapter 8 sketches how some of the ideas we have studied in the plane and on the sphere can be studied on more general surfaces. It gives us a first chance to study some spaces other than plane regions and spheres. In particular, we see how the "global" topology of the surface puts restrictions on the "local" data of indices of a vector field. We use this to discuss the Euler characteristic of a surface. Some of the arguments in this section will depend on geometric constructions that will only be sketched, usually by drawing pictures showing how to deform one surface into another. Later in the chapter we discuss briefly what it would take to make these arguments rigorous, and later in the book we take up the study of surfaces more systematically.

CHAPTER 7

Indices of Vector Fields

7a. Vector Fields in the Plane

We want to look at continuous vector fields on an open set U in the plane, but allowing them to have a finite number of *singularities*. A singularity will be a point at which either the vector field is not defined, or a point where it is defined and is zero. A *vector field on U with singularities in Z* will therefore be a continuous mapping

$$V\colon\ U \setminus Z \to \mathbb{R}^2 \setminus \{0\},$$

where Z is a finite set in U. For vector fields arising from flow of a fluid, the singularities may arise from "sources," where fluid is entering the system, or "sinks," where it is leaving, or some other discontinuity.

Given such a vector field V, to each point P in U one can define an integer called the *index of V at P*, denoted $\mathrm{Index}_P V$. To do this, take a disk D_r of some radius r about P that does not meet Z at any point except (perhaps) P. Let C_r be the boundary of this disk. The restriction $V|_{C_r}$ of V to C_r is a mapping from this circle to $\mathbb{R}^2 \setminus \{0\}$, so it has a winding number (by §3d). Define the index to be this winding number:

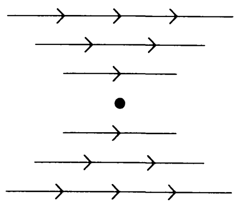
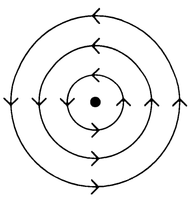
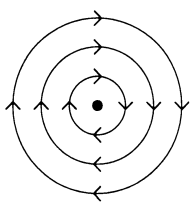
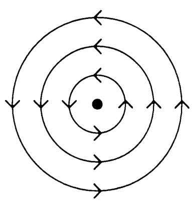
$$\mathrm{Index}_P V \;=\; W(V|_{C_r}, 0)\,.$$

Lemma 7.1. (*a*) *This definition is independent of choice of r.*

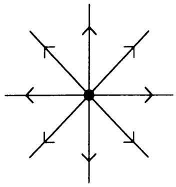
(*b*) *If P is not in Z, then* $\mathrm{Index}_P V = 0$.

Proof. The proof of (a) is the same as that of Lemma 3.28, and (b) follows from Proposition 3.20. □

From the definition, the index at P depends only on the restriction of V to an arbitrarily small neighborhood of P. Here are some examples of vector fields with singularities at the origin, with the corresponding indices; the calculations are left as exercises. Instead of drawing vectors at many points, it is more useful to draw some flow lines, i.e., curves that are tangent to the vector field at each point.

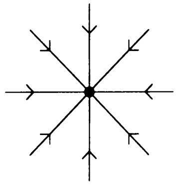
Vector field V(x, y)		*Index at* 0
(1) $(x^2+y^2, 0)$		0
(2) $(-y, x)$		1
(3) $(y, -x)$		1
(4) $\left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right)$		1

(5) (x,y)



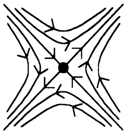
1

(6) $(-x,-y)$



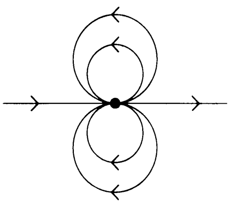
1

(7) (y,x)



-1

(8) $(x^2-y^2, 2xy)$



2

Note that (2) and (3) are opposite as vector fields, as are (5) and (6), but have the same index. Similarly, (4) is a positive multiple of (2). These are special cases of Exercise 7.4, which shows that the magnitude and sign of the vectors does not affect the index (which explains why the flow lines, without even their sense of direction, determine the indices).

Exercise 7.2. Construct, for each integer n, a vector field with a singularity at the origin with index n.

Exercise 7.3. Let V_0 and V_1 be continuous vector fields in a punctured neighborhood U of P, and suppose there is a continuous mapping $H: U \times [0,1] \to \mathbb{R}^2$ such that

$$H(Q \times 0) = V_0(Q), \qquad H(Q \times 1) = V_1(Q) \qquad \text{for all } Q \text{ in } U;$$
$$H(Q \times t) \neq 0 \qquad \text{for all } Q \text{ and all } 0 \leq t \leq 1.$$

Show that V_0 and V_1 have the same index at P.

Exercise 7.4. Show that if ρ is any continuous function defined in a punctured neighborhood of P that is always positive or always negative in this neighborhood, then the index at P of $\rho \cdot V$ is the same as the index at P of V.

The following proposition relates the behavior of a vector field along the boundary of a region to the indices of the vector field at singular points inside:

Proposition 7.5. *Let V be a vector field with singularities in U. Suppose γ is a closed 1-chain in U whose support does not meet the singular set Z of V, such that $W(\gamma, P) = 0$ for all P not in U. Then*

$$W(V_* \gamma, 0) \;=\; \sum_{P \in Z} W(\gamma, P) \cdot \mathrm{Index}_P V.$$

Proof. Let $Z = \{P_1, \ldots, P_r\}$, and let $D_1, \ldots, D_r$ be disjoint closed disks centered at the points $P_1, \ldots, P_r$, all contained in U. Let γ_i be the standard counterclockwise path around the boundary of D_i, and let $n_i = W(\gamma, P_i)$. Then γ and $n_1\gamma_1 + \ldots + n_r\gamma_r$ have the same winding number around every point outside $U \setminus Z$. It follows from Proposition 6.15 that

$$W(V_* \gamma, 0) \;=\; n_1 W(V \circ \gamma_1, 0) + \ldots + n_r W(V \circ \gamma_r, 0),$$

and this is the assertion to be proved. □

Corollary 7.6. *If $W(V_* \gamma, 0) \neq 0$, then V must have at least one nonvanishing index at a point P with $W(\gamma, P) \neq 0$.* □

The simplest case of the proposition is:

Corollary 7.7. *If U contains a closed disk D, and V has no singularities on the boundary circle C of D. Then*

$$W(V|_C, 0) \;=\; \sum_{P \in D} \mathrm{Index}_P V.$$

If $W(V|_C, 0) \neq 0$, V must have a singularity with nonvanishing index inside D. □

Problem 7.8. Generalize this discussion to allow the singularity set Z to be infinite but *discrete*, i.e., Z is a closed subset of U such that each P in Z has a neighborhood U_P in U such that $U_P \cap Z = \{P\}$. Show that the sum in Proposition 7.5 is automatically finite.

Problem 7.9. Let f be a $\mathscr{C}^\infty$ function defined in U, defining a gradient vector field $V = \mathrm{grad}(f)$ on U. A point P is a *critical point* for f if the gradient vanishes at P, and P is called *nondegenerate* if the "Hessian" at P,

$$\frac{\partial^2 f}{\partial x^2}(P)\cdot\frac{\partial^2 f}{\partial y^2}(P) - \left(\frac{\partial^2 f}{\partial x\,\partial y}(P)\right)^2$$

is not zero. (a) Show that, if P is a nondegenerate critical point, then

$$\mathrm{Index}_P(V) = \begin{cases} 1 & \text{if } f \text{ has a local maximum or minimum at } P, \\ -1 & \text{otherwise (when } P \text{ is a saddle point).} \end{cases}$$

(b) Suppose D is a disk in U with boundary C, and f has only nondegenerate critical points in D, with none on the boundary, and f is constant on C. Show that the number of local maxima plus the number of local minima is one more than the number of saddle points. "On a circular island, the number of peaks plus the number of valleys is one more than the number of passes."

7b. Changing Coordinates

Later we want to define the index of a vector field on a surface other than an open set in the plane. To do this, the essential point is to know that a change of coordinates does not change the index. This result, which is intuitively obvious from pictures of vector fields, takes some care to state and a little work to prove. To state it, suppose $\varphi\colon U \to U'$ is a diffeomorphism from one open set in the plane onto another; that is, φ is $\mathscr{C}^\infty$, one-to-one, and onto, and the inverse map $\varphi^{-1}\colon U' \to U$ is also a $\mathscr{C}^\infty$ mapping. At any point P in U, we have the Jacobian matrix

$$J_{\varphi,P} = \begin{bmatrix} \dfrac{\partial u}{\partial x}(P) & \dfrac{\partial u}{\partial y}(P) \\ \dfrac{\partial v}{\partial x}(P) & \dfrac{\partial v}{\partial y}(P) \end{bmatrix},$$

where $\varphi(x, y) = (u(x, y), v(x, y))$ in coordinates. This gives a linear mapping from vectors in $\mathbb{R}^2$ to vectors in $\mathbb{R}^2$ (see Appendix C). If V is a continuous vector field in U, define the vector field $\varphi_* V$ in U'

by the formula

$$(\varphi_* V)(P') = J_{\varphi,P}(V(P)),$$

where P is the point in U mapped to P' by φ, i.e., $P = \varphi^{-1}(P')$. If V has singularities in the set Z, V' will have singularities in $\varphi(Z)$.

Lemma 7.10. *With V and $\varphi_* V$ as above, then, for any P in U,*

$$\mathrm{Index}_{\varphi(P)}(\varphi_* V) = \mathrm{Index}_P V.$$

The proof of this lemma is given in Appendix D. We will also want to compare the indices of two different vector fields on a surface. The following lemma will be used to reduce to the case where they agree in a neighborhood of some point. We say that two vector fields V and W *agree* on a set A if $V(P) = W(P)$ for all P in A. The proof is also in Appendix D.

