V.G. Boltyanskiĭ

V.A. Efremovich

Intuitive Combinatorial Topology

Translated from the Russian by Abe Shenitzer

With the Editorial Assistance of John Stillwell

Springer

V.G. Boltyanskiĭ
CIMAT
A.P. 402
Guanajuato, Gto.
36000 Mexico
boltian@fractal.cimat.mx

V.A. Efremovich
(deceased)

Abe Shenitzer (translator)
Department of Mathematics
York University
North York, Ontario M3J IP3
Canada
shenitze@mathstat.yorku.ca

Mathematics Subject Classification (2000): 54-01, 14P25, 57N05, 55-01, 76A

Library of Congress Cataloging-in-Publication Data
Boltianskii, V. G. (Vladimir Grigor'evich), 1925–
Intuitive combinatory topology / V.G. Boltyanskii, V.A. Efremovich; translated by Abe Shenitzer.
p. cm. — (Universitext)
Includes bibliographical references and index.
ISBN 0-387-95114-8 (alk. paper)
1. Combinatorial topology. I. Efremovich, V. A. II. Title. III. Series.
QA612.B65 2001
514′.22—dc21 00-057420

Printed on acid-free paper.

Translated from the Russian, *Naglyadnaya topologiya* © 1982 Nauka.

Production managed by Michael Koy; manufacturing supervised by Erica Bresler.
Typeset by Archetype Publishing, Inc., Monticello, IL.
Printed and bound by R.R. Donnelly and Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-95114-8 SPIN 10777104

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

Translator's note. I wish to thank John Stillwell, Robert Burns, and David Kramer for eliminating a number of linguistic infelicities and technical flaws. Thanks are also due to Michael Koy, production editor, and to Ina Lindemann, editor, who saw to it that everything went as smoothly as possible.

Abe Shenitzer

Introduction by the Editor of the Russian Original

The elementary ideas of topology are based on direct observation of the world around us. It is clear that the geometric properties of a figure are not exhausted by its metric properties (such as lengths, angles, and so on); there are things outside the bounds of traditional geometry. Thus a curve (a rope, a wire, a long molecule) cannot be described by its length alone. Indeed, it can be closed or not, and if closed it can be knotted in complicated ways. Two or more closed curves can be linked in a variety of ways. Solids and their surfaces can have holes. What characterizes such properties of solids is that they are unaltered by deformations resulting from arbitrary distortions that do not involve tearing. Such properties are called topological. In addition to elementary geometric figures, many purely mathematical objects have topological properties, and it is this that determines their importance.

It is easier to determine the existence of topological properties of figures than it is to create a "calculus" of such properties, that is, to develop a branch of mathematics with exact concepts, rigorous rules and methods, as well as mathematical formulas for the representation of topological magnitudes.

The earliest important insights and exact topological relations are due to Euler, Gauss, and Riemann. But it is no exaggeration to say that topology as an independent discipline was created at the end of the nineteenth century by Henri Poincaré. The evolution of topology and the solution of its intrinsic problems turned out to be difficult and prolonged; in fact, it extended over seventy to eighty years. Many deep discoveries were made, which led in a number of cases to a revision of its foundations. Some of the greatest mathematicians, including Russians, took part in this process of development. (In the 1920s P.S. Uryson and P.S. Alexandrov established in Moscow the Soviet school of topology.) Until the end of the 1950s mathematicians in other areas regarded topology as a beautiful but useless plaything. I freely admit that as a student in the 1950s I chose topology as my future area of research because I was captivated by its beauty and "otherness" (compared with the traditional areas of mathematics), and that for a long time, until the late 1960s, I was dissatisfied with the nature of its development, marked as it was by a paucity of applications. Nonetheless, it is important to note that many beautiful topological results had by then been obtained in areas such as function theory and

complex analysis, qualitative theory of dynamical systems and partial differential equations, operator calculus, and even in algebra.

However, it was not until the early 1970s that topological methods began to penetrate strongly the apparatus of modern physics. Today their importance for different areas of physics is beyond doubt. In particular, topological methods are used in field theory and general relativity, in the physics of the anisotropy of solid media, in the physics of low temperatures, and in modern quantum theory. This makes it necessary to publish sufficiently elementary popular books on topology and its applications, accessible (at least in part) to high school seniors and beginning undergraduates interested in the natural sciences and technology.

The two eminent authors of this book, V.G. Boltyanskiĭ and V.A. Efremovich, have devoted many years to the popularization of topological ideas. The book includes a supplementary chapter dealing with an interesting application of topology to the theory of nematic liquid crystals. Its author, V.P. Mineev, has contributed significantly to the introduction of topological methods into theoretical physics.

I hope that this book will turn out to be very useful to a wide circle of readers.

S.P. Novikov

Introduction by the Authors

Topology is a relatively young and very important branch of mathematics. The famous German mathematician Hermann Weyl said that "the angel of topology and the devil of abstract algebra fight for the soul of each individual mathematical domain." He thus pointed out the remarkable subtlety and beauty of topology as well as the extent to which all of modern mathematics is interlaced in a remarkable way with the ideas of topology and algebra. In recent years topology has penetrated more and more into physics, chemistry, and biology. The reader will find an example of the use of topological ideas in physics in V.P. Mineev's supplement. However, it is difficult to enter the magical world of topology. Just as the scaffolding surrounding an unfinished building prevents one from perceiving the beauty of its design, so too the many tiresome details of the theory that fill books on topology prevent one from seeing with the mind's eye this beautiful mathematical structure. Even professional mathematicians often give up rather than face the difficulties barring the way to the mastery of topology (especially algebraic topology, whose elements are dealt with in the third chapter of this book).

All this makes it imperative to write popular books on topology. A first book of this kind was written in our country in the 1930s (P.S. Alexandrov and V.A. Efremovich, *Short Survey of the Fundamental Ideas of Topology*. Moscow, 1936). Then, beginning in 1957, our book *Short Survey of the Fundamental Concepts of Topology* appeared in installments in issues 2, 3, 4, and 6 of the Soviet journal *Mathematical Education* (translations of the book were published in Poland, Japan, and Hungary). Both of these books became bibliographical rarities long ago. Part of the material from our *Short Survey* is contained in the present book and represents V.A. Efremovich's contribution (he was the moving spirit behind the creation of the *Short Survey* and the popularization of topology). The major part of the text is new and was written by me. This gave me the opportunity to include a few recent results. I have also added more than 200 problems, for I think that the study of a scientific, or popular-scientific, book is useful only if the reader reflects on the issues it deals with.

I would like to take this opportunity to thank S.P. Novikov for his valuable remarks. I also wish to thank in advance all readers who take the trouble to express their opinions and to comment on the book.

V.G. Boltyanskiĭ

This book was prepared for publication by V.G. Boltyanskiĭ, who reworked and supplemented the material in our *Short Survey*. I wish to take this opportunity to express my deep thanks to him. I also wish to thank S.P. Novikov for his valuable comments and support.

V.A. Efremovich

Contents

Translator's Note v

Introduction by the Editor of the Russian Original vii

Introduction by the Authors ix

1 Topology of Curves 1
1.1. The Concept of Continuity 1
1.2. What Is Topology Concerned With? 4
1.3. The Simplest Topological Invarients 8
1.4. The Euler Characteristic of a Graph 11
1.5. Intersection Index 15
1.6. The Jordan Curve Theorem 19
1.7. What Is a Curve? 22
1.8. Peano Curves 28

2 Topology of Surfaces 31
2.1. Euler's Theorem 31
2.2. Surfaces 33
2.3. The Euler Characteristic of a Surface 38
2.4. Classification of Closed Orientable Surfaces 42
2.5. Classification of Closed Nonorientable Surfaces 48
2.6. Vector Fields on Surfaces 55
2.7. The Four Color Problem 60
2.8. Coloring Maps on Surfaces 62
2.9. Wild Spheres 66
2.10. Knots 70
2.11. Linking Numbers 76

3 Homotopy and Homology 81
3.1. Periods of Multivalued Functions 81

3.2. The Fundamental Group .....
3.3. Cell Decompositions and Polyhedra .....
3.4. Coverings .....
3.5. The Degree of a Mapping and the Fundamental Theorem of Algebra .....
3.6. Knot Groups .....
3.7. Cycles and Homology .....
3.8. Topological Products .....
3.9. Fiber Bundles .....
3.10. Morse Theory .....

Appendix A: Topological Objects in Nematic Liquid Crystals

A.1. Nematics .....
A.2. Disclination in the Nematic .....
A.3. Disclination and Topology .....
A.4. Singular Points .....
A.5. What Else Is There? .....

Bibliography

Index

1

Topology of Curves

1.1. The Concept of Continuity

The evolution of every area of mathematics begins with a fundamental idea, a fundamental concept that permeates the whole structure and determines its character. The fundamental concept of topology is *continuity*. We encounter it already in analysis. But there, as a result of its subordination to other concepts of analysis, its development is insignificant. It is in topology that continuity has been developed fully and in all possible ways. We give two examples that illustrate its application.

Example 1 We show that the cubic equation

$$x^3 + ax^2 + bx + c = 0 \tag{1}$$

with positive real coefficients a, b, and c has at least one real root.

We write equation (1) (for $x \neq 0$) in the form

$$x^3\left(1 + \frac{a}{x} + \frac{b}{x^2} + \frac{c}{x^3}\right) = 0. \tag{2}$$

For very large $|x|$ the fractions a/x, b/x^2, c/x^3 are very small. Hence the expression in parentheses differs very little from 1 and thus is positive. It follows that for very large $|x|$ the left side of equation (2) has the same sign as x^3, i.e., the same sign as x. In other words, the left side of equation (1) is negative for large negative x (x_0 in Figure 1) and positive for large positive x (x_1 in Figure 1). The graph of the function is a *continuous curve*. It follows that when it crosses the x-axis (from p_0 to p_1) it must intersect it in at least one point. The point x' of intersection of the graph with the x-axis yields a root of equation (1).

Problems

1. Show that every equation of odd degree with real coefficients has at least one real root.
2. Show that for negative c equation (1) has at least one positive root.

Example 2 We show that it is possible to circumscribe a square about every closed curve K.

To this end we draw two parallel straight lines l and l' so that the curve K is contained in the strip between them. Then we translate them, parallel to their original positions, until they touch K. The resulting parallel straight lines m and

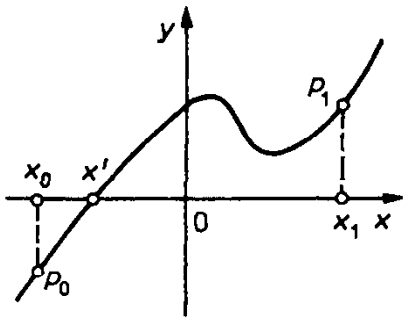


FIGURE 1.

m' (Figure 2a) are called *support lines* of K. Now we draw two more support lines perpendicular to l (Figure 2b) and obtain a rectangle $ABCD$ circumscribed about K. It remains to show that for a proper choice of direction of the straight line l the rectangle becomes a square.

Let $h_1(l)$ be the length of the side AD parallel to l, and $h_2(l)$ the length of the side AB perpendicular to l. The rectangle becomes a square if $h_1(l) - h_2(l) = 0$.

Let l^* be a straight line perpendicular to l. The circumscribed rectangle with sides parallel and perpendicular to l^* coincides with the rectangle $ABCD$, but now the side AB is parallel to l^* and the side AD is perpendicular to l^*, i.e., $h_1(l^*) = |AB| = h_2(l)$ and $h_2(l^*) = |AD| = h_1(l)$. Hence

$$h_1(l^*) - h_2(l^*) = -(h_1(l) - h_2(l)). \tag{3}$$

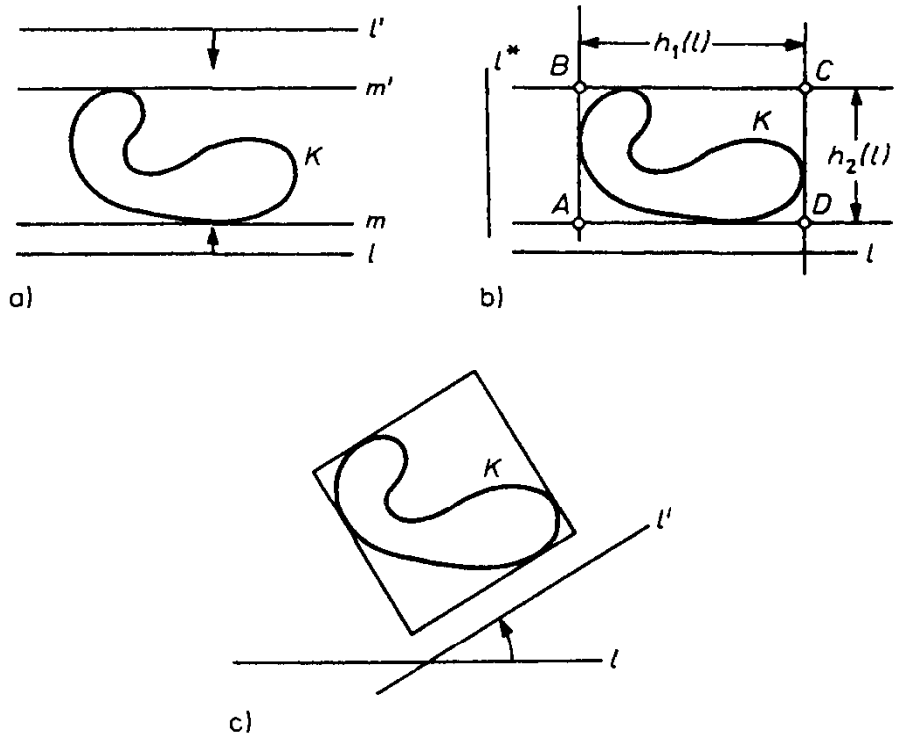


FIGURE 2.

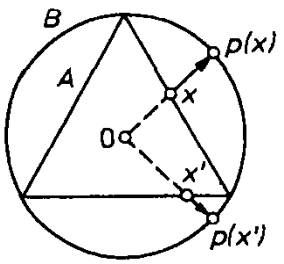


FIGURE 3.

Now we rotate l until it coincides with l^*. As we do this, the circumscribed rectangle changes continuously. The difference $h_1(l)-h_2(l)$ depends continuously on l. But when we change from l to l^*, this difference changes sign (see (3)). It follows that during its continuous change the difference takes on the value 0 (for a suitable l), i.e., the rectangle becomes a square (Figure 2c).

Problems

3. Show that it is possible to circumscribe a rhombus with angle 60° about every closed curve K.
4. Show that if the diameter of a plane figure does not exceed a certain value d (i.e., the distance between any two of its points does not exceed d), then there exists a regular hexagon with distance d between opposite sides that contains this figure.
5. Show that if the diameter of a figure in space is less than or equal to d, then there exists a regular octahedron that contains the figure and has the property that the distance between its opposite faces is d.

In topology we consider functions of the most general kind. To prescribe a function is to associate with every point x of a set A (the *domain* of the function) a suitably defined point $f(x)$ of a set B. In this case we also say that a *mapping* f of the set A into the set B is given and write briefly $f\colon A \to B$.

Example 3 Let A be the boundary of an equilateral triangle and B its circumcircle (Figure 3). The *central projection* p of the points of A onto the circle is a mapping $p\colon A \to B$.

A function $f\colon A \to B$ is said to be *continuous* at a point $x_0 \in A$ if whenever x is "close" to (differs by"little" from) x_0, $f(x)$ is "close" to (differs by "little" from) $f(x_0)$.

More precisely, *a function f is said to be continuous at x_0 if for every number $\epsilon > 0$ there is a number $\delta > 0$ such that whenever x differs from x_0 by less than δ, $f(x)$ differs from $f(x_0)$ by less than ϵ*. Of course, for this definition to make sense notions of distance between points must be defined in the sets A and B.

To understand better the meaning of continuity of a mapping we consider an example of a break, i.e., of the undoing of its continuity. We take a rubber ring and stretch it until it breaks at one of its points. One part of Figure 4a shows the

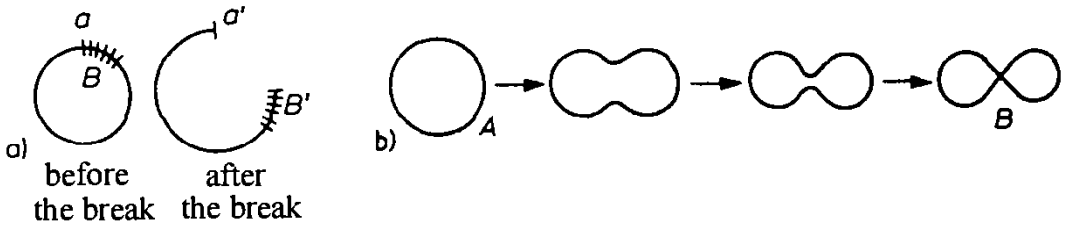


FIGURE 4.

unbroken ring, the breaking point a, and a portion B of the ring initially "close" to a (at a distance 0 from it), while the other part shows the broken ring and the new positions a' and B' of a and B respectively. We can describe the breaking event by saying that while B was close to a (we write $B\delta a$) before the ring broke, B' is *not* close to a'. Now the following definition becomes understandable:

Let $f\colon x \to x'$ be a mapping of some figure. Let B denote any part of the mapped figure and let $B' = f(B)$ denote the *image* of B. The mapping f is said to be continuous at a point a if $B\delta a$ implies $B'\delta a'$.

One can show that this definition is equivalent to the previous one.

If a mapping $f\colon A \to B$ is continuous at every point x_0 of A, then we say that f is *continuous*. Intuitively, we can think of a continuous mapping as a mapping that takes close points of A to close points of B, i.e., f does not disturb the cohesion of A. Note that different points of A may be taken to the same point of B (they may be "glued together" as in Figure 4b, and so on).

Problems

6. Show that the mapping in Example 3 is continuous.

7. For an arbitrary real a let $f(a)$ denote the largest root of the equation $x^3 - 3x + a = 0$. Is the function $f(x)$ continuous?

1.2. What Is Topology Concerned With?

A mapping $f\colon A \to B$ is said to be *one-to-one and onto*, or, briefly, *bijective*, if the preimage of every point of B is exactly one point of A. This means two things: First, no two points of A are mapped on the same point of B (i.e., the mapping does not "glue points together"), and second, the mapping associates with every point of B a point of A (i.e., A is mapped onto all of B rather than onto a part of it). For a bijective mapping $f\colon A \to B$ we can define the inverse mapping $f^{-1}\colon B \to A$ (that associates with every point $y \in B$ the point in A mapped by f on y).

A mapping $f\colon A \to B$ is said to be *homeomorphic* (or a *homeomorphism*) if it is both bijective and bicontinuous, i.e., f as well as the inverse mapping f^{-1} are continuous.

Intuitively, one can think of a homeomorphism as a mapping of a set on another

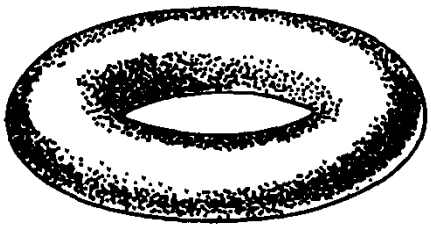


FIGURE 5.

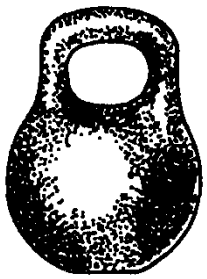


FIGURE 6.

set that involves no tearing and no gluing together. We can think of the figures A and B as made up of very strong and elastic material. We are allowed to stretch and bend each figure but not to tear it or glue it together. If, using the permissible modifications, we can make the (modified) figures coincide, then we say that they are homeomorphic. Thus the boundary of a triangle (or, more generally, of any polygon) is homeomorphic to a circle.

Translator's note. "No tearing" means that points initially together must not come apart (this is the continuity condition), and "no gluing" means that points initially apart should not come together (this is the bijective condition). It is, however, possible to cut points apart, do some deformation, then *rejoin* the separated points. See comment on p. 11.

Example 4 The surfaces of a ball, cube, and cylinder are homeomorphic to one another but not to a torus (which can be thought of intuitively as the surface of an automobile tire (Figure 5)). The surface of the weight in Figure 6 *is* homeomorphic to a torus.

Example 5 Consider some of the letters of the Latin alphabet. The letters **I, J, L, M, N, S, U, V, W, Z** are homeomorphic. The letters **E, F, T, Y** are homeomorphic to one another but not to any of the letters in the first group. Finally, the letter **Q** is not homeomorphic to any other letter of the Latin alphabet.

Example 6 Consider Figure 7 below. A is a semicircle centered at o without its endpoints m and n, and B is tangent to A and parallel to the diameter mn. The central projection p: $A \to B$ with center o is a homeomorphism. The semicircle is homeomorphic to an open segment, i.e., a segment without its endpoints (it can be straightened out). It follows that *a straight line is homeomorphic to an open segment.*

Problems

8. Show that the surface of a cylinder without its top (a "cup") is homeomorphic to a disk.

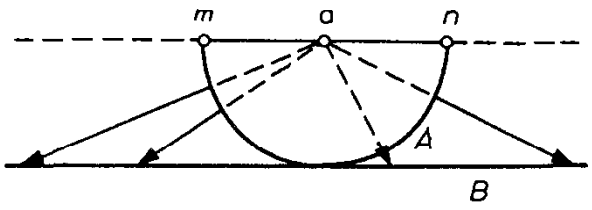


FIGURE 7.

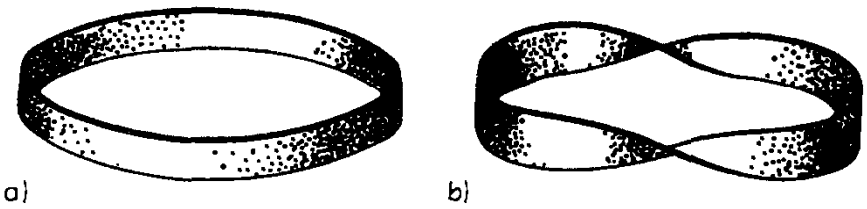


FIGURE 8.

9. Show that a plane is homeomorphic to an open disk (i.e., a disk without its circular boundary).

10. Show that the figures in Figure 8 (a band homeomorphic to the lateral surface of a cylinder and a band with two twists) are homeomorphic.

[*Note.* The figures in Figure 8 (as well as in others) seem to have thickness—seem to have been made of some material. The reader must keep in mind that this done for intuitive appeal. We are actually dealing with "mathematical" surfaces without thickness.]

It is instructive to compare the notions of homeomorphism and congruence of figures. In geometry we study mappings that preserve distances between points. These mappings are called (rigid) *motions*. A motion transfers a figure from place to place without changing the distances between its points. Two figures that can be made to coincide by means of motions are said to be *congruent*. They are regarded as copies of each other, i.e., as being the same from a geometric standpoint. In topology we study more general mappings, namely homeomorphisms. Two homeomorphic figures are regarded as copies of each other, i.e., as being the same from a topological standpoint. Properties of figures unchanged by homeomorphisms are called topological properties, or *topological invariants*. Topology investigates topological properties of figures.

Problems

11. A figure A can consist of finitely many or of infinitely many points. In the first case $n(A)$ will denote the number of its points. In the second case we will write $n(A) = \infty$. Is $n(A)$ a topological invariant?

12. If a figure is homeomorphic to a plane figure, then we say that it is "embeddable in a plane." Is the property of being embeddable in a plane a topological invariant?

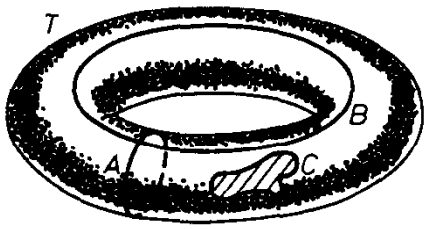


FIGURE 9.

FIGURE 10.

It would be incorrect to think that it is possible to bring any two homeomorphic figures in space into coincidence by bending stretching and moving (without cutting and gluing together). For example, this cannot be done in the case of the figures represented in Figure 8, for the two figures are differently embedded in space. In order to bring them into coincidence by permissible modifications we must first cut the figure in 8a, twist it twice, and glue together the points that were originally together. This procedure (cutting and appropriate regluing after stretching and after modifying the positions of parts of a figure) is often used in topology to demonstrate the homeomorphism of two figures.

The sameness of the disposition of two figures in space (or in a figure that contains them) is made precise by the concept of *isotopy*. We say of two figures A and B that they are *isotopic* in a figure P that contains them (or *topologically equally disposed* in P) if there is a homeomorphism of P that takes A to B. The bands in Figure 8 are homeomorphic but not isotopic in space (this will be proved later). We can speak of embedding properties if we are given two figures of which one contains the other. Topology is also concerned with the investigation of embedding properties (i.e., with the investigation of topological invariants of pairs of figures).

Problems

13. The curve A in Figure 9 does not divide the torus into two parts, whereas the curve C does. Are A and C isotopic in the figure T? Are they isotopic in three-dimensional space?

14. Show that the *meridian* A and the *parallel* B on the torus T in Figure 9 are isotopic in T.

15. Show that any two points of the figure eight (Figure 10) other than the point x are isotopic.

1.3. The Simplest Topological Invariants

We mentioned in Example 4 that the sphere is not homeomorphic to a torus. But how can one prove that two figures are not homeomorphic? Obviously, the mere fact that someone has been unable to find a homeomorphism between these two figures does not imply that such a homeomorphism does not exist.

To prove that two figures are not homeomorphic one uses topological invariants. For example, if there is a rule that associates with every figure a number such that the numbers associated with homeomorphic figures are always the same, then this number is linked with some property of figures unchanged by a homeomorphism, and thus is a topological invariant. If the numbers associated with two figures are different, then these figures are not homeomorphic.

Example 7 Each of the two lowercase letters **i** and **j** consists of two disconnected pieces, while each of the remaining letters of the Latin alphabet consists of a single connected piece. The number of connected pieces a figure is made up of (the *number of components* of a figure) is a topological invariant; homeomorphic figures have the same number of components. For example, the letters **i** and **b** are not homeomorphic.

Example 8 The figure eight represented in Figure 10 contains a point x whose removal results in a disconnected figure, i.e., a figure with more than one component (see Figure 11). A point with this property is called a *cut point* of the figure; no point x' other than x has this property (see Figure 12).

The property of being, or of not being, a cut point is topological: if x is a cut point of A and $f\colon A \to B$ is a homeomorphism, then $f(x)$ is a cut point of B. It follows that the number of cut points of a figure is a topological property, and so is the number of points that are not cut points.

Problems

16. For each (capital) letter of the Latin alphabet give the number of cut points as well as the number of points that are not cut points. Show that no two of the letters **D**, **I**, **E**, **P** are homeomorphic.

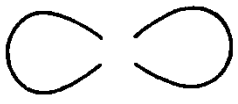


FIGURE 11.

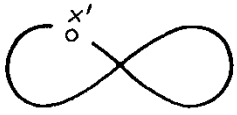


FIGURE 12.

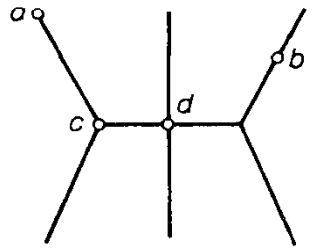


FIGURE 13.

17. Show that for every natural number n there is a figure with just n cut points as well as a figure with exactly n points that are not cut points.

Example 9 Let A be a figure made up of a finite number of arcs and let x be one of its points. The number of arcs issuing from x is called the *degree* of x in A. Consider the figure shown in Figure 13. The respective degrees of the points a, b, c, d in this figure are $1, 2, 3, 4$. The numbers of points of A with degrees $1, 2, 3, 4$, and so on, are different topological invariants of A.

Problems

18. Show that the letters **E** and **K** are not homeomorphic.

19. Give a necessary and sufficient condition for a figure made up of a finite number of arcs to be homeomorphic to a circle.

Figures consisting of a finite number of arcs are called in topology *finite graphs*. A finite graph contains a finite number of *vertices* some of which may be connected by arcs free of self-intersections, called the *edges* of the graph. Two vertices may be connected by more than one edge. *Loops*, i.e., edges that begin and end at the same vertex, are admissible.

Problems

20. Let G be a finite graph. Let $a_k(G)$ denote the number of vertices with degree k. Show that the number of edges is $\frac{1}{2}(a_1(G) + 2a_2(G) + 3a_3(G) + \cdots)$.

21. Show that in every finite graph the number of vertices with odd degree is even.

Example 10 A graph is called *Eulerian* if it can be "drawn in a single sweep," i.e., if it can be traversed in a continuous way without going over any of the edges twice. It is clear that for a graph the property of being Eulerian is topological. It turns out, however, that this is not a new invariant, and that it can be expressed in terms of the concept of the degree of a point (see Problem 24).

Problems

22. Show that if every vertex of a finite graph has degree at least 2, then the graph contains a curve homeomorphic to a circle and consisting of edges only.

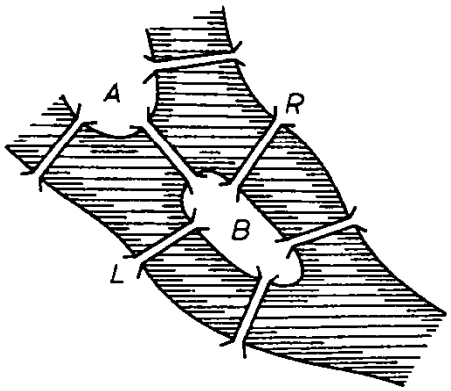


FIGURE 14.

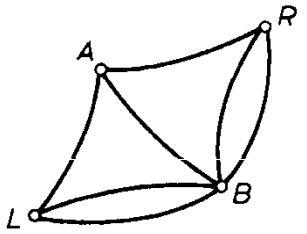


FIGURE 15.

23. Show that if all the vertices of a finite connected graph have even degree, then it is Eulerian and can be traversed beginning and ending at any of its vertices.

24. Show that a connected graph is Eulerian if and only if it contains at most two vertices with odd degree.

There is a close connection between Eulerian graphs and the *Königsberg bridge problem* first investigated by Euler. At that time there were in Königsberg (today's Kaliningrad) seven bridges (Figure 14) over the river Pregel. The question was whether it was possible to take a walk in town and cross each bridge exactly once. To answer this question we associate with the map of the town a graph: L and R denote the left and right shores of the river and A and B the islands. The edges of the graph correspond to the bridges (Figure 15). All four vertices of the graph have odd degree, and thus it is not Eulerian. It follows that one cannot walk and cross each bridge exactly once.

Problems

25. Show that if one added one more bridge that fits the city map in Figure 14, then the resulting map would be such that one could cross each bridge exactly once.

26. A *complete graph* is a finite graph without loops in which every two vertices are connected by exactly one edge. When is a complete graph Eulerian?

1.4. The Euler Characteristic of a Graph

Each graph can be built up by successive addition of edges. For example, we can number the edges of the required graph, then draw the first, second, etc., edges.

Example 11 We want to construct the graph in Figure 16. Its edges are numbered (some of the edge lines are broken to suggest their possible location in space).

The numbering of the edges in Figure 16 is such that at each step of drawing the edges in the order indicated by the numbering we obtain a *connected* graph. But if we reversed the numbering order, then, at a certain step, we would obtain a disconnected graph consisting of three isolated edges. This disconnected graph would become connected only after we drew additional edges. The following is a natural question: Given a connected graph, does there always exist a numbering of its edges such that we always get a connected graph when we draw the edges in the order indicated by the numbering?

There is an affirmative answer to this question (see Problem 28). In other words, *every connected graph can be obtained as follows*: *We choose an edge, then we add another edge so that the resulting graph is connected, then another edge* (*so that the resulting graph is connected*), etc. This insight can be referred to as the theorem on drawing connected graphs.

Problems

27. Show that every connected graph can be drawn "in a single sweep," provided we stipulate that each edge should be traversed exactly twice.

28. Use Problem 27 to obtain a proof of the theorem on drawing connected graphs.

29. Show that any two vertices of a connected graph can be joined by a *simple trail* of edges, i.e., a trail (of edges) whose union is homeomorphic to a segment.

Hint. If the trail joining a and b passes twice through a vertex c then it contains a *closed trail* (one that begins and ends at c) which can be eliminated.

30. Show that if it is possible to join any two vertices of a graph G by at least two simple trails, then this graph has no vertex of degree 1. Is the converse also true?

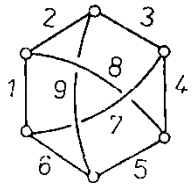


FIGURE 16.

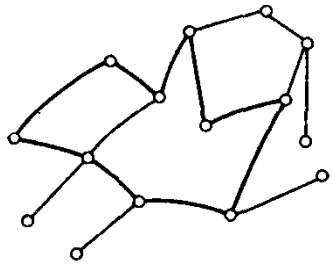


FIGURE 17.

A *circuit* in a graph is a closed trail that is homeomorphic to a circle (Figure 17). A connected graph without circuits is called a *tree*. We show that for a tree with V vertices and E edges we have:

$$V - E = 1. \tag{4}$$

The proof is by induction on the number E of edges. Equation (4) holds for $E = 1$, i.e., when the tree consists of one edge and two vertices. Suppose (4) holds for $E = n$. Let G be a tree with $n + 1$ edges. Since G is connected, it can be obtained from a connected graph G' with n edges by adding to it an extra edge r (this follows from the theorem on drawing connected graphs). Since G' has no circuits it is a tree. Since it has n edges, equation (4) holds for it. But then G' has $n + 1$ vertices. Only one vertex of the added edge r is a vertex of G' (otherwise, we could choose in G' a simple trail joining a and b, add to it the edge r, and thus obtain a circuit in G; see Figure 19). Thus adding r to G' has added just one new vertex (see Figure 20). But now G has $n + 2$ vertices and $n + 1$ edges, i.e., equation (4) holds for it. Since (4) holds for $E = n + 1$, it holds for every tree.

Let G be a graph with V vertices and E edges. We call the difference $V - E$ the *Euler characteristic* of G and denote it by $\chi(G)$. Thus the *Euler characteristic of any tree is* 1.

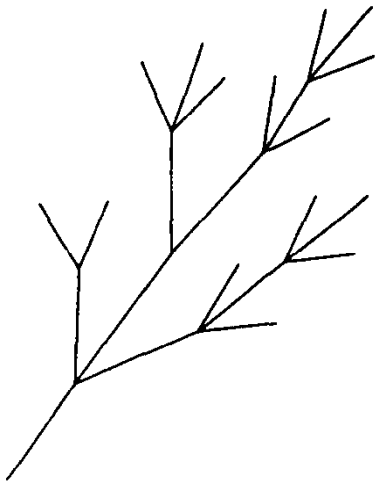


FIGURE 18.

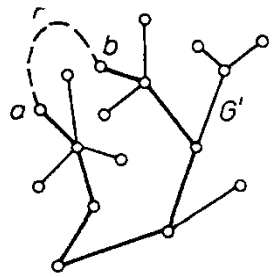


FIGURE 19.