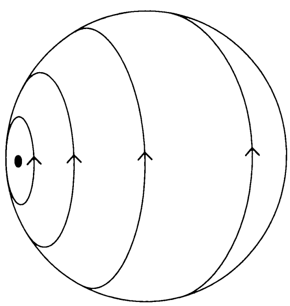
Lemma 7.11. *Suppose V and W are continuous vector fields with no singularities on an open neighborhood U of a point P. Let $D \subset U$ be a closed disk centered at P. Then there is a vector field $\tilde{V}$ with no singularities on U such that: (i) $\tilde{V}$ and V agree on $U \setminus D$; and (ii) $\tilde{V}$ and W agree on some neighborhood of P.*

7c. Vector Fields on a Sphere

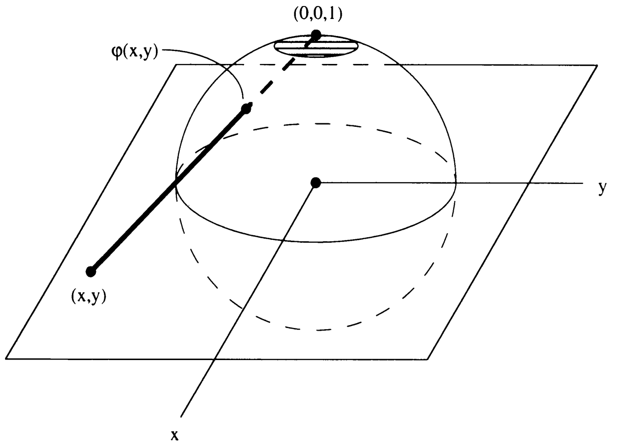
A vector field V on a sphere S assigns to each point P in S a vector $V(P)$ in the tangent space $T_P S$ to S at P, the mapping from P to $V(P)$ being continuous. If $S = S^2$ is the standard sphere, and $P = (x, y, z)$, the tangent space is

$$T_P S = \{(a, b, c) \colon (a, b, c) \cdot (x, y, z) = ax + by + cz = 0\}.$$

As before, we may allow a finite set Z of singularities. In fact, one of our goals is to show that any vector field on sphere *must* have singularities.



We want to flatten out the sphere, say by stereographic projection from a point on the sphere, so that the vector field determines a corresponding vector field on the plane, and we will use what we know about vector fields on the plane.



We need the following, whose proof is left as an exercise:

Lemma 7.12. *The inverse φ of this polar coordinate mapping takes (x, y) in the plane to*

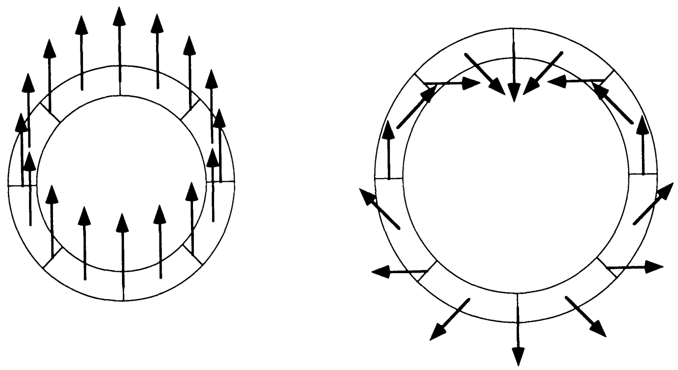
$$\varphi(x,y) = \left(\frac{2x}{x^2+y^2+1}, \frac{2y}{x^2+y^2+1}, \frac{x^2+y^2-1}{x^2+y^2+1}\right) \quad \text{in} \quad S^2.$$

The Jacobian matrix $J_{\varphi,P}$ of φ at $P = (x, y)$ maps $\mathbb{R}^2$ one-to-one onto the tangent space to S^2 at the point $\varphi(P)$. If V is a vector field on S^2, and φ^*V is the vector field on $\mathbb{R}^2$ defined by the equation

$$J_{\varphi,P}((\varphi^*V)(P)) = V(\varphi(P)),$$

then φ^*V is continuous at P if V is continuous at $\varphi(P)$.

Now suppose V is a continuous vector field with no singularities on S^2. Then φ^*V is a vector field on $\mathbb{R}^2$ with no singularities. Let C_r be a large circle centered at the origin in the plane. We know from Corollary 7.7 that the winding number of φ^*V around C_r must be zero. To get a contradiction, we must use the fact that V is continuous and nonzero at the north pole. For r large, C_r can be thought of as a small circle around the north pole. The winding number of φ^*V around such a circle is not zero, even though it comes from a vector field that is not zero in the disk near the north pole. This can be seen by unraveling what happens to a vector field on the sphere when the sphere is flattened out. Think of the vectors as attached to a small band, which is turned over:



The winding number of this vector field around this circle is 2, which shows that no such vector field can exist.

Problem 7.13. Give an analytic proof of the fact that this winding number is 2.

This shows more than the fact that a vector field on a sphere must have singularities. If V is a vector field on S^2 with singularities in a finite set Z that does not include the north pole, we can define the index $\mathrm{Index}_P V$ at a point P of Z to be the index of $\varphi^* V$ at the corresponding point $\varphi^{-1}(P)$. Then the proof shows that, for any such vector field V on S^2,

$$\sum_{P \in Z} \mathrm{Index}_P V = 2 .$$

In order to include the north pole in these considerations, we also consider a stereographic projection from the south pole. To make the orientations match in our two charts, we first reflect in the x-axis. So we define $\psi \colon \mathbb{R}^2 \to S^2$ by

$$\psi(x, y) = \left(\frac{2x}{x^2 + y^2 + 1}, \frac{-2y}{x^2 + y^2 + 1}, \frac{-x^2 - y^2 + 1}{x^2 + y^2 + 1} \right).$$

Exercise 7.14. Show that the composite $\varphi^{-1} \circ \psi$ takes $z = x + iy$ to $1/z$.

Now for any vector field V with a finite number of singularities on S^2, and any point P, we can define $\text{Index}_P V$ as either the index of φ^*V or of ψ^*V at the corresponding point. It follows from Lemma 7.10 that these indices agree, if both are defined. In fact, we could use a stereographic projection from any point; all the coordinate transformations as above are $\mathscr{C}^\infty$.

Proposition 7.15. *For any vector field with singularities V on S^2,*

$$\sum_{P \in Z} \text{Index}_P V = 2.$$

Proof. Instead of arguing as we did above, we can argue in two steps:

Step 1. There is a vector field V on S^2 with $\Sigma_{P \in Z} \text{Index}_P(V) = 2$. For example, if W is the vector field on $\mathbb{R}^2$ given in (8), i.e., $W(x,y) = (x^2 - y^2, 2xy)$, then $V = \varphi_* W$ is a vector field on the complement of the north pole with one singularity of index 2 at the south pole. A short calculation shows that V extends continuously to the north pole, with value there the vector $(-2, 0, 0)$.

Step 2. We show that the sum of the indices of any two such vector fields V and W on S^2 is the same. Let P be a point where neither has a singularity. By Lemma 7.11, replacing V by another vector field $\tilde{V}$ with the same indices as V, we can assume that V and W agree in some neighborhood of P. Then using stereographic projection from P, one has two vector fields on the plane that agree outside some large disk that contains all the singularities of either vector field. Taking a larger circle C_r, their winding numbers around C_r will be the same, and an application of Corollary 7.7 shows that the sum of their indices is the same. □

Exercise 7.16. Give an alternative proof of Step 1 by finding a vector field on S^2 with two singular points, each with index 1.

Problem 7.17. (a) If $f\colon S^2 \to \mathbb{R}^3$ is a continuous mapping, show that there is some point P in S^2 and some real number λ so that $f(P) = \lambda P$. (b) If $f\colon S^2 \to S^2$ is a continuous mapping, show that there is some point P in S^2 such that $f(P) = P$ or $f(P) = -P$.

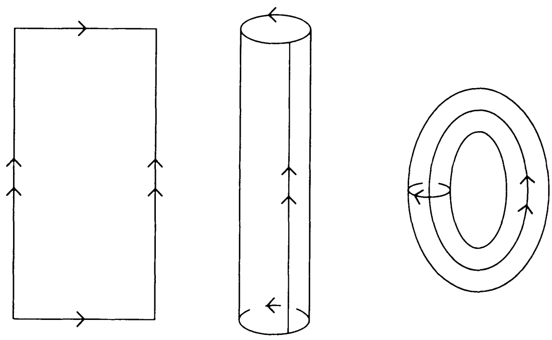
Problem 7.18. Give a mathematical formulation and proof of the statement: "On a spherical planet, the number of peaks plus the number of valleys is two more than the number of passes."

CHAPTER 8

Vector Fields on Surfaces

8a. Vector Fields on a Torus and Other Surfaces

Let us look next at a surface X which is a *torus*, i.e., the surface of a doughnut. This can be realized concretely in several ways: as a surface of revolution, by an explicit equation in 3-space, as a Cartesian product $S^1 \times S^1$, or by taking a square or rectangle and identifying opposite edges:



It is clear by any of these descriptions that there are vector fields on X that have no singularities at all. In order to state the analogue of Proposition 7.15 for a torus, we need to define the index of a vector field at a point on X. One way to do this is to realize X as a quotient space of the plane $\mathbb{R}^2$, identifying two points if their difference is in the lattice $\mathbb{Z}^2$. This amounts to identifying the opposite sides of the unit square $[0,1] \times [0,1]$. We have a mapping

$$p\colon \mathbb{R}^2 \to X = S^1 \times S^1,$$

and giving a vector field V on X is the same as giving a vector field $\tilde{V}$ on $\mathbb{R}^2$ that is unchanged by translation by any vector in $\mathbb{Z}^2$. We can define the index $\mathrm{Index}_P V$ to be the index of the corresponding vector field $\tilde{V}$ at any point of $\mathbb{R}^2$ that maps to P. As before, we allow a finite set Z of singularities, and require V to be continuous outside Z. Then we have the analogous proposition:

Proposition 8.1. *For any vector field with singularities V on a torus X,*

$$\sum_{P \in Z} \mathrm{Index}_P V = 0.$$

Proof. Take a square $R = [a, a+1] \times [b, b+1]$ so that the image in X of the boundary ∂R does not hit the singularity set Z. Look at the corresponding vector field $\tilde{V}$ on a neighborhood of R. By Proposition 7.5 the winding number of $\tilde{V}$ around ∂R is the sum of the indices of $\tilde{V}$ inside R. This winding number is zero, since the vector field is the same on opposite sides of the square. And the indices inside are the indices of V on X. □

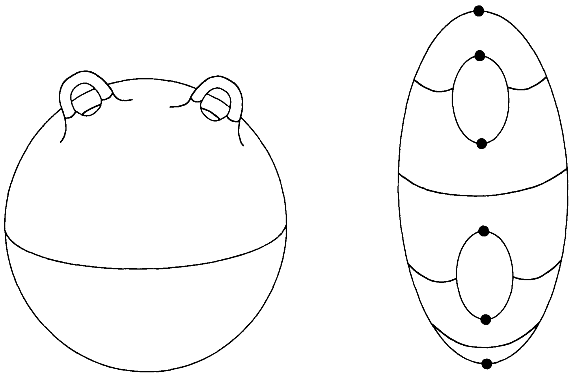
For example, realizing X as a surface of revolution in space as in the above picture, the projection of a vertical vector $(0,0,1)$ onto the tangent space at each point gives a vector field with singularities at the four points with horizontal tangents. The indices at the points at the top and bottom (where the height has local maximum and minimum) are $+1$, while those at the two saddle points are -1.

Exercise 8.2. Show that the following formula defines a diffeomorphism of $\mathbb{R}^2/\mathbb{Z}^2 = S^1 \times S^1$ with a surface of revolution in $\mathbb{R}^3$. Take $0 < r < R$ and define the map by

$$(x,y) \mapsto (R + r\cos(2\pi y)) \cdot (0, \cos(2\pi x), \sin(2\pi x)) \\ + r\sin(2\pi y) \cdot (1,0,0).$$

Find the vector field on the plane corresponding to the above vector field, and verify that the indices are 1, -1, -1, and 1.

We want to generalize what we have just seen from the sphere and torus to other surfaces, in particular to the surface of a doughnut with g holes, or a "sphere with g handles":

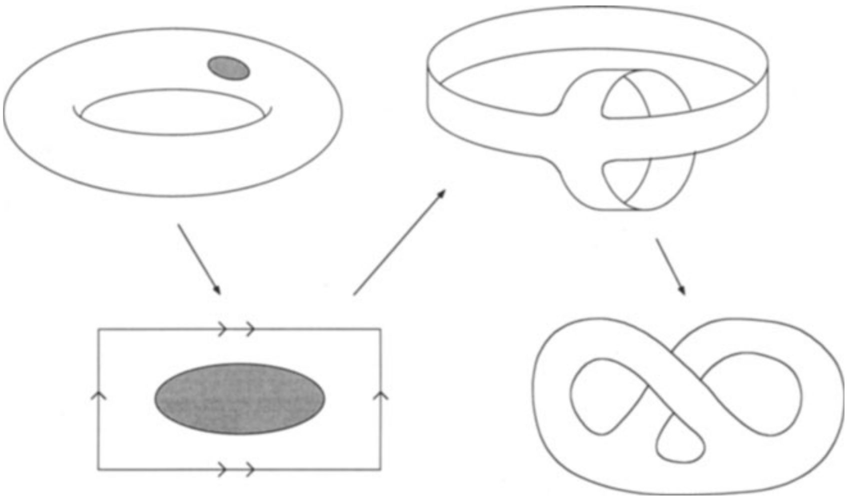


We will argue geometrically and rather loosely for now, and postpone until later a discussion of what needs to be done to make the arguments precise and rigorous. From the second picture, taking V to be the vector field with $V(P)$ the projection on the tangent space T_PX of a vertical vector $(0, 0, 1)$ as before, we see that there are two points with index 1 (at the top and bottom), and $2g$ points with index -1 (at the saddles). This gives one vector field the sum of whose indices is $2 - 2g$. The claim is that this is always the case.

Theorem 8.3 (Poincaré–Hopf). *Let X be a sphere with g handles. For any vector field V with singularities on X, the sum of the indices of V at the singular points is $2 - 2g$.*

Proof. Having seen one such vector field, it is enough to show that any two vector fields have the same sum of indices. By Lemma 7.11, we can take a disk in X where both have no singularities, and modify one so that they agree on such a disk. The idea of the proof is to mimic the proof for a sphere: take a circle C in such a disk, and punch out a smaller disk D (inside the circle) from the surface, and spread the complement $X \setminus D$ out on the plane. The two vector fields will then have the same winding number around the curve C, so they will have the same sum of indices by Proposition 7.5.

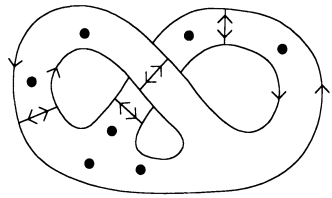
For $g > 0$, however, this complement is not diffeomorphic (or homeomorphic) to a plane domain. However, it can be realized with a mapping $\varphi\colon X \setminus D \to \mathbb{R}^2$ that is a local diffeomorphism, i.e., every point P in $X \setminus D$ has a neighborhood that is mapped diffeomorphically (with a diffeomorphic inverse) onto its image. To visualize this, look first at the torus. The complement of a disk is formed of two bands joined together. The mapping φ from $X \setminus D$ can be visualized by picturing the bands over the plane.



The image of the circle C goes around near the boundary. We can define the winding number $W(V|_C, 0)$ by cutting C up into pieces, each of which is mapped one-to-one by φ into the plane, and using the usual definition on these pieces. We can also define the index of V at a point P of $X \setminus D$ by using the local diffeomorphism φ_* to identify V with a vector field $\varphi_* V$ near $\varphi(P)$, and defining $\mathrm{Index}_P V$ to be the index of $\varphi_* V$ at $\varphi(P)$. To finish the proof, it is enough to show that

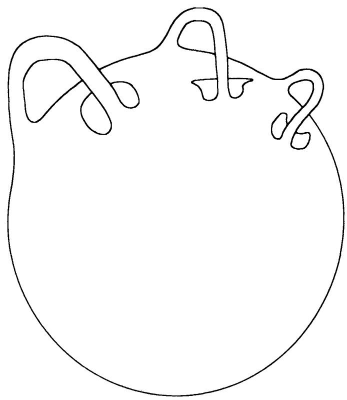
$$\sum_{P \in X \setminus D} \mathrm{Index}_P V = W(V|_C, 0).$$

To see this, one can add some crosscuts, being careful not to go through any singularities, and apply Proposition 7.5 to the restriction of V to each of the pieces.

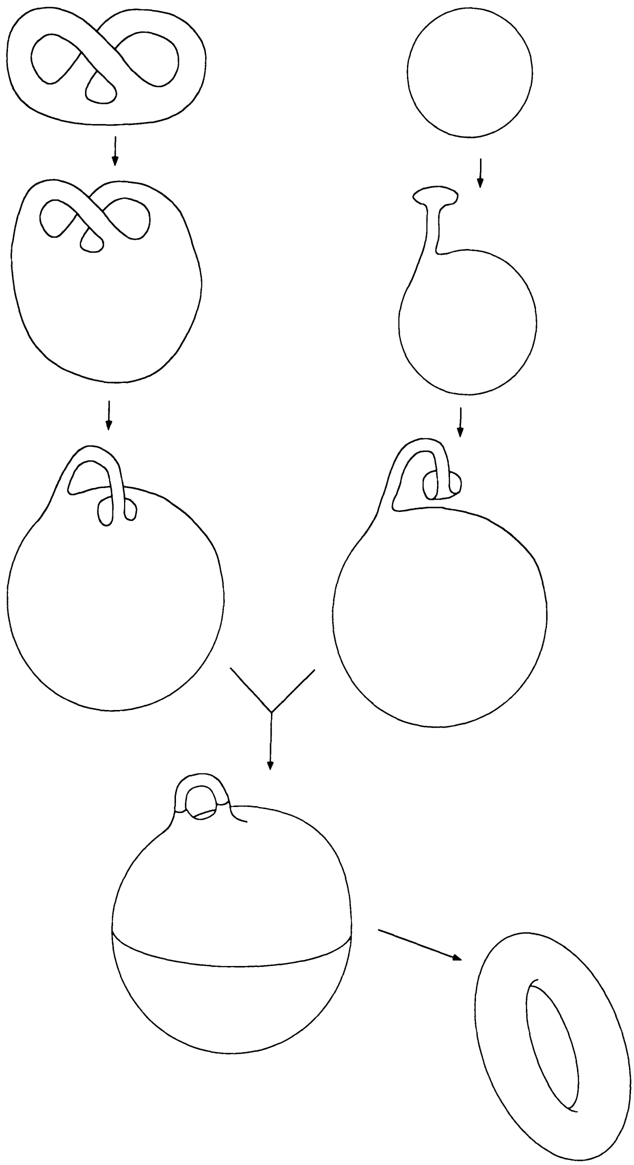


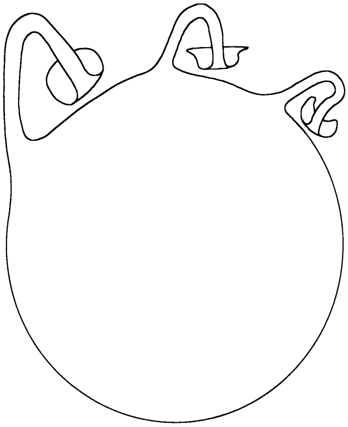
As usual, the added pieces get counted twice with opposite signs, so they cancel, and one is left with the displayed equation.

The same works for any genus g, although the visualization is a little harder. We claim that, when a closed disk D is removed from X, the complement can be realized over the plane as indicated (for $g = 3$):



Once we have this, the same argument completes the proof of the theorem. To realize $X \setminus D$ this way, an essential point is that, as a larger and larger disk is removed from X, the complements are all diffeomorphic. To aid in visualization, the situation for $g = 1$ is redone this way on a separate page. The picture after that shows a disk with "fingers," which can be sewn to the back of the above figure, giving a sphere with g handles.





Exercise 8.4. Describe a vector field on X that has exactly $2g-2$ singular points, each with index -1.

It is time to discuss what it might take to make this sort of argument rigorous. First, of course, one should give precise definitions of all the objects involved: surfaces, a sphere with g handles, a vector field on a surface, and the index of a vector field at a point. This not being a course on manifolds, we will not go through all the details (but see Appendix D for some of them). A key point, however, is that a surface is covered by the images of coordinate charts $\varphi_\alpha\colon U_\alpha \to X$, that are homeomorphisms from open sets U_α in the plane to open sets $\varphi_\alpha(U_\alpha)$ in X, and the change of coordinate mappings $\varphi_\beta^{-1} \circ \varphi_\alpha$, defined from part of U_α to part of U_β, should be $\mathscr{C}^\infty$. For the sphere, for example, stereographic projections from the two poles gave coordinate charts φ and ψ, and for the torus, the mapping $p\colon \mathbb{R}^2 \to X = S^1 \times S^1$, restricted to small open sets in the plane, gives charts on X.

A vector field V on X then determines a vector field V_α on U_α, and these are related by the condition that $(\varphi_\beta^{-1} \circ \varphi_\alpha)_*(V_\alpha) = V_\beta$ on the open subsets where both are defined. In fact, a vector field on X can be defined as a collection of such vector fields V_α, related by these compatibilities under changes of coordinates. The index of V at a point P in X can be defined as the index of V_α at the point P_α, if $\varphi_\alpha(P_\alpha) = P$. The key Lemma 7.10 implies that this is independent of choice of chart near P. Note that the surface X is assumed to have a differentiable structure, but that the vector field is only assumed to be continuous in the complement of a finite set.