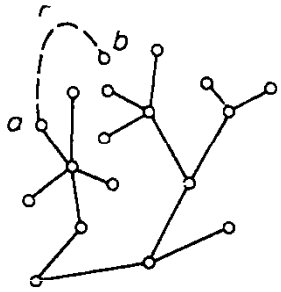


FIGURE 20.

Problems

31. A graph without circuits is called a *forest*. Show that the number of trees that "grow" in a forest G (i.e., the number of components of G) is $\chi(G)$.

32. Show that any two vertices of a tree can be joined by just one simple trail. Is the converse assertion also true?

Let G be a connected graph that is not a tree. Then G contains a circuit. Let r_1 in Figure 21 be an edge of this circuit. If we erase r_1, then the resulting graph G' is still connected (the end vertices of r_1 in G' are connected by the simple trail that results when we erase r_1 from the circuit). G and G' have the same vertices. If G' in Figure 22 is not yet a tree, i.e., if it contains a circuit, then we erase an edge r_2 of this circuit and obtain a connected graph G'' with the same vertices as G, and so on. After erasing edges $r_1, r_2, \ldots, r_k$ (we will occasionally refer to them as erasable edges) we end up with a graph G^* that has the same vertices as G and is acyclic, and thus a tree. We call G^* a *spanning tree* of G.

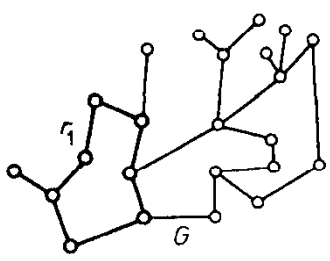


FIGURE 21.

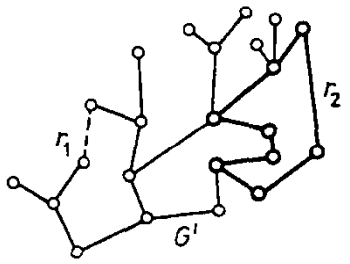


FIGURE 22.

Let V be the number of vertices of G, and thus also of G^*. Equation (4) implies that G^* has $V - 1$ edges. But then G has $V - 1 + k$ edges. Thus

$$\chi(G) = V - (V - 1 + k) = 1 - k. \tag{5}$$

Since $k \geq 1$, $\chi(G) \leq 0$. It follows that *if G is a connected graph, then $\chi(G) \leq 1$; $\chi(G) = 1$ if and only if G is a tree.*

According to equation (5), the number of erasable edges is $k = 1 - \chi(G)$. Thus to obtain G we add to one of its spanning trees $1 - \chi(G)$ erasable edges, each of which joins two vertices (possibly coincident in case of a loop) of the spanning tree.

Problems

33. If it is possible to obtain a connected graph G by adding loops to a suitable tree, then the spanning tree of G is uniquely determined and coincides with this tree. Is the converse statement true?

34. Show that if a graph G consists of l components, then $\chi(G) \leq l$. When does the equality $\chi(G) = l$ hold?

35. We say that a current is defined on a graph if each of its edges is assigned a direction and a nonnegative number (its current) such that the Kirchhoff vertex theorem holds at each vertex: The sum of the incoming currents is equal to the sum of the outgoing currents. Show that the only current on a tree is the trivial current, i.e., the current all of whose edge currents are 0.

36. Let G be a connected graph and $r_1, r_2, \ldots, r_k$ the edges erased in going from G to its spanning tree G^*. Define arbitrary currents on $r_1, r_2, \ldots, r_k$. Show that it is possible to define currents on the remaining edges so that we end up with a current on G and that this can be done in just one way.
Hint. Let r be any erasable edge. Then there is a unique circuit that contains this edge and no other erasable edge. With each edge of this circuit we associate a current whose magnitude and direction agree with the magnitude and direction of the current on r. Now we assign to each edge the sum of all circuit currents of all circuits that contain the edge in question. This yields the required current on G. (For each edge we form the algebraic sum of the currents involved. The direction of the edge is determined by the sign of this sum.) If there were two different currents that

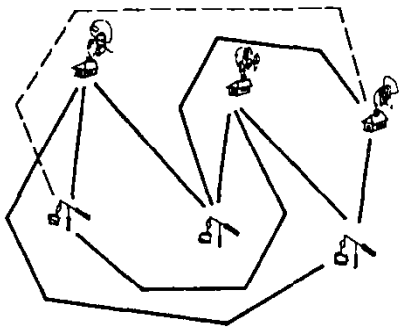


FIGURE 23.

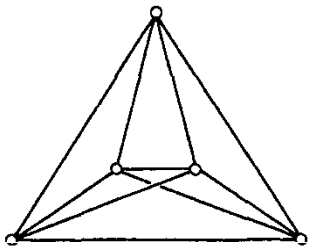


FIGURE 24.

agreed on the erasable edges, then their difference would yield a nontrivial current on the spanning tree G^*.

1.5. Intersection Index

In the next two examples we consider graphs that cannot be embedded in the plane.

Example 12 ("houses and wells"). H_1, H_2, H_3 (houses) and B_1, B_2, B_3 (wells) are six points on the plane. Is it possible to join each house to each well by noncrossing paths on the plane? The answer is no; if all but one path have been "built" (Figure 23), there is "no space" on the plane for the last path. Thus the graph in Figure 23 is nonplanar.

Example 13 Let P_2 be a complete graph with five vertices. One of the edges in Figure 24 is broken; there is "no space" on the plane for it. Thus P_2 is also nonplanar.

It is of interest that the graphs P_1 and P_2 are test graphs for whether a graph can or cannot be embedded in the plane. Specifically, the Polish mathematician K. Kuratowski proved that *if a graph cannot be embedded in the plane, then it contains a graph that is homeomorphic to P_1 or to P_2.*

Problems

37. Show that the graph in Example 11 (Figure 16) is nonplanar.

38. Let the edges of a graph be the sides and shortest diagonals (raised above the plane to avoid crossing) of a regular n-gon. Show that the graph is planar for even n and nonplanar for odd $n > 3$.

39. Let the edges of a graph be the sides and longest diagonals (raised above the plane to avoid crossing) of a regular $2n$-gon. Show that while this graph is nonplanar for $n \geq 3$, it can placed on a torus.

40. Let the edges of a graph be the sides and longest diagonals (raised above the plane to avoid crossing) of a regular $(2n + 1)$-gon. Show that this graph is nonplanar for $n \geq 2$. Can it be placed on a torus?

Of course, the remarks in Examples 12 and 13 (to the effect that at some point there is "no more space" on the plane) were meant to make the relevant results plausible. A rigorous proof that P_1 and P_2 are not embeddable in the plane will be given below.

Let a and b be two intervals on the plane such that neither contains the endpoints of the other. If a and b intersect, then we write $J(a, b) = 1$. Otherwise, we write $J(a, b) = 0$. We call $J(a, b)$ the *intersection index of a and b*.

We will call a finite set of intervals on the plane a *chain*. The elements of a chain are its *members*, and their endpoints are its *vertices*.

Let x and y be chains such that neither contains vertices of the other. Let $a_1, \ldots, a_m$ be the members of x and $b_1, \ldots, b_n$ the members of y. We write $J(x, y) = 0$ if the sum $\sum_{i,j} J(a_i, b_j)$ is even and $J(x, y) = 1$ if it is odd. We call the number $J(x, y)$ the *intersection index of the chains x and y* (more precisely, the intersection index modulo 2).

A chain each of whose vertices is incident with an even number of members is called a *cycle* (modulo 2). We will show that *the intersection index of two cycles is always* 0.

Since every vertex of a cycle has degree at least 2, the cycle must contain a part that is homeomorphic to a circle (see Problem 22). If we remove this part from the cycle, then what is left is still a cycle (for each vertex has even degree). In this cycle we can again find a part that is homeomorphic to a circle, and so on. Thus a cycle can be represented as the union of a finite number of parts each of which is homeomorphic to a circle (and no two such parts share intervals).

We see that in order to show that the intersection index of two cycles x and y on the plane is always 0 it suffices to prove this for the case where x and y are both homeomorphic to a circle. By moving x and y ever so slightly (so as not to change their intersection index) we can achieve that no member of x is parallel to a member of y. At this stage, we choose a straight line l that is not parallel to any straight line that joins a vertex of the cycle x to a vertex of the cycle y.

Now we move the cycle x (thought of as a rigid body) parallel to l (Figure 25). The only time the intersection index $J(x, y)$ can change is when a vertex of one cycle ends up on a side of the other cycle (the choice of l guarantees that no vertex of one cycle can coincide with a vertex of the other cycle). But when a member a of the cycle x passes through a vertex q of the cycle y, the parity of the number of intersections does not change (see Figures 26–28). A similar statement holds if

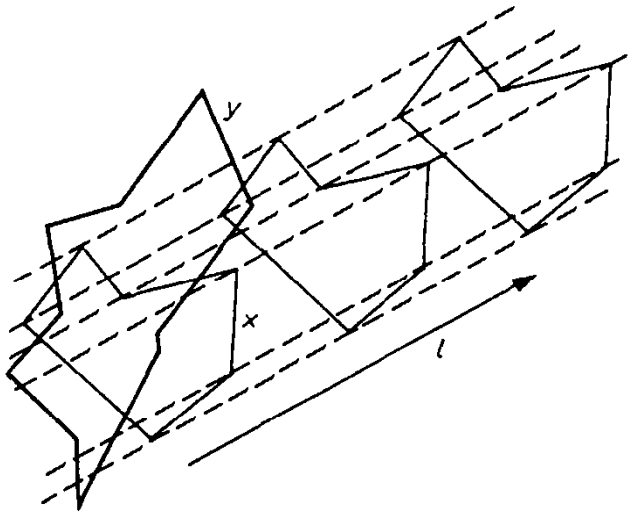


FIGURE 25.

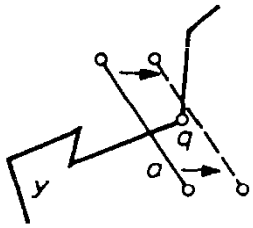


FIGURE 26.

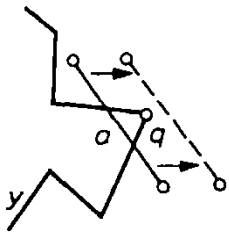


FIGURE 27.

a vertex of the cycle x passes through a segment of the cycle y. It follows that the intersection index $J(x, y)$ does not change. Finally, as a result of the motion of x, there comes a moment when the cycles x and y are disjoint (Figure 25), i.e., their intersection index is 0. But then, initially, we must also have had $J(x, y) = 0$.

We are now in a position to prove that the graph P_1 (Example 12) is not embeddable in a plane. Call two paths *separate* if they lead from different houses to different wells. Without barring intersections, we draw all needed paths (in the form of polygonal trails) and denote by I the total number of points of intersection of all pairs of separate paths. We will show that I is odd for all locations of the paths.

Suppose we vary the location of, say, the path H_1B_1. We denote this path in its initial location by x and in its new location by x' (see Figure 29). The four paths that join the houses H_2 and H_3 to the wells B_2 and B_3 are separate from H_1B_1 and

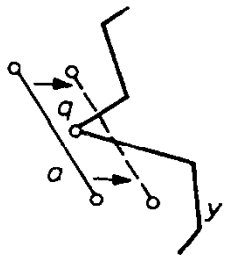


FIGURE 28.

form the cycle $H_2B_2H_3B_3H_2$, to be denoted by y. Together, the paths x and x' also form a cycle. Since the intersection index of two cycles is always 0, it follows that $J(x, y) = J(x', y)$. In other words, the number of crossing points of the path x and the cycle y (i.e., the paths separate from x) has the same parity as the number of crossing points of the path x' and the cycle y. Thus replacing x by x' does not alter the parity of I.

By now it is clear that *the number I has the same parity for all locations of the paths on the plane.* Indeed, replacing one path of the first representation by the corresponding path of the second representation, then the next path, and so on, is tantamount to the stepwise replacement of the first representation by the second, and—as was proved—the parity of I remains unchanged in the process.

Since there is just one crossing point in Figure 23, it follows that I is odd for all locations of the paths. Thus it is not possible to place the paths so as to avoid crossings (if this could be done, then we would have $I = 0$). We have proved that the graph P_1 is not embeddable in the plane.

Problems

41. Show that the graph P_2 in Example 13 is not embeddable in the plane.

42. Show that (just as on the plane) the intersection index of two cycles on a sphere is 0. Show that there are two cycles on a torus whose intersection index is 1.

Let a and b be two oriented intervals neither of which contains a vertex of the other. Suppose that we traverse a in the direction that agrees with its orientation. We write $J(a, b) = 1$ or $J(a, b) = -1$ according as the interval b crosses a from right to left or from left to right. If the two intervals do not cross then we write $J(a, b) = 0$. The number $J(a, b)$ is called the intersection index of the oriented intervals a and b.

We called the chains considered earlier chains modulo 2. Now we consider finite sets of oriented intervals on the plane and call such a set a chain (more precisely, an *integral chain*). Let $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$ be the oriented intervals that make up the integral chains x and y respectively. Then, as before, we define *the intersection index of the ordered pair of integral chains x and y* by the formula

$$J(x, y) = \sum_{i,j} J(a_i, b_j).$$

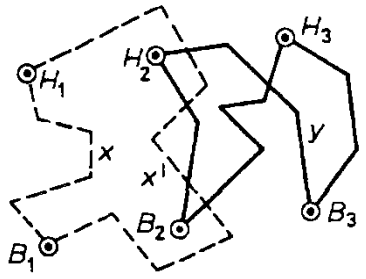


FIGURE 29.

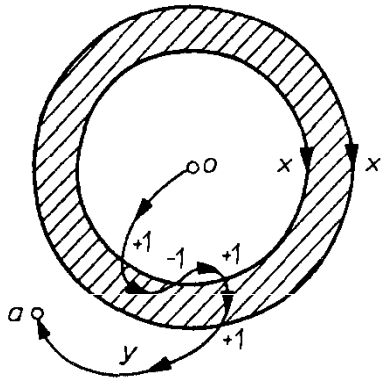


FIGURE 30.

Finally, we call a chain a *cycle* (more precisely, an *integral cycle*) if at each vertex the number of oriented intervals that enter it is equal to the number of oriented intervals that leave it.

Problems

43. We call a closed polygonal trail an *oriented boundary* if it is homeomorphic to a circle and if each of its segments is oriented by an arrow (so that at each vertex one segment leads in and one segment leads out). An oriented boundary is a cycle. Show that every (integral) cycle can be represented as the union of finitely many oriented boundaries no two of which share a segment.

44. Show that the intersection index of two integral cycles is 0.

45. The cycle x in Figure 30 consists of two disjoint oriented boundaries. Show that a point a is in the exterior of the annulus if and only the condition $J(x, y) = 2$ holds for every oriented polygonal trail that joins o to a. When is a in the crosshatched region of the annulus?

1.6. The Jordan Curve Theorem

We showed that the intersection index of two cycles on the plane is 0. The reader may have thought about the following, simpler, proof of this fact: At every intersection point the closed polygonal trail x goes either from the interior of the closed

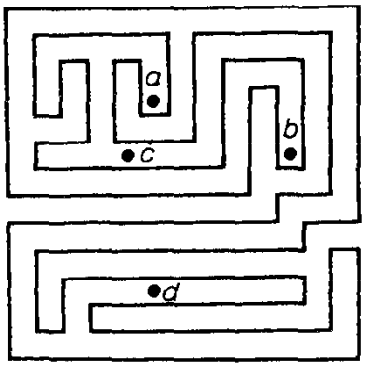


FIGURE 31.

polygonal trail y to its exterior or, conversely, from its exterior to its interior. Since x gets into the interior of y as many times as it gets out of it (these occurrences alternate), the number of crossings is even.

For this proof to be correct, what must be presupposed is a clear definition of the concept of the *interior of a region*. But this concept is not as simple as it appears to be at first sight. It is elucidated in the next paragraph.

By a *simple closed curve* we mean a closed curve homeomorphic to a circle. The Jordan curve theorem states that *every simple closed curve on the plane divides it into two regions* (its interior and exterior). To clarify the content of the theorem we consider two points p and q not on the simple closed curve. If we can join p and q by a polygonal trail that does not meet l, then we say that *p and q lie in the same region relative to l*. If every polygonal trail joining p and q meets l, then we say that p and q lie in different regions. The apparent obviousness of the Jordan curve theorem is due to the fact that the first curves one tends to think of are curves such as a circle, the boundary of a convex polygon, and so on.

Example 14 Consider the points a, b, c, d and the simple closed polygonal trail in Figure 31. At first sight it is anything but obvious which of these points lies in which of the two regions.

We will now prove the Jordan curve theorem for the case of a simple closed polygonal trail.

Let $b_1, b_2, \ldots, b_n$ be the edges, in this order, of the simple closed polygonal trail l in Figure 32. Let p and p' be two points symmetric with respect to b_1. Beginning at p we draw a segment parallel to b_1 just long enough to meet the bisector of the angle between b_1 and b_2. Beginning at that point we draw a segment parallel to b_2 just long enough to meet the bisector of the angle between b_2 and b_3, and so on. In this way we obtain a polygonal trail x whose edges are equidistant from the corresponding edges of l. If the distance $|pp'|$ is short enough, then x does not meet l, and if we traverse x beginning at p, then we end up at p or at p'. We claim that x cannot lead to p': If it did, then we could add the segment pp' to x and obtain a cycle that intersects the cycle l in just one point. This would mean that the intersection index of x and l is 1, which is impossible. Thus x is a closed polygonal trail that begins at p, goes around l once, and returns to p. Similarly,

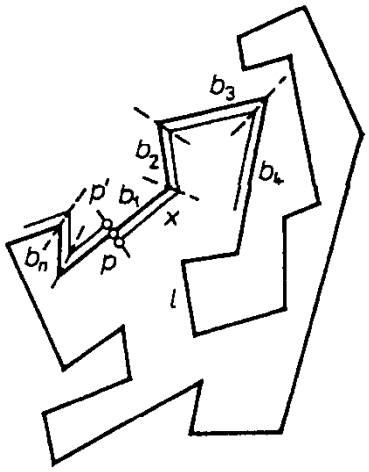


FIGURE 32.

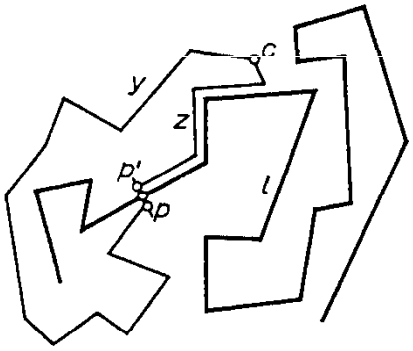


FIGURE 33.

we obtain a closed polygonal trail x' that begins at p', goes around l once, and returns to p'.

Now let c be a point not on l. We can join c to p or p' without meeting l. To do this we draw a ray from c that intersects x or x', traverse the ray from c to the first of these points of intersection (with either x or x'), and continue on the appropriate one of the curves x and x' until we reach p or p'.

It is not difficult to see that two different polygonal trails y and z which begin at c, do not meet l, and end at p or p', end at the same point. Indeed, if they ended at different points, as in Figure 33, then $y \cup z$ and the segment pp' would form a cycle whose intersection index with l would be 1, which is impossible.

Let U be the set of points on the plane that can be joined to p without meeting l and let V be the set of points on the plane that can be joined to p' without meeting l. U and V are the two regions into which, in accordance with the Jordan curve theorem, l divides the plane. If c_1 and c_2 lie in the same region (say U), then there are polygonal trails y_1 and y_2 that join them to p without meeting l. Their union is a polygonal trail that joins c_1 and c_2 without meeting l. Thus two points in the same region can be joined by a polygonal trail that does not meet l. If c_1 and c_2 are in different regions, then they cannot be joined by a polygonal trail that does

not meet l (otherwise, as before, we could obtain a cycle whose intersection index with l is 1).

Note that the "distant" points of the plane belong to the same region with respect to l. Hence one of the two regions into which l divides the plane is unbounded, and the other is bounded. The first of these regions is called the *exterior* of l and the second, its *interior*.

Problems

46. In case of a complicated polygonal trail l, such as the one in Figure 31, it is difficult to decide whether a point c is in the interior or in the exterior of l. Show that a ray from a point c that avoids the vertices of l intersects l in an even number of points if c is in the exterior of l and in an odd number of points if c is in the interior of l.

47. Show that a simple closed curve on a sphere divides it into two regions.

48. Suppose that each of k polygonal trails on the plane joins two given points p and q. If these trails have only p and q in common, then they divide the plane into k regions.

We wish to point out (without proof) that *any two simple closed curves l_1 and l_2 on the plane are isotopic*, i.e., that there is a homeomorphism of the plane that takes l_1 to l_2. This theorem says more than the Jordan curve theorem. Indeed, let l_1 be a circle and l_2 a simple closed curve on the plane. A homeomorphism of the plane that takes l_1 to l_2 takes the exterior of l_1 to the exterior of l_2 and the interior of l_1 (i.e., an open disk) to the interior of l_2. It follows that the union of a simple closed curve and its interior is homeomorphic to a (closed) disk. All the Jordan curve theorem asserts is the existence of two regions—the interior and exterior of a simple closed curve.

1.7. What Is a Curve?

Euclid defined a curve as "a length without width." Of course, this is an intuitive description of a curve and not a definition. The example that follows shows that this description is anything but satisfactory.

Example 15 Consider a square of area 1 (Figure 34a). We remove from it a cross (Figure 34b) in which the width of the two "beams" is such that its area is $\frac{1}{4}$. From each of the remaining 4 squares we again remove a cross (Figure 34c) such that the combined area of the 4 removed crosses is $\frac{1}{8}$. From each of the remaining 16 squares we again remove a cross (Figure 34d) such that the combined area of the 16 removed crosses is $\frac{1}{16}$, and so on. Let A_n be the figure that remains after n steps of our procedure and let A denote the "limiting figure" $A_1 \cap A_2 \cap \cdots \cap A_n \cap \cdots$. In spite of the fact that the squares are getting ever smaller, this figure turns out to have positive area! In fact, since the area of the removed squares is given by the

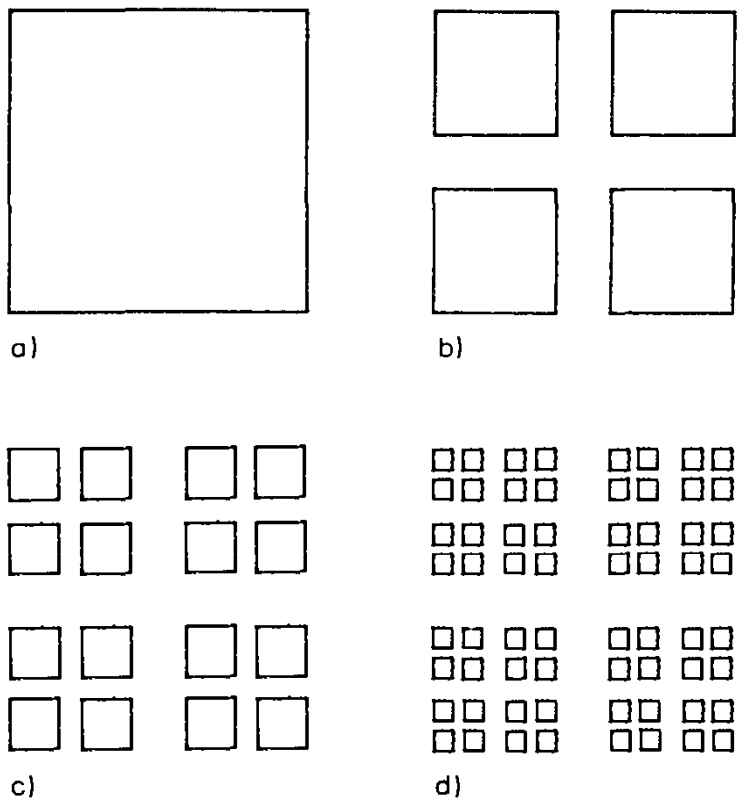


FIGURE 34.

sum of the geometric progression $\frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = \frac{1}{2}$, it follows that the area of A is $1 - \frac{1}{2} = \frac{1}{2}$.

Now we construct a simple arc, i.e., a figure homeomorphic to a segment, that passes through each of the points of A. To do this we start with a bent strip containing the 4 squares of step one in the previous construction (Figure 35a). Then we narrow the strip and introduce enough bends so that it contains all squares of step two in the previous construction (Figure 35b), then all squares of step three in the previous construction (Figure 35c), and so on.

After n steps of our present procedure we obtain a strip B_n contained in the previous strips and containing A_n (and thus also A). Let B denote the "limiting figure" $B_1 \cap B_2 \cap \cdots$, i.e., the intersection of all strips. Since B contains A, its area is at least $\frac{1}{2}$. Figure 35 shows that B is an extremely convoluted curve. This curve has positive area, and so is hardly "length without breadth."

Euclid gives yet another description of a curve as the "boundary of a surface." But, as we will soon see, the notion of boundary conceals many surprises.

We are used to the idea that the plane abuts on every segment of a plane curve "on two sides." For example, if l is a simple closed curve, then the two regions U and V determined by l abut on it throughout its extent (i.e., for every $x \in l$ there are points of U and V arbitrarily close to x); l is their simultaneous boundary.

Our intuition tells us that a plane curve cannot simultaneously be the boundary of more than two regions. But in this our intuition deceives us!

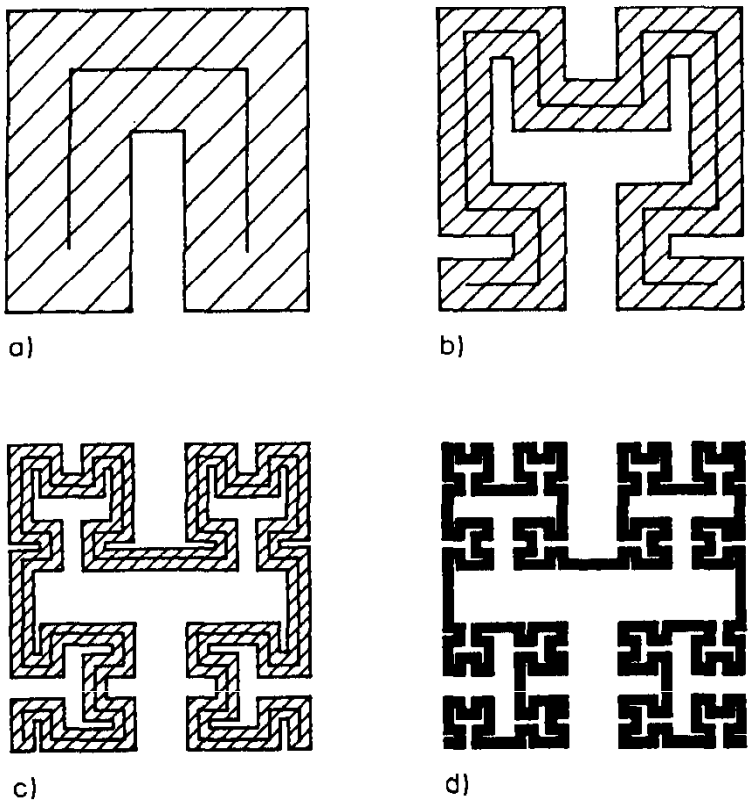


FIGURE 35.

Example 16 We show that there is a curve on the plane that is *simultaneously the boundary of three regions.* This curve was discovered by the Japanese mathematician Wada.

Consider an island in the ocean with two lakes, one cold and the other warm. We build canals to bring water from the sea and the two lakes to the island's interior. During the first day we build a canal from the warm lake that makes no contact with either the ocean or the cold lake such that each point on land is less than 1 away from warm water. On the second day we build a canal from the cold lake that makes no contact with either the ocean or the warm lake or the canal built on the previous day such that each point on land is less than 1 away from cold water. On the third day we build a corresponding canal from the sea. Thus after three days each point on land is less than 1 away from warm, cold, and salt water (see Figure 36).

During the next three days the canals are further extended, so that each point on land is now less than $\frac{1}{2}$ away from the three kinds of water. After three more days each point on land is less than $\frac{1}{4}$ away from the three kinds of water, and so on. Note that at the end of each day of work the land remains connected, so that as time passes the network of canals gets ever denser.

In the limit we obtain three disjoint networks of canals with warm, cold, and salt water. What remains of the land is a curve each of whose points is arbitrarily close to all three kinds of water. In other words, three domains abut on our curve throughout its extent!

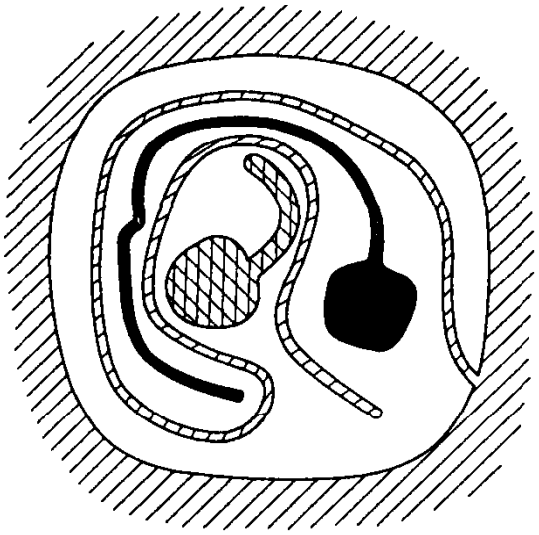


FIGURE 36.

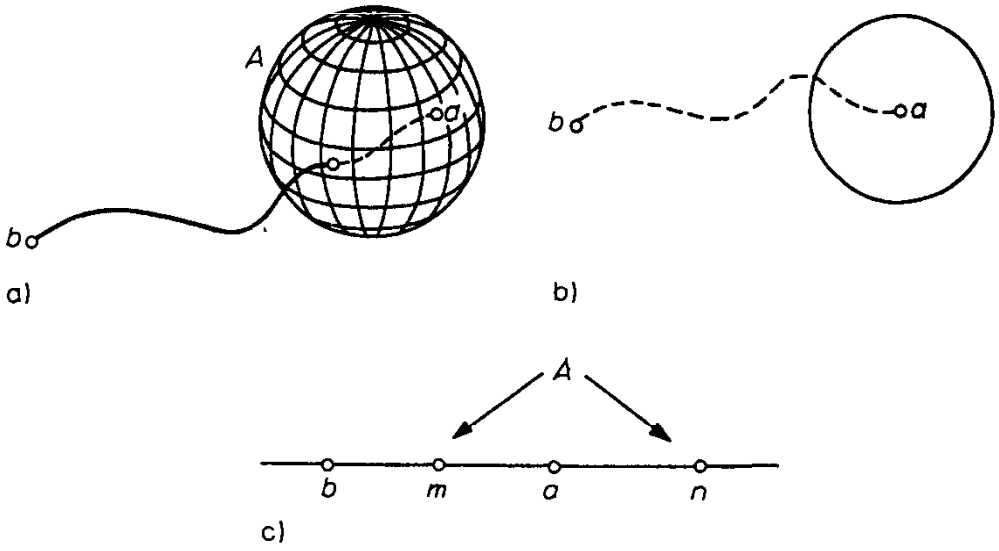


FIGURE 37.

Euclid gave a third description of a curve: A surface has two extensions, a curve has one extension, and a point has no extension. Many mathematicians tried to define the number of extensions, or the *dimension* of a figure. The final clarification of this concept and the justification of dimension theory is due to P.S. Uryson (1898–1924) and K. Menger (1902–1985).

We say that a set A, embedded in a figure X, separates a point a from a point b if there is no connected set in X that contains a and b and does not intersect A. Thus the spherical boundary of a (closed) ball separates the interior points of the ball from its exterior points (Figure 37a). This example shows that in three-dimensional space we can separate points by two-dimensional figures. On the plane (which is a two-dimensional figure) we can separate a point and nearby points from the other points by a one-dimensional figure, i.e., a curve (Figure 37b). Finally, on a straight line (a one-dimensional figure) we can separate a point and nearby points

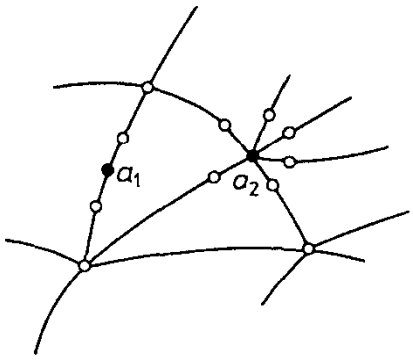


FIGURE 38.

from the other points by a pair of points (the points m, n in Figure 37c), i.e., a zero-dimensional figure.

These examples suggest that in a figure with n extensions (or in what we call an *n-dimensional* figure) we can separate a point and nearby points from the other points by a figure with fewer extensions. This suggests that we should define a zero-dimensional figure, use the notion of a zero-dimensional figure to define a one-dimensional figure (i.e., a curve), use the notion of a one-dimensional figure to define a two-dimensional figure, and so on.

We will say that *a figure X is zero-dimensional if it contains no connected figure with more than one point.* Thus a figure with a finite number of points is zero-dimensional and so, too, is the figure A in Example 15.

Suppose that we have already defined the notion of an $(n-1)$-dimensional figure. Then we say that a figure is n-dimensional if it is not $(n-1)$-dimensional and if it is possible to separate each of its points together with nearby points from the rest of the figure by a figure of dimension $n-1$ or lower. This is precisely Uryson's definition of dimension.

Example 17 A graph is a one-dimensional figure, i.e., a curve. Indeed, a point x and nearby points can be separated from the rest of the graph by a finite (i.e., zero-dimensional) set: Specifically, the separating set contains two points if a is an interior point of an edge (e.g., a_1 in Figure 38) and k points if a is a vertex of degree k (e.g., a_2 in Figure 38).

Example 18 The Polish mathematician Sierpiński constructed an interesting curve. We divide a square into nine equal squares and discard the central square (Figure 39a). Then we divide each of the remaining squares into nine equal squares and again discard the central square (Figure 39b). After one more such operation we arrive at the figure shown in Figure 39c. In the limit we obtain a one-dimensional figure C, i.e., a curve (known as the *Sierpiński carpet*).