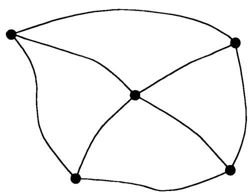
The description of surfaces via cutting and pasting, which we have indicated in pictures, could be done explicitly in coordinates, as a large and not very pleasant exercise. For example, one can verify that removing a slightly larger disk from a surface gives diffeomorphic complements. The tools for doing this sort of argument properly are developed systematically in the subject of differential topology, see Milnor (1965), Wallace (1968), and Guillemin and Pollack (1974).

Exercise 8.5. What can you say about the number of peaks, valleys, and passes on a planet shaped like a sphere with g handles?

Exercise 8.6. Does a sphere with g handles have the fixed point property?

8b. The Euler Characteristic

Suppose X is a compact surface, and we have a *triangulation* of X. This means that X is cut into pieces homeomorphic to triangles, fitting together along the edges. We will have a certain number v of vertices (points), a number e of edges (homeomorphic to closed intervals), and a number f of faces (homeomorphic to closed triangles). These homeomorphisms are assumed to take the ends of intervals to two distinct vertices, and the three boundary pieces of a triangle onto three distinct edges.

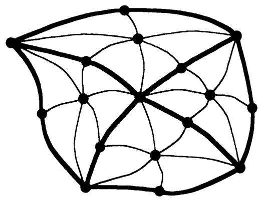


As you may have discovered in an exercise in the Preface, the number $v - e + f$ is independent of the triangulation:

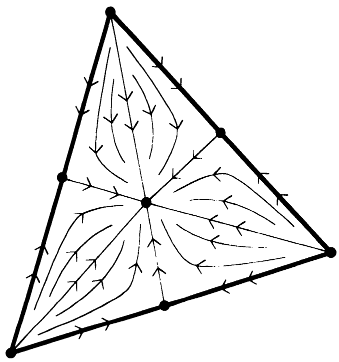
Proposition 8.7. *For any triangulation of a sphere X with g handles,*

$$v - e + f = 2 - 2g.$$

Proof. The idea is to construct a vector field V on X with one singularity of index 1 for each vertex, one of index -1 for each edge, and one of index 1 for each face. Then the proposition follows from the Poincaré–Hopf theorem. To do this, do a "barycentric subdivision": put a new vertex in each edge, and one in each face, and connect them as shown.



Then construct a vector field on X, so that in each triangle it looks like:



Construct it first along the edges: on the old edges pointing from the old vertices to the new ones added in the middle of the edges, and on the new edges pointing toward the new vertices in the faces; then fill in over the new triangles to make it continuous. If this is done, one has a vector field V whose singularities are at the vertices, and whose indices are: $+1$ if P is an old vertex; -1 if P is a new vertex along an edge; and $+1$ if P is a new vertex in a face. There are v of the first, e of the second, and f of the third, so by the Poincaré–Hopf theorem, $v \cdot (+1) + e \cdot (-1) + f \cdot (+1) = 2 - 2g$. □

The number $v - e + f$, for any triangulation, which is the same as the sum of the indices of any vector field, is called the *Euler characteristic* of the surface.

Exercise 8.8. Construct a triangulation on the sphere with g handles, and verify the proposition for this triangulation.

Problem 8.9. (a) Show that, for any triangulation of a sphere with g handles:

(i) $2e = 3f$;
(ii) $e \leq {}^1\!/_2 v \cdot (v-1)$; and
(iii) $v \geq {}^1\!/_2(7 + \sqrt{49 - 24(2-2g)})$.

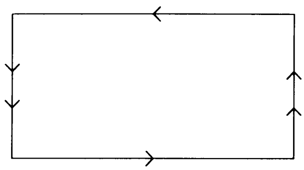
(b) Find lower bounds for v, e, and f for $g = 0$ and $g = 1$, and construct triangulations achieving these lower bounds.

If N is the largest integer less than or equal to the number ${}^1\!/_2(7 + \sqrt{49 - 24(2-2g)})$, it is a fact that any map on X can be colored with N colors, and N is the smallest number for which this is true. See Rademacher and Toeplitz (1957) and Coxeter (1989). Surprisingly, this is *much* easier for $g > 0$ than for $g = 0$.

Exercise 8.10. Generalize Proposition 8.7 to allow arbitrary convex polygons for faces in place of triangles.

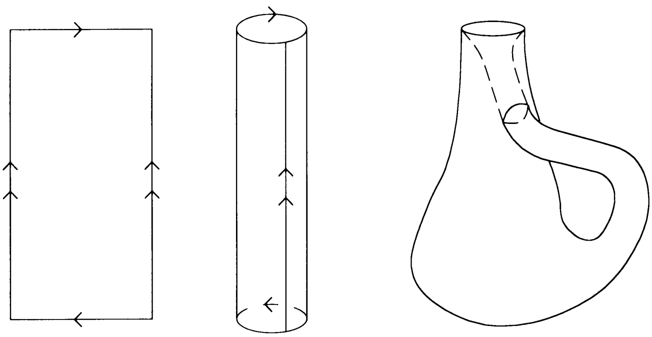
Problem 8.11. Suppose a sphere with g handles is decomposed as in the preceding exercise, but with each polygon having the same number p of edges, and assume that each vertex lies on the same number q of edges. (a) Show that $1/p + 1/q = {}^1\!/_2 + (1-g)/e$. (b) When $g = 0$, find all positive integers p, q, and e that satisfy this equation, and show that all possibilities are realized by the boundaries of the five platonic (regular) solids.

A sphere with g handles can be *oriented*, i.e., one can coherently (continuously) define a notion of counterclockwise (or "which way is up") in the neighborhood of any point. (See Appendix D for a precise definition.) It is a fact that the only compact surfaces that can be oriented are diffeomorphic to spheres with g handles. At least with the assumption that the surface can be triangulated we will prove this in Chapter 17. There are also compact surfaces that cannot be oriented. One example is the *projective plane* $\mathbb{RP}^2$, which can be realized as a quotient space of the sphere S^2, by identifying each point with its antipodal point. The projection from S^2 to $\mathbb{RP}^2$ is two-to-one, but a local diffeomorphism. One can also get $\mathbb{RP}^2$ by taking the upper hemisphere, and identifying opposite points on the boundary equator, which is the same as identifying opposite points on the boundary of a disk. Hence the projective plane can also be realized by identifying the opposite sides of a rectangle as shown:



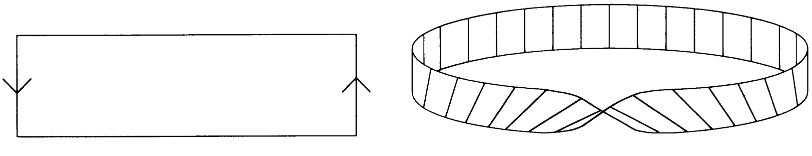
Exercise 8.12. Construct a vector field on $\mathbb{RP}^2$ that has one singular point with index 1. Show that the sum of the indices of any vector field on $\mathbb{RP}^2$ is 1. Triangulate $\mathbb{RP}^2$ and compute its Euler characteristic.

Another nonorientable surface is the *Klein bottle*, which can be realized by identifying the opposite sides of a rectangle as shown:

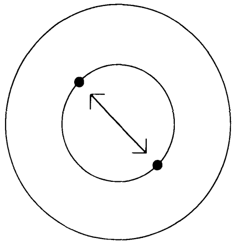


Exercise 8.13. What is the Euler characteristic of the Klein bottle?

The *Moebius band* is a nonorientable surface whose boundary is a circle:

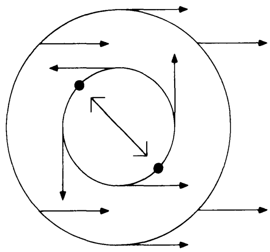


Or one may construct a Moebius band by taking an annulus, and identifying opposite points of one of the boundary circles:



Exercise 8.14. (a) Show that these two descriptions agree. (b) What do you get when you sew two Moebius bands together along their boundary circles?

Exercise 8.15. Find a vector field on the Moebius band with the boundary behavior shown and one singular point.

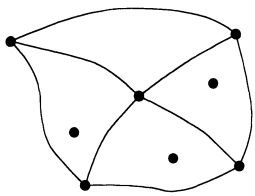


What is the index at the singular point?

Starting with any surface, one can cut out a disk, and paste back in a Moebius band, by identifying points on the boundary circles. This is called a *crosscap*.

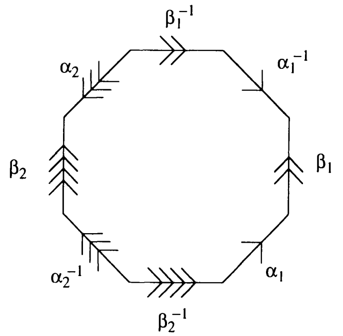
Project 8.16. Investigate the surfaces that arise this way, especially the sums of indices of vector fields and the notion of Euler characteristic. What do you get when you do this to a sphere? What happens to the Euler characteristic of a surface when this is done to it? What is the Euler characteristic of the surface obtained by punching h disjoint disks from a sphere with g handles, and sewing h Moebius bands onto their boundaries. Can you realize the Klein bottle this way? Can two of these be homeomorphic, if you start with a different g and h? When?

There is a beautiful proof of the Poincaré–Hopf theorem for a compact surface X that is oriented, using an inner product, varying continuously, on each tangent space. (If the surface is embedded in 3-space, one can identify each tangent space $T_P X$ as a subspace of $\mathbb{R}^3$, and use the standard inner product on $\mathbb{R}^3$.) To show that two vector fields V and W on X have the same sum of indices, triangulate X so all the singularities of V or W are inside triangles, and so that each triangle contains at most one singularity.



If P is a singular point in a triangle T, one can see that the number $\mathrm{Index}_P V - \mathrm{Index}_P W$ is the change in angle of the vector field $V - W$ around the boundary of T, divided by 2π; this change in angle is zero if there is no singularity in T. Adding over all the triangles, noting the cancellations along the edges, the theorem follows. (See Hopf (1983).)

There is a related proof, closer to our first proof for a torus. As we will see in Chapter 17, the surface X can be realized by identifying sides as indicated on a plane polygon R with $4g$ sides:



Exercise 8.17. Show that the resulting surface is a sphere with g handles.