The figure C is a *universal* plane curve: *If a curve l can be embedded in the plane, then it can be embedded in the Sierpiński carpet*, i.e., there is a curve $l' \subseteq C$ that is homeomorphic to l. It is clear that a curve that cannot be embedded in the

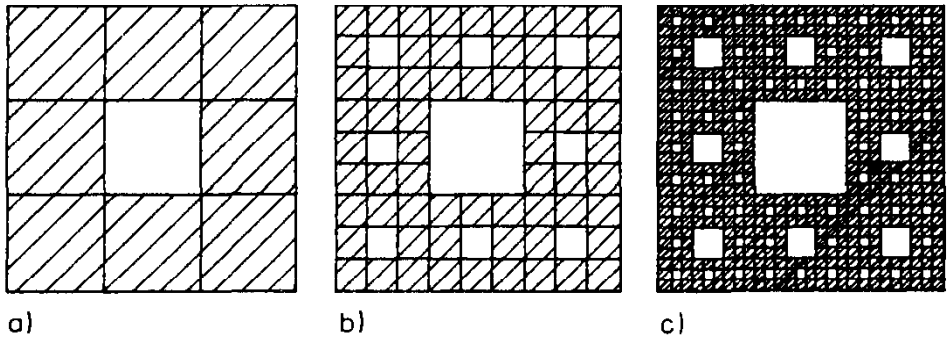


FIGURE 39.

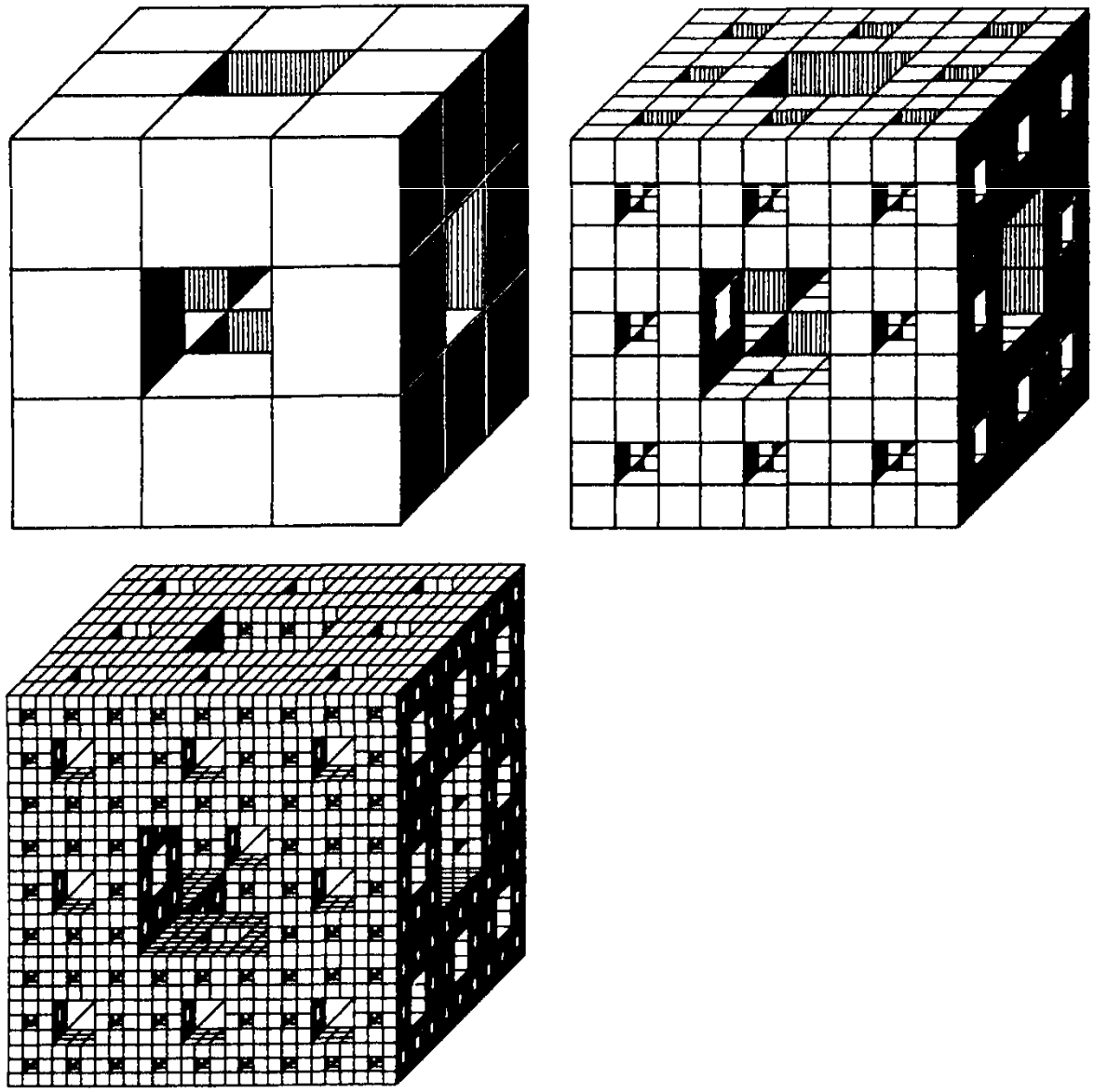


FIGURE 40.

plane cannot be embedded in the Sierpiński carpet either. But as was shown by the Austrian mathematician Menger, there is a curve in space, analogous to the Sierpiński carpet (Figure 40), in which any curve can be embedded.

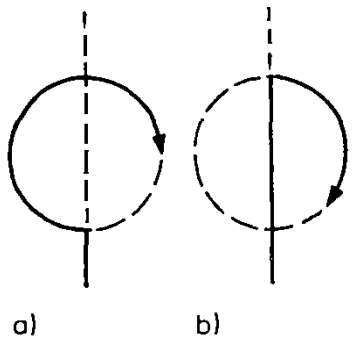


FIGURE 41.

Problems

49. Is there a plane curve that is simultaneously the boundary of 20 regions?

50. Show that the diagonal of a square in which the Sierpiński carpet is constructed intersects it in a zero-dimensional set. Deduce from this that the Sierpiński carpet is a one-dimensional figure, i.e., a curve.

51. Show that the property of being a curve is a topological invariant of a figure.

1.8. Peano Curves

A frequent intuitive description of a curve is that it is the track of a moving point.

Example 19 Suppose that a moving point traverses the figure shown in Figure 41 in two different ways (the solid line represents the path run through by a certain time, and the broken line represents the future path). In both cases the moving point traverses the same figure, i.e., the *track* is the same, but the paths are different.

We give an exact definition of the concept of a *path*. Consider a figure A and a point moving from time $t = 0$ to time $t = 1$. Suppose that we know the position $a(t) \in A$ for each $t \in [0, 1]$, i.e., that with every point $t \in [0, 1]$ there is associated a point $a(t) \in A$. In this way we obtain a mapping of the interval $[0, 1]$ into A, and this mapping is continuous, i.e., $a(t)$ changes continuously with t. This mapping is said to represent a path. Hence the definition: *We call a continuous mapping of the interval* $[0, 1]$ *into a figure* A *a path* (in A).

A simple arc can be viewed as a path (for it is the homeomorphic image of an interval, and a homeomorphic mapping is continuous). In particular, the curve (of positive area) in Example 15 can be viewed as the track of a moving point. This alone shows that the notion of a path is not simple. This is confirmed by the following example.

Example 20 We show that it is possible to construct a path that passes through every point in a square. In other words, we show that there is a continuous mapping of an interval onto a square. Such a path is called a *Peano curve*.

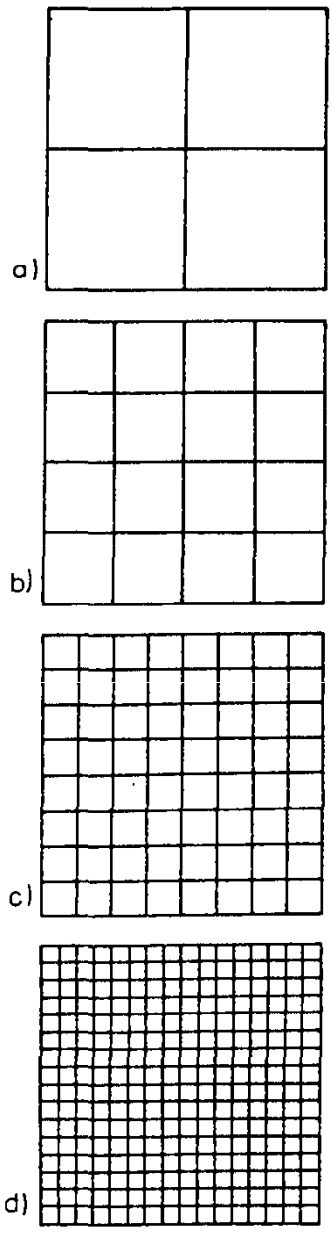


FIGURE 42.

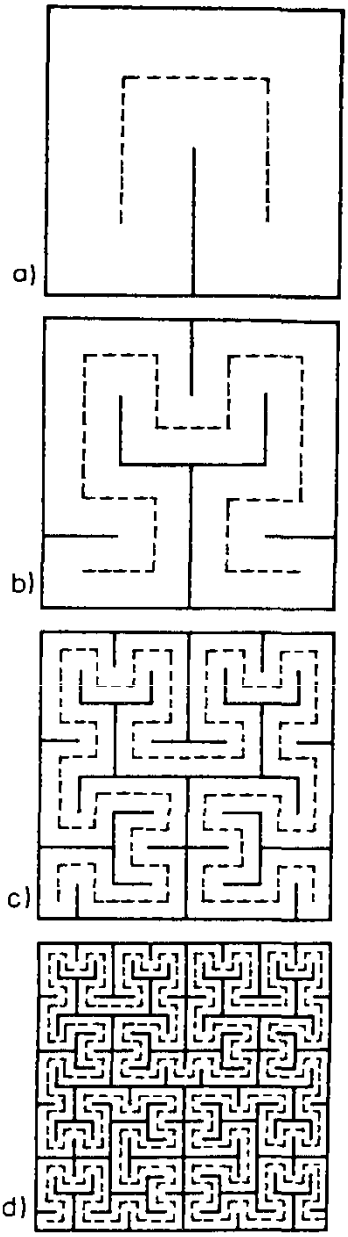


FIGURE 43.

To obtain a Peano curve we construct in a square Q ever more winding strips and take the limit of their midlines. This is a shorthand description of the following procedure: We divide the square into 4, 16, 64, ..., 4^n, ... congruent squares (Figure 42). At each stage of subdivision we remove some of the sides of the "subsquares" of the square. The leftover sides form non-removable partitions that determine the successive strips. In each strip we introduce its midline. (The first few winding strips and their midlines are shown in Figure 43; the midlines are the broken lines.) The limit of these midlines is a path that fills the whole square, i.e., a Peano curve. A more technical description of this procedure follows.

We consider a continuous mapping of $[0, 1]$ on the first broken line (Figure 43a). Here $\left[0, \frac{1}{4}\right]$ is mapped on the part of the line in the lower left quarter of the large square, $\left[\frac{1}{4}, \frac{1}{2}\right]$ on the part of the line in the upper left quarter of the large square,

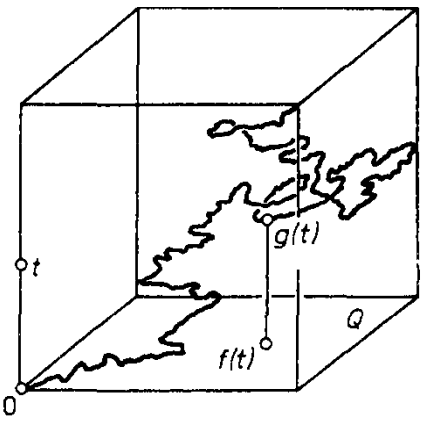


FIGURE 44.

$\left[\frac{1}{2}, \frac{3}{4}\right]$ on the part of the line in the upper right quarter of the large square, and $\left[\frac{3}{4}, 1\right]$ on the part of the line in the lower right quarter of the large square; we denote this mapping by $f_1(t), t \in [0, 1]$. By $f_2(t)$ we denote the mapping of $[0, 1]$ on the broken line shown in Figure 43b. In this, the second stage of our construction, the intervals $\left[0, \frac{1}{16}\right], \left[\frac{1}{16}, \frac{2}{16}\right], \ldots, \left[\frac{1}{16}, 1\right]$ are mapped on the successive parts of the broken line in the 16 squares of the large square. By $f_3(t)$ we denote the mapping of $[0, 1]$ on the broken line of the third stage (Figure 43c), and so on. The limit of the sequence of functions $f_1(t), f_2(t), f_3(t), \ldots$ is a mapping f: $[0, 1] \to Q$, i.e., a path in the square Q, which is also a Peano curve. It is easy to see that this limit exists. Consider, for example, the point $\frac{1}{3} \in [0, 1]$. Since $\frac{1}{3} \in \left[\frac{1}{4}, \frac{1}{2}\right]$, $f_1(\frac{1}{3})$ is in the left upper square in Figure 42a. Since $\frac{1}{3} \in \left[\frac{5}{16}, \frac{6}{16}\right]$, $f_1(\frac{1}{3})$ is in the sixth of the squares traversed by the broken line in Figure 43b, i.e., in the left upper square in Figure 42b. Since $\frac{1}{3} \in \left[\frac{21}{64}, \frac{22}{64}\right]$, $f_1(\frac{1}{3})$ is in the 22nd of the squares traversed by the broken line in Figure 43c, i.e., in the left upper square in Figure 42c, and so on. The limit of this nested sequence of decreasing squares, in our case the left upper corner of the square, is $f(\frac{1}{3})$. The point $f(t)$ is determined in a similar way for every $t \in [0, 1]$.

We note that the Peano curve is not a simple arc: It has infinitely many points that are "glued together," (i.e., the square contains infinitely many points that are traversed by $f(t)$ more than once).

Problems

52. Show that there are points in the square Q through which the Peano curve $f(t)$ passes four times but no points through which it passes five times.

53. Do there exist Peano space curves, i.e., paths that fill an octant?

54. We place a square in a horizontal plane and consider a Peano curve on this square. Let $g(t)$ be the point in space that is t above the point $f(t)$ (Figure 44). Show that as t traverses $[0, 1]$, $g(t)$ traverses a path in space that is a simple arc. Show that the projection of this simple arc on the horizontal plane fills the whole square Q. In other words, we have constructed a curve, a simple arc, that turns out to form an intricate "roof" over all of Q.

We saw that there are complicated paths. This example shows that there are complicated simple arcs as well.

2

Topology of Surfaces

2.1. Euler's Theorem

The table below lists the number of edges, vertices, and faces of the five Platonic solids.

Name of polyhedron	Number V of vertices quotation	Number E of edges	Number F of faces
tetrahedron	4	6	4
cube	8	12	6
octahedron	6	12	8
dodecahedron	20	30	12
icosahedron	12	30	20

The table shows that for each Platonic solid we have the relation

$$V - E + F = 2. \tag{6}$$

It is easy to show that relation (6) is also valid for pyramids, prisms, and other polyhedra. Euler was the first to recognize and prove this important property of polyhedra.

We will give a precise formulation of Euler's theorem. To do this we note that each face of any of the polyhedra under consideration is homeomorphic to a circular disk, and that the surface of each of these polyhedra (more generally, of each convex polyhedron) is homeomorphic to a sphere (to prove the latter assertion note that if P is one of our polyhedra and S is a sphere in the interior of P centered at o, then the projection with center o of the surface of P to S yields the required homeomorphism). It follows that *relation* (6) *holds for any polyhedron whose surface is homeomorphic to a sphere and each of whose faces is homeomorphic to a disk.*

This theorem can be given a purely topological formulation. To this end note that the vertices and edges of each of our polyhedra form a connected graph that divides its surface into surface elements, i.e., into pieces homeomorphic to disks. This implies the following theorem (which is somewhat more general than Euler's theorem):

> *Suppose that a connected graph with V vertices and E edges on a sphere (or on a surface homeomorphic to a sphere) divides it into F regions ("faces"). Then V, E, and F are related by* (6).

The idea for a proof of this result is found in Problem 55.

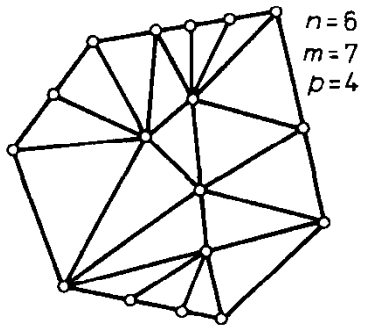


FIGURE 45.

Problems

55. Let G be a connected graph on a sphere. Let G^* be a spanning tree of G and k the number of omitted edges (i.e., edges of G not in G^*). Show that G^* defines on the sphere just one region (piece) and that therefore (6) holds for it. Show that addition of any one of the omitted edges increases the number of pieces by 1 and deduce from this Euler's theorem.

56. Show that relation (6) holds for any connected graph on the plane (provided that the surrounding unbounded region is counted as a "piece").

57. Let G be a graph that can be embedded in the plane. Show that for every embedding G divides the plane into $r - V + E + 1$ regions; here r denotes the number of components of G, V the number of vertices, and E the number of edges.

58. Let a convex n-gon be divided into triangles by m points on its sides and p points in its interior. We assume that if two triangles share a segment, then it is a common side (Figure 45). Show that the n-gon is divided into $m + n + 2p - 2$ triangles.

59. Let n_3 be the number of triangular faces of a convex polyhedron, n_4 the number of its quadrangular faces, and so on. Show that

$$3n_3 + 2n_4 + n_5 \geq 12 + n_7 + 2n_8 + 3n_9 + 4n_{10} + \cdots.$$

When does the inequality become an equality?

60. A connected graph on a sphere is said to define a *topologically regular decomposition of the sphere* if each face of the decomposition is an n-gon (i.e., if it is bounded by a closed chain of n edges) and if k faces meet at each vertex. Show that if E is the number of edges, then

$$\frac{1}{n} + \frac{1}{k} = \frac{1}{2} + \frac{1}{E}.$$

Deduce from this that in addition to the decompositions that are topologically equivalent to the five Platonic solids, there are just two types of topologically regular decompositions; the latter are shown in Figure 46.

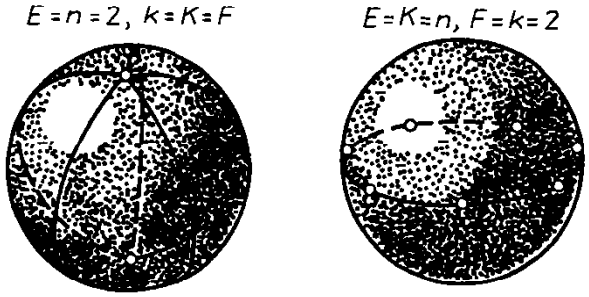


FIGURE 46.

2.2. Surfaces

Example 21 Figure 47 depicts a "book with three pages." This figure is differently structured in the vicinity of each of the points x, y, and z. The neighborhood of y has the form of a semidisk with y on its boundary; in this case we say that y lies on the boundary of the figure. The neighborhood of z consists of three semidisks with common diameter; in this case we say that the figure is *ramified* at z (i.e., three or more sheets of the figure abut on a single curve). Finally, the neighborhood of x is a disk with x in its interior; thus x is not on the boundary and the figure is not ramified at x.

A figure is called a *surface* if each of its points x has a neighborhood homeomorphic to a disk (with x in its interior). A surface has neither boundary nor ramifications. A sphere and a torus are surfaces. We consider also *surfaces with boundary*. They have boundaries but are not ramified. A disk is a surface with boundary. A sphere with round holes (Figure 48) is likewise a surface with boundary.

Example 22 A torus with a round hole is a surface with boundary. Such a surface is called a *handle* (Figure 49).

Example 23 In their works published between 1862 and 1865 the German mathematicians Möbius and Listing described an interesting example of a surface with boundary. It is obtained by taking a rectangular strip (Figure 50a), twisting it once

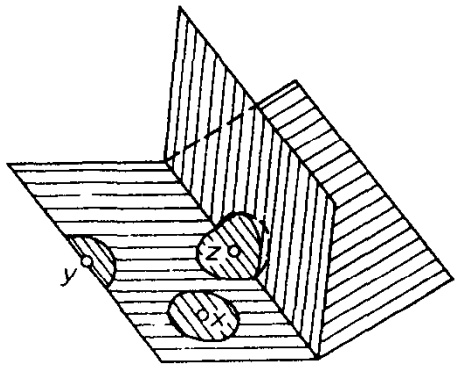


FIGURE 47.

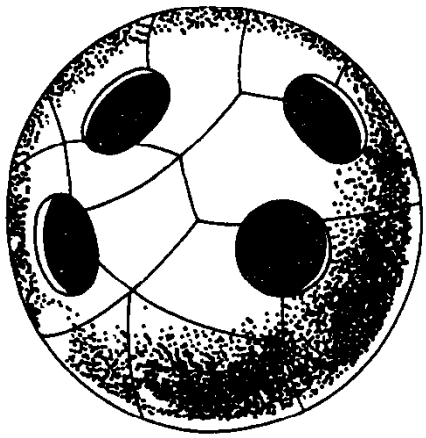


FIGURE 48.

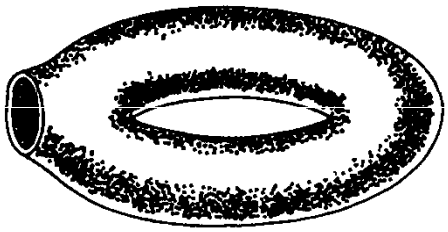


FIGURE 49.

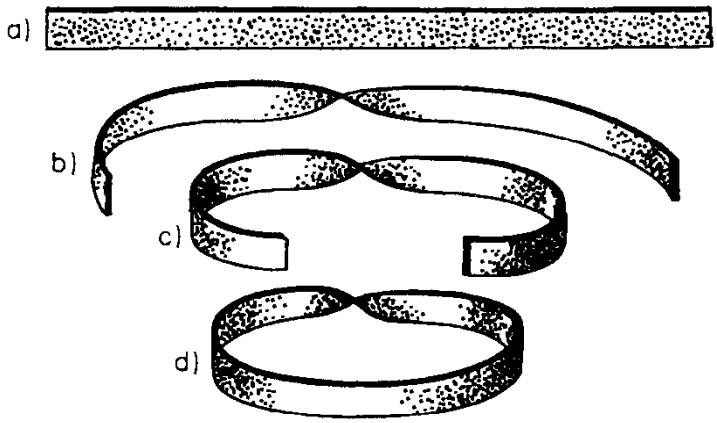


FIGURE 50.

(Figure 50b, c), and gluing its ends together. The resulting surface (Figure 50d) with boundary is called a *Möbius strip*. This surface is one-sided. If we go over a Möbius strip with a paintbrush, then we return to the starting point on the opposite side. If we continue, then we paint the whole Möbius strip and see that it has indeed just one side.

Of course, for a one-sided surface to be intuitively describable in terms of coloring it must have thickness, and so must be made of some material. But its mathematical version has no thickness. Hence the need for a different description of one-sidedness.

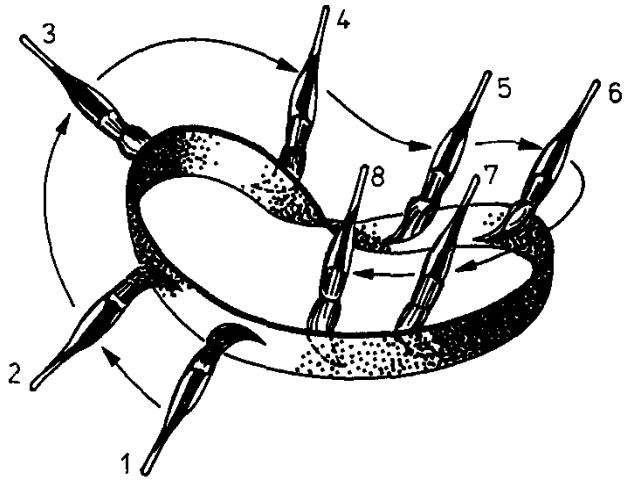


FIGURE 51.

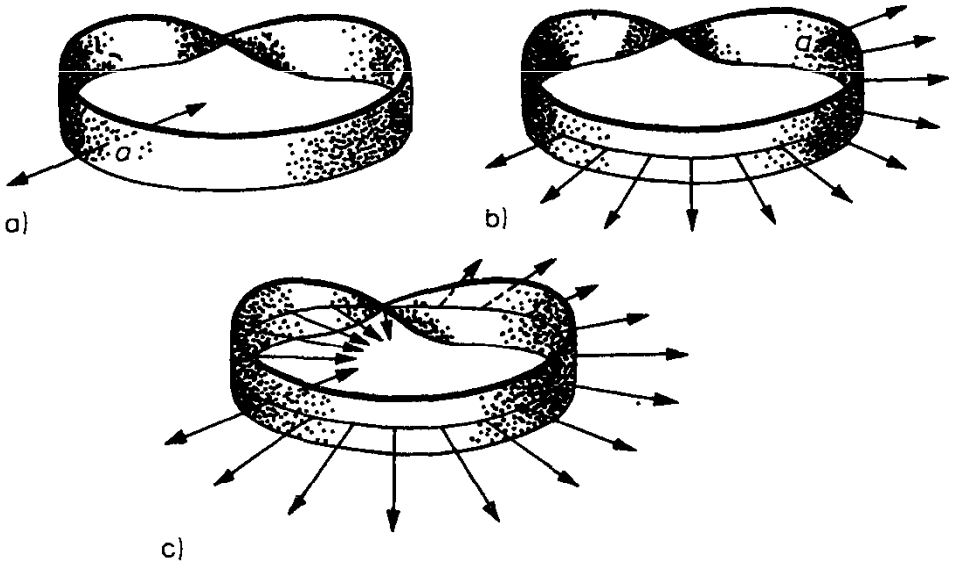


FIGURE 52.

At each point a of a Möbius strip we can draw two perpendicular oppositely directed vectors. They are called the *normals* to the Möbius strip at a. We choose one of them and move it together with a (Figure 52b). By the time a has gone around the whole Möbius strip, the normal attached to it points in a direction opposite to its initial direction (Figure 52c). This shows that *there is a closed path on a Möbius strip whose traversal results in the normal changing its direction.* Such surfaces are said to be *one-sided.*

The use of normals involves not only the surface but also its disposition in space. Hence the need for an intrinsic definition of one-sidedness of a surface. Draw a small circle around each point a that is not a boundary point and orient it (by means of arrows) counterclockwise when viewed from the end of a normal at a (Figure 53a). When a moves, so do the normal at a and the oriented circle around a. As a result of traversing the whole Möbius strip the orientation of the circle is reversed (for the normal changes its orientation; Figure 53b). Thus there is

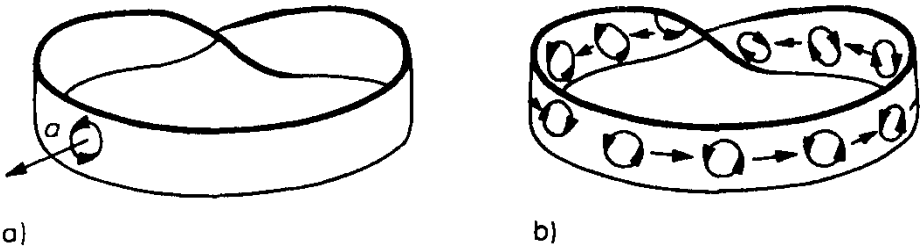


FIGURE 53.

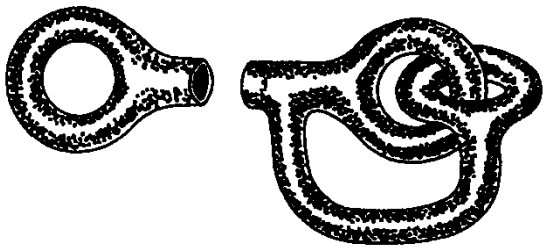


FIGURE 54.

a closed path on the Möbius strip whose traversal reverses the orientation of an oriented circle. Such a path is said to be *an orientation-reversing circuit.*

A surface without an orientation-reversing circuit is called *orientable* (or *two-sided*) and a surface with such a circuit is called *nonorientable* (or *one-sided*). Intuitively, a surface is orientable if it can be covered with small circles that can be oriented so that neighboring circles have the same orientation.

Let Q_1 and Q_2 be two surfaces with boundaries each of which is homeomorphic to a circle (Figure 54). By gluing these boundaries together we obtain a new surface. We say that we have closed the hole in Q_1 with Q_2 (or conversely).

Example 24 Consider a sphere with p round holes and close each of them with a handle. The resulting surface (Figure 55a) is called a *sphere with p handles.* A sphere with one handle (Figure 55b) is homeomorphic to a torus, and a sphere with two handles is homeomorphic to a pretzel (it is obtained by gluing together two handles; Figure 55c).

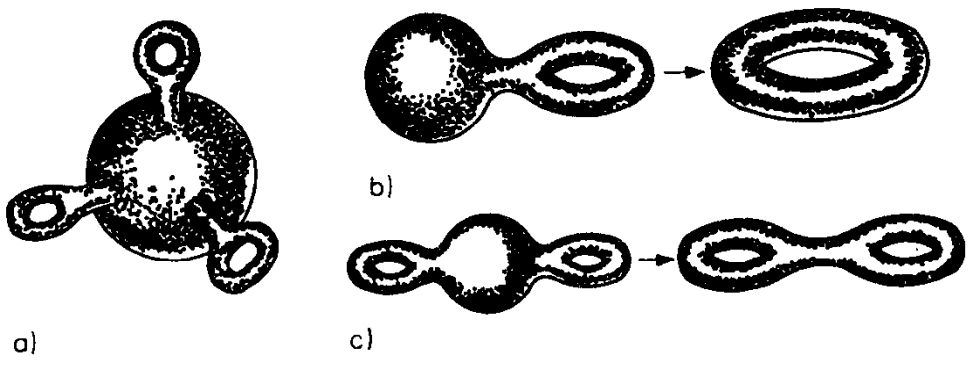


FIGURE 55.

Problems

61. Show that the graph "houses and wells" (Example 12) can be put on a Möbius strip without self-intersections.

62. Consider the serrated figure in Figure 56a with four pairs of segments. In each of the four pairs, twist one of the two segments and glue them together (Figure 56b). Show that the resulting surface is one-sided and that its boundary is homeomorphic to a circle.

63. Drill three disjoint cylindrical holes through a ball. Show that the surface of the resulting solid is homeomorphic to a sphere with three handles.

64. Drill three cylindrical holes through a ball so that their axes pass through the center of the ball. Show that the surface of the resulting solid is homeomorphic to a sphere with five handles.

65. Consider the square in Figure 57a. If we glue together its opposite sides and observe the indicated directions, then we obtain a torus (see Figures 57b, c, and d). What surface do we obtain if we glue together the opposite sides of the figure in Figure 58, observing the indicated directions (the side c is left unglued)?

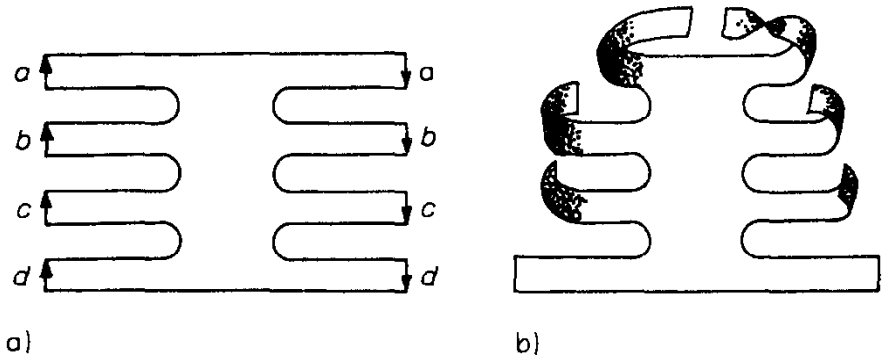


FIGURE 56.

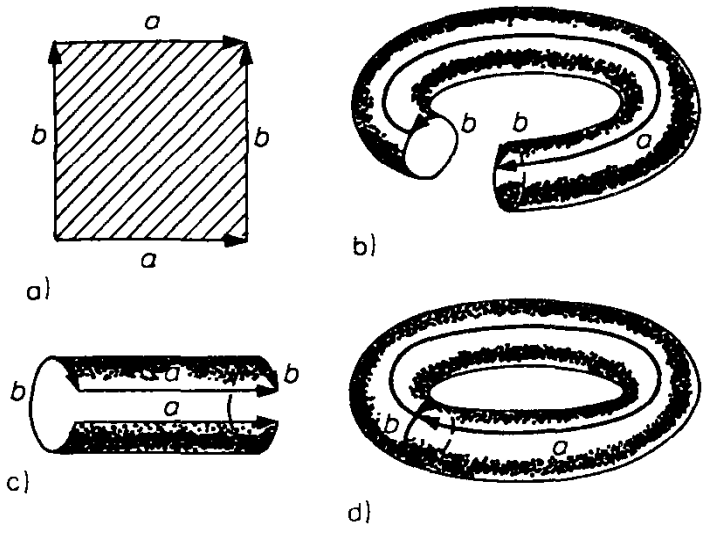


FIGURE 57.

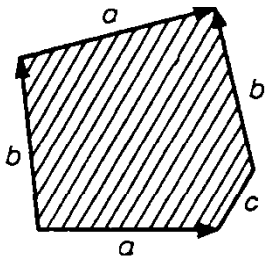


FIGURE 58.

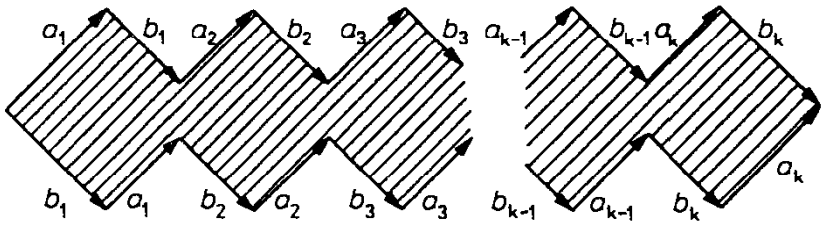


FIGURE 59.

66. Consider the $4k$-gon in Figure 59. What surface do we obtain if we glue together the sides with the same lettering, observing the indicated directions?

We are now ready to formulate the theorem on the *topological classification of surfaces* discovered in the nineteenth century by Möbius and the French mathematician Jordan. We consider only closed surfaces (i.e., surfaces without boundary that can be divided into a finite number of polygons). For example, the plane is not a closed surface. Indeed, a finite graph on the plane does not divide it into regions each of which is homeomorphic to a disk. The problem of topological classification of surfaces is *to give pairwise nonhomeomorphic closed surfaces such that every closed surface is homeomorphic to one of the given surfaces.*

First we state the solution of this problem for orientable surfaces. Let P_0 denote the sphere and P_k the sphere with k handles. It turns out that *the surfaces*

$$P_0, P_1, P_2, \ldots, P_k, \ldots \tag{7}$$

give a complete classification of closed orientable surfaces, i.e., this sequence is a complete list of all topologically different closed orientable surfaces. The proof of this assertion is given in the next two sections.