Given vector fields V and W on X, one can find such a realization of X so that none of the edges go through singularities of V or W. Then V and W determine vector fields $\widetilde{V}$ and $\widetilde{W}$ on the polygon, and

$$\sum_{P\in X} \mathrm{Index}_P V - \mathrm{Index}_P W = \sum_{P\in R} \mathrm{Index}_P \widetilde{V} - \mathrm{Index}_P \widetilde{W}$$
$$= \frac{1}{2\pi}(\text{change in angle of } \widetilde{V} - \widetilde{W} \text{ around } \partial R) = 0,$$

the last since the changes over identified edges of the boundary cancel.

We saw that the Euler characteristic is $2-2g$ by looking at the vector field of a fluid flowing down the surface, or the gradient of the height function for the surface sitting nicely in space. That picture also shows how one can build up the topology of X by looking at the portion of X whose height is at most h, and seeing how the topology changes as h increases. One sees that changes occur only when the height crosses the singularities, and that the change there is controlled by the indices of these singularities. This is the beginning of the beautiful subject of *Morse theory*. See Milnor (1963).

There is an important method for reducing some problems about nonorientable surfaces to the case of orientable surfaces. For a nonorientable surface X, there is an orientable surface $\widetilde{X}$ and a two-to-one mapping $p\colon \widetilde{X} \to X$, which is a local diffeomorphism. The two points in $\widetilde{X}$ over a point P in X correspond to the two ways to orient X near P. For example, if X is the projective plane, then $\widetilde{X}$ is the sphere. We will discuss this in more detail when we come to covering spaces, see §16a.

Problem 8.18. Show that a vector field V on X determines a vector field $\widetilde{V}$ on $\widetilde{X}$, and that the sum of the indices of $\widetilde{V}$ is twice the sum of the indices of V. Deduce that if X is any compact surface, the sum of indices of all vector fields with singularities on X is the same. If X is the surface constructed by sewing h Moebius bands to a sphere with g handles with h disks removed, show that $\widetilde{X}$ is a sphere with $2g+h-1$ handles.

PART V

COHOMOLOGY AND HOMOLOGY, II

In Chapter 9 the first homology group $H_1 U$ of a plane region U is computed for a plane region "with n holes." The notion of the integral of a closed $\mathscr{C}^\infty$ 1-form over any continuous path or any 1-chain is defined. We show that these integrals are the same over homologous 1-chains. This leads to useful methods for computing integrals and winding numbers, and relating the two notions; these are described in the third section.

The last section of Chapter 9, which is optional, takes a look at how these ideas are used in complex analysis. This is written as a (very) short course in complex analysis. It is self-contained, except for some calculations left as exercises, but in practice it will probably be most useful to those who have seen some of it before. For example, if you have seen Cauchy's formula and the residue theorem for regions such as disks and rectangles, this will show how the ideas of topology lead to the appropriate generalizations involving winding numbers. (This section includes all the analysis that will be needed when we study Riemann surfaces in Part X.)

The basic theme of the Mayer–Vietoris story in Chapter 10 is to find relations among the homology and cohomology groups of two open sets and their union and intersection. This possibility of comparing different homology groups H_k, for different k as well as different spaces, is a salient feature of algebraic topology. For cohomology, the beginnings of this story were seen in Chapter 5. This chapter proves analogous results for homology, and then completes the cohomology story. This package of results, called the Mayer–Vietoris theorem, gives a powerful tool for calculating homology and cohomology groups.

It will be evident that the cohomology and homology groups behave in a similar, or more precisely, dual fashion. This duality will be made explicit in Part VIII. In fact, the proof of the full Mayer–Vietoris theorem for cohomology of plane domains will depend on this duality.

CHAPTER 9

Holes and Integrals

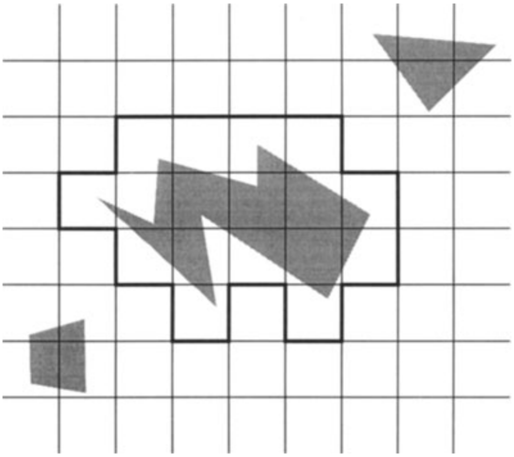
9a. Multiply Connected Regions

Let U be an open set in the plane, and let $A = \mathbb{R}^2 \setminus U$. We know that for a fixed closed chain γ in an open set U, the function $P \mapsto W(\gamma, P)$ is constant on connected components of $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$. It is therefore constant on connected components of A, since each connected component of A is contained in some connected component of $\mathbb{R}^2 \setminus \mathrm{Supp}(\gamma)$. If A is a connected component of $\mathbb{R}^2 \setminus U$, we write $W(\gamma, A)$ for the value of $W(\gamma, P)$ for all P in A, and call it the *winding number of* γ *around* A. Note that the connected components of a closed set such as $\mathbb{R}^2 \setminus U$ are closed, but they need not be path-connected (see Exercise 9.5).

Our goal is to calculate the first homology group of U, at least if this complement A is not too complicated. The idea is to find a closed path or 1-chain that "goes once" around each "piece" of A, and to show that these give a free basis for H_1U. The following lemma gives a precise statement and rigorous proof that this is possible:

Lemma 9.1. *Suppose A is a disjoint union of two closed sets B and C, with B bounded. Then there is a closed 1-chain γ in U such that $W(\gamma, P) = 1$ for all P in B and $W(\gamma, P) = 0$ for all P in C.*

Proof. By the compactness of B, there is a positive ε so that every point of B is at least distance ε away from any point of C. Take a grid G so that its bounded (closed) rectangles cover B, but such that none of these rectangles meets both B and C, and so that none of the infinite rectangles meets B. This can be achieved by taking an infinite grid so that the distances between parallel lines is less than $\varepsilon/\sqrt{2}$, and letting G be the collection of lines that hit B.



The idea is to define γ to be the sum of the boundaries of the (closed) bounded rectangles in the grid that meet B:

$$\gamma = \sum_{R_k \cap B \neq \varnothing} \partial R_k .$$

Clearly γ is a closed 1-chain, since it is the sum of closed 1-chains. We claim next that γ is a 1-chain in U. To see this, suppose an edge σ occurs with a nonzero coefficient in γ, but that σ is not contained in U. We know that σ cannot meet C, since it is in a rectangle that meets B, and none of the rectangles meets both. So σ must meet B. But then each of the two (bounded) rectangles that σ separates meets B, so they occur in the displayed sum. Their contributions to σ therefore cancel, which shows that σ cannot occur in γ.

Next we show that $W(\gamma, P) = 1$ for P in B. Any point P in B is in one of the closed bounded rectangles, say R_l, and we let Q be a point in the interior of R_l. Using the fact that P and Q are in the same component of the complement of the support of γ, we have

$$W(\gamma, P) = W(\gamma, Q) = \sum_k W(\partial R_k, Q) = W(\partial R_l, Q) = 1.$$

Similarly, if $P \in C$, then $W(\gamma, P) = \Sigma_k W(\partial R_k, P) = 0$. □

A 1-chain γ having the property of the lemma is certainly not unique, but any other 1-chain with this property would have the same winding numbers around all points not in U, and it follows from Theorem 6.11 that it would be homologous to γ.

We next describe what it means for U to have "n holes." First we describe the "infinite" part of the complement A, denoted by A_∞, which will be a closed set with the property that winding numbers of closed 1-chains in U around points in A_∞ are always zero. In most examples it is obvious what A_∞ should be, but the general definition is a little complicated. Certainly A_∞ should contain any unbounded connected component of A; to assure that we get a closed set, we define A_∞ by the following:

$$A_\infty = \{P \in A\text{: for any } \varepsilon > 0 \text{ there is a connected subset } C \text{ of } A \text{ containing a point within distance } \varepsilon \text{ of } P \text{ and a point farther than } 1/\varepsilon \text{ from the origin}\}.$$

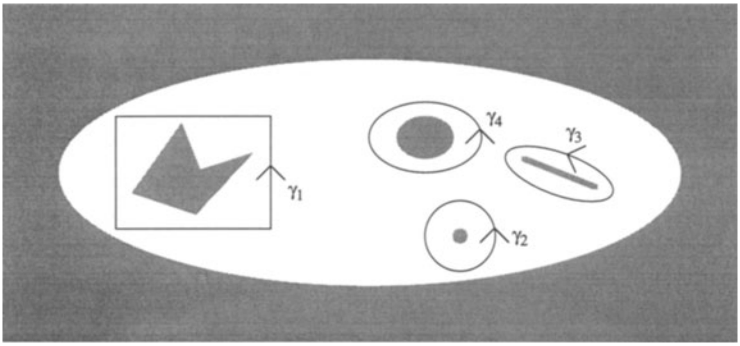
Lemma 9.2. *The set A_∞ is a closed subset of A, and $W(\gamma, P) = 0$ for all closed 1-chains γ in U and all points P in A_∞.*

Proof. If P is in the closure of A_∞, take a sequence P_n in A_∞ approaching P, and a connected set C_n as in the definition for P_n for some ε_n, with $\varepsilon_n \to 0$ as $n \to \infty$. It follows immediately that P is in A_∞. For any compact subset K of U, it follows from the definition that any point of A_∞ is in the unbounded component of $\mathbb{R}^2 \setminus K$. Note that the support of any 1-chain is compact. By Proposition 6.8 the winding number of any closed 1-chain around a point of A_∞ must be zero. □

Note that A_∞ can have more than one connected component, or it can be empty. Now suppose that

$$\mathbb{R}^2 \setminus U = A_1 \cup A_2 \cup \ldots \cup A_n \cup A_\infty,$$

where each A_i, $1 \le i \le n$, is a connected, closed, bounded, nonempty set, and these $n+1$ sets are disjoint. Such an open set U is sometimes called *multiply connected*, or $(n+1)$*-connected*. Not every U has such a complement, see Exercise 9.5.



By Lemma 9.1, for each i between 1 and n there is a closed chain γ_i in U so that $W(\gamma_i, A_i) = 1$, and $W(\gamma_i, A_j) = 0$ for $j \neq i$. Such 1-chains are not unique, but any other choices are homologous to these.

Proposition 9.3. *Any closed* 1*-chain* γ *on* U *is homologous to the* 1*-chain* $m_1\gamma_1 + \ldots + m_n\gamma_n$, *where* $m_i = W(\gamma, A_i)$.

Proof. With $m_i = W(\gamma, A_i)$, γ and $\Sigma m_i \gamma_i$ have the same winding numbers around all points not in U, and the result follows from Theorem 6.11. □

The integers $m_i = W(\gamma, A_i)$ are uniquely determined by the homology class γ, again by Theorem 6.11. In other words:

Corollary 9.4. *The homology classes of the closed chains* $\gamma_1, \ldots, \gamma_n$ *form a free basis of* H_1U, *giving an isomorphism* $H_1U \cong \mathbb{Z}^n$.

Exercise 9.5. (a) Let U be the set of points (x, y) such that $|y| < 1$ or $2n < x < 2n + 1$ for some integer n. Show that A_∞ is the union of an infinite number of connected components. (b) Let A be the union of the interval $\{(0, y)\colon 0 \leq y \leq 1\}$ and the set $\{(x, \sin(1/x))\colon x > 0\}$. Show that A is closed and connected, but not path-connected. (c) Let A be the union of the origin and the points $(1/n, 0)$ for n a positive integer. Show that each point of A is a connected component of A, but the origin has no neighborhood disjoint from the other components.

Exercise 9.6. Show that if $U \subset U'$ and U is n-connected, and U' is n'-connected, and $n > n'$, then there is no retract from U' onto U.

Problem 9.7. (a) Show that if U is the complement of the set $\mathbb{N}$ of nonnegative integers (identify n with $(n, 0)$), then H_1U is isomorphic to the free abelian group on the points in $\mathbb{N}$. In particular, H_1U is not finitely generated. (b) Compute H_1U when U is the complement of the set in Exercise 9.5(c).

Problem 9.8. (a) If U is bounded, show that A_∞ is a connected component of $\mathbb{R}^2 \setminus U$. (b) If the plane is identified with the complement of the north pole in a sphere S (by stereographic projection, see §7c), show that the union of A_∞ and the north pole is the connected component of the complement of U in S that contains the north pole.

Problem 9.9. Generalize Corollary 9.4 as follows. Suppose U is any open set in the plane, and K is a compact subset of U that has n connected components $K_1, \ldots, K_n$. There is a homomorphism

$$H_1(U \setminus K) \to H_1U \oplus \mathbb{Z} \oplus \ldots \oplus \mathbb{Z} \cong H_1U \oplus \mathbb{Z}^n,$$

defined as follows. Given a closed 1-chain γ on $U \setminus K$, the first component of this map takes the class of γ in $H_1(U \setminus K)$ to the class of γ in H_1U, and the other components take the class of γ to the winding numbers of γ around the components $K_1, \ldots, K_n$. Show that this homomorphism is an isomorphism.

9b. Integration over Continuous Paths and Chains

We want to define the notion of the integral $\int_\gamma \omega$ of a closed $\mathscr{C}^\infty$ 1-form ω on an open set U over an arbitrary closed path γ in U. We cannot use calculus, but the idea we used for defining the winding number works perfectly well. If γ is defined on an interval $[a, b]$, subdivide the interval into $a = t_0 \le t_1 \le \ldots \le t_n = b$, so that each subinterval $[t_{i-1}, t_i]$ is mapped by γ into some open rectangle U_i contained in U. Such a subdivision exists by the Lebesgue lemma, since each point in the image of γ is contained in some such rectangle. The restriction of ω to U_i is exact (by Proposition 1.12), so we may find a $\mathscr{C}^\infty$ function f_i on U_i such that $df_i = \omega$ on U_i. Let $P_i = \gamma(t_i)$, $0 \le i \le n$. Define the integral $\int_\gamma \omega$ by

$$\int_\gamma \omega = (f_1(P_1) - f_1(P_0)) + (f_2(P_2) - f_2(P_1)) + \ldots + (f_n(P_n) - f_n(P_{n-1})).$$

The proof that this definition is independent of choices is almost the same as that of Proposition 3.1. Since the intersection of two rectangles containing $\gamma([t_{i-1}, t_i])$ is connected, and f_i is unique up to adding a constant on a connected set, the sum is independent of choices of U_i and f_i. The rest of the proof is identical, by observing that refining a subdivision by adding a point doesn't change the answer.

If γ is a $\mathscr{C}^\infty$ path, it follows also from Proposition 1.16 that this definition of $\int_\gamma \omega$ agrees with that using calculus.

We extend the notion to all 1-chains $\gamma = n_1\gamma_1 + \ldots + n_r\gamma_r$, with γ_i paths, by linearity:

$$\int_\gamma \omega = n_1 \int_{\gamma_1} \omega + \ldots + n_r \int_{\gamma_r} \omega .$$

Note that the winding number is a special case of integral:

$$W(\gamma, P) = \int_\gamma \omega_P ,$$

where, if $P = (x_0, y_0)$,

$$\omega_P = \frac{1}{2\pi}\omega_{P,\vartheta} = \frac{1}{2\pi} \cdot \frac{-(y-y_0)\,dx + (x-x_0)\,dy}{(x-x_0)^2 + (y-y_0)^2} .$$

Exercise 9.10. If $\omega = df$ is exact, and γ is a 1-chain with boundary $\partial\gamma = \Sigma m_j P_j$, show that $\int_\gamma \omega = \Sigma m_j f(P_j)$. In particular, $\int_\gamma df = 0$ if γ is closed.

Proposition 9.11. *If γ and δ are homologous 1-chains in U, then*

$$\int_\gamma \omega = \int_\delta \omega .$$

Proof. The proof is the same as the proof of Lemma 6.7, which refers to Theorem 3.6. One simply changes winding numbers to integrals, sectors $U_{i,j}$ to open rectangles, and angle functions ϑ_{ij} to arbitrary functions $f_{i,j}$ with $df_{i,j} = \omega$ on $U_{i,j}$. □

It follows in particular that the integral of ω is the same over two paths that are homotopic closed paths, or homotopic paths with the same endpoints. It also follows that the integrals are unchanged by the reparametrization of paths.

Corollary 9.12. *Given two closed 1-chains γ and δ on U, the following are equivalent:*

(1) $W(\gamma, P) = W(\delta, P)$ *for all P not in U;*
(2) $\int_\gamma \omega = \int_\delta \omega$ *for all closed 1-forms ω on U; and*

(3) *$\gamma - \delta$ is a boundary* 1*-chain in* U.

Proof. The implications (3)$\Rightarrow$(2)$\Rightarrow$(1) have just been seen, and (1)$\Leftrightarrow$(3) is Theorem 6.11. □

Corollary 9.13. *Let U be any open set in the plane. Then the following are equivalent*:

(1) $W(\gamma, P) = 0$ *for all closed chains γ in U and all $P \notin U$*;
(2) $\int_\gamma \omega = 0$ *for all closed chains γ in U and all closed* 1*-forms ω on U*;
(3) *every closed* 1*-chain is a boundary*: $H_1 U = 0$; *and*
(4) *every closed* 1*-form ω on U is exact*: $H^1 U = 0$.

These conditions hold whenever U is 1*-connected, i.e.,* $\mathbb{R}^2 \setminus U = A_\infty$.

Proof. The equivalence of the first three conditions follows from the preceding corollary. We have (4)$\Rightarrow$(2) by Exercise 9.10. Conversely, (2) implies that $\int_\gamma \omega = \int_\delta \omega$ whenever γ and δ are segmented paths with the same endpoints, and ω is any closed 1-form; we saw in Chapter 1 that this makes ω exact, which is (4). □

We know that homotopic closed paths are homologous.

Problem 9.14. (a) Show that homologous closed paths must be homotopic when U is an open rectangle or any convex open set. (b) Do the same when U is the complement of a point, or an annulus

$$U = \{(x, y) \colon r_1{}^2 < (x - x_0)^2 + (y - y_0)^2 < r_2{}^2\}.$$

(c) *Challenge*. What if U is the complement of two points?

Problem 9.15. Suppose $\gamma = \Sigma n_i \gamma_i$ is a closed 1-chain in an open set U such that each γ_i is a $\mathscr{C}^\infty$ path. If γ is homologous to zero, show that one can write $\gamma = \Sigma m_j(\partial \Gamma_j)$, where each Γ_j is a $\mathscr{C}^\infty$ map from the unit square to U.

Problem 9.16. If X is any closed, connected subset of the plane, and U is a bounded connected component of $\mathbb{R}^2 \setminus X$, show that every closed 1-form on U is exact.

Problem 9.17. (a) If $\omega_t = p(x, y, t)\,dx + q(x, y, t)\,dy$ is a continuously varying family of closed $\mathscr{C}^\infty$ 1-forms on U, i.e., the functions p and q are continuous on $U \times [a, b]$, show that the function $t \to \int_\gamma \omega_t$ is a continuous function of t. (b) Use this to give another proof of Proposition 3.16.

9c. Periods of Integrals

With U an $(n+1)$-connected region as above, and ω any closed 1-form on U, define the *period* (or *module of periodicity*) of ω *around* A_i to be the integral of ω along γ_i, with γ_i as defined just before Proposition 9.3. By Proposition 9.11, this is independent of choice of γ_i. Denote this period by $\mathfrak{p}_i(\omega)$ or $\mathfrak{p}(\omega, A_i)$:

$$\mathfrak{p}_i(\omega) = \mathfrak{p}(\omega, A_i) = \int_{\gamma_i} \omega .$$

These numbers determine integrals of ω along any closed path, provided one knows the winding number of the path around each A_i:

Proposition 9.18. *For any closed* 1*-chain* γ *and closed* 1*-form* ω,

$$\int_\gamma \omega = W(\gamma, A_1)\,\mathfrak{p}_1(\omega) + W(\gamma, A_2)\,\mathfrak{p}_2(\omega) + \ldots + W(\gamma, A_n)\,\mathfrak{p}_n(\omega) .$$

Proof. Since γ and $\Sigma_i W(\gamma, A_i) \cdot \gamma_i$ are homologous, this follows from Proposition 9.11. □

Applying Corollary 9.12, we have:

Corollary 9.19. *A closed* 1*-form* ω *on* U *is exact if and only if all of its periods* $\mathfrak{p}_i(\omega)$ *are zero.*

For integrals along paths that are not closed, if γ and δ are two paths with the same endpoints, applying the proposition to $\gamma - \delta$, we have

$$\int_\gamma \omega - \int_\delta \omega = m_1\,\mathfrak{p}_1(\omega) + m_2\,\mathfrak{p}_2(\omega) + \ldots + m_n\,\mathfrak{p}_n(\omega) ,$$

with $m_1, \ldots, m_n$ integers. This means that the integral is determined up to adding integral combinations of the periods. For example, when $U = \mathbb{R}^2 \setminus \{P\}$ and $\omega = \omega_{P,\vartheta}$ (see Problem 2.10), there is only one period, which is 2π, and we recover the fact that the integral is determined up to integral multiples of 2π.