2.3. The Euler Characteristic of a Surface

Let Q be a surface (with or without boundary, one- or two-sided) that admits *decomposition into polygons*, i.e., it is possible to "draw" a graph on the surface that divides it into finitely many pieces each of which is homeomorphic to a disk. Let V be the number of vertices of the graph, E the number of its edges, and F

the number of polygons into which the graph divides Q. The number

$$\chi(Q) = V - E + F \tag{8}$$

is called the *Euler characteristic* of the surface Q. Note that $\chi(Q)$ seems to depend on the particular decomposition of Q into polygons and not on Q itself. On the other hand, Euler's theorem shows that if Q is homeomorphic to a sphere, then the Euler characteristic of Q is independent of the choice of its decomposition into polygons: $\chi(Q) = 2$ (see (6)). We will show that *the Euler characteristic of any surface Q is independent of its decomposition into polygons and is determined by the surface alone*. Thus the Euler characteristic of a surface is a topological invariant.

Let G_1 and G_2 be two graphs on the surface Q each of which divides it into polygons homeomorphic to disks. We denote the respective numbers of vertices, edges, and faces associated with the graph G_1 by V_1, E_1, and F_1, and the corresponding numbers associated with the graph G_2 by V_2, E_2, and F_2. While it can happen that the number of points of intersection of the two graphs is infinite, it can be reduced to finitely many by "slightly moving" G_1.

If the graph $G_1 \cup G_2$ is not connected, then by a slight motion of G_1 and G_2 we can achieve that they have common points, and that therefore their union is connected. In other words, we can assume a priori that the number of points of intersection of G_1 and G_2 is finite and that the union of the two graphs is connected. We regard all points of intersection of the two graphs as well as their vertices as the new vertices. Then $G_1 \cup G_2$ is a finite connected graph (its edges are pieces of the edges of G_1 and G_2 resulting from the division of the "old" edges by points of $G_1 \cup G_2$).

Let V be the number of vertices of $G_1 \cup G_2$, E the number of its edges, and F the number of faces into which it divides Q. We want to prove the equalities

$$\begin{aligned} V_1 - E_1 + F_1 &= V - E + F, \\ V_2 - E_2 + F_2 &= V - E + F. \end{aligned} \tag{9}$$

They imply that $V_1 - E_1 + F_1 = V_2 - E_2 + F_2$. Both equalities in (9) are proved in the same way. We prove the first.

Let M be one of the polygons (faces) determined by the graph G_1. We denote the number of vertices and edges of $G_1 \cup G_2$ in the interior of M (not on its boundary) by V' and E' respectively, and the number of vertices of $G_1 \cup G_2$ (including those on its edges) on the boundary of M by q. Finally, we denote the number of faces defined by $G_1 \cup G_2$ and contained in M by F'. For example, in Figure 60 we have $V' = 4$, $E' = 12$, $F' = 9$, $q = 15$.

Now we cut out from the surface Q the polygon M (together with the part of $G_1 \cup G_2$ on it). Since M is homeomorphic to a disk and thus to a hemisphere, we can supplement it with a hemisphere and thus obtain a surface homeomorphic to a sphere (Figure 61). On this sphere we now have a connected graph with $V' + q$ vertices, $E' + q$ edges, and $F' + 1$ faces (there are F' faces in M, and there is the

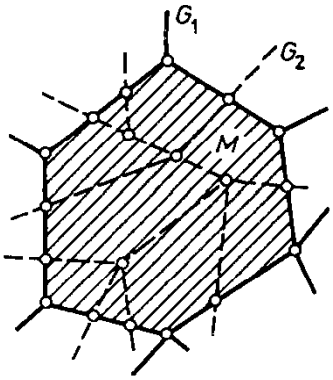


FIGURE 60.

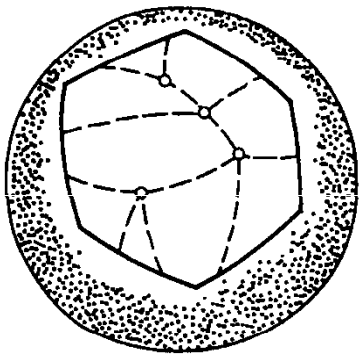


FIGURE 61.

additional hemisphere). We thus obtain

$$(V' + q) - (E' + q) + (F' + 1) = 2,$$

i.e.,

$$V' - E' + F' = 1. \tag{10}$$

We now go back to the surface Q in which the graph $G_1 \cup G_2$ was embedded. If we remove from $G_1 \cup G_2$ the part that was in the interior of M, then we obtain a new graph with the same value of $V - E + F$ as $G_1 \cup G_2$. This is so because, instead of V' vertices, E' edges, and F' faces, we now have 0 vertices, 0 edges, and one face (the polygon itself), i.e., the number $V' - E' + F'$ goes over into the number $0 - 0 + 1$, which is the same as the number in (10).

It is now clear that if we remove from $G_1 \cup G_2$ the part consisting of the interiors of all polygons defined by $G_1 \cup G_2$, then we obtain a new graph G^* for which the value of $V - E + F$ is the same as that for $G_1 \cup G_2$. In other words,

$$V^* - E^* + F^* = V - E + F; \tag{11}$$

here V^* and E^* denote, respectively, the number of vertices and edges of G^*, and F^* denotes the number of faces determined by G^*.

Finally, it is easy to see that G^* can be obtained from G_1 by placing some new vertices on its edges. But addition of a vertex increases the number of edges by 1

(because a certain edge is thereby divided into two). If G^* is obtained from G_1 by the addition of k new vertices, then $V^* = V_1 + k$ and $E^* = E_1 + k$. Moreover, $F^* = F_1$ (for G^* and G_1 determine the same faces). It follows that

$$V^* - E^* + F^* = (V_1 + k) - (E_1 + k) + F_1 = V_1 - E_1 + F_1.$$

This equality and (11) yield the first of the equalities in (9).

We see that the Euler characteristic of a surface does not depend on its subdivision into polygons but is determined by the surface itself. Moreover, the Euler characteristic is a topological invariant, i.e., if two surfaces Q_1 and Q_2 are homeomorphic, then $\chi(Q_1) = \chi(Q_2)$. This is so because, under a homeomorphism $f\colon Q_1 \to Q_2$, a graph G_1 embedded in Q_1 goes over into a graph G_2 embedded in Q_2, and on each of the two surfaces the number of vertices, edges, and faces is the same.

Problems

67. Show that a sphere with q holes has Euler characteristic $2 - q$.

68. Let Q_1 and Q_2 be two surfaces with boundaries, each of which is homeomorphic to a circle. Show that if we glue the boundaries together as in Figure 54, then the Euler characteristic of the resulting surface is $\chi(Q_1) + \chi(Q_2)$.

69. What are the Euler characteristics of a disk, a handle, and a Möbius strip?

70. Show that the Euler characteristic of the surface P_k is $2 - 2k$.

71. A torus admits regular topological decompositions (cf. Problem 60). Show that each face of such a decomposition is either a triangle, a quadrangle, or a hexagon and give examples for each of these posibilities.

72. Draw a graph with V vertices and E edges on a closed surface Q. The graph divides the surface into F surface pieces (of which none need be homeomorphic to a disk). Show that $V - E + F \geq \chi(Q)$.
Hint. In order to divide the surface Q into polygons (homeomorphic to a disk) it suffices to carry out one or more of the following operations:

(a) addition of a new vertex on an edge;
(b) addition of an edge that has just one vertex in common with the graph already drawn;
(c) addition of an edge connecting two vertices of the graph already drawn.

Show that each of these operations can only increase $V - E + F$.

73. A graph with V vertices and E edges is embedded in a closed surface Q. The graph divides the surface into F regions. Show that if each of these regions has on its boundary at most k edges, then $(k - 2)F \leq kV - \chi(Q)$.

2.4. Classification of Closed Orientable Surfaces

The surfaces P_0, P_1, P_2, ... have different Euler characteristics (Problem 70) and thus are pairwise nonhomeomorphic. This being so, all we need do to complete the proof of the theorem formulated at the end of Section 2.2 is show that any closed orientable surface is homeomorphic to one of the surfaces P_0, P_1, P_2, ... We do this in a number of steps.

A) Let Q be a connected closed orientable surface. We put on it a connected graph G that divides it into regions homeomorphic to a disk. About each vertex of G we take a small disk, to be referred to in the sequel as a *cap*. About each edge of G we take a *strip* that connects the caps at its endpoints. If we remove all caps and strips, then what is left of each region is a piece homeomorphic to a disk (to be referred to in the sequel as its *core*). In Figure 62, which depicts a piece of Q, the caps are hatched, the strips are dotted, and the cores are left unmarked. We divide Q into caps, strips, and cores, reassemble it—by gluing—out of these pieces, and note at each step of this process what has been obtained.

We remove from Q all cores of all faces and denote what is left by Q_0. The boundary of Q_0 consists of the contours of the cores.

B) We choose in the graph G a spanning tree (see Figure 63). We cut all strips corresponding to the omitted edges (i.e., edges not in the tree) in the middle. We call the segments a_1b_1, a_2b_2, ..., a_pb_p along which the strips have been cut

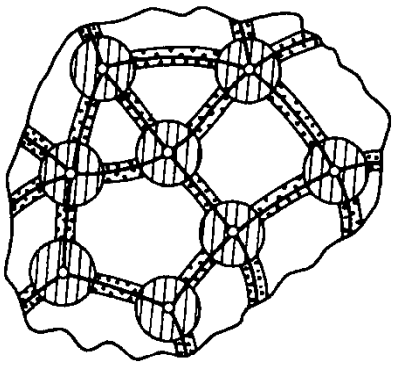


FIGURE 62.

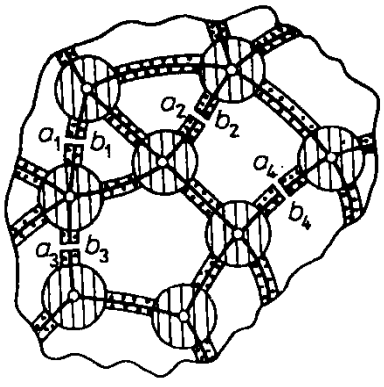


FIGURE 63.

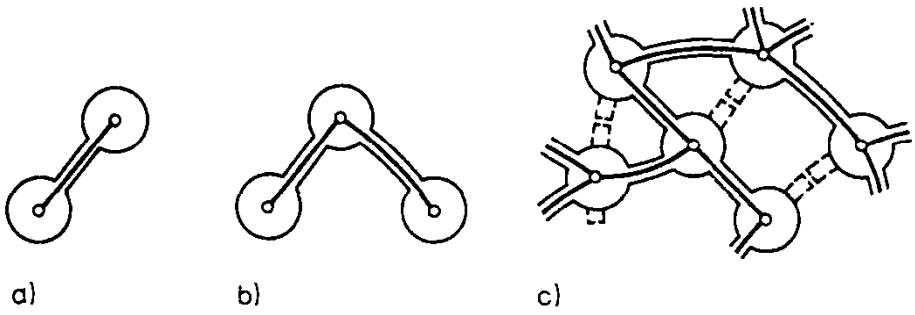


FIGURE 64.

chords. We proceed step by step. Cutting along the chord a_1b_1 changes Q_0 into a surface Q_1. Cutting along a_2b_2 in Q_1 yields Q_2. We continue in this way, and finally, by cutting along a_pb_p in Q_{p-1}, we obtain a surface Q_p. To get Q_0 out of Q_p we must reglue along the chords.

C) Before regluing we show that the surface Q_p is homeomorphic to a disk. We draw the spanning tree of the graph G by adding an edge at a time, but do this so that at each step we have a tree. The strip and the two caps corresponding to the first edge and its endpoints form a surface homeomorphic to a disk (Figure 64a). By adding the strip and cap corresponding to the second edge we again obtain a surface homeomorphic to a disk (Figure 64b). Each step of the construction involves the gluing of a strip and a cap to a surface homeomorphic to a disk, and the result is again a surface homeomorphic to a disk. After the whole spanning tree has been drawn we obtain a surface homeomorphic to a disk and consisting of all caps and strips corresponding to the edges of the spanning tree. To obtain the surface Q_p all we need do is glue on the half-strips obtained from the remaining strips after they were cut along the chords (the half-strips are drawn in Figure 64c with little strokes). The surface resulting from the gluing on of a half-strip continues to be homeomorphic to a disk.

Next we show that *each of the surfaces* $Q_p, Q_{p-1}, \ldots, Q_0$ *is homeomorphic to a sphere with a finite number of holes some of which may be closed with handles.* This is obvious in the case of Q_p, which is homeomorphic to a disk, i.e., a sphere with one hole.

D) For each $i = 1, \ldots, p$ we consider the transition from Q_{i-1} to Q_i (i.e., the cutting along the chord a_ib_i) and the converse transition from Q_i to Q_{i-1}. There are two possibilities: the points a_i and b_i lie either on the same component of the boundary of Q_{i-1} or on different components.

If a_i and b_i lie on different components of the boundary of Q_{i-1}, then cutting along the chord a_ib_i reduces the number of holes by 1 (Figures 65a and 65b). This means that transition from Q_i to Q_{i-1} increases the number of holes by 1. If Q_i can be obtained from a sphere by cutting in it a few holes and closing some of them with handles, then the same is true for Q_{i-1}.

E) Now consider the case where the endpoints of the chord a_ib_i lie on the same component of the boundary of Q_{i-1} (Figure 66). The surface Q_i, obtained as a result of cutting (Figure 67), is homeomorphic to the surface (Figure 68) obtained

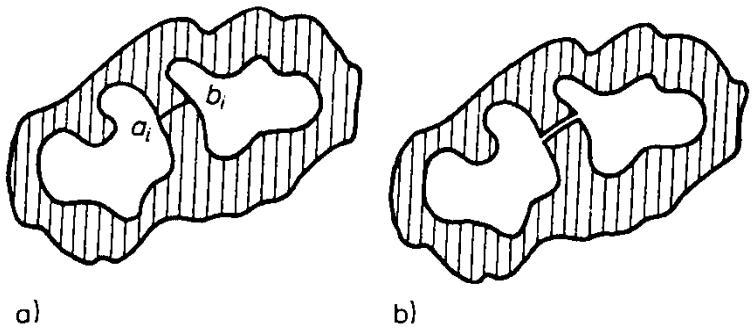


FIGURE 65.

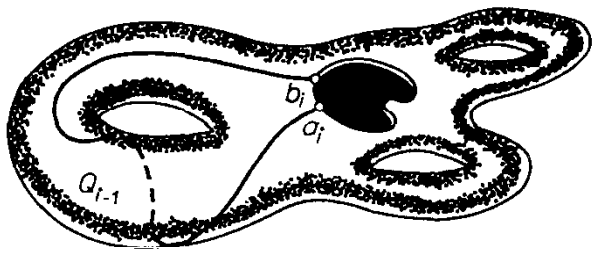


FIGURE 66.

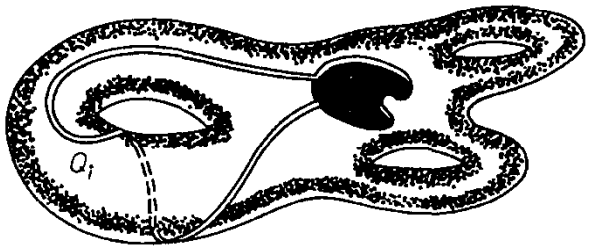


FIGURE 67.

from Q_{i-1} by two cuts: a cut along the closed curve l (Figure 69) that does not intersect the boundary of Q_{i-1} (this yields the intermediate surface Q_i^* shown in Figure 69), followed by a cut along the chord $a_i c_i$ whose endpoints lie on different components of the boundary of Q_i^*. As we saw in D), transition from Q_i (see Figure 68) to Q_i^* (Figure 69) creates a hole. It remains to investigate the transition from Q_i^* to Q_{i-1}.

Now Q_i^* was the result of cutting Q_{i-1} along the curve l that does not meet the boundary of Q_{i-1}. If instead of making this cut we remove from Q_{i-1} a narrow strip L that contains in its interior the curve l (Figure 70), then we obtain a surface homeomorphic to Q_i^*. The strip L is homeomorphic either to a Möbius strip or to the lateral surface of a cylinder. If we cut it (Figure 71), then we can flatten it into a rectangular strip. What must be ascertained is whether or not gluing it together involves a twist.

Since the initial surface Q (as well as the surfaces $Q_0, Q_1, \ldots, Q_p$) are orientable and a Möbius strip contains an orientation-reversing closed path, the strip L cannot be homeomorphic to a Möbius strip. But then L must be homeomorphic to the lateral surface of a cylinder, and when cut along the curve l, it yields two

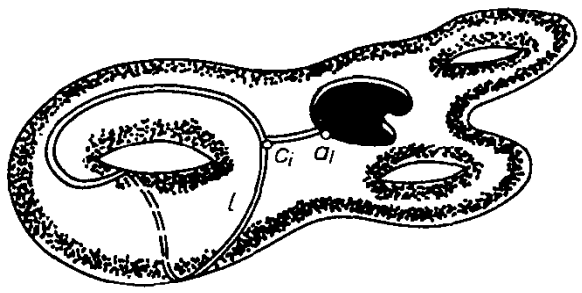


FIGURE 68.

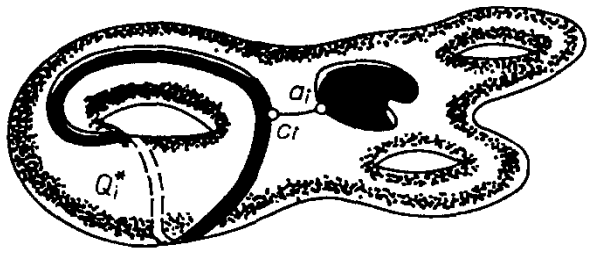


FIGURE 69.

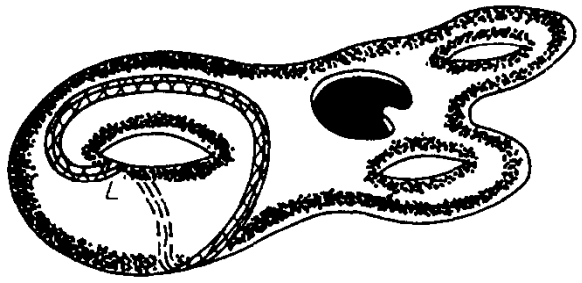


FIGURE 70.

parts. Thus Q_i^* gains two boundary components l_1 and l_2. The transition from Q_i^* to Q_{i-1} consists in gluing together the curves l_1 and l_2, which are the boundaries of holes in Q_i^*. A detailed description of this step follows.

We surround the curves l_1 and l_2 by narrow ringlike strips and connect the latter by a strip. In this way, we obtain on Q_i^* a figure (an "eyeglass frame"; Figure 72) homeomorphic to a disk with two holes (Figure 73). When gluing together the contours l_1 and l_2 we must make sure that they have opposite orientations, for otherwise, as shown in Figure 74, the (hatched) strip would become a Möbius strip—a possibility ruled out by the orientability of Q_{i-1}. All this shows that gluing together the curves l_1 and l_2 is equivalent to closing a hole in Q_i^* with a handle (Figure 75). We conclude that transition from Q_i to Q_i^* produces a hole, and transition from Q_i^* to Q_{i-1} results in decreasing the number of holes by 1 and in gluing in a handle. If Q_i can be obtained from a sphere with holes by closing some of the holes with handles, then this applies as well to Q_{i-1}.

F) We show by induction that Q_0 is obtained from a sphere by making in it $k+r$ holes ($k \geq 0, r > 0$) and by closing k of them with handles. It remains to add that in going over from Q_0 to the initial surface Q we must again glue in all cores in

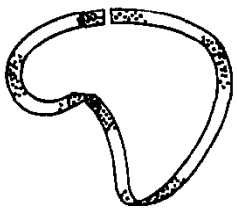


FIGURE 71.

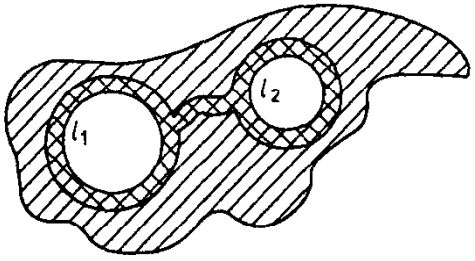


FIGURE 72.

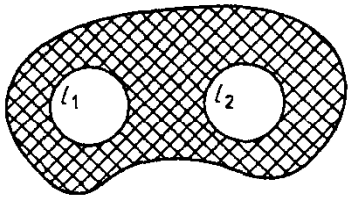


FIGURE 73.

Q_0, i.e., we must close each of the k holes in Q_0 with a disk. Thus Q is obtained from a sphere by making k holes in the latter and closing them with handles, i.e., Q is homeomorphic to one of the surfaces $P_0, P_1, P_2, \ldots$.

Problems

74. Formulate and prove a theorem on the topological classification of orientable surfaces with boundary.

75. Let q nonintersecting closed curves be given on a surface P_k. Suppose that P_k remains connected if we cut it along these curves. Show that $q \leq k$.

76. Let Q be a closed surface with a regular topological decomposition such that each face is a pentagon and four faces meet at each vertex. Show that if the number of faces is not a multiple of 8, then the surface is not orientable.

77. Three curves p, q, r are given on a closed surface Q. The curves are homeomorphic to segments, have common endpoints, and are otherwise pairwise disjoint. If Q remains connected when cut along one of the curves $p \cup q$, $p \cup r$, and $q \cup r$, then this is also true of at least one of the two other curves.

78. Consider a regular dodecahedron (Figure 76a). If we extend all edges of a face to intersection, then we obtain a regular five-pointed star (Figure 76b). Two such stars, constructed for neighboring faces, share a segment; in

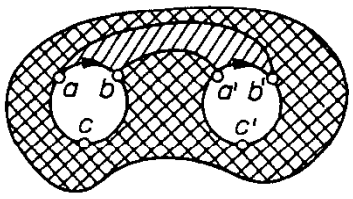


FIGURE 74.

FIGURE 75.

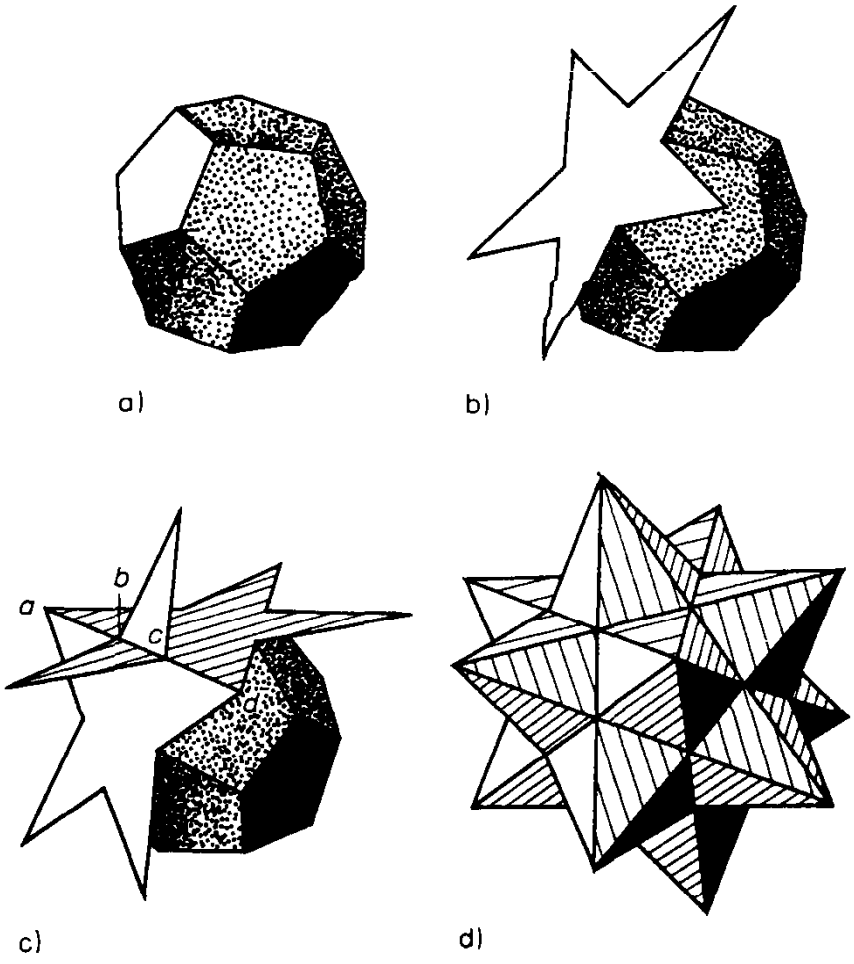


FIGURE 76.

the case of the neighboring faces selected in Figure 76c it is the segment ad. We stipulate that the stars adjoin only at the segments ab and cd, and that the segment bc is their redundant intersection due to their unsuitable embedding in space. If we construct similar stars for all faces of the dodecahedron (Figure 76d), then we obtain a surface Q embedded in space

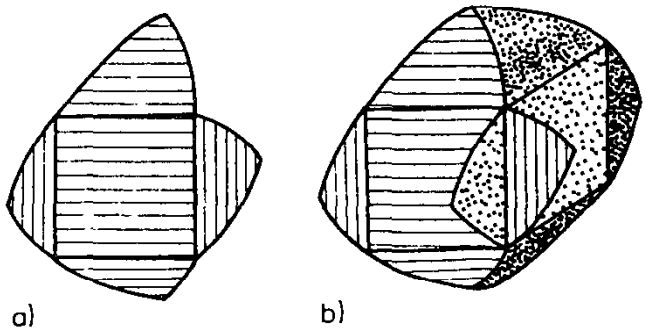


FIGURE 77.

with self-intersections (the intersection lines are the edges of the initial dodecahedron). Show that this surface is orientable and that its Euler characteristic is -16, i.e., that it is homeomorphic to a sphere with nine handles.

Starting with a dodecahedron, we can construct yet another surface. We add to the contour of each star segments such as bc (Figure 76c) and obtain in this way a five-element closed polygonal line (with self-intersections). Then we smooth these polygonal lines so that the self-intersections are eliminated (and so that we still have segments such as ad as sides). In each of these polygonal lines we hang a face (a pentagon). In this way we obtain a surface with 12 pentagons in which the number of vertices (such as a and d) is also 12 and the number of edges is 30. Show that this surface is orientable and that its Euler characteristic is -6, i.e., that it is homeomorphic to a sphere with four handles.

79. Over each of the faces of a cube we set up a "four-spiked star" (with crooked spikes; see Figure 77a) so that neighboring stars touch at the boundaries of their spikes (Figure 77b). In this way we obtain in space a surface with self-intersections (the lines of self-intersection are the edges of the cube). Show that this surface is homeomorphic to P_3.

80. Suppose that we modify the construction in Problem 79 by using "three-spiked stars" and apply it to a tetrahedron, an octahedron, and an icosahedron respectively. Describe the three resulting surfaces.

2.5. Classification of Closed Nonorientable Surfaces

It is impossible to embed a closed nonorientable surface in space without self-intersections.

Example 25 Figure 78a shows a surface with boundary l, and Figure 78b shows a cut across its "throat." If we close the hole l with a disk, then we obtain a closed surface with self-intersection (Figure 78c). The self-intersection is due to our construction of the surface. Later we will see that it is impossible to embed this surface in three-space without self-intersection. Outside three-space we can double

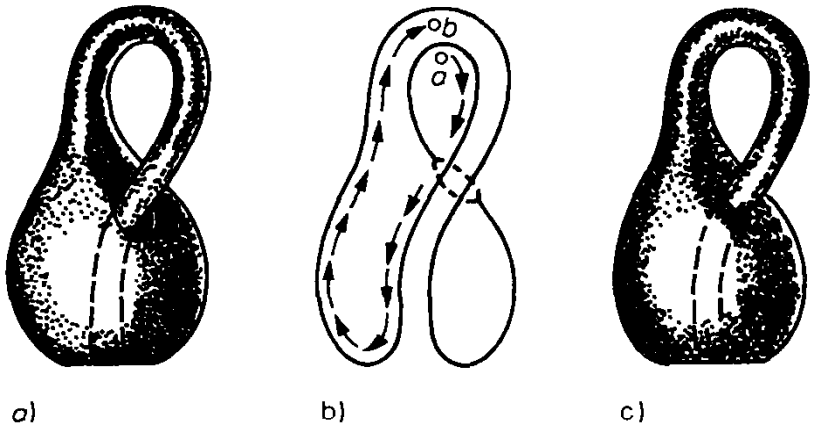


FIGURE 78.

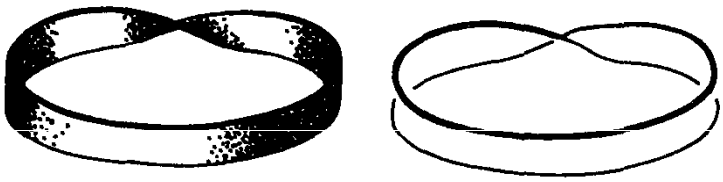


FIGURE 79.

the points of the section curve and in this way eliminate the self-intersection. The resulting surface is called a *Klein bottle*. It is one-sided. By appropriately moving a point located outside the throat we can move it into its interior (Figure 78c).

Example 26 Since the boundary of a Möbius strip is homeomorphic to a circle (Figure 79), we could try to glue it along its boundary to the boundary of a hole in some surface. Figure 80a shows a Möbius strip (a ring with a twist), and Figure 80b shows a piece of a surface Q with a hole cut in it. If we untwist the inner "shovel blade" of Q, then it is easy to see (Figure 80c) that the hole cut in the surface is homeomorphic to a disk. Since the surfaces in Figures 80a and 80b have equal boundaries, we can glue them together along these boundaries, i.e., *we can attach a Möbius strip to a circular hole cut in a surface* Q. It is true that the Möbius strip intersects Q but we can take it that this intersection is solely the result of its "awkward" embedding in space.

There is another description of the closing of a hole with a Möbius strip. This description calls for some preliminaries.

We ask: what figure do we get by cutting a Möbius strip along its midline, i.e., by gluing together the narrow sides of a twisted rectangular strip to obtain a Möbius strip and by cutting it along its midline mnp (Figure 81a)? To answer this question we change the order of the above steps, i.e., we cut a rectangle with reversely oriented short sides along the line mnp (Figure 81b) and glue together the narrow sides of the two halves of the rectangle in accordance with their orientations. To do the latter we flip the lower half of the cut rectangle (Figure 81c) and place the two halves as shown in Figure 81d. Now the required gluing is easy (Figure 81e).

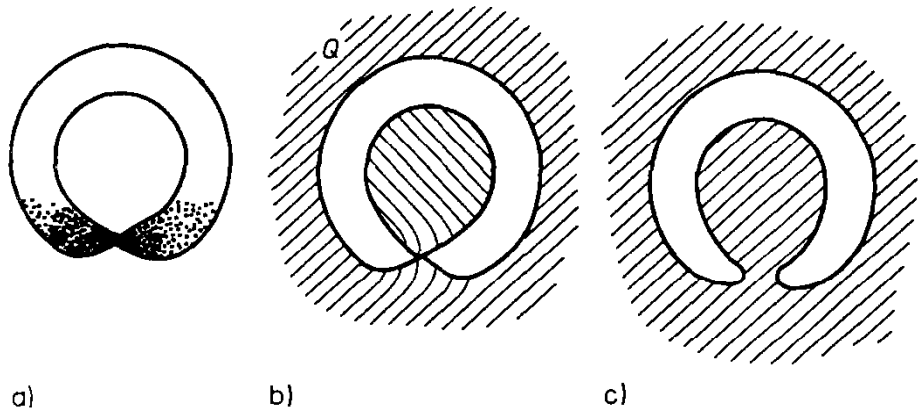


FIGURE 80.

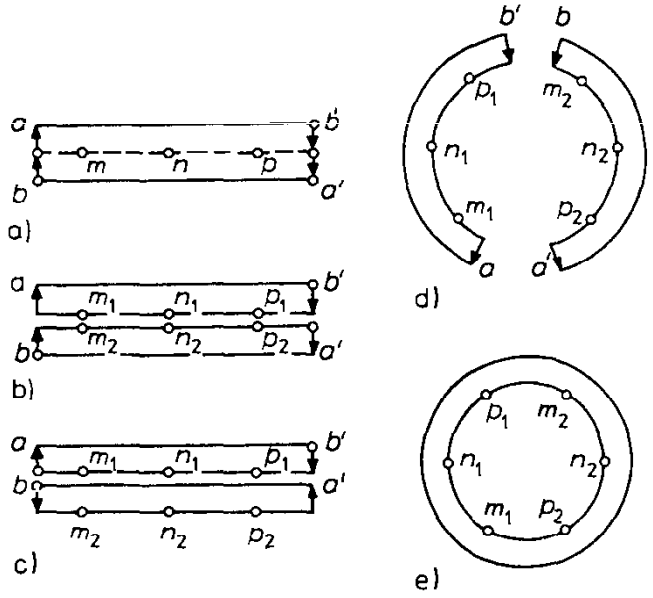


FIGURE 81.

We see that *cutting a Möbius strip along its midline yields a figure homeomorphic to an annulus*. Reversing our steps changes an annulus into a Möbius strip. More specifically, *if on one of the two circles of an annulus we glue together diametrically opposite points* (see Figure 81e), *then we obtain a Möbius strip*.

Now let l be the boundary of a hole on a surface Q. We cut from Q a narrow strip (an annulus) around the hole l with outer boundary l' (Figure 82). As a result we obtain a surface homeomorphic to Q (with a larger hole l') and a detached ring. We now glue together the diametrically opposite points on the boundary l of the annulus. This turns the annulus into a Möbius strip, which we glue in the hole l'. In effect, we have glued a Möbius strip in the surface Q (strictly speaking, in a surface homeomorphic to Q). But cutting the surface along l' and the subsequent gluing are unnecessary steps: It suffices to glue together the diametrically opposite points of l. Thus *gluing together of the diametrically opposite points on the boundary of a hole is equivalent to gluing in it a Möbius strip*.

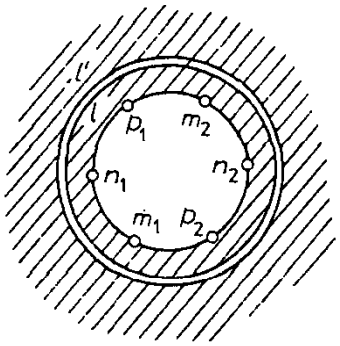


FIGURE 82.

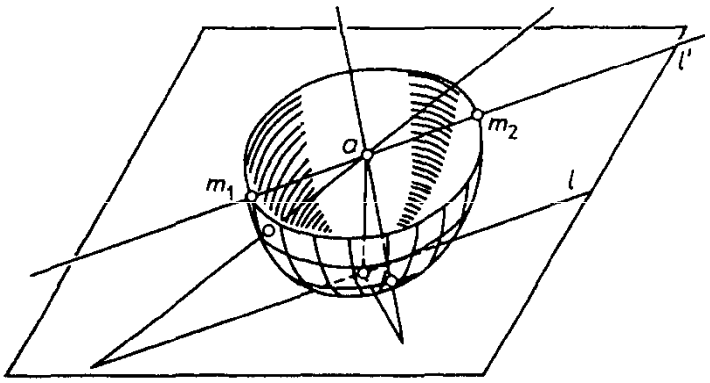


FIGURE 83.