Exercise 9.20. Compute the integral $\int_\gamma \omega$, where ω is the 1-form

$$\omega = \sum_{n=1}^{17} \frac{-y\,dx + (x-n)\,dy}{(x-n)^2 + y^2} ,$$

and $\gamma(t) = (t\cos(t), t\sin(t))$, $0 \le t \le 6\pi$.

Exercise 9.21. Show that, given U as above, for any real numbers $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ there is a closed 1-form ω with these periods.

This means that the linear mapping from the vector space of closed 1-forms on U to $\mathbb{R}^n$, $\omega \mapsto (\mathfrak{p}_1(\omega), \ldots, \mathfrak{p}_n(\omega))$, is surjective, and by Corollary 9.19 the kernel is the space of exact 1-forms. This sets up an isomorphism (see §C1)

$$\{\text{closed 1-forms on } U\}/\{\text{exact 1-forms on } U\} \;\cong\; \mathbb{R}^n,$$

i.e., the De Rham group H^1U is an n-dimensional vector space. If P_i is any point in A_i, $1 \le i \le n$, the classes $[\omega_{P_i}]$ form a basis for $H^1(U)$.

Problem 9.22. (a) Suppose $n = 2$, and the periods of ω are $\mathfrak{p}_1$ and $\mathfrak{p}_2$. Let P and Q be two fixed points in U. Show that if the periods are not zero, and the ratio $\mathfrak{p}_1/\mathfrak{p}_2$ is rational, there is a number r so that if γ and δ are any two paths from P to Q, then $\int_\gamma \omega - \int_\delta \omega$ is an integer times r. (b) Show that, if $\mathfrak{p}_1/\mathfrak{p}_2$ is not rational, and U is connected, there is no such r.

9d. Complex Integration

The plane $\mathbb{R}^2$ can be identified with the complex numbers $\mathbb{C}$, the pair (x, y) being identified with $z = x + iy$. Functions on open sets in the plane will be written as functions of z.

A *complex 1-form* ω on an open set U in the plane is given by a pair of ordinary (real) 1-forms ω_1 and ω_2, written in the form

$$\omega \;=\; \omega_1 + i\,\omega_2.$$

Define $d\omega = d\omega_1 + i\,d\omega_2$, and for a function $f = u + iv$, with u and v real-valued functions on U, set $df = du + i\,dv$. The form ω is *closed* if $d\omega = 0$, and *exact* if $\omega = df$. For example, we have the 1-form dz defined by $dz = dx + i\,dy$. A complex 1-form can by multiplied by a complex-valued function: if ω is as above, and $f = u + iv$, then $f \cdot \omega$ is the complex 1-form

$$f \cdot \omega \;=\; (u + iv) \cdot (\omega_1 + i\omega_2) \;=\; (u\omega_1 - v\omega_2) + i(u\omega_2 + v\omega_1).$$

Exercise 9.23. (a) If $f = u + iv$, with u and v $\mathscr{C}^\infty$ functions, show that the 1-form $f(z)\,dz$ is closed if and only if u and v satisfy the Cauchy–Riemann equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

(b) Show that if the conditions in (a) are satisfied, then the complex derivative $f'(a) = \lim_{z\to a}(f(z) - f(a))/(z-a)$ exists at each point a in U.

Let us call f *analytic* if its real and imaginary parts satisfy the Cauchy–Riemann equations. For example, if f is locally expandable in a power series, then f is analytic, so $f(z)\,dz$ is closed. Indeed, if $f(z) = \Sigma_{n=0}^{\infty} a_n(z-z_0)^n$ in a disk around z_0, then $f = dg$ on that disk, with

$$g(z) = \sum_{n=0}^{\infty} \frac{a_n}{n+1}(z-z_0)^{n+1}.$$

For an example that is closed but not exact, consider dz/z on the complement of the origin:

$$\begin{aligned}\frac{dz}{z} &= \frac{dx + i\,dy}{x+iy} = \frac{(x-iy)(dx+i\,dy)}{x^2+y^2}\\ &= \frac{x\,dx + y\,dy}{x^2+y^2} + i\frac{-y\,dx + x\,dy}{x^2+y^2} = d(\log(r)) + i\omega_\vartheta\end{aligned}$$

Similarly, $dz/(z-a) = d(\log(|z-a|)) + i\omega_{a,\vartheta}$, with $\omega_{a,\vartheta}$ as defined in Chapter 2.

If γ is a path or chain in U, the integral of $\omega = \omega_1 + i\omega_2$ along γ is defined by

$$\int_\gamma \omega = \int_\gamma \omega_1 + i\int_\gamma \omega_2.$$

For example, if $\gamma(t) = a + r\cdot e^{it}$, $0 \le t \le 2\pi$, is a circle around the point a, then

$$\int_\gamma \frac{dz}{z-a} = i\int_\gamma \omega_{a,\vartheta} = 2\pi i.$$

In general, if γ is any chain not containing a in its support, we see similarly that

$$\int_\gamma \frac{dz}{z-a} = i\int_\gamma \omega_{a,\vartheta} = 2\pi i\cdot W(\gamma, a).$$

The main fact about complex integrals is

Theorem 9.24 (Cauchy Integral Theorem). *If γ is a closed chain in U whose winding number around any point not in U is zero, and f is an analytic function in U, then for any a in U that is not in the* *support of* γ,

$$W(\gamma, a)\cdot f(a) = \frac{1}{2\pi i}\int_\gamma \frac{f(z)}{z-a}\,dz.$$

Proof. Look at $F(z) = (f(z) - f(a))/(z-a)$, which is analytic on $U \setminus \{a\}$. Using the formula proved just before the statement of the theorem, the displayed formula is equivalent to the formula $\int_\gamma F(z)\,dz = 0$. Let $\delta_r(t) = a + re^{2\pi i t}$, $0 \le t \le 1$, with r small enough so the disk of radius r around a is contained in U. Let $n = W(\gamma, a)$. Since $W(\gamma, P) = n \cdot W(\delta_r, P)$ for all $P \notin U \setminus \{a\}$, we know by Corollary 9.12 that $\int_\gamma F(z)\,dz = n \cdot \int_{\delta_r} F(z)\,dz$, so it suffices to prove that $\lim_{r\to 0}\int_{\delta_r} F(z)\,dz = 0$. But by Exercise 9.23(b), $F(z)$ has a limit as z approaches a, and the fact that $\int_{\delta_r} F(z)\,dz$ approaches zero as the radius goes to zero follows easily (see Exercise B.8). □

The simplest form of Cauchy's formula is when γ is a circle about a. Since the right side of Cauchy's formula is expandable in a power series in a neighborhood of a, this implies in particular the fact that any analytic function is locally expandable in a power series.

Exercise 9.25. (a) Let U be an open set containing two concentric circles and the region between them. For a in the region between the circles, show that

$$f(a) = \frac{1}{2\pi i}\int_{\gamma_r} \frac{f(z)}{z-a}\,dz - \frac{1}{2\pi i}\int_{\gamma_\varepsilon} \frac{f(z)}{z-a}\,dz,$$

where γ_r and γ_ε are counterclockwise paths around the larger and smaller circle. (b) Deduce Riemann's theorem on removable singularities: if f is analytic and bounded in a punctured neighborhood of a point b, then f extends to an analytic function in a full neighborhood of b.

A related application is to the general residue theorem. If f is analytic in a punctured neighborhood of a point a (i.e., in some $U \setminus \{a\}$, for U a neighborhood of a), the *residue* of f at a, denoted $\mathrm{Res}_a(f)$, is defined by

$$\mathrm{Res}_a(f) = \frac{1}{2\pi i}\int_\delta f(z)\,dz,$$

where δ is a small counterclockwise circle around a. By the Cauchy integral theorem, this is independent of the circle chosen.

Exercise 9.26. If f is given by a converging series $\Sigma_{n=-m}^{\infty} c_n(z-a)^n$ in a punctured neighborhood of a (i.e., f has at most a *pole* at a), show that $\mathrm{Res}_a(f) = c_{-1}$.

Theorem 9.27 (Residue Theorem). *If f is analytic in $U \setminus \{a_1, \ldots, a_r\}$, and γ is a closed 1-chain in $U \setminus \{a_1, \ldots, a_r\}$ such that $W(\gamma, P) = 0$ for all P not in U, then*

$$\frac{1}{2\pi i}\int_\gamma f(z)\,dz = \sum_{i=1}^{r} W(\gamma, a_i)\cdot \mathrm{Res}_{a_i}(f).$$

Proof. Take disjoint small circles δ_i around a_i. Then, with $n_i = W(\gamma, a_i)$, γ and $\Sigma n_i\delta_i$ have the same winding number around all points not in $U \setminus \{a_1, \ldots, a_r\}$. It follows from Corollary 9.12 that the integral of the closed 1-form $f(z)\,dz$ around γ and around $\Sigma n_i\delta_i$ gives the same answer. □

Problem 9.28. Extend the residue theorem to allow f to be analytic outside any discrete set S in U (i.e., each point in U has a neighborhood containing at most one point of S).

Exercise 9.29. Suppose U is an $(n+1)$-connected open set as in the first section, with $\mathbb{R}^2 \setminus U = A_1 \cup \ldots \cup A_n \cup A_\infty$, and f is analytic on U; let $\mathfrak{p}_i$ be the period of $f(z)\,dz$ around A_i. Show that, for a closed chain γ on U,

$$\int_\gamma f(z)\,dz = W(\gamma, A_1)\,\mathfrak{p}_1 + \ldots + W(\gamma, A_n)\,\mathfrak{p}_n.$$

Exercise 9.30. Deduce the Cauchy Integral Theorem directly from the Residue Theorem.

Suppose now f is *meromorphic* in U, i.e., near every point a in U one can write $f(z) = (z-a)^m \cdot h(z)$, with h analytic at a, $h(a) \neq 0$, and m an integer. The *order of f at a*, denoted $\mathrm{ord}_a(f)$, is defined to be this integer m.