Example 27 In projective geometry one adds to the points of the Euclidean plane "points at infinity." With each straight line on the Euclidean plane we associate a point at infinity. The points at infinity associated with parallel straight lines are the same (i.e., "parallel straight lines intersect at infinity"), and the points at infinity associated with nonparallel straight lines are different. The plane enlarged by the addition of points at infinity is called the *projective plane*.

To clarify the topological structure of the projective plane we consider a hemisphere centered at o that touches the plane and is disposed so that the plane determined by the boundary of the hemisphere is parallel to the (original) plane (Figure 83). The central projection with center o is a homeomorphism of the open hemisphere (i.e., the hemisphere minus its boundary) to the whole Euclidean plane.

Draw a straight line l through the point of contact of the hemisphere with the plane and a straight line l' through o parallel to l. The straight lines l and l' intersect at the same point at infinity. This means that our projection maps the points m_1 and m_2 in which l' meets the boundary of the hemisphere to the point at infinity on l. But then the projection of the hemisphere with the boundary onto the projective plane is not one-to-one. To make the projection one-to-one (and therefore a homeomorphism) we must identify diametrically opposite points on the boundary of the hemisphere. In other words, the projective plane is homeomorphic to a hemisphere with a Möbius strip attached to its boundary (or to a sphere with

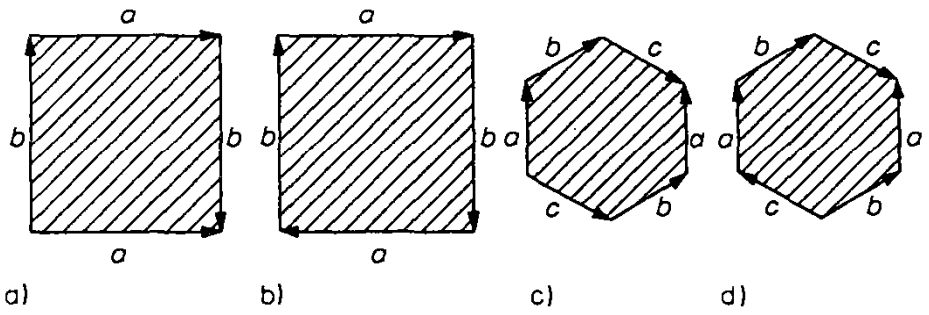


FIGURE 84.

a hole closed with a Möbius strip). It follows that unlike the Euclidean plane, the projective plane is one-sided.

We can now formulate the second half of the theorem of Möbius and Jordan on the classification of surfaces. We will list all topologically different types of closed nonorientable surfaces. Let N_q be the surface obtained from a sphere with q holes each of which is closed with a Möbius strip. It turns out that *the surfaces*

$$N_1, N_2, \ldots, N_q, \ldots \tag{12}$$

yield a complete topological classification of all closed nonorientable surfaces.

Problems

81. Show that if one cuts a hole in the surface N_q, then one obtains a surface that can be embedded in three-space without self-intersections.
Hint. The resulting surface is homeomorphic to the surface studied in Problem 62.

82. Show that the Euler characteristic of the surface N_q is $2 - q$.

83. Consider a sphere with $m + n + p$ holes. If m of them are closed with handles and n with Möbius strips, show that the Euler characteristic of the resulting surface is $2 - 2m - n - p$.

84. Show that the graph "4 houses and 4 wells" (whose edges are paths joining different houses to different wells) cannot be put on a projective plane without self-intersections but can be so put on a torus.

85. Suppose that one can draw the graph "m houses and n wells" on a surface Q. Show that $\chi(Q) \leq m + n - mn/2$.

86. Which of the surfaces $N_1, N_2, N_3, \ldots$ is homeomorphic to a Klein bottle and which is homeomorphic to a projective plane?

87. Consider the figures a, b, c, d in Figure 84. What surfaces do we obtain if, minding orientations, we glue together the sides marked with the same letters?

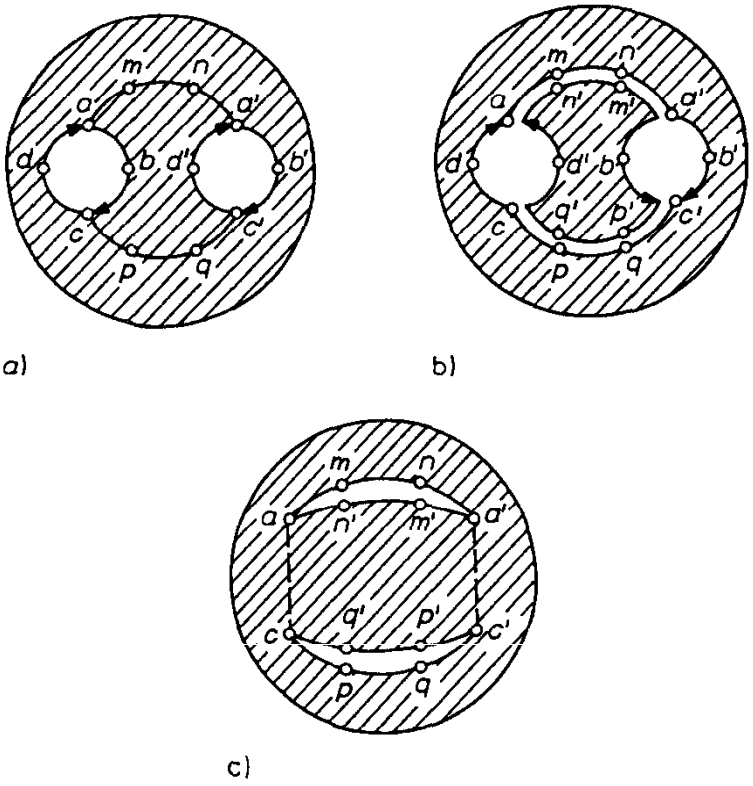
88. Consider a Möbius strip in three-dimensional space R^3 and a point p in R^4 that contains R^3. Add to the Möbius strip all segments that join its boundary points to p. Show that the resulting surface is homeomorphic to a projective plane. Show also that every surface N_q can be embedded in R^4 without self-intersections.

The surfaces $N_1, N_2, N_3, \ldots$ are pairwise nonhomeomorphic for they have different Euler characteristics (Problem 82). Thus in order to show that these surfaces yield a complete topological classification of all closed nonorientable surfaces it suffices to show that every closed nonorientable surface is homeomorphic to one of the surfaces $N_1, N_2, N_3, \ldots$. This proof is similar to the proof in Section 2.4. One difference is that now the strip L (see Figure 70) can be homeomorphic to a Möbius strip (we are dealing with nonorientable surfaces). In that case the surface Q_i^* resulting from the removal of L has just one component (for the boundary of L, i.e., of a Möbius strip, is homeomorphic to a circle). Conversely, Q_{i+1} arises out of Q_i^* if L is glued to one of the boundary components of Q_i^*, i.e., Q_{i-1} arises out of Q_i^* *when a hole is closed with a Möbius strip.* Another difference is that now the gluing together of the curves l_1 and l_2 shown in Figures 72 and 73 can be carried out regardless of whether they have the same or opposite orientations (in the latter case, the gluing together is equivalent to gluing in a handle; see Figure 75). In the first case Q_{i-1} arises out of Q_i^* as a result of two holes being closed with Möbius strips (Problem 89). Thus the argument in Section 2.4 shows that *every closed nonorientable surface Q can be obtained from a sphere with $k + q$ holes by closing k of the holes with handles and q of them with Möbius strips.* Here $q \geq 1$; if q were 0, then we would obtain the orientable surface P_k. It remains to mention that if we have glued in at least one Möbius strip in the surface, then gluing in a handle is equivalent to gluing in two Möbius strips (Problem 90). It follows that the surface obtained from a sphere with $k + q$ holes by closing k of the holes with handles and q of them with Möbius strips ($q \geq 1$) is homeomorphic to the surface obtained from a sphere with $2k + q$ holes by closing all of them with Möbius strips. In other words, Q is homeomorphic to one of the surfaces $N_1, N_2, N_3, \ldots$.

Problems

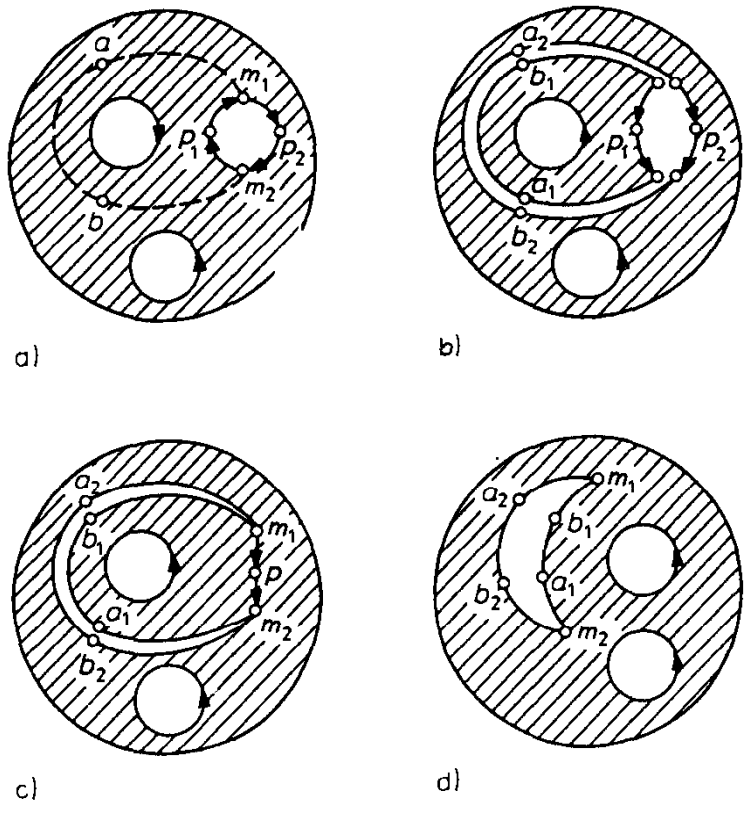
89. Two holes are cut in a disk. Their boundaries l_1 and l_2 are given the same orientations and are glued together. Show that this is equivalent to closing each of the holes with a Möbius strip.
Hint. Make two auxiliary cuts along the curves $amna'$ and $cpqc'$ (Figure 85a) and flip the cut-out piece (Figure 85b). Now l_1 and l_2 can be glued together (Figure 85c). One must also close the two auxiliary cuts, i.e., glue together the diametrically opposite points on each of the two boundaries.

90. Cut three holes in a disk. Close one of them with a Möbius strip. Give the boundaries of the two remaining holes opposite orientations and glue them together (this gives a handle). Show that all this is equivalent to closing all three holes with Möbius strips.
Hint. Make the auxiliary cut m_1abm_2 and flip the cut-out piece (Figure 86). The resulting hole has the shape of a sickle. Glue together the diametrically opposite points on its boundary. Also, give the the same orientations to the boundaries of the remaining two holes and glue them together.



a) b) c)

FIGURE 85.



a) b) c) d)

FIGURE 86.

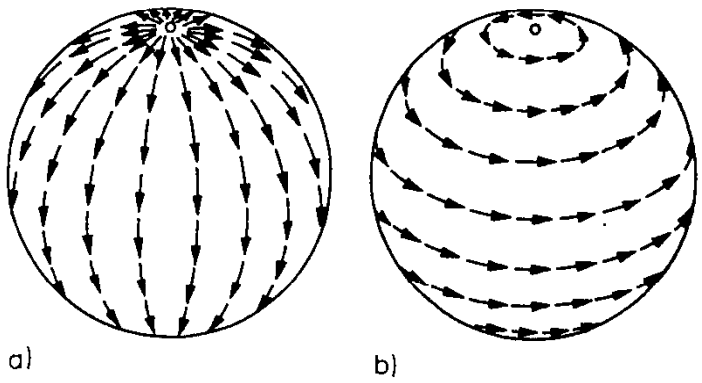


FIGURE 87.

91. Formulate and prove a theorem on the topological classification of nonorientable surfaces with boundary.

2.6. Vector Fields on Surfaces

In this section we investigate the following problem: Can one construct a continuous *direction field* on a given orientable surface Q, i.e., can one choose at each of its points a nonzero tangent vector that changes continuously when we move from point to point?

Example 28 The directions from north to south on a sphere (Figure 87a) have *singular points* at the poles; at these points the vectors have different orientations and the continuity is destroyed. The same is true of directions from west to east (Figure 87b). We will see later that, quite generally, there is no continuous direction field on the whole sphere. Another formulation of the same fact is known as the hedgehog theorem: If a spine (a nonzero not necessarily tangent vector) grows at each point of a sphere and the direction of the spines changes continuously, then there is at least one spine that is perpendicular to the sphere. If this were false, then, by projecting the spine $\overrightarrow{aq}$ to the tangent plane at a in the direction parallel to the radius at a, we would obtain a continuous field of nonzero tangent vectors, and this is impossible.

Figures 89a and 89b show the vector fields of Example 28 near the north pole. Figure 89c shows a more complicated singular point (known as a *saddle point*). If we go around all three singular points once—say counterclockwise—then the direction vector also rotates once counterclockwise in the first two cases and clockwise in the third case (see Figures 90a, 90b, and 90c). We say that the singular point has index $+1$ in the first two cases and index -1 in the third case.

The French mathematician Henri Poincaré (1854–1912) proved the following result. *Let Q be a closed orientable surface. If a vector field of nonzero tangent vectors is given on Q, and if that vector field is continuous with the exception of a finite number of singular points, then the sum of the indices of all singular points is $\chi(Q)$.*

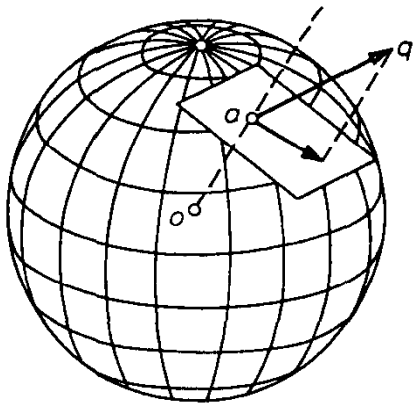


FIGURE 88.

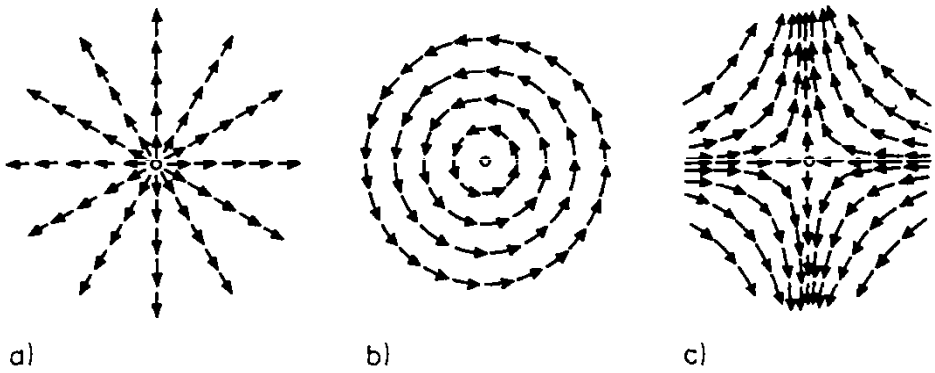


FIGURE 89.

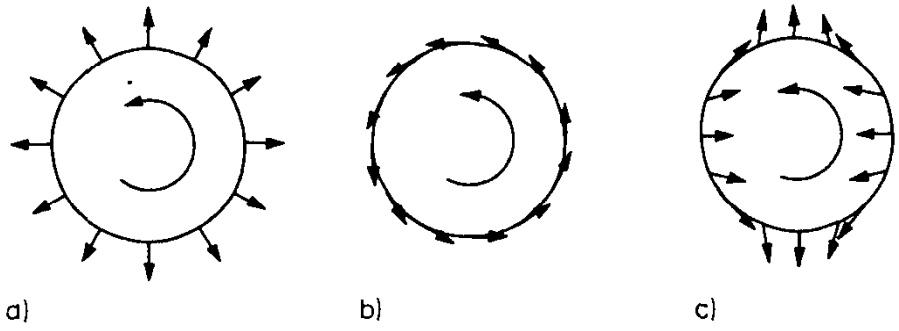


FIGURE 90.

Example 29 Since $\chi(P_k) = 2 - 2k$, $\chi(P_k) \neq 0$ for $k \neq 1$. It follows that a continuous vector field of nonzero tangent vectors on an orientable surface other than a torus P_1 must have singular points. (The vectors on parallel great circles on a torus form a continuous vector field of nonzero tangent vectors that has no singular points.)

Our proof of the Poincaré theorem involves two steps. First we show that the index sums of two vector fields are equal, and then we construct a vector field for which this sum is easy to compute.

Consider two different vector fields of nonzero vectors on a surface Q each of which has finitely many singular points. Let $v_1(x)$ be the vector of the first field at

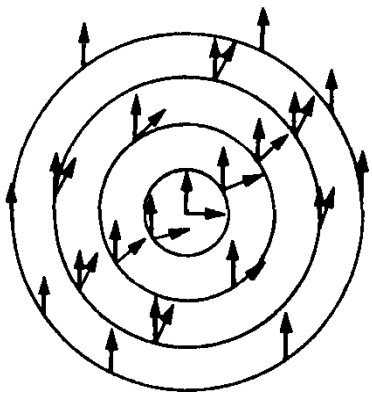


FIGURE 91.

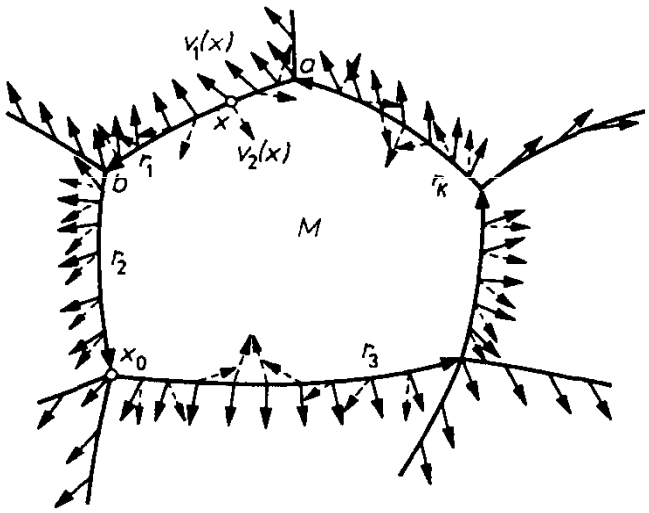


FIGURE 92.

x and $v_2(x)$ the vector of the second field at x. We divide Q into small polygons such that each of them contains at most one singular point of each of the two fields and all singular points are in the interior of these polygons.

We note that if x is a nonsingular point of the vector field v_1, then it is possible to turn the vectors near it so that the resulting vector field is also continuous and $v_1(x)$ goes over into a preassigned vector (Figure 91: With increasing neighborhood radius the vectors turn less and less). We take advantage of this and turn the vectors of the field v_1 near the vertices of the polygonal net so that the vectors $v_1(x)$ and $v_2(x)$ coincide at every vertex (Figure 92).

Since Q is orientable, we can specify on it a positive direction for measuring angles (say counterclockwise, as determined by an observer outside the surface).

Now we consider an edge r_1 and distinguish on it a direction (say, from vertex a to vertex b). When we go from a to b in this direction we observe the vector $v_1(x)$, and when we return from b to a we observe the vector $v_2(x)$. When we traverse the edge r_1 there and back, the observed vector changes continuously and returns to the initial position (for $v_1(a) = v_2(a)$ and $v_1(b) = v_2(b)$). We denote the number of revolutions (with respect to a specified direction for measuring angles) of the observed vector by $d(r_1)$. In Figure 92 we have $d(r_1) = 1$, $d(r_2) = 0$, and

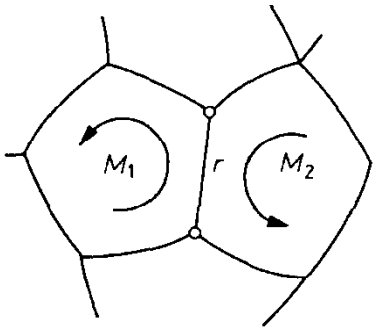


FIGURE 93.

$d(r_3) = -1$. If we specify the opposite direction on r_1 (from b to a), then $d(r_1)$ changes its sign (for the observed vector turns in the opposite direction).

Let M be one of our polygons. If we traverse its contour (in the positive direction), then each of the vectors $v_1(x)$ and $v_2(x)$ revolves a certain number of times. We denote these numbers by $z_1(M)$ and $z_2(M)$ respectively.

Let $r_1, r_2, \ldots, r_k$ be the edges of the polygon M oriented in accordance with the positive orientation of its boundary (see Figure 92). Beginning at a, we traverse the boundary in the positive direction and observe the vector $v_1(x)$. After returning to a we traverse the boundary in the opposite direction and observe the vector $v_2(x)$. The total number of revolutions of the observed vector is $z_1(M) - z_2(M)$. But we can observe the rotations of the vectors "in small portions": We observe $v_1(x)$ when traversing r_1 and $v_2(x)$ when traversing r_1 in the opposite direction. Then we observe $v_1(x)$ when traversing r_2 and $v_2(x)$ when traversing r_2 in the opposite direction, and so on. In this case we count $d(r_1) + d(r_2) + \cdots + d(r_k)$ revolutions. Since the total rotation is independent of the order in which we add the angles of rotation of the vectors on each edge, it follows that

$$z_1(M) - z_2(M) = d(r_1) + d(r_2) + \cdots + d(r_k). \tag{13}$$

From (13) we can easily deduce the relation

$$\sum z_1(M) = \sum z_2(M), \tag{14}$$

where we are summing over all polygons. Indeed, let us sum equation (13) over all polygons and consider the sum on the right-hand side of the resulting equation. Since each edge r belongs to two polygons M_1 and M_2 (see Figure 93), it appears in that sum twice. However, when the boundary of M_1 is traversed in the positive direction r obtains a certain orientation, and when the boundary of M_2 is traversed in the positive direction r obtains the opposite orientation. Hence $d(r)$ and $-d(r)$ appear exactly once on the right-hand side. Since this is true for every edge, it follows that $\sum z_1(M) - \sum z_2(M) = 0$.

Let M be a polygon and x_0 a singular point of the field $v_1(x)$ in that polygon. We imagine a system of simple closed curves in M that circle x_0 (Figure 94). Since the field $v_1(x)$ is continuous, it follows that if we go from one of two close curves of our system to another, then the number of revolutions of the vector $v_1(x)$ changes

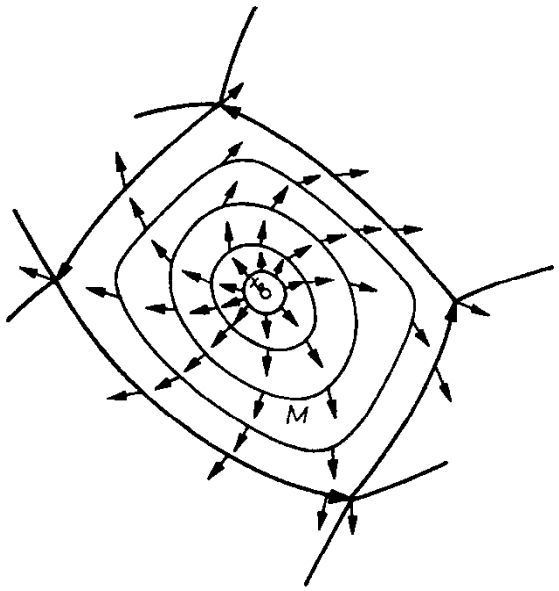


FIGURE 94.

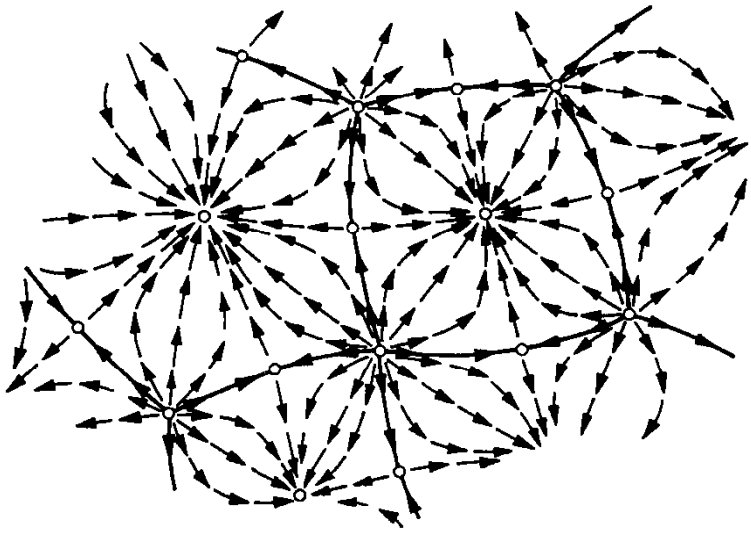


FIGURE 95.

by a small amount. But this number is an integer, which means that the only small change is a zero change; i.e., the number in question is unchanged when we go over from one curve to the next. If we go around the boundary of M, then the number of revolutions involved is $z_1(M)$. If we go around a circle centered at x_0, then the number of revolutions involved is equal to the index of x_0. This means that $z_1(M)$ is equal to the index of x_0 (if there are no singular points in the interior of M, then $z_1(M) = 0$). It follows that $\sum z_1(M)$ is equal to the sum of the indices of all singular points of the field $v_1(x)$. Similarly, $\sum z_2(M)$ is equal to the sum of the indices of all singular points of the field $v_2(x)$. This and (14) show that *the sum of the indices is the same for both fields*. This proves the first part of Poincaré's theorem.

We now choose in each polyhedron a "center" and on each edge its "midpoint." Then we construct a vector field of the kind shown in Figure 95. On the edges the

FIGURE 96.

vectors are oriented from the vertices to the midpoints. Also, vectors issue from the vertices and enter the centers. The singular points of this field on the surface Q are the vertices, the centers, and the midpoints. The index of a vertex or a center is $+1$, and the index of a midpoint of an edge is -1 (it is a saddle point) (Figure 96). It follows that for this field—and thus for every field—the sum of the indices of all singular points is equal to $E \cdot (+1) + V \cdot (-1) + F \cdot (+1) = \chi(Q)$. (*Translator's note.* For a remarkable application of this result see Section **3.10.** (Morse theory))

Problems

92. Show that on every closed surface there is a vector field with just one singular point.

93. Show that on every surface with boundary there is a vector field without singular points (the vectors at the boundary points are tangent to the surface but not necessarily to the boundary).

94. Show that Poincaré's theorem holds for orientable surfaces with boundary, provided the boundary vectors are tangent to the boundary.

95. Prove Brouwer's theorem: *Let D be a (closed) disk. Every continuous mapping $f\colon D \to D$ has a fixed point, i.e., a point $x \in D$ such that $f(x) = x$.*

Hint. Suppose there is no fixed point. If we associate with each point x the vector from x to $f(x)$, then we obtain a nonzero continuous vector field without singularities.

2.7. The Four Color Problem

Call the regions into which a finite graph G divides the plane *countries*. Countries A and B in Figure 97 are neighbors (they share an edge), and so are countries B and C (they share two edges). Countries A and C are not neighbors; they share a vertex but not an edge.

We want to color the countries with different colors so that the the resulting map is "political," in the sense that neighbor countries are colored differently. In order to economize on the number of colors nonneighbor countries may be colored the same color. What is the least number of colors needed to color any map on the plane?

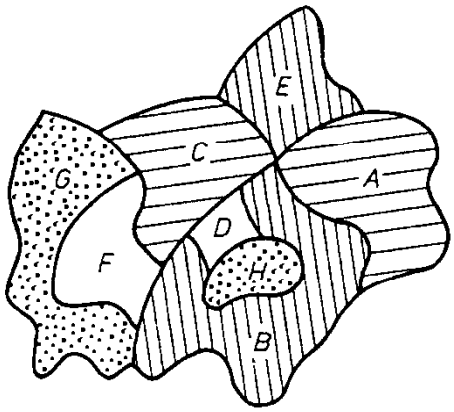


FIGURE 97.

This problem was posed in 1852 by a student in London named Guthrie who was struck by the fact that four colors sufficed for coloring the counties on a map of England. Guthrie conjectured that *four colors suffice for coloring any map.* After close to forty years the English mathematician Heawood showed that five colors suffice for coloring any planar map. The four color problem attracted ever greater attention. In 1968, Ore and Stemple showed that a map with no more than 40 countries could be colored with four colors.

At present, we believe that the four-color-problem conjecture has been proved. The reason for saying "we believe" is that, so far, all proofs of the conjecture have made use of computers, and have involved so many computations that their verification is a practical impossibility.

The first machine solution was obtained in 1976 by the Americans K. Appel and W. Haken. With the aid of a computer (which "helped them" to improve stepwise an initial program) they were able to reduce all possible maps to some 2000 (precisely given) types and developed a computer program for their investigation. Except for three types of maps (which it could not handle and which were verified "by hand"), the computer used the program to solve the following problem: *Consider a particular type of map. Can you find a map of this type that cannot be colored with four colors*? After some ten billion arithmetical and logical operations the computer answered "no," went over to the next type of map, and so on. After getting a "no" answer for all types of maps, Appel and Haken declared that they had obtained a computer-aided solution of the four color problem.

But the correctness of this machine solution cannot be guaranteed. For it is conceivable that for some type of map—say type 17—the "no" answer was the result of a flaw in the electronics (a rather common phenomenon) and not of faultless analysis. Since the computer operators do not know this, they ignore type 17 and continue with map types 18, 19, and so on. Nor could we guarantee correctness if we were willing to repeat the machine experiment (and thus delay matters for months); it is conceivable that a mistake can occur when the "new" computer carries out the millions of computations involved in checking the answer for maps of type 17.

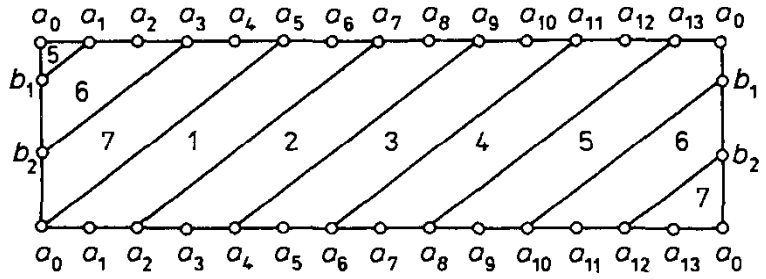


FIGURE 98.

To find out about the present status of the four color problem see the paper by Robin Thomas, An update on the four color problem, *Notices of the AMS*, August 1998, 848–859. (Trans.)

Problems

96. Consider a graph on the plane (or on the sphere) all of whose vertices have even index. Show that the resulting map can be colored with two colors. *Hint.* Use the intersection index.

97. Show that every map in the plane (or on the sphere) can be colored with five colors.

98. Consider a graph on a surface such that at least one of any two neighboring countries is a triangle. Show that such a map can be colored with four colors.

99. Consider two concentric circles connected by partitions. How many colors are needed to color the resulting map?

2.8. Coloring Maps on Surfaces

Example 30 Heawood showed that *an arbitrary map on a torus can be colored with seven colors* (this follows from equation (16) proved below). He also gave an example that showed that fewer colors do not suffice. Specifically, if we join the opposite sides of the rectangle in Figure 98, then it becomes a torus with seven countries any two of which are neighbors (Figure 99), i.e., each country must be colored a different color.

If n colors suffice for coloring every map on a surface Q and there is a map that cannot be colored with fewer than n colors, then n is called the *chromatic number* of Q; it is denoted by $\operatorname{col}(Q)$. As noted earlier, for the sphere $\operatorname{col}(P_0) = 4$ and for the torus $\operatorname{col}(P_1) = 7$. Except for the Klein bottle N_2, for which $\operatorname{col}(N_2) = 6$, the chromatic number of a closed surface is given by the *Heawood formula*

$$\operatorname{col}(Q) = \left[\frac{7 + \sqrt{49 - 24\chi(Q)}}{2}\right]. \tag{15}$$

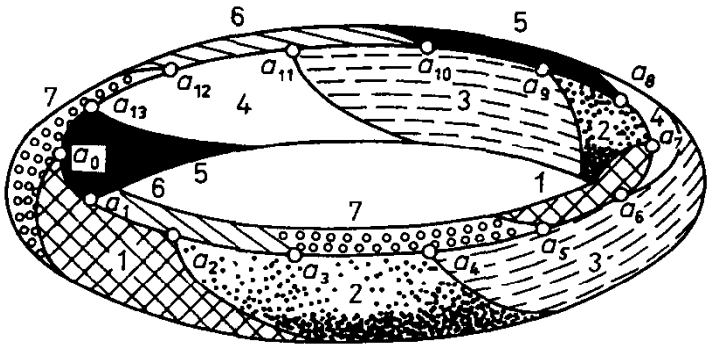


FIGURE 99.

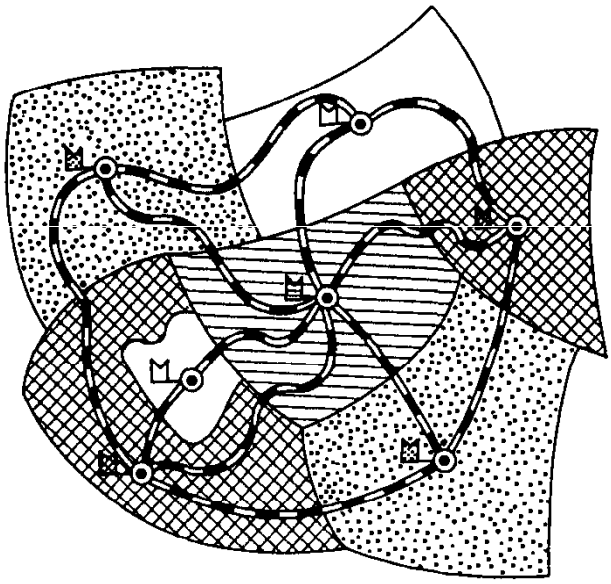


FIGURE 100.

Here the brackets indicate the integral part of the fraction.

We owe these results to a number of generations of mathematicians. Heawood was able to prove the inequality

$$\operatorname{col}(Q) \leq \left[\frac{7+\sqrt{49-24\chi(Q)}}{2}\right]. \tag{16}$$

It thus remained to show that there is a map on Q that cannot be colored with fewer colors than given in (15). At first such maps were given for a few particular orientable and nonorientable surfaces. The existence of the required maps for arbitrary nonorientable surfaces was proved by Ringel (1954), and for arbitrary orientable surfaces by Ringel and Youngs (1968).

We will prove the inequality (16). Consider a map on Q whose coloring requires $c = \operatorname{col}(Q)$ colors. We choose a point (the "capital") in the interior of each country. On the territory of any two neighboring countries we build a "railway line" connecting their capitals (Figure 100) so that the different railway lines do not cross. Instead of coloring a country a particular color we set up a flag of this color in its capital. If two countries are connected by a railway line, (i.e., if they

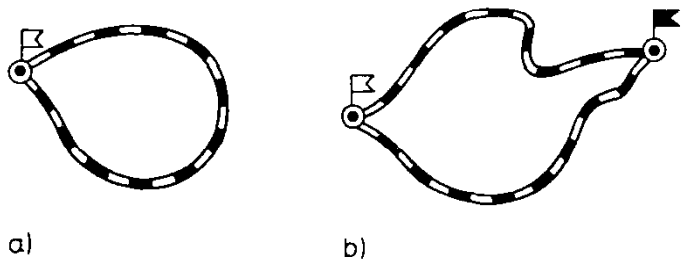


FIGURE 101.

are neighbors), then the colors of their flags must be different. It follows that we must color the vertices of the graph G^* (whose edges correspond to the railway lines) so that any two neighboring vertices (i.e., vertices connected by an edge) are colored differently. It is clear that the *chromatic number* of G^*, i.e., the smallest number of colors needed for coloring it in the indicated way, is c.

We remove from G^* an arbitrary vertex together with all edges that abut on it. Call the resulting graph G'. If the chromatic number of G' is not less than c, then we can take G' instead of G^*. It can happen that we can repeat this step, i.e., we can again remove from G' a vertex and the edges abutting on it, and so on. Finally, we obtain a graph G^{**} in G^* that cannot be further simplified, i.e., the chromatic number of G^{**} is c but the chromatic number is decreased if we remove from it any vertex and the edges abutting on it. We denote the number of vertices of G^{**} by V, the number of its edges by E, and the number of faces determined by this graph on Q by F. Then (see Problem 72)

$$V - E + F \geq \chi(Q). \tag{17}$$

The number of edges terminating at any vertex of G^{**} is at least $c - 1$. Suppose not, i.e., suppose that the edges terminating at a vertex $b \in G^{**}$ are $[bq_1], \ldots, [bq_k]$ and $k < c - 1$. We remove b and these edges from G^{**} and obtain in this way a graph G'' whose chromatic number is less than c. We color this graph with $c - 1$ colors. Since $k < c - 1$, it follows that at least one of these colors was not used when coloring the vertices $q_1, \ldots, q_k$. If we color b with the unused color, then we obtain a coloring of G^{**} with $c - 1$ colors. Of course, this contradicts the choice of G^{**}.

Thus the number of edges abutting on each vertex of G^{**} is at least $c - 1$. It follows (see Problem 20) that

$$(c - 1)V \leq 2E. \tag{18}$$

We claim that each of the regions determined by G^{**} has at least three edges. Indeed, a "one-edge" (Figure 101a) would imply the existence of a railway line from a capital to the same capital (without a detour through other capitals) and a "two-edge" (Figure 101b) would imply that two capitals are connected by two railway lines; but we did not build such railway lines.

If we count the edges of all F regions, then we see that this number is at least $3F$; here we counted each edge twice (for it abuts on two regions). Thus $3F \leq 2E$, i.e.,

$\frac{2}{3}E - F \geq 0$. If we add this inequality to (17), then we obtain $V - \frac{1}{3}E \geq \chi(Q)$, or $2E \leq 6V - 6\chi(Q)$. Using (18), we obtain $(c-1)V \leq 6V - 6\chi(Q)$, i.e.,

$$c - 1 \leq 6 - \frac{6\chi(Q)}{V}. \tag{19}$$

Now it is not difficult to complete the proof. Suppose that the surface Q is homeomorphic to a sphere: $Q = P_0$. Then $\chi(Q) = 2$, i.e., the inequality to be proved, namely (16), is of the form $\text{col}\,(Q) \leq 4$. This inequality is correct, for the four color problem has been solved.

Now let $Q = N_1$, i.e., $\chi(Q) = 1$. Then (16) has the form $\text{col}\,(Q) \leq 6$. This inequality is correct, for (19) implies that $c - 1 \leq 6 - \frac{6}{V}$, so that $c - 1 \leq 5$ (for $c - 1$ is an integer).

Finally, let Q be different from P_0 and N_1. Then $\chi(Q) \leq 0$ (see Problems 70 and 82). Since $V \geq c$ (otherwise, we could color the graph with $c - 1$ colors), we have $-6\chi(Q)/V \leq -6\chi(Q)/c$, and thus, in view of (19), $c - 1 \leq 6 - 6\chi(Q)/c$, i.e., $c^2 - 7c + 6\chi(Q) \leq 0$. This means that c lies in the interval bounded by the zeros of the polynomial $x^2 - 7x + 6\chi(Q)$ (the zeros are real, for $6\chi(Q) \leq 0$). Hence c is not greater than the greater of the two zeros, i.e.,

$$c \leq \frac{1}{2}\left(7 + \sqrt{49 - 24\chi(Q)}\right),$$

which shows that (16) holds in this case as well.

Problems

100. A surface Q has been obtained from a sphere with $k + q$ holes by closing k of the holes with handles. Show that $\text{col}\,(Q) = \text{col}\,(P_k)$.

101. A surface Q has been obtained from a sphere with $k + q$ holes by closing q of the holes with Möbius strips. Show that $\text{col}\,(Q) = \text{col}\,(N_q)$. In particular, the chromatic number of the Möbius strip is 6.

102. Find a map on the projective plane (or on the Möbius strip) that cannot be colored with five colors.

103. Determine the chromatic number of a graph whose vertices and edges are, respectively, the vertices and edges of an n-gon.

104. Determine the chromatic number of the graph "m houses and n wells."

105. Show that for every surface there is a graph that cannot be embedded in it.

106. Suppose that a graph with chromatic number 2 has n vertices. What is the largest number of edges this graph can have?

107. Show that if we draw on a given surface a map with sufficiently small countries, then we can color it with seven colors.

108. Show that if we can draw on a surface Q a graph with chromatic number c, then $\text{col}\,(Q) \geq c$.

109. Show that it is possible to draw on a Klein bottle a complete graph with six vertices. Deduce from this that $\text{col}\,(N_2) \geq 6$.

2.9. Wild Spheres

In this section we study the *intersection index in space* and deal with problems connected with the space version of the Jordan curve theorem.

Let Q be a surface (possibly with boundary) made up of plane polygons, and let G be a graph whose edges are rectilinear segments. G and Q are said to be *in general position* if the vertices of the graph are not in Q and its edges have no points in common with the edges of the polygons in Q. If the number of points of intersection of the graph with the surface is even, then we put $J(G, Q) = 0$, and if this number is odd, then we put $J(G, Q) = 1$. The number $J(G, Q)$ is called the *intersection index* of the graph G with the surface Q.

As in Section 1.5, we show that *if the surface Q has no boundary and the graph G is a cycle* (i.e., *the number of edges at each of its vertices is even*), *then* $J(G, Q) = 0$.

Problems

110. The union of finitely many polygons in space is called a *two-dimensional cycle* (*modulo* 2) if these polygons have no interior points in common and the number of polygons at each edge is even. Show that if G is a one-dimensional cycle, Q a two-dimensional cycle, and G and Q are in general position, then $J(G, Q) = 0$.

111. Let Q be a two-dimensional cycle modulo 2. Think of it as made up of metal polygons. Show that it is possible to fill the regions into which Q divides space with two liquids of different color so that the colors of the liquids on each side of any polygon are different.

112. Show that there is a "spatial map" whose coloring (or filling with fluids of different colors) calls for at least 2000 colors.

113. Consider finitely many oriented polygons in space that have no common interior points (but may have common edges or vertices). They are said to form a *two-dimensional integral cycle* if at each oriented edge the number of polygons that abut on it positively (exemplified by M_1 and M_4 in Figure 102) is equal to the number of polygons that abut on it negatively (exemplified by M_2 and M_3 in Figure 102). Show that the intersection index

$$J(G, Q) = \sum_{i,j} J(r_i, M_j)$$

is 0 if G is a one-dimensional integral cycle and Q is a two-dimensional integral cycle in space. Here the sum extends over all oriented segments $r_1, \ldots, r_k$ that form the cycle G and over all oriented polygons that form the cycle Q. Moreover, $J(r_i, M_j) = +1$ if the orientations agree with the right-hand rule (Figure 103a) and $J(r_i, M_j) = -1$ otherwise (Figure 103b).

114. Let Q be a two-dimensional cycle and let p and q be two points that are not incident with it. Let x be a directed polygonal line from p to q. Show

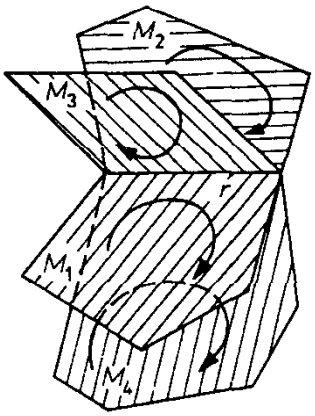


FIGURE 102.

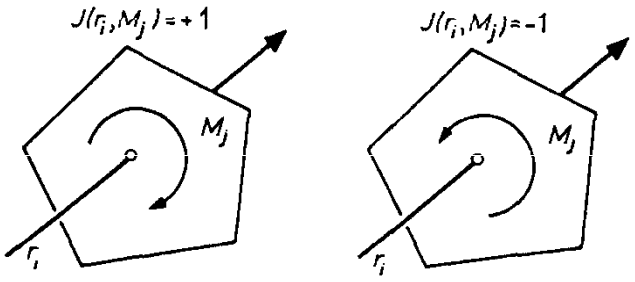


FIGURE 103.

that the intersection index $J(x, Q)$ depends on the disposition of p and q but not on the choice of x.

115. Let Q be a two-dimensional cycle modulo 2. Show that its polygons can be oriented so that it becomes an integral cycle.

116. Let Q be a two-dimensional integral cycle. Show that if there exists a directed polygonal line x (that is not closed) for which $J(x, Q) = n$, then Q divides space into at least $n + 1$ regions. Is the converse true?

Using the notion of intersection index modulo 2 one can prove (as in Section 1.6) the space analogue of the Jordan curve theorem: *A closed surface without self-intersections disposed in three-space divides it into two parts.* One of these two parts is bounded and is called its *interior*, while the other part is unbounded and is called its *exterior*. It is easy to see that a surface without self-intersections disposed in three-space must be two-sided (the statement of the theorem makes no reference to this), for it divides space into two parts. This confirms the fact that closed one-sided surfaces cannot be placed in space without self-intersections.

Problems

117. Let Q be a surface made up of plane polygons and let l be a ray issuing from a point c in space. Show that c is in the interior of Q if and only if l intersects Q in an odd number of points.

118. Does there exist a set in space that is the common boundary of three regions?

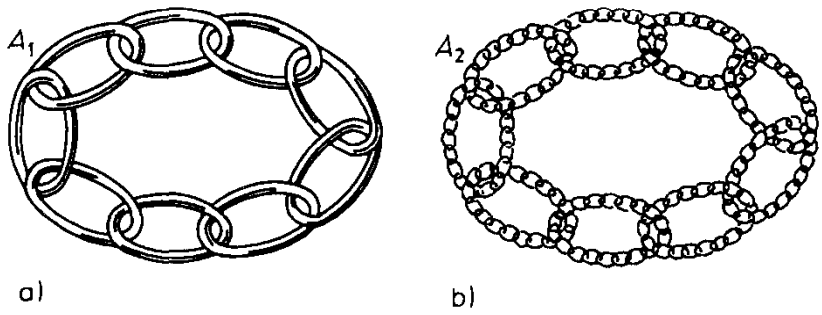


FIGURE 104.

At the end of Section 1.6, when proving the plane version of the Jordan curve theorem, we noted that the union of a simple closed curve and its interior is homeomorphic to a disk. It seems obvious that the spatial analogue of this result, i.e., the statement that the union of a surface homeomorphic to a sphere and its interior must be homeomorphic to a ball, is true. This is indeed the case for simple surfaces, such as convex polyhedra, but not in general—here our intuition has deceived us. In other words, there is a surface in three-space homeomorphic to a sphere such that its union with its interior is not homeomorphic to a ball. The construction of such a "wild sphere" is connected with the work of two mathematicians, the Frenchman Antoine and the American Alexander.

Before describing a wild sphere we talk about *contractibility of a curve.* A simple closed curve in a figure M is said to be *contractible* if it can be contracted (in M) to a point. We can visualize this by considering a system of concentric circles on a disk.

Problems

119. Let M be an open ball. Show that a simple closed curve in M is contractible.

120. Let A be a figure in R^3 consisting of a finite number of points. Show that a circle l in the exterior of A is contractible in the exterior of A.
Hint. Consider a ball containing the circle l. The ball can be deformed by depressions so that the resulting solid continues to be homeomorphic to a ball and contains in its interior the curve l but none of the points of A.

The assertion in Problem 120 can be explained as follows. If l is a circle that bypasses the (finitely many) points of A, then the disk spanning l can be "removed" from the points of A it passes through (if any) by a small deformation. This may give the impression that a circle in the complement of any zero-dimensional set A is contractible. But this is false.

Example 31 We consider a closed chain A_1 with a number of solid links (Figure 104a). We embed a similar closed chain in the interior of each link of A_1 and obtain in this way a set $A_2 \subset A_1$ (Figure 104b). We apply the same procedure to the links of A_2 and obtain $A_3 \subset A_2$. By continuing this process we end up with a nested sequence of sets $A_1 \supset A_2 \supset A_3 \supset \cdots$. The intersection of these sets is *Antoine's necklace* A^*.

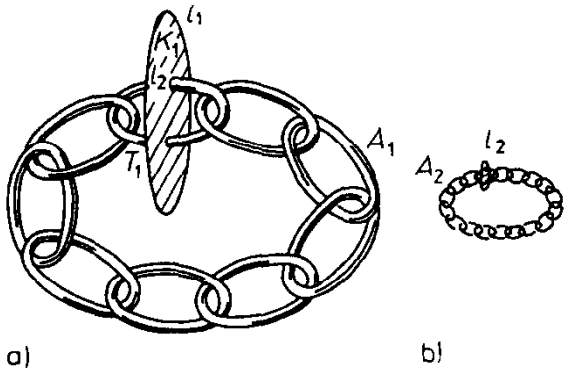


FIGURE 105.

The set A^* is zero-dimensional. Indeed, the diameters of the chain links that make up A_n decrease indefinitely with increasing n, so that there is no connected set in A^* containing more than one point.

Now let l_1 be a circle that entwines the initial chain A_1 and let K_1 be the disk bounded by l_1. In Figure 105a, K_1 intersects the torus T_1, the surface of one of the solid chain links, in two circles (meridians); we denote one of them by l_2. The part of the disk K_1 bounded by l_2 is a smaller disk K_2. This smaller disk is in the same position with respect to the part of A_2 in the interior of T_1 as K_1 was with respect to A_1 (Figure 105b). One can imagine that K_2 (much like K_1) intersects one of the tori (that form the boundary of the solid chain links of A_2) in two circles, one of which we denote by l_3. Continuing in this way we see that the intersection of K_1—which we can also think of as a membrane spanning l_1—with each of the sets $A_1, A_2, A_3, \ldots$ is nonempty. Hence K_1 has a nonempty intersection with the limit set A^*. Obviously, instead of K_1 we can take any other membrane that is the result of a continuous deformation of a disk and spans l_1. The preceding argument shows that any such membrane has a nonempty intersection with A^* (this will be proved in Section 3.6). It follows that *the circle l_1 is not contractible in the exterior of Antoine's necklace A^**. The latter has no connected part consisting of more than one point and so is zero-dimensional, and yet no membrane that is the result of a continuous deformation of a disk and spans l_1 can bypass it.

Problems

121. Show that it is possible to span l_1 by a membrane homeomorphic to a handle that is completely in the exterior of A^*.

122. Construct a simple closed curve that passes through all points of Antoine's necklace.

Example 32 We are now in a position to describe a wild sphere. Let S be a sphere that contains in its interior the set A_1 as well as the circle l_1 (Figure 105a). Now

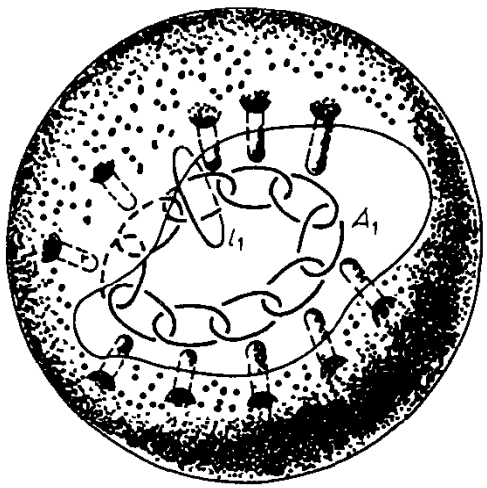


FIGURE 106.

imagine that the sphere has been depressed so that the depressions lead into its interior and come close to each of the tori bounding A_1 (Figure 106). This can be done so that the depressions do not touch the circle l_1. The resulting surface S_1 (a sphere with tubelike depressions) is homeomorphic to a sphere and its interior U_1 is homeomorphic to an open ball and contains l_1. Now we extend each tubelike depression by forming at its end finer depressions that run in the interiors of the tori bounding A_1 and come close to the links of A_2. We again obtain a surface S_2 whose interior U_2 is homeomorphic to an open ball and contains l_1. Then we form still finer depressions that come close to the links of A_3, and so on. At each step of this construction we obtain a surface S_n homeomorphic to a sphere whose interior U_n is homeomorphic to an open ball and contains l_1. The depressions added at the successive steps become ever shorter. This being so, the surface is modified ever more slightly and the limit surface S^* is homeomorphic to a sphere. Since the "tentacles" of the surfaces S_n come ever closer to A^*, S^* contains A^*. It follows that the interior U^* of S^* does not intersect A^*, for the latter lies on the boundary S^* of the region U^*. Thus all of U^* lies in the exterior of A^*. This being so, l_1, which lies in U^*, is not contractible, for it is not contractible in the exterior of A^*. It follows (see Problem 119) that U^* cannot be homeomorphic to an open ball. As a result, the union of S^* and its interior is not homeomorphic to a closed ball.

2.10. Knots

A knot in a string can be undone if the ends of the string are free. That is why the knots considered in topology always involve closed curves.

Example 33 Figure 107 shows a *simple knot* (sometimes also referred to as a "trefoil knot").

Example 34 A double knot (Figure 108a) must not be confused with the the so-called sailor knot (Figure 108b); the former is easily undone, and so is con-

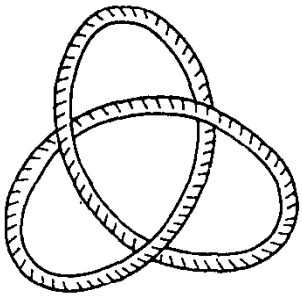


FIGURE 107.

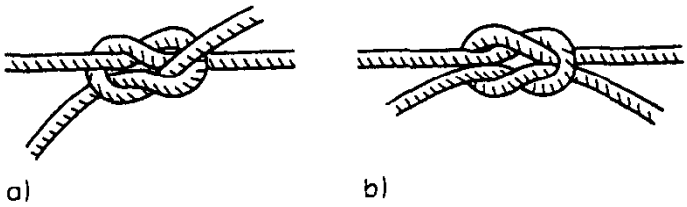


FIGURE 108.

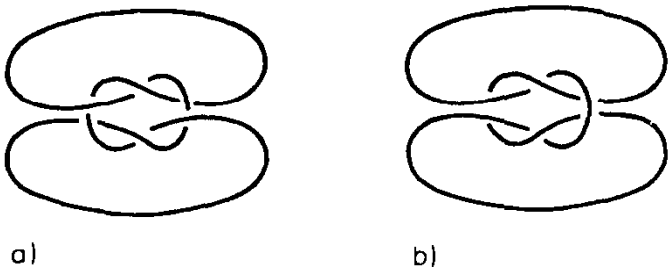


FIGURE 109.

temptuously referred to by sailors as the "granny knot." The topological schemes of these knots are shown in Figure 109.

From the topological viewpoint, a *knot* is a curve in three-space homeomorphic to a circle. Two knots are regarded as different if they are not isotopic. It is intuitively clear that the knotted and knot-free curves in Figure 110 are topologically differently disposed, i.e., they are nonisotopic. We will prove this in Section 3.6.

We think of a knot as realized by a simple closed polygonal line l and consider its projection on a horizontal plane. This projection can be self-intersecting. We assume that each point of intersection is double, i.e., that the number of intersecting edges is just two. If necessary, this can be achieved by a slight shifting of edges. Figure 111a is a so-called *normal projection of a knot*, which means that an edge passing under another edge is represented by a broken line. While adhering to the same convention, we can also represent a knot as a smooth plane curve (Figure 111b).

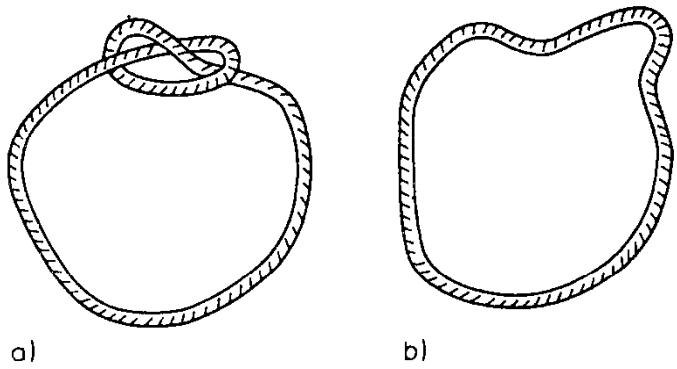


FIGURE 110.

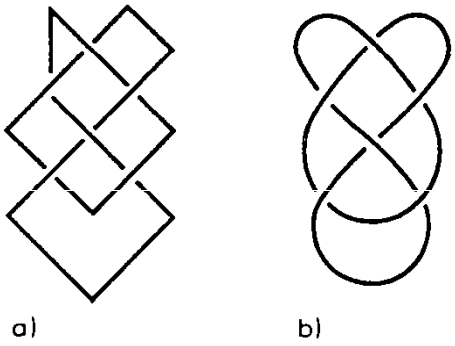


FIGURE 111.

Problems

123. Show that every integral one-dimensional cycle can be represented as the union of orientable knots, i.e., as the union of closed oriented curves in space any two of which can have common vertices but no other common points.

124. Show that the triply twisted strip in Figure 112 is homeomorphic to a Möbius strip and that its boundary is isotopic to a trefoil knot.

125. Show that the "twisted buckle" in Figure 113 is homeomorphic to a handle and that its boundary is isotopic to a trefoil knot.

126. Show that an orientable or nonorientable surface that is not homeomorphic to a disk and whose boundary is homeomorphic to a circle can be so embedded in three-space that its boundary becomes a trefoil knot.

Problems 124–126 give rise to the following question: *Given a knot L, is there always a membrane spanning L, i.e., a surface (without self-intersections) with*

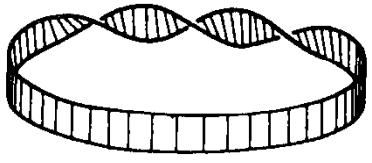


FIGURE 112.

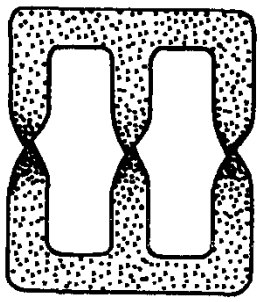


FIGURE 113.

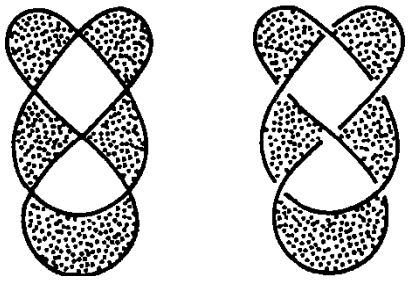


FIGURE 114.

L as boundary. An affirmative answer to this question was given by F. Frankl. His elegant argument is that the normal projection of a knot L divides the plane into regions such that the resulting map can be colored with two colors, say white and red. The possibility of such a chessboard coloration is implied by Problem 96, for there are exactly four edges at every vertex of the graph resulting from the projection; here it is assumed that the exterior (unbounded) country is colored white. In Figure 114a the curve of each knot is shown without breaks, so that the countries are clearly visible. If we put back the breaks in the projection of L, i.e., if we use a normal projection, then the drawing becomes "spatial," i.e., the red regions seem to flow into one another at the crossings (see Figure 114b). This gives the desired surface with L as boundary. We note that in general, this surface is not orientable (see Figure 112). With more care, we can construct for every knot an orientable surface without self-intersections whose boundary is this very knot (see Problems 130–132).

Problems

127. Using the indicated procedure, construct for each of the knots in Figure 115 a surface with this knot as boundary.

128. Show that the Frankl surface is nonorientable if and only if there exists a simple closed curve that passes through only the red regions and has an odd number of transitions from country to country at the double intersection points (Figure 116).

129. A *link* is the union of a number of simple closed curves in space that are pairwise disjoint (Figure 117). Show that for every link there is a closed

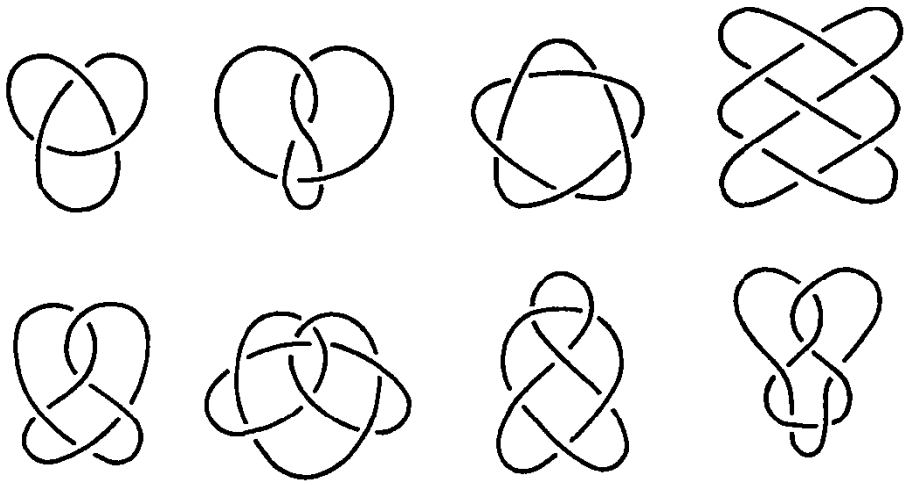


FIGURE 115.

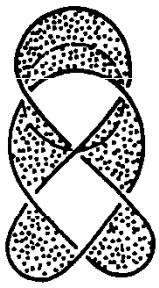


FIGURE 116.

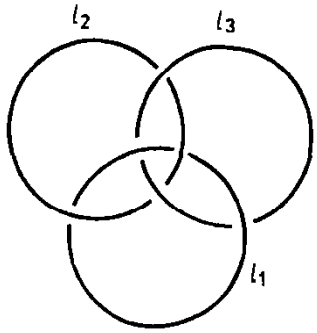


FIGURE 117.

surface without self-intersections whose boundary is this very link.

130. Let L be a link. Choose an orientation for each of the curves of L and denote by z the one-dimensional integral cycle that results from the normal projection of L. z divides the plane into countries. Choose a point in each country and denote by o the point in the exterior country. With each of these points we associate a number $k(M)$ equal to the intersection index $J(x, z)$, where x is a directed polygonal line leading from o to the distinguished point of M (Figure 118). Let M_1 be a country for which $|k(M)|$ takes on its maximum value. Show that the orientations of the parts of the cycle z that make up the boundary of M_1 are such that they

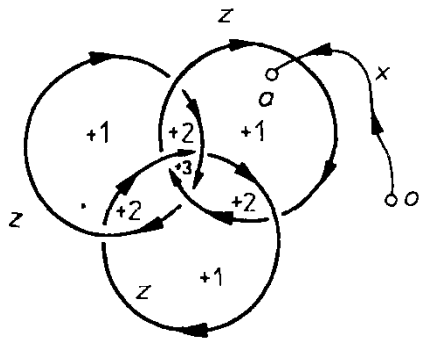


FIGURE 118.

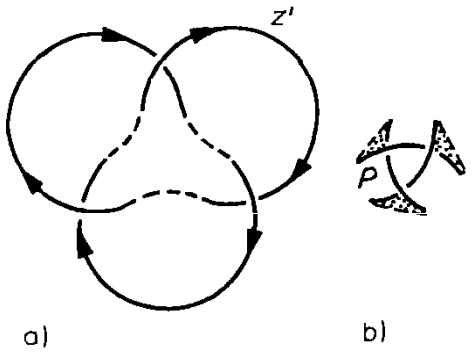


FIGURE 119.

form a (clockwise or counterclockwise) circuit of M_1.

131. Let z and M_1 be defined as in Problem 130. If we remove from z the parts that form the boundary of M_1 and replenish the gaps with "dikes" (drawn as dotted lines in Figure 119a), then we obtain a cycle z' with a smaller number of countries. By applying a similar construction to the link L we obtain a link L' whose projection yields the cycle z'. The removed parts are replenished by dikes so that it is possible to span the resulting curve by a surface P homeomorphic to a disk that resembles in the vicinity of the dikes a twisted shovel blade (see Figure 119b). Prove the following result: If an orientable membrane Q spans L', then gluing the surface P to Q (at the dikes) yields an orientable surface spanning the cycle z.

132. Use the two preceding problems to prove the following result: *If L is a link in space and if each of its curves is given an orientation, then there exists an orientable surface with boundary L such that one of its orientations yields the orientations of the curves of the link.*

Hint. One must see to it that at each step of the construction the boundary of the surface in question is be seen from above.

133. Suppose that the largest value of $|k(M)|$ (see Problem 130) for the cycle z is n. If Q is a surface spanning the link L in accordance with the construction presented in Problems 130–132, then we have the inequality $\chi(Q) \geq n - q$, where q is the number of double intersection points of the cycle z.

134. Prove that if the link L has l components, the number of double intersection points of its normal projection z is q, and the largest value of $|k(M)|$ for the cycle z is n, then we can span this link by a membrane homeomorphic to a sphere with k handles and l circular holes. Moreover, $k \leq 1 + (q - n - l)/2$.

135. Show that one can span a sailor knot (as well as a granny knot) by a membrane homeomorphic to a sphere with three holes two of which are closed with handles.

2.11. Linking Numbers

For two disjoint oriented circles x and y in space (x the first and y the second) we make the following definition of their *linking number.*

Consider the normal projection of the link $x \cup y$ on a ("horizontal") plane. Let a be a double point (double intersection point) of the projection at which x passes under y. If we move toward a following the direction on x, then we see (in the projection) that y cuts x either from left to right (Figure 120a) or from right to left (Figure 120b). In the first case we associate with a the number $+1$ and in the second case, the number -1. As for other double points, i.e., self-intersection points of a circle or points at which x passes over y, we associate with them the number 0. The number obtained by summing these numbers over all double points of the projection is called the *linking number* of x and y and is denoted by $\varpi(x, y)$.

Example 35 For the two intertwined chain links in Figure 121a the linking number is $+1$ (see Figure 121b). For the circles in Figure 122 we have $\varpi(x, y) = 3$.

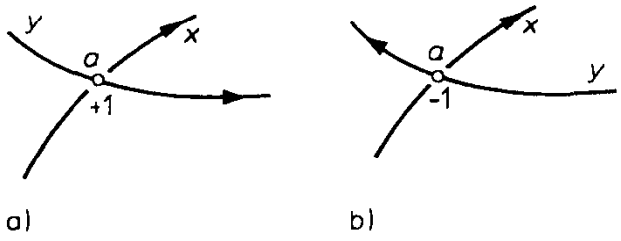


FIGURE 120.

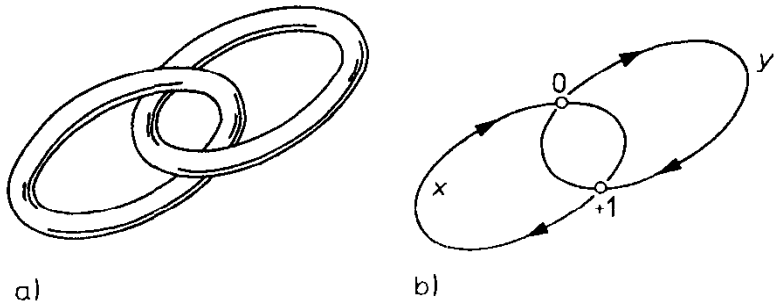


FIGURE 121.

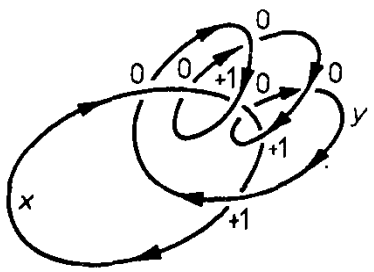


FIGURE 122.

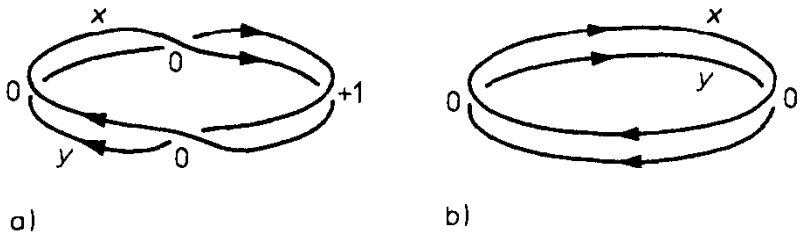


FIGURE 123.

Later we will see that the linking number depends solely on the disposition of the circles and not on the type of projection. Moreover, $\mathfrak{w}(x, y)$ does not change if the circles x and y are subjected to a continuous deformation that preserves their disjointness throughout the process of deformation. It turns out that the linking number $\mathfrak{w}(x, y)$ is an *isotopy invariant*, i.e., if f is a homeomorphism of three-space, then

$$\mathfrak{w}(f(x), f(y)) = \pm\mathfrak{w}(x, y).$$

Example 36 At the end of Section 1.2 we mentioned that a strip with two twists and an untwisted strip (see Figure 8) are homeomorphic but that these two figures are not isotopic in space. Now we can prove this result. To this end we consider the linking numbers of the boundaries of the two strips. For the strip with two twists the value of this number is either $+1$ or -1 (depending on the way the strip is twisted), and for the untwisted strip its value is 0 (Figure 123). That is why the strip with two twists cannot go over into the untwisted strip under a homeomorphism of space. To put it more precisely, this cannot happen because during the deformation the boundaries of the strips remain disjoint, which is why their linking number does not change.