Exercise 9.31. Show that, with f and h as above, if δ is the boundary of a disk about a such that h is nowhere zero in the disk, then

$$\mathrm{ord}_a(f) = W(f\circ\delta, 0).$$

Theorem 9.32 (Argument Principle). *Suppose f is meromorphic in*

U, and γ is a closed 1-chain in U not passing through any zero or pole of f, such that $W(\gamma, P) = 0$ for all P not in U. Then

$$W(f\circ\gamma, 0) = \sum_a W(\gamma, a)\cdot \mathrm{ord}_a(f),$$

where the sum is over the (finitely many) zeros or poles a for which $W(\gamma, a) \neq 0$.

Proof. If $\{a_1, \ldots, a_r\}$ are the zeros and poles around which γ has a nonzero winding number (see Problem 9.28), take a small circle δ_i around a_i as in the proof of the Residue Theorem. Applying Proposition 6.15 to the mapping $f\colon U\setminus\{a_1, \ldots, a_r\}\to U' = \mathbb{C}\setminus\{0\}$, we get the formula $W(f\circ\gamma, 0) = \Sigma_{i=1}^{r} W(\gamma, a_i)\cdot W(f\circ\delta_i, 0)$. □

The following problems give the analytic interpretation, and a typical application, of the Argument Principle:

Problem 9.33. (a) If f is meromorphic at a, show that

$$\mathrm{ord}_a(f) = \mathrm{Res}_a\left(\frac{f'}{f}\right).$$

(b) Under the conditions of the Argument Principle, use the Residue Theorem to show that

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}dz = \sum_a W(\gamma, a)\cdot \mathrm{ord}_a(f).$$

(c) Show directly that

$$\frac{1}{2\pi i}\int_\gamma \frac{f'(z)}{f(z)}dz = W(f\circ\gamma, 0)$$

by interpreting the integrand as $d(\log(f(z)) = d(\log|f(z)|) + i\cdot d(\arg(f(z)))$, with $\log|f(z)|$ a well-defined function, and $\arg(f(z))$ the multivalued angle function.

Problem 9.34. (a) Let U, f, and γ be as in the Argument Principle. If g is analytic function on U, show that

$$\frac{1}{2\pi i}\int_\gamma g(z)\frac{f'(z)}{f(z)}dz = \sum_a W(\gamma, a)\cdot g(a)\cdot \mathrm{ord}_a(f).$$

(b) Suppose the restriction of f to a closed disk D around z_0 in U is one-to-one, and let γ be the boundary of D. Show that the function that takes w to $(1/2\pi i) \int_\gamma z f'(z)/(f(z) - w)\, dz$ defines an inverse function to f in a neighborhood of $f(z_0)$.

The following is Rouché's theorem:

Problem 9.35. Suppose the chain γ is homologous to zero in U, and suppose f and g are analytic functions in U such that

$$|f(z) - g(z)| < |f(z)| + |g(z)|$$

for all z in the support of γ. Show that

$$\sum_a W(\gamma, a) \cdot \mathrm{ord}_a(f) = \sum_a W(\gamma, a) \cdot \mathrm{ord}_a(g).$$

In particular, if the winding number of γ is 1 for points a in some region V, and the winding number is zero elsewhere, then f and g have the same number of zeros in V, counting multiplicities.

CHAPTER 10

Mayer–Vietoris

10a. The Boundary Map

For open sets U and V in the plane (or any topological space) we will define a homomorphism

$$\partial\colon H_1(U \cup V) \to H_0(U \cap V),$$

called the *boundary map*.[5] To do this we need a lemma.

Lemma 10.1. *If γ is a 1-cycle on $U \cup V$, there are 1-chains γ_1 on U and γ_2 on V such that $\gamma_1 + \gamma_2$ is homologous to γ on $U \cup V$.*

Proof. We know from Lemma 6.4(b) that if a path is subdivided, it is homologous to the sum of the paths into which it is divided. By the Lebesgue lemma, each path occurring in γ can be subdivided so that the image of each piece is in U or in V. So γ is homologous to a sum $\Sigma n_i \tau_i$, where each τ_i is a path in U or in V (or both). Then γ_1 can be taken to be the sum of those $n_i\tau_i$ for which τ_i is a path in U, and γ_2 can be the sum of the others. □

Construction of the boundary map $\partial\colon H_1(U \cup V) \to H_0(U \cap V)$. Recall that $H_1(U \cup V) = Z_1(U \cup V)/B_1(U \cup V)$ is the group of 1-cycles,

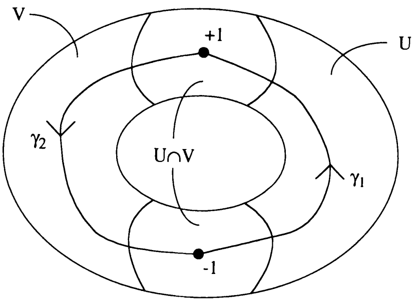
[5] This symbol ∂ is certainly overworked in this subject. So far we have used it for the boundaries of 1-chains and for maps of rectangles. Rather than introducing different notations for the different uses, we try for clarity by saying in each case to what sort of object the "∂" is being applied.

modulo the subgroup of 1-boundaries, on $U \cup V$. And $H_0(U \cap V) = Z_0(U \cap V)/B_0(U \cup V)$ is the group of 0-cycles modulo 0-boundaries on $U \cap V$. We will write $[\gamma]$ for the homology class in $H_1(U \cup V)$ defined by a 1-cycle γ on $U \cup V$, and we write $[\zeta]$ in $H_0(U \cap V)$ for the class defined by a zero cycle ζ on $U \cap V$.

Given a homology class α in $H_1(U \cup V)$, by the lemma, we may choose a 1-cycle that represents α and has the form $\gamma_1 + \gamma_2$, where γ_1 is a 1-chain on U and γ_2 is a 1-chain on V; in symbols, $[\gamma_1 + \gamma_2] = \alpha$. Since the boundary of this 1-cycle is zero, we have $\partial(\gamma_1) = -\partial(\gamma_2)$. This 0-cycle $\partial(\gamma_1) = -\partial(\gamma_2)$ is a 0-cycle on U and on V, so it is a 0-cycle on $U \cap V$. We define $\partial\colon H_1(U \cup V) \to H_0(U \cap V)$ by sending α to the class of this 0-cycle:

$$\partial(\alpha) \;=\; \partial([\gamma_1 + \gamma_2]) \;=\; [\partial(\gamma_1)] \;=\; -[\partial(\gamma_2)].$$

Note that, although the zero cycle $\partial(\gamma_1) = -\partial(\gamma_2)$ is a boundary on U and on V, it need not be a boundary on $U \cap V$:



The hardest part of our task is to show that this class $[\partial\gamma_1]$ is well defined.

Lemma 10.2. *The class of $\partial\gamma_1$ in $H_0(U \cap V)$ is independent of the choice of γ_1 and γ_2.*

Proof. Before giving the proof, we need a more canonical way to subdivide our 1-chains, by cutting each one exactly in half. For any path $\gamma\colon [0, 1] \to U$, define the 1-chain $S(\gamma)$ by the formula $S(\gamma) = \sigma + \tau$, where σ and τ are the restrictions of γ to the two halves of the interval, but rescaled as in Lemma 6.4(b). Extend this operator S linearly to all 1-chains $\gamma = \Sigma\, n_i\gamma_i$ by setting $S(\gamma) = \Sigma\, n_i S(\gamma_i)$. It follows from this definition that the boundary of the 1-chain $S(\gamma)$ is the same as the boundary of γ.

If $\Gamma\colon [0,1]\times[0,1]\to U$ is a map of the unit square, we can similarly subdivide Γ into four pieces $\Gamma^{(1)}$, $\Gamma^{(2)}$, $\Gamma^{(3)}$, and $\Gamma^{(4)}$ as indicated:

$\Gamma^{(3)}$	$\Gamma^{(4)}$
$\Gamma^{(1)}$	$\Gamma^{(2)}$

Each restriction is rescaled to be a map from $[0,1]\times[0,1]$ to U. In formulas,

$$\Gamma^{(1)}(t,s)=\Gamma(^1\!/_2t,\, ^1\!/_2s), \qquad \Gamma^{(2)}(t,s)=\Gamma(^1\!/_2+{}^1\!/_2t,\, ^1\!/_2s),$$
$$\Gamma^{(3)}(t,s)=\Gamma(^1\!/_2t,\, ^1\!/_2+{}^1\!/_2s), \qquad \Gamma^{(4)}(t,s)=\Gamma(^1\!/_2+{}^1\!/_2t,\, ^1\!/_2+{}^1\!/_2s).$$

It is clear from the picture, and easy to verify from the formulas, that this subdivision is compatible with taking the boundary, i.e.,

$$S(\partial\Gamma) \;=\; \partial\Gamma^{(1)}+\partial\Gamma^{(2)}+\partial\Gamma^{(3)}+\partial\Gamma^{(4)},$$

the inside boundaries canceling as usual.

Now we prove the lemma. Suppose the class α is also represented by the cycle $\gamma_1'+\gamma_2'$ for 1-chains γ_1' on U and γ_2' on V. Since both sums represent γ, we know that $(\gamma_1+\gamma_2)-(\gamma_1'+\gamma_2')$ is a boundary on $U\cup V$, so we can write

$$(\gamma_1+\gamma_2)-(\gamma_1'+\gamma_2') \;=\; \sum n_i\partial\Gamma_i$$

for some maps Γ_i from rectangles to $U\cup V$. We must show that $[\partial\gamma_1]=[\partial\gamma_1']$ in $H_0(U\cap V)$, i.e., that $\partial\gamma_1-\partial\gamma_1'$ is a boundary of some 1-chain on $U\cap V$. We apply the subdivision operator S to each side of the displayed equation. On the left, each of the four 1-chains is replaced by another with the same support and the same boundary. On the right, each $\partial\Gamma_i$ is replaced by a sum of the boundaries of the four subdivisions of Γ_i. So we have an equation of the same form, but with all the Γ_i's cut into quarters. The operator S can be applied again, which subdivides each of these quarters into quarters, and so on, dividing the smaller squares into quarters. When applied p times, we have an equation

$$S^p(\gamma_1)+S^p(\gamma_2)-S^p(\gamma_1')-S^p(\gamma_2') \;=\; \sum n_iS^p(\partial\Gamma_i).$$

By the Lebesgue lemma, the restrictions of Γ_i to small enough portions of the rectangles must be mapped into U or into V. It follows that for some large p the right side of this equation can be written in the form $\tau_1+\tau_2$, where τ_1 is a 1-boundary on U and τ_2 is a 1-boundary