Example 37 A constant current I flowing through a rectilinear conductor P generates a magnetic field whose intensity H at a distance r from the conductor has the value $H = 2I/r$. The potential of the magnetic field is defined as the amount of work required to transport a charged particle with unit charge from a certain fixed point x_0 (the point of *zero potential*) to a given point. In the case under consideration, the potential W of the magnetic field is multivalued. Figures 124a and 124b show two ways of transporting the charged particle from x_0 to a. The second way calls for an additional amount of work. Since the required force is $\frac{2I}{r}$

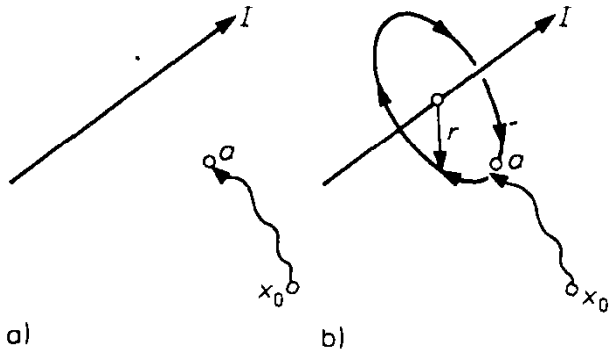


FIGURE 124.

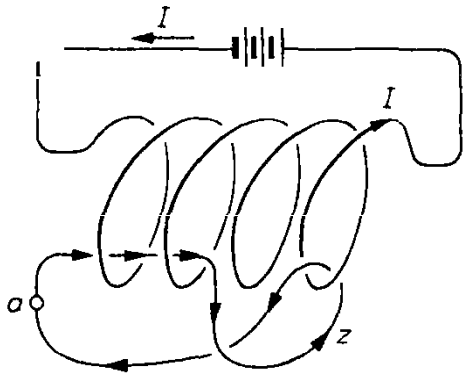


FIGURE 125.

and the additional path length is $2\pi r$, the additional amount of work is $4\pi I$. We see that going around the conductor once (along any path) changes the magnetic potential $W(a)$ by $4\pi I$. If we go around the conductor m times (m any integer), then the potential changes by $4\pi Im$. The same expression for the change of potential applies to arbitrary (not just rectilinear) conductors (see Figure 125). The number of turns (windings) of the path around the conductor is the negative of the linking number of the conductor P and the path z, i.e., if the closed path z around the conductor involves m turns, then the magnetic potential change is $4\pi Im$, and $m = -\text{ɯ}(x, y)$. If the current is measured in amperes, then Im is also called the ampere linking number.

Problems

136. Show that interchanging the circles does not change the linking number, i.e., $\text{ɯ}(x, y) = \text{ɯ}(y, x)$.

137. Show that if oriented circles x' and y' are symmetric with respect to some plane to oriented circles x and y (including orientations), then $\text{ɯ}(x', y') = -\text{ɯ}(x, y)$.

138. What is the linking number of the boundary of a Möbius strip (Figure 50d) and its midline?

139. Let Q be a surface homeomorphic to a Möbius strip (for example, the surface in Figure 126). Let x be its boundary and y its midline (i.e.,

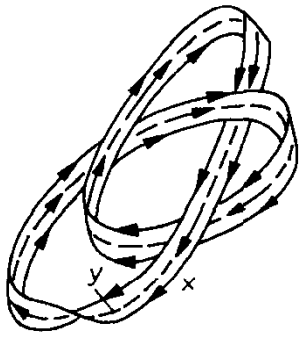


FIGURE 126.

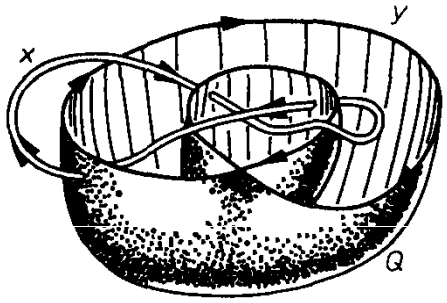


FIGURE 127.

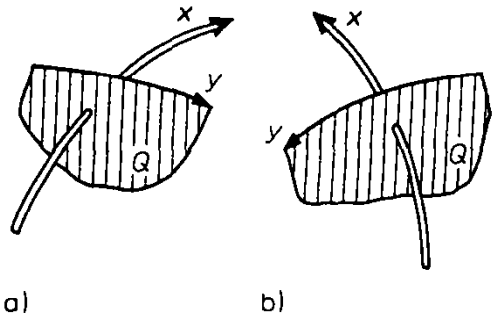


FIGURE 128.

the image of the midline of the Möbius strip in Figure 50d under this homeomorphism). Show that $\varpi(x, y)$ is odd.

We will now give an equivalent definition of linking number. Imagine that the circles x and y lie "almost completely" in the normal projection plane, so that one of the circles lies somewhat lower than the other only in the vicinity of the double points. Next consider an orientable membrane Q spanning y as described in Problem 132 (the boundary is fully visible if you look at Q from above). Near the boundary the membrane should be almost vertical (Figure 127). At points at which x is above y it is also above Q, so that x and Q do not intersect there. On the other hand, they do intersect at points at which y is above x. The intersection index of x and Q is $+1$ if y runs from left to right (Figure 128a) and -1 if it runs from right to left (Figure 128b); we are looking in the direction of the orientation of x. By comparing the definitions of intersection index and linking number we

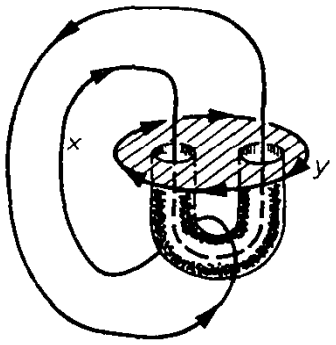


FIGURE 129.

see that

$$\mathfrak{w}(x, y) = J(x, Q), \tag{20}$$

where Q is the two-dimensional membrane that spans y and is oriented in accordance with the orientation of y.

Equation (20) holds for an arbitrary membrane Q. Indeed, let Q and Q' be two membranes spanning y (and having the same orientations as y). Consider the difference of Q and Q', i.e., the union of Q and Q' with the orientation of Q' reversed. This difference is a two-dimensional integral cycle (even if Q and Q' intersect). Since the intersection index of the integral cycle x and this two-dimensional cycle is 0, it follows that $J(x, Q) = J(x, Q')$.

Equation (20) implies that the linking number, initially defined by a normal projection, does not depend on the disposition of the projection plane. It also implies other properties of the linking number (see above).

Problems

140. Suppose that a membrane that is in general position with respect to a circle x spans a circle y. Assume that the membrane has just one point in common with x. Prove that any membrane spanning x has at least one point in common with y.

141. Let x and y be oriented circles in space. Equation (20) implies that if an orientable membrane can span y so that it does not intersect x (Figure 129), then $\mathfrak{w}(x, y) = 0$. Prove the converse proposition.

142. Prove that if $\mathfrak{w}(x, y)$ is even, then there exists a (not necessarily orientable) membrane spanning y that has no points in common with x.

143. Check that the linking number of any two circles in Figure 117 is 0. Construct a membrane homeomorphic to a handle that spans one of these circles and does not intersect the other two.

3

Homotopy and Homology

3.1. Periods of Multivalued Functions

Let h be a path in a figure X leading from an initial point x_0 to an endpoint x_1. In other words, let h: $[0, 1] \to X$ be a continuous mapping satisfying the conditions $h(0) = x_0$ and $h(1) = x_1$. We will deform this path within the figure X but keep x_0 and x_1 fixed.

Two paths h_1 and h_2 in a figure X both of which have the same endpoints are said to be *homotopic* in X if there is a continuous deformation (within X) that takes h_1 to h_2 (see Figure 130; the fine lines are intermediate paths). In symbols, $h_1 \sim h_2$.

Example 38 Two paths on a disk with the same endpoints are always homotopic. In intuitive terms, we can justify this claim by thinking of the path as a stretched rubber thread within the disk. If the endpoints of the thread are x_0 and x_1 and if it is free to move, then, whatever its initial form, it ends up as the segment with endpoints x_0 and x_1. Thus any path on the disk is homotopic to the segment connecting its endpoints. This implies that any two paths on the disk with endpoints x_0 and x_1 are homotopic.

Example 39 Let X be an annulus with boundary circles centered at o. We choose a point x_0 in X. For each point $x \in X$ we denote the magnitude of the angle $\angle x_0 o x$ by $\varphi(x)$ (Figure 131). The function $\varphi(x)$ is multivalued (it is determined to within a summand of the form $2k\pi$, k an integer). We assume that an arbitrary one of these values has been chosen at $x = x_0$ and we denote it by φ_0. If x is moved within X, then the angle $\varphi(x)$ changes continuously. Hence to each path h in X leading from x_0 to x_1 there corresponds a uniquely determined value φ_1 of the multivalued function $\varphi(x)$. To obtain this value we choose values of $\varphi(x)$ such that, beginning with φ_0 at x_0, the change in the value of $\varphi(x)$ along h is continuous. Regardless of whether we reach x_1 by moving along h or along a path homotopic to h, we end up with the same value of $\varphi(x)$. Indeed, under a continuous deformation of h, each particular value of $\varphi(x)$ on h also changes continuously. But then it must stay constant, for the difference between two values of $\varphi(x)$ for the same x is $2k\pi$.

Example 40 Recall Example 37. The choice of a point x_0 of zero potential determines in the exterior of the conductor P a multivalued function $W(x)$ (the magnetic potential function). If we move from x_0 to x_1 along a path h, then we obtain at x_1 a uniquely determined value of the magnetic potential. While homotopic paths leading from x_0 to x_1 yield the same value of the magnetic potential

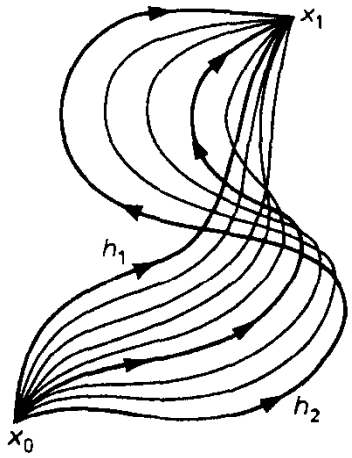


FIGURE 130.

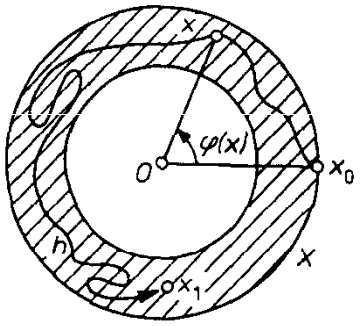


FIGURE 131.

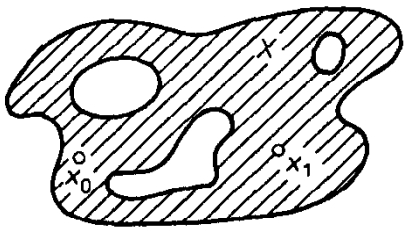


FIGURE 132.

at x_1, nonhomotopic paths with these endpoints may yield very different values at that point.

Problems

144. Define a multivalued function on the set X in Figure 132 that takes on infinitely many values at x_0 including the values 0, 1, and $\sqrt{5}$.

145. Define a multivalued function on the set X in Figure 132 with the following property: There are nonhomotopic paths from x_0 to x_1 that yield the same functional value at x_1.

Example 41 Consider the figure X below (see Figure 133). Let $\varphi_1(x)$, $\varphi_2(x)$, and $\varphi_3(x)$ denote the magnitudes of the angles $\angle a_1 o_1 x$, $\angle a_2 o_2 x$, and $\angle a_3 o_3 x$

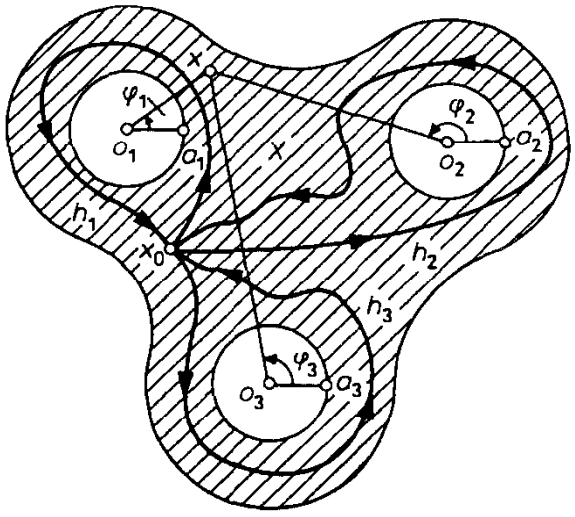


FIGURE 133.

respectively. The function $f(x) = \varphi_1(x) + \sqrt{2}\varphi_2(x) - \sqrt{3}\varphi_3(x)$ is multivalued on X. If we go from x_0 to x_0 along the path h_1, then the value of $\varphi_1(x)$ at x_0 changes by 2π while the values of $\varphi_2(x)$ and $\varphi_3(x)$ at x_0 remain unchanged. It follows that the value of $f(x)$ at x_0 changes by 2π. Similarly, going along h_2 changes the value of $f(x)$ at x_0 by $2\pi\sqrt{2}$ and going along h_3 changes it by $-2\pi\sqrt{3}$. We can call 2π, $2\pi\sqrt{2}$, and $-2\pi\sqrt{3}$ the *periods* of $f(x)$ corresponding to the paths h_1, h_2, and h_3 respectively.

Let h_1 and h_2 be paths beginning and ending at x_0. By the *product* h_1h_2 of these paths we mean the path obtained by traversing first h_1 and then h_2. It is clear that traversing h_1h_2 changes the value of $f(x)$ at x_0 by $2\pi + 2\pi\sqrt{2}$. Similarly, traversing h_3h_1 changes the value of $f(x)$ at x_0 by $-2\pi\sqrt{3} + 2\pi$. More generally, we can say that multiplication of paths (which begin and at x_0) results in the addition of the corresponding periods of $f(x)$.

Traversal of homotopic paths leads to the same value of $f(x)$. In this sense homotopic paths are indistinguishable. Therefore, it is natural to form *classes of mutually homotopic paths*. We denote the class of all paths homotopic to a path h by $[h]$ and the set of all such classes by $\pi(X)$. These classes can be multiplied. Specifically, to multiply the classes $[h]$ and $[k]$ we form the product hk of $h \in [h]$ and $k \in [k]$ and define the *product* of these classes as the class containing the product hk: $[h] \cdot [k] = [hk]$.

The idea behind the introduction of classes of paths is clear. To each class there corresponds a period of the multivalued function $f(x)$, and when two classes are multiplied their periods are added.

3.2. The Fundamental Group

Given a figure X we can investigate its classes of homotopic paths and their products. We limit ourselves to paths in X that begin and end at the same point

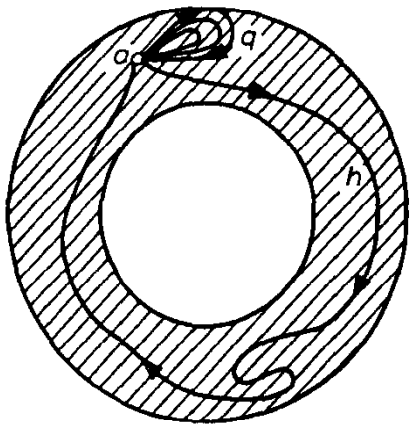


FIGURE 134.

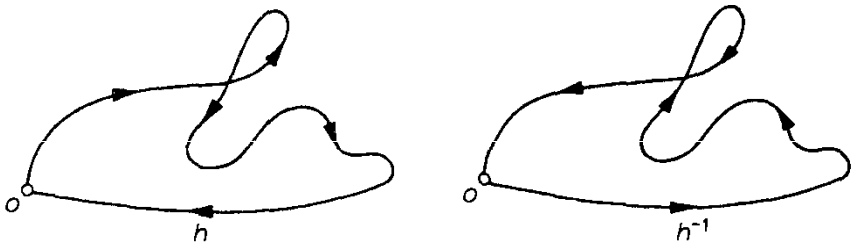


FIGURE 135.

$x_0 \in X$. Two such paths can be multiplied. All mutually homotopic paths are put in a single class. If a is such a class and h is one of its elements, then we say that h is a *representative* of a and write $a = [h]$. The set of all classes is denoted by $\pi(X)$. Multiplication of classes is defined as described in the previous section (see Figure 133): If a and b are two classes of paths (that begin and end at x_0) and h and k are their representatives, i.e., if $a = [h]$ and $b = [k]$, then we define their product as the class represented by the path hk, i.e., we put $ab = [hk]$. The product just defined is independent of the choice of class representatives. It turns out that under this operation of multiplication the set $\pi(X)$ forms a group.

We sketch a proof of this assertion. Let h be a path in a class a and let q be a path contractible to a point. Then $qh \sim h$ (Figure 134) and $hq \sim h$ ($\sim$ stands for "is homotopic to"). Denote by 1 the class of paths contractible to a point. (These paths are sometimes referred to as *null homotopic paths*). For any class $a \in \pi(X)$ we have $1a = a$ and $a1 = a$, i.e., 1 is the identity element for the multiplication defined on $\pi(X)$.

Now let h be a representative of a class a. Let h^{-1} be the path that is the result of traversing h in the opposite direction (Figure 135). Each of the paths hh^{-1} and $h^{-1}h$ is contractible to a point; Figure 136 shows how hh^{-1} can be contracted to a point. If we denote the class containing h^{-1} by a^{-1}, then we have $aa^{-1} = 1$ and $a^{-1}a = 1$, i.e., each element of $\pi(X)$ has an inverse.

It is not difficult to show that the multiplication on $\pi(X)$ is associative. Hence $\pi(X)$ is a group. It is called the *fundamental group of the figure* X (with base point x_0).

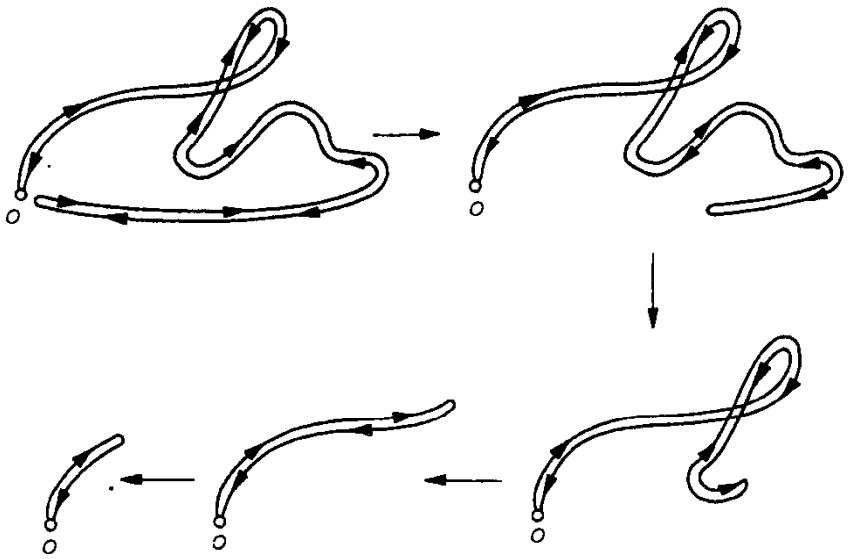


FIGURE 136.

If two points x_0 and x_0' can be connected by a path in X, then the fundamental groups with these base points are isomorphic (see Problem 148). Thus if X is path-connected (and this is the only case we will consider), then we can speak of the fundamental group of the figure X without specifying the point at which it was based. *The fundamental group is a topological invariant*, i.e., two homeomorphic figures X and Y have isomorphic fundamental groups $\pi(X)$ and $\pi(Y)$. Fundamental groups were introduced by Poincaré.

Problems

146. If the group $\pi(X)$ is trivial (i.e., if it consists of just the identity element), then we say that the figure X is *simply connected.* In other words, a figure X is simply connected if every closed path in it is contractible to a point. Show that every convex figure (for example, a straight line, a segment, a disk, a ball, a convex n-gon, a polyhedron) is simply connected.

147. Show that a sphere is simply connected.
Hint. Any path on the sphere (including a sphere-filling curve similar to a Peano curve) can be deformed to a "smooth" path that does not cover the whole sphere.

148. Let w be a path connecting two points x_0 and x_0' of a figure X. With each closed path h with initial point x_0 associate the path $h^* = w^{-1}hw$ with initial point x_0' (Figure 137). Show that this association determines an isomorphism between the fundamental groups of X based at the points x_0 and x_0'.

Example 42 We will show that *the fundamental group of a circle is a free cyclic group*, i.e., that it is isomorphic to the additive group of the integers. To this end, we denote a path that goes once around a circle B uniformly (i.e., without change of direction) in the positive direction by a and the oppositely oriented path by

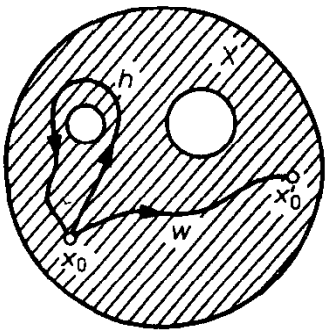


FIGURE 137.

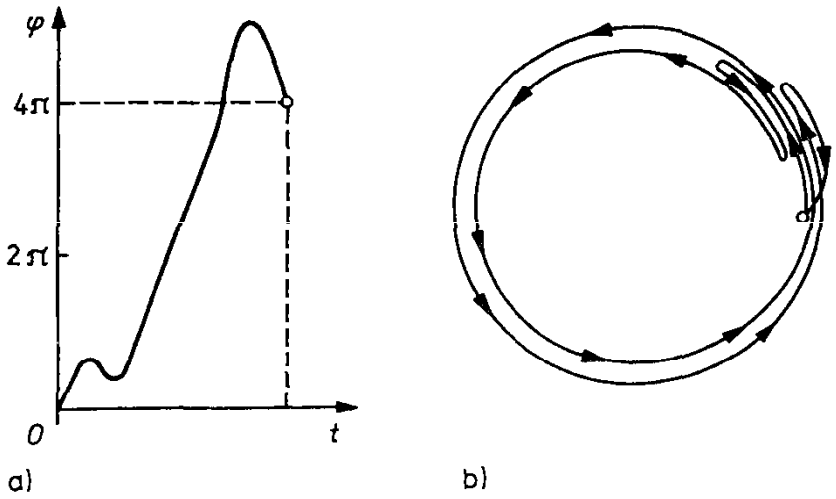


FIGURE 138.

a^{-1}. Then a^n denotes the path that goes around the circle n times: in the positive direction if $n > 0$ and in the negative direction if $n < 0$ (the path a^0 coincides with the initial point x_0).

With each path we can associate a curvilinear image that determines the position of a point moving on the path by the value t of a parameter (for example, time), t in the unit interval $0 \leq t \leq 1$. The position of the point moving on the circle is also determined by its angle coordinate φ on B (counted from the point x_0). If t is laid off on the axis of abscissas and φ on the axis of ordinates, then we obtain a graphic relation $\varphi(t)$ (with $\varphi(0) = 0$).

If a uniformly moving point goes around the circle n times, then we obtain the path a^n whose curvilinear image is the segment joining the points $(0, 0)$ and $(1, 2n\pi)$. But the point can also move on the circle with many changes of direction. Figure 138a shows the curvilinear image of a path represented schematically in Figure 138b. The curvilinear image of any closed path on the circle always joins $(0, 0)$ and $(1, 2n\pi)$, n an integer. Regardless of the course of the path we always return to the point x_0, so that the angle coordinate [of the endpoint of the path] is always an integral multiple of 2π. The number n is called the *winding number* of the path.

Any path f with winding number n is homotopic to the path a^n. To see this note that we can move each point of the curvilinear image of f parallel to the

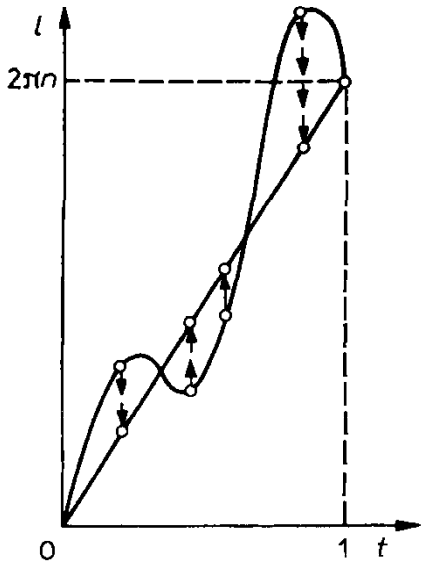


FIGURE 139.

axis of ordinates so that it ends up on the curvilinear image of a^n. If we do this simultaneously for all points (Figure 139), then the curvilinear image of f goes over into the segment that is the curvilinear image of a^n. The deformation of the curvilinear image of f yields a deformation of f. This implies the homotopy of the paths f and a^n. It follows that all paths with winding number n are homotopic to the path a^n, i.e., belong to a single class of paths. Paths with different winding numbers are not homotopic. Thus *there is a one-to-one correspondence between the fundamental group of a circle B and the integers.* When paths are multiplied, their winding numbers are added, i.e., the group $\pi(B)$ is isomorphic to the additive group of the integers.

Problems

149. Show that the fundamental group of an annulus is the free cyclic group.

150. Let X be the plane with a point removed. Show that $\pi(X)$ is the free cyclic group.

151. Show that the interior of a simple closed plane curve l is simply connected. If the boundary of a region G of the plane consists of more than one closed curve (see Figure 132), then G is not simply connected.

3.3. Cell Decompositions and Polyhedra

We have frequently considered a surface Q and a graph G on Q that divided it into parts homeomorphic to a disk. These were examples of *cell decompositions.* A surface can be represented as the union of pairwise disjoint *cells*: *0-dimensional*, *1-dimensional*, and *2-dimensional.* [We will usually omit the word "dimensional"]. The 0-cells are points—the vertices of the graph. The 1-cells are the edges of the graph (without their endpoints); every 1-cell is homeomorphic to a segment (without its endpoints). The 2-cells are the pieces of the surface resulting from

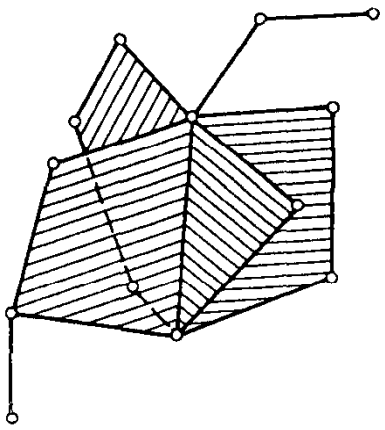


FIGURE 140.

cutting it along the edges of the graph G. Every 2-cell is homeomorphic to an open disk.

We can also consider cell decompositions in which three, four, or more 2-cells meet at an edge (a 1-cell), and not just two or one, as in the case of surfaces with boundary. It can also happen that no 2-cells meet at an edge (see Figure 140). If a cell decomposition is made up of just 0- and 1-cells, then it is a graph. In topology we consider cell decompositions of arbitrary dimension. For example, a 3-dimensional cell decomposition consists of cells of dimension 0, 1, 2, and 3. If we remove from this cell decomposition all cells of dimension 0, 1, and 2, then it reduces to just 3-cells each of which is homeomorphic to an open ball. A figure with a cell decomposition is called a *polyhedron*. The figures considered in Examples 16, 18, and 31 are not polyhedra.

Example 43 The sphere P_0 can be represented as a cell decomposition consisting of a 0-cell and a 2-cell. This cell decomposition has no 1-cells. If we remove from the sphere a point, then the remaining part τ is homeomorphic to an open disk.

Example 44 Figure 141 shows a cell decomposition consisting of a disk and its circular boundary, with the latter divided into two semicircles r_1 and r_2. If we glue together the diametrically opposite points of the circle, then the two edges r_1 and r_2 become a single edge r and the disk becomes a projective plane. In this way we obtain a cell decomposition of the projective plane consisting of a vertex, an edge r, and a 2-cell τ.

Example 45 We draw on a torus a parallel a and a meridian b that intersect at o (see Figure 57d). In this way we obtain a cell decomposition of the torus consisting of a vertex o, two edges a and b, and a 2-cell τ. Actually, by cutting a torus along a meridian and a parallel we obtain a square (see Problem 65), i.e., a piece homeomorphic to a disk.

Problems

152. Show that a handle can be represented as a cell decomposition consisting of a vertex, three edges, and a 2-cell (see Figure 58).

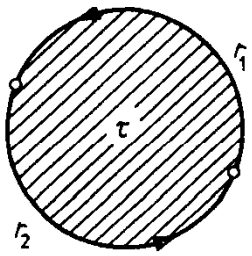


FIGURE 141.

153. Show that a sphere with k handles can be represented as a cell decomposition consisting of a vertex, $2k$ edges, and a 2-cell (see Figure 59).

We will now explain what is meant by the "directed traversal of the boundary of a face" (a 2-dimensional cell). If the surface is homeomorphic to a disk, then the meaning of the term is obvious. In more complicated cases we define it as follows. Cut the face along all edges (Figure 142a). This yields a piece homeomorphic to a disk (Figure 142b). Traverse the boundary of this disk once in a definite direction. If we reglue the disk to get back the initial face, then our directed traversal of the boundary of the disk determines a "directed traversal of the boundary of the face." We can also proceed differently, namely, we can move in the interior of the cell, "very close" to its boundary, without ever crossing it (see Figure 142c).

Now consider a face of a cell decomposition. Orient the edges it borders on (i.e., choose an arbitrary direction on each edge) and mark them with letters $a, b, c, \ldots$. Next traverse the boundary of the face and write down a certain *monomial*. Specifically, if the first traversed edge is marked with a, write the monomial a or a^{-1} according as the direction on a agrees with, or is opposite to, the direction of traversal. If the next traversed edge is, say, d, write to the right of the monomial already written down d or d^{-1} according as the direction on d agrees with, or is opposite to, the direction of traversal, and so on. The monomial obtained upon completion of the traversal is sometimes referred to as the *edge path*.

The edge path depends on the direction of traversal of the boundary of the face and on the initial edge traversed. With each face we associate one of its edge paths. For example, if, as in Figure 143, we traverse the boundaries of the cells τ_1, τ_2, τ_3,

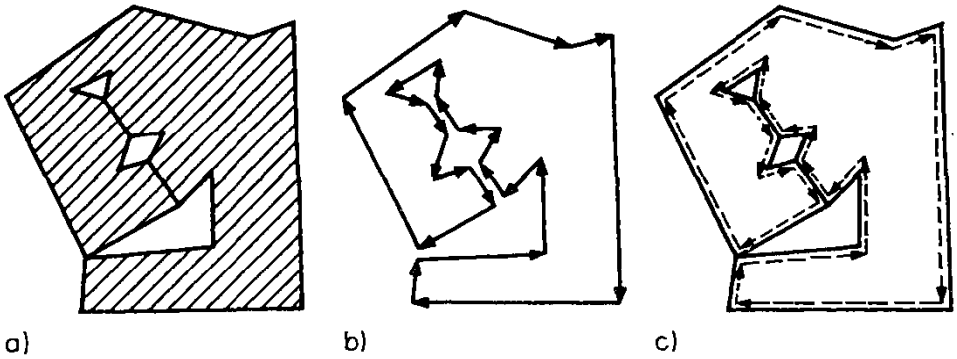


FIGURE 142.

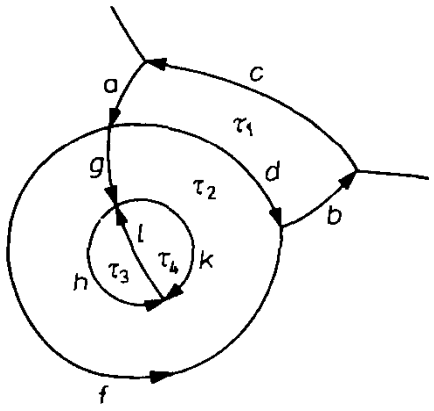


FIGURE 143.

and τ_4 counterclockwise, then we obtain the monomials $adbc$, $kh^{-1}g^{-1}fd^{-1}g$, hl, and $l^{-1}k^{-1}$; note that the edge g appears twice in the edge path of τ_2.

Now we describe (without proof) a method for computing the fundamental group of a connected polyhedron X. Choose one of the cell decompositions of X and denote by G the graph formed by its vertices and edges. Choose a spanning tree in G and assign the number 1 to all its edges. Orient arbitrarily the remaining edges (dams) and mark them with letters $a, b, c, \ldots$. For each of the faces of the cell decomposition write down its edge path, ignoring the edges marked with 1. Finally, we construct the group with generators $a, b, c, \ldots$ (which correspond to the dams) whose defining relations are the equalities obtained by putting the edge paths equal to the identity. *The resulting group is isomorphic to the fundamental group of the polyhedron X.*

Example 46 In the cell decomposition of the projective plane considered in Example 44 every spanning tree consists of a single vertex. Hence the edge r is the generator of the fundamental group. Further, the edge path of the 2-cell τ (see Figure 141) is rr. It follows that the fundamental group of the projective plane is generated by r, which satisfies the defining relation $r^2 = 1$, i.e., this group has order 2.

Example 47 Consider the cell decomposition of the torus described in Example 45. The edge path of the 2-cell τ (see Figure 141) is $aba^{-1}b^{-1}$ (the boundary in Figure 57a must be traversed counterclockwise). It follows that the fundamental group of the torus has two generators a and b connected by the single defining relation $aba^{-1}b^{-1} = 1$, i.e., $ab = ba$. Hence this group is the free abelian group on two generators.

Problems

154. Use the cell decomposition in Problem 152 to show that the fundamental group of a handle is a group on three generators a, b, and c connected by the single defining relation $ba = cab$. This group is not commutative; for example, $ba \neq ab$.

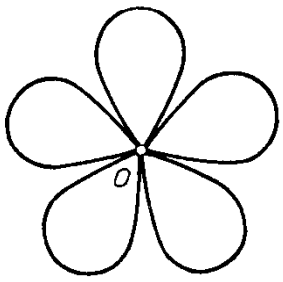


FIGURE 144.

155. Use the result in Problem 153 to show that the group $\pi(P_k)$ has as generators the $2k$ elements $a_1, b_1, a_2, b_2, \ldots, a_k, b_k$ connected by the single defining relation $a_1b_1a_1^{-1}b_1^{-1}a_2b_2a_2^{-1}b_2^{-1}\cdots a_kb_ka_k^{-1}b_k^{-1} = 1$. For $k \geq 2$ this group is not Abelian (for example, $a_1b_1 \neq b_1a_1$).

156. Show that the group $\pi(N_q)$ has as generators the q elements $c_1, c_2, \ldots, c_k$ connected by the single defining relation $c_1^2c_2^2\cdots c_k^2 = 1$.

157. Show that two closed surfaces (without boundary) are homeomorphic if and only if their fundamental groups are isomorphic.

158. By a bouquet B_k^1 of circles we mean the union of k circles that share a point o but are otherwise pairwise disjoint (Figure 144). Show that $\pi(B_k^1)$ is the free group on k generators.

159. Let X be a region on the plane with a single exterior boundary and k interior boundaries (see Figure 132). Show that $\pi(X)$ is the free group on k generators.

3.4. Coverings

Example 48 Let B be a circle centered at o. Fix an initial point x_0 on B and denote by $\varphi(x)$, $x \in B$, the magnitude of the central angle $\angle x_0ox$. $\varphi(x)$ is well-defined to within a multiple of 2π. The curvilinear image E of this multivalued function can be constructed on the lateral surface of an infinite cylinder. It has the form of a helix with pitch 2π (Figure 145). Let p denote the projection of E on the boundary circle B of the base of the cylinder. For every $x \in B$ we choose a small neighborhood U of this point. The part of E projected on U consists of disjoint pieces $\ldots, V_{-1}, V_0, V_1, \ldots$ (Figure 146). The projection p maps each of these pieces homeomorphically onto U.

This property leads us to the notion of a *covering*. Let p be a continuous mapping of a figure E on a figure B. Let p have a property analogous to that described in Example 48, i.e., for every $x \in B$ there is a neighborhood U such that the preimage $p^{-1}(U)$ (all points of E mapped by p on points of U) consists of disjoint pieces mapped by p homeomorphically onto U. Under these conditions E is called a *covering* of B. The parts of $p^{-1}(U)$ mapped homeomorphically onto U are called *sheets* of the covering. Depending on the number of sheets, we distinguish two-sheeted, three-sheeted, etc., coverings. The covering of a circle by a helix described above is infinitely sheeted.

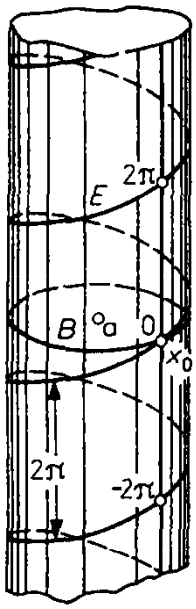


FIGURE 145.

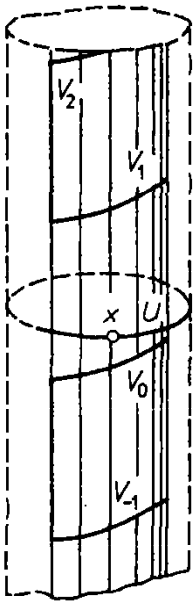


FIGURE 146.

Example 49 Every one-sided surface has a particular two-sheeted covering by a certain two-sided surface P. To obtain it, we embed the surface N in space without corners (but with self-intersections). At each point $x \in N$ we introduce two segments, xx' and xx'', of small length ϵ. Both segments are perpendicular to N at x but go from x to different sides of N. If N were two-sided, then x' and x'' would describe two surfaces "parallel" to N. Since N is one-sided, we obtain just one surface P. To see that this is so note that when the normal xx' traverses a suitable path on the one-sided surface N it changes its direction, i.e., it goes over into xx''. It follows that x' and x'' belong to the same piece of the surface P. The following is an intuitive description of our construction: Imagine that N is made of a burnable material of a certain thickness. We paint all of it with a thin layer of fireproof paint. If we burn the surface N, then we are left with a thin layer of paint that forms the surface P and provides a two-sheeted covering of N. Moreover, P is two-sided, for one of its sides is directed towards the (burned) surface N and the other is directed away from it.

For example, imagine a Möbius strip made of a burnable material of a certain thickness. We paint all of it with a layer of fireproof paint. If we burn the Möbius strip, then we are left with a strip that is homeomorphic to the lateral surface of a cylinder and provides a two-sheeted covering of the Möbius strip. A model shows that the covering strip has four twists.

Problems

160. Show that if E is a k-sheeted covering of a polyhedron B, then $\chi(E) = k\chi(B)$.

161. Show that a two-sheeted covering of the surface N_q is a sphere with $q-1$ handles.

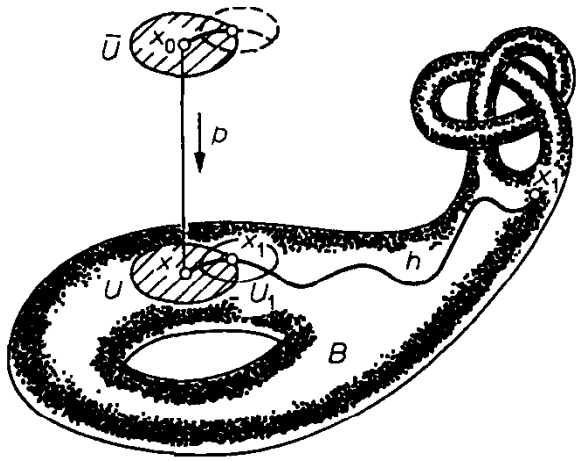


FIGURE 147.

Let E be a covering of B and p: $E \to B$ be the corresponding projection. Let h be a path on B with initial point x_0 and let $\overline{x}_0 \in E$ be a point "above" x_0 such that $p(\overline{x}_0) = x_0$. Then there is a (uniquely determined) path $\overline{h}$ that begins at $\overline{x}_0$ and is mapped by p on h. It is called a *covering path*. To construct it, let $\overline{U}$ be the sheet of the covering that contains $\overline{x}_0$. Since p: $\overline{U} \to U$ is a homeomorphism, we can lift the piece of h in U in a unique way to the sheet $\overline{U}$ (Figure 147). If x_1 is the endpoint of the lifted piece of the path, then there is a suitable neighborhood U_1 of x_1 and a suitable sheet of the covering such that we can extend the covering path $\overline{h}$ a bit further, and so on.

Using the notion of covering paths we can formulate a theorem on the connection between coverings and the fundamental group. We state it (without proof) in the following simplified form: *If a connected polyhedron E is a k-sheeted covering of B and if the order* (i.e., the number of elements) *of its fundamental group $\pi(E)$ is n, then the order of the fundamental group $\pi(B)$ is kn.*

A covering E of B is called *universal* if it is simply connected. The theorem just stated implies that the number of sheets of a universal covering of B is equal to the order of $\pi(B)$; any other covering has fewer sheets.

The covering of the projective plane by a sphere (see Problem 161) is universal because a sphere is simply connected. A sphere is also its own universal covering. We prove that *the plane is a universal covering of every closed surface other than the sphere and the projective plane*. Note that every one-sided surface N has a two-sheeted covering by a two-sheeted surface P, so that a universal covering of P is also a universal covering of N. Thus we need only consider two-sided surfaces other than the sphere.

We divide the plane into congruent squares by two systems of parallel lines. Appropriate gluing turns each square into a torus. We map the plane onto a torus by mapping the points that occupy the same positions in the system of squares (tori) in Figure 148 on the point in the same position in the square (torus) in Figure 149. Since the plane is simply connected, the covering of the torus by the plane is universal.

We call the squares *fundamental regions*. Each fundamental region is a connected piece of the covering (the plane) and can be mapped in a one-to-one way

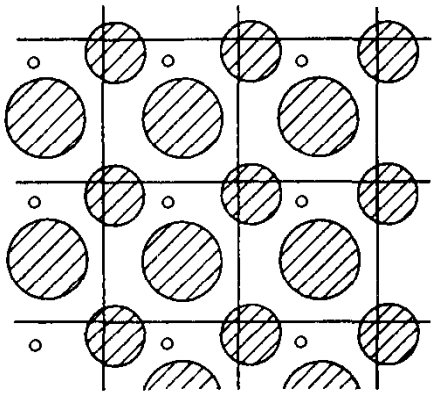


FIGURE 148.

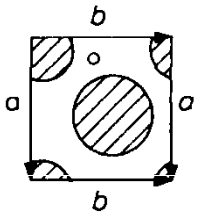


FIGURE 149.

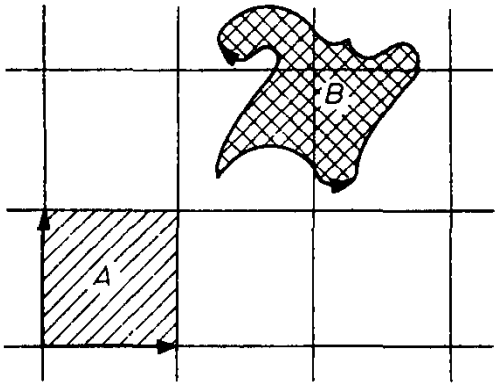


FIGURE 150.

onto the torus. Figure 150 shows that a fundamental region is not uniquely defined.

We now describe a decomposition into fundamental regions that can be glued together so as to yield other two-sided surfaces such as, say, P_2. Such a decomposition can be conveniently realized using *hyperbolic geometry*. In this geometry the sum of the angles of a polygon is smaller than in *Euclidean geometry*. Also, the sum of the angles decreases as the area of the polygon increases. For example, there is a regular octagon with angles $\pi/4$. If such octagons are arranged so that that their sides touch in pairs, then we obtain a tessellation of the hyperbolic plane in which eight octagons meet at each vertex. Figure 151 shows such a tessellation of the *Poincaré model* of the hyperbolic plane. The octagons involved are fundamental regions. (This model is homeomorphic to an open disk and therefore also to the Euclidean plane.) By suitably gluing together the sides of an octagon we obtain P_2 (see Figure 59). Thus the hyperbolic plane provides a universal covering

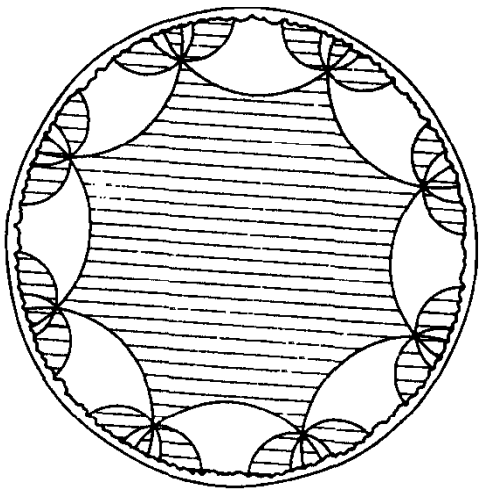


FIGURE 151.

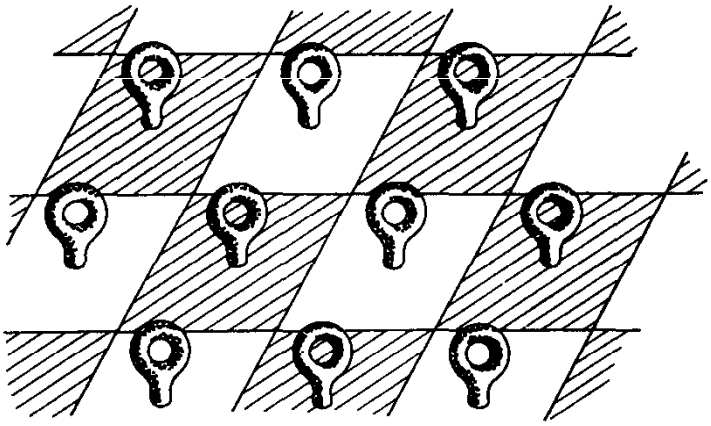


FIGURE 152.

of P_2. It is possible to construct analogous decompositions of the hyperbolic plane for every surface P_k $(k \geq 2)$.

Problems

162. Figure 152 depicts a plane with infinitely many handles. Show that it can provide a covering for every surface P_k $(k \geq 2)$.

163. Show that the surface depicted in Figure 153 can provide a covering for every surface P_k $(k \geq 2)$.

164. Construct a universal covering of the figure consisting of a sphere and a circle that touches it.

3.5. The Degree of a Mapping and the Fundamental Theorem of Algebra

Figure 154 shows a continuous mapping f of a circle P on a circle Q. Two parts of P are mapped on a neighborhood of $y \in Q$ with orientation preserved (the two

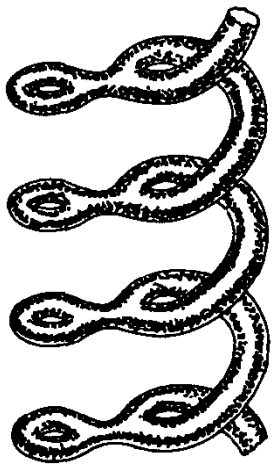


FIGURE 153.

parts of P and the neighborhood of y have the same orientation). We say that the mapping has degree 2 at y. The degree of the mapping at x is also 2: True, there are four parts of P that are mapped on a neighborhood of $x \in Q$, but three of them are mapped positively and one negatively. If f maps p sheets of P positively on a neighborhood of $z \in Q$ and n sheets negatively, then we say that its *degree* at z is $p - n$. The degree of f is constant at all points of Q (and equal to 2); for example, at x we have $p - n = 3 - 1 = 2$.

The concept of the degree of a mapping can also be used for mappings of surfaces. Let P and Q be two closed orientable surfaces with orientations. Let $f\colon P \to Q$ be a continuous mapping. Think of P as (possibly) multisheeted and as (possibly) having folds. When sheets of P are mapped on a neighborhood of a point $z \in Q$, some of them may be mapped positively (i.e., with preservation of orientation, as in Figure 155a) and some negatively (as in Figure 155b). If all sheets of P have been mapped homeomorphically on a neighborhood of $z \in Q$, p of them positively and n negatively, then the difference $p - n$ is called the *degree* of f at z.

It is not difficult to see that the degree of f stays constant in the neighborhood

$$p - n = 3 - 1 = 2.$$

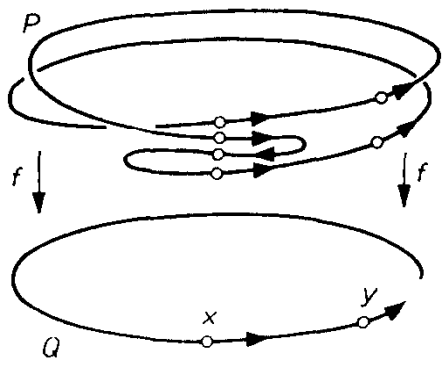


FIGURE 154.

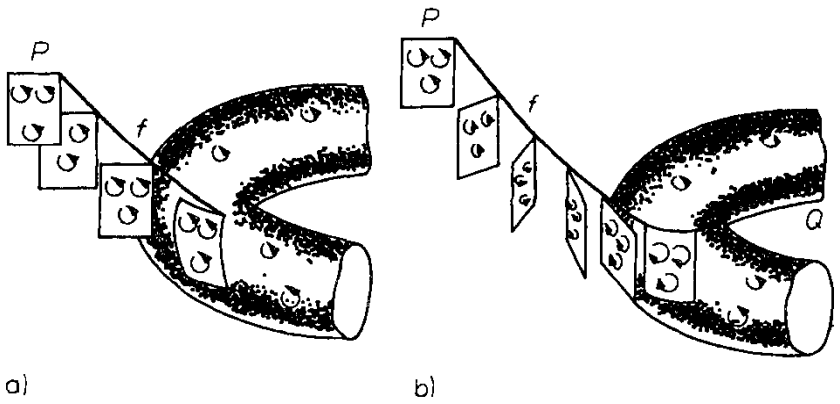


FIGURE 155.

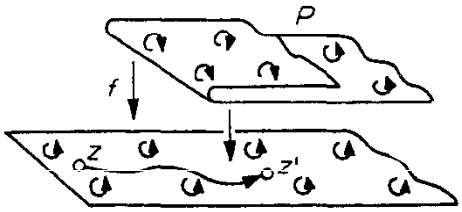


FIGURE 156.

of any point of Q. This is so because the only time the numbers p and n change is when one goes past a fold. Still, the difference $p - n$ remains unchanged (Figure 156). Also, the degree of f is unchanged when f is subjected to a continuous deformation. This is so because the forming or smoothing of folds has no effect on the degree of a mapping.

The concept of the degree of a mapping can be used to give an elegant proof of the *fundamental theorem of algebra*, which asserts that *every polynomial*

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

of degree $m \geq 1$ *with complex coefficients* $a_1, \ldots, a_m$ *has at least one root.*

Consider a sphere S that touches the plane at the origin of the coordinate system. We call this point the *south pole* of S and the diametrically opposite point n of S its *north pole* (Figure 157). A complex number $x + iy$ is represented on the plane as the point with coordinates x and y (Figure 158). We regard the point in which the ray nz penetrates S the *representation* of z on S. Conversely, a point a on S is the representation of the complex number in which the straight line na intersects the plane. The north pole n represents no complex number. We agree to associate with it the "infinite" complex number, or point at infinity, denoted by ∞. This is motivated by the fact that the farther a point z on the plane is from the origin, the closer its representation on S is to n. S is called the *complex*, or *Riemann, sphere*. In contrast to the projective plane (see Figure 83), we can obtain the sphere S from the plane by adding to it the single point at infinity.

We will represent the values z on one complex sphere S_1 and the values of the polynomial $f(z)$ on another sphere S_2. To each finite point z_1 on S_1 there

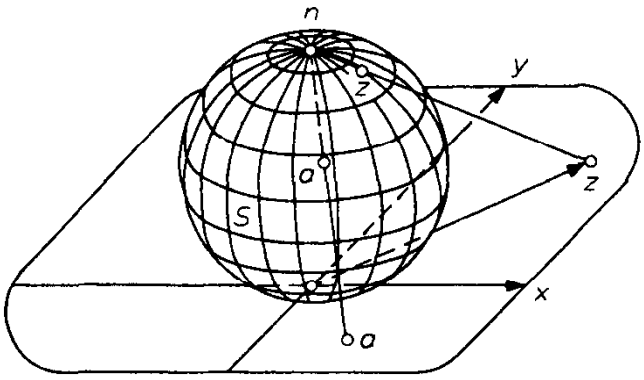


FIGURE 157.

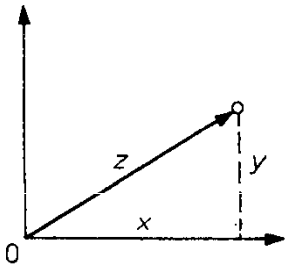


FIGURE 158.

corresponds the finite point $f(z)$ on S_2. If z tends to ∞, so too does $f(z)$ on S_2. Indeed,

$$f(z) = z^m \left(1 + \frac{a_1}{z} + \frac{a_2}{z^2} + \cdots + \frac{a_{m-1}}{z^{m-1}} + \frac{a_m}{z^m}\right).$$

If $z \to \infty$ (i.e., if $|z|$ increases indefinitely), then the expression in parentheses tends to 1 and $z^m \to \infty$. It follows that if we put $f(\infty) = \infty$, then f yields a continuous mapping of S_1 to S_2.

To prove the fundamental theorem of algebra we must prove that there is a point $z \in S_1$ such that $f(z) = 0$, i.e., that the point 0 on S_2 is the image of at least one point $z \in S_1$. Suppose this is not the case, i.e., suppose that the point 0 on S_2 is not covered by the image $f(S_1)$ of S_1. This means that the degree of the mapping $f\colon S_1 \to S_2$ near $0 \in S_2$ is 0. Since the degree of a mapping is constant near every point, the degree of f is 0. Thus to prove the fundamental theorem of algebra it suffices to prove that the degree of f is not 0. We will show that the degree of f is m, i.e., that it coincides with the degree of the polynomial $f(z)$ (this was the reason for the introduction of the concept of the degree of a mapping).

We change the values of the coefficients of the polynomial by letting them tend to 0; the polynomial changes and the mapping $f\colon S_1 \to S_2$ is continuously deformed. As a result we obtain the polynomial $f_1(z) = z^m$. Since a continuous deformation does not affect the degree of a mapping, f and f_1 have the same degree. The degree of f_1 is easy to compute.

We divide the plane by rays issuing from 0 into m congruent angles (Figure 159). Raising a number z to the mth power multiplies its argument by m. This means that the image of each of the angular sectors under under f_1 is all of S_2. But then

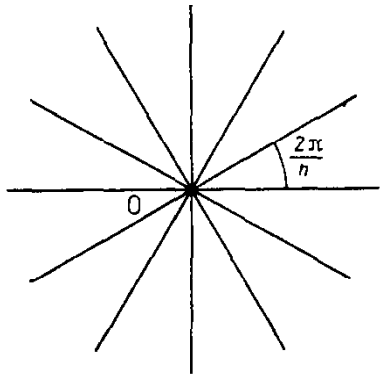


FIGURE 159.

the image of S_1 under f_1 is an m-tuple (positive) covering of S_2. This means that the degree of f_1—and therefore also of f—is m. This proves the fundamental theorem of algebra.

There are now many proofs of the fundamental theorem of algebra, all of them topological. This means that all of them make use of continuity in one form or another. This theorem cannot be proved without the use of topology. In other words—strange as this may sound—it is possible to show that the fundamental theorem of algebra is not an algebraic theorem.

Problems

165. Show that if $q \geq mk$, then there is a mapping f: $P_q \to P_k$ of degree m.

166. Show that if P and Q are orientable surfaces and the mapping f: $P \to Q$ is a k-sheeted covering, then the degree of f is $\pm k$.

167. Show that if $f(z)$ is a polynomial of degree $m > 1$, then there is at least one (complex or real) c for which the equation $f(z) = c$ has at most $m - 1$ different solutions.
Hint. If the number of different solutions were m for every c, then the mapping f: $S_1 \to S_2$ would be a covering, and thus a homeomorphism.

168. For $k \geq 1$ show that for every mapping f: $P_0 \to P_k$ and for every point $Q \in P_k$ there is a mapping F: $P_0 \times I \to P_k$ such that $F(R, 0) = f(R)$ and $F(R, 1) = Q$ for every $R \in P_0$, i.e., that the degree of f is 0.
Hint. Use the universal covering of P_k.

3.6. Knot Groups

Let L_1 and L_2 be two knots in three-space. Let D_1 be the complement of L_1 (the set of points in space that are not on L_1) and D_2 the complement of L_2. If L_1 and L_2 are equal (isotopic), i.e., if there is a homeomorphism f of space onto itself that takes L_1 to L_2, then $f(D_1) = D_2$ and the complement spaces are homeomorphic. It follows that the groups $\pi(D_1)$ and $\pi(D_2)$ are isomorphic, i.e., *the fundamental group of the complement space of a knot is an invariant of the knot.* This invariant

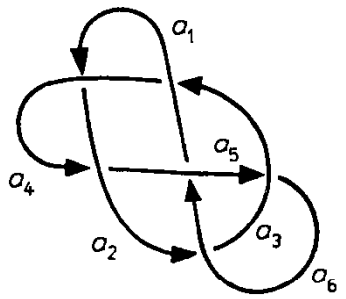


FIGURE 160.

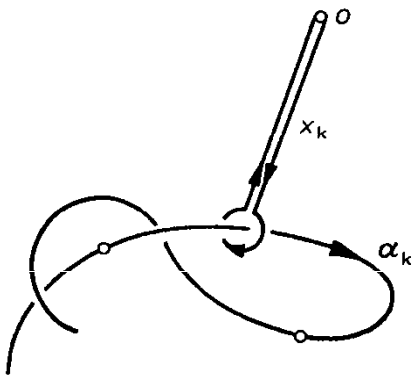


FIGURE 161.

is called the *knot group*. We will denote it by G, i.e., $G(L_1) = \pi(D_1)$. If two knots L and L' have different knot groups $G(L)$ and $G(L')$, then the knots L and L' are not isotopic.

We now give a procedure (without proof) for computing the knot group. Consider the normal projection of a knot L (see Section 2.10). The breaks decompose the normal projection of L into N arcs $a_1, a_2, \ldots, a_n$. We choose an orientation on L and mark it with arrows on $a_1, a_2, \ldots, a_n$ (Figure 160). To describe the group $G(L)$ we choose in space a point o above the curve L. Now we form a closed path x_k that begins at o, goes around the arc α_k located above a_k, and is oriented clockwise when looked at in the direction of the orientation of α_k (Figure 161). The homotopy classes of the paths x_k $(k = 1, \ldots, n)$ are the generators of the knot group.

Now we consider a double point of the projection, go around it (clockwise) on a small circle l, and write down a monomial as follows: We write x_i if the arc a_i leads into the interior of the little circle and x_i^{-1} otherwise. After going around l we will have written down, from left to right, a product of four factors. We set this product equal to 1. For example, for the double point in Figure 162 we obtain the relation

$$x_i x_k^{-1} x_j^{-1} x_k = 1.$$

It is not difficult to see that the path $x_i x_k^{-1} x_j^{-1} x_k$ is homotopic to 0 in the complement space: in Figure 163 we see a membrane, homeomorphic to a disk, that spans this path. It turns out that if we write down an equation analogous to the one above for

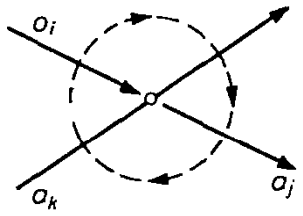


FIGURE 162.

each of the double points, then we obtain a complete system of defining relations for the knot group. This procedure is applicable to any link.

Before investigating specific knots and links we consider the following algebraic example.

Example 50 We show that the group on three generators x_1, x_2, x_3 with defining relations

$$x_2x_1x_3^{-1}x_1^{-1} = 1, \qquad x_3x_2x_1^{-1}x_2^{-1} = 1, \qquad x_1x_3x_2^{-1}x_3^{-1} = 1$$

is not abelian. For proof we consider the group G' of symmetries of an equilateral triangle. This group consists of six elements: three rotations about the center of the triangle through angles 0, $\frac{2\pi}{3}$, and $\frac{4\pi}{3}$, and three reflections x_1', x_2', x_3' whose axes are shown in Figure 164. It is easy to show that the relations just given hold for x_1', x_2', and x_3'. Since G' is not Abelian, the same is true of G.

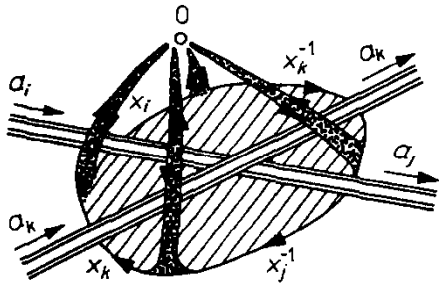


FIGURE 163.

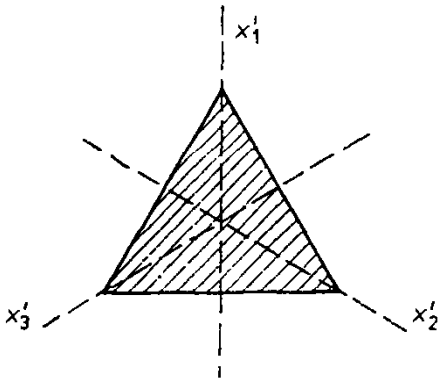


FIGURE 164.

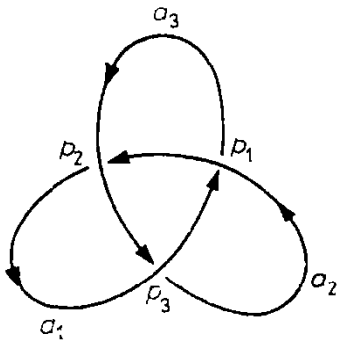


FIGURE 165.

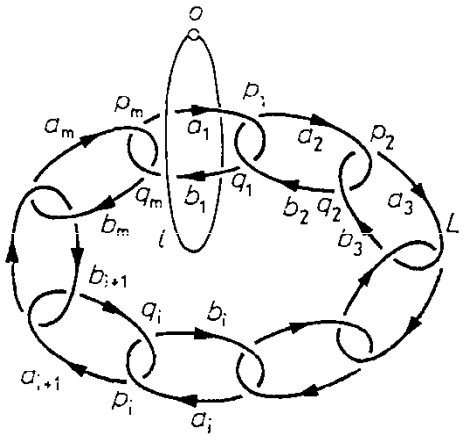


FIGURE 166.

Example 51 Figure 165 shows the projection of a trefoil knot L. The relations (formed at the double points p_1, p_2, p_3) connecting the generators x_1, x_2, x_3 coincide with those in Example 50. Hence the group $G(L)$ of this knot is not abelian. But then L cannot be isotopic to a circle (for which the fundamental group of the complement space is a free cyclic—and thus Abelian—group). It follows that the knot L cannot be unknotted without cutting the thread.

Example 52 Figure 166 shows a link L formed by the midlines of the tori that make up the set A_1 in Figure 104a. The group $G(L)$ of this link has $2m$ generators $x_1, \ldots, x_m, y_1, \ldots, y_m$. We obtain it by considering the paths that loop the arcs $a_1, \ldots, a_m, b_1, \ldots, b_m$ in Figure 166. The generators are connected by the $2m$ relations (arising at the double points p_i, q_i)

$$x_i x_{i+1}^{-1} x_i^{-1} y_{i+1} = 1, \qquad x_i y_{i+1}^{-1} y_i^{-1} y_{i+1} = 1 \qquad (i = 1, \ldots, m),$$

where we stipulate that $x_{m+1} = x_1$, $y_{m+1} = y_1$. The circle l in Figure 166 represents a path in the complement space of the link L whose homotopy class is represented by $x_1^{-1} y_1$ (Figure 167). We show that the path l is not homotopic to 0 in the complement space, i.e., if we contract l to a point, then it necessarily cuts the link L.

For proof, let G' denote the group of symmetries of a regular m-gon. This group consists of rotations through $0, 2\pi/m, 4\pi/m, \ldots, 2(m-1)\pi/m$ about the center

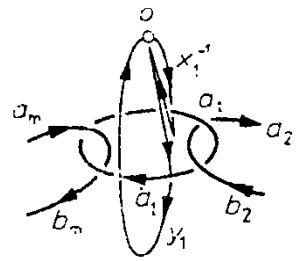


FIGURE 167.

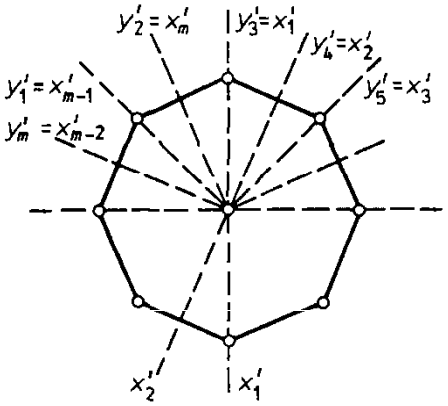


FIGURE 168.

of the m-gon and reflections about the axes $x_1', \ldots, x_m'$ shown in Figure 168. We denote these reflections by the same symbols x_i'. Next we define

$$y_1' = x_{m-1}', \qquad y_2' = x_m', \qquad y_3' = x_1', \ldots, y_m' = x_{m-2}'.$$

It is easy to see that the elements x_i', y_i' of the group G' satisfy all the previous relations given in terms of x_i, $+y_i$ ($(x_i')^{-1} = x_i'$; hence $x_i'x_{i+1}'$ is a rotation through $2\pi/m$). Also, the element $(x_i')^{-1}y_1'$ (which represents a rotation through $4\pi/m$) is different from the identity element of G' (i.e., from the identity mapping). But then the element $(x_i)^{-1}y_1$ is different from the identity element of G. In other words, the circle l generates in the complement space a path that is not homotopic to 0.

In much the same way we can show that the path l is not homotopic to 0 in the complement space associated with the union of the midlines of the tori that make up the set A_2 (see Figure 104b), etc. This provides another justification of the claim that Antoine's necklace has the properties ascribed to it in Example 31.

Problems

169. Show that it is impossible to "split" the link in Figure 117 without cutting one of its curves.

Hint. Show that the circle l_1 defines a nontrivial element of the group $G(L)$, where L is the link determined by the two other circles. To this end show that $G(L)$ is a free group on two generators.

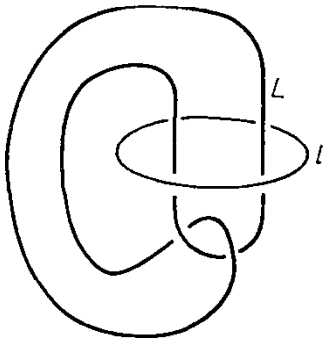


FIGURE 169.

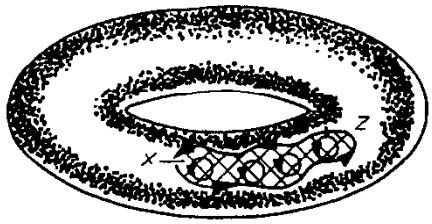


FIGURE 170.

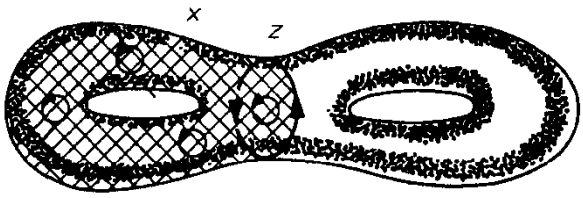


FIGURE 171.

170. Show that it is impossible to remove the circle in Figure 169 from the curve L; this means that there is no membrane in the complement space of L that is homeomorphic to a disk and spans l. Show also that there is a membrane homeomorphic to a handle that spans l and lies in the complement space of L.

3.7. Cycles and Homology

In Figures 170 and and 171 the cycle z (drawn as an unbroken line) bounds a region x on the surface. This region (a membrane spanning z) has the same orientation as the cycle. In what follows, we refer to *boundaries* (i.e., cycles that can be spanned by a membrane) as trivial or *homologous to zero.*

Figure 172a shows two cycles, z_1 and z_2. Their union is denoted by $z_1 + z_2$. The difference $z_1 - z_2$ (i.e., the sum of z_1 and $-z_2$, where $-z_2$ is obtained from z_2 by a change of orientation) is shown in Figure 172b. As the boundary of x, $z_1 - z_2$ is homologous to zero; we say that z_1 and z_2 are *homologous.*

Homology groups, introduced by Poincaré, are important topological invariants.

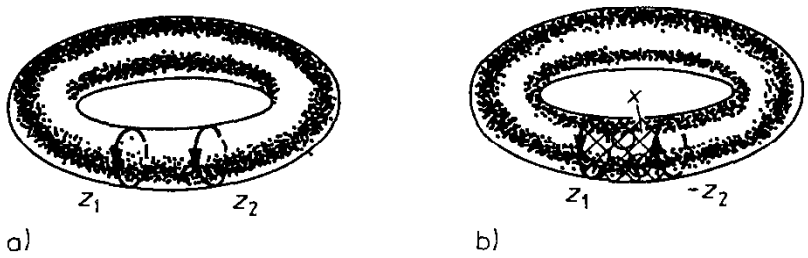


FIGURE 172.

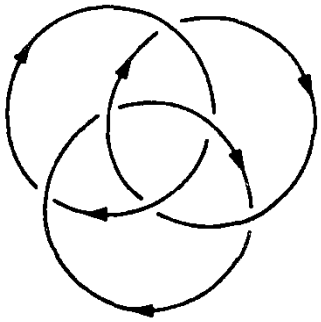


FIGURE 173.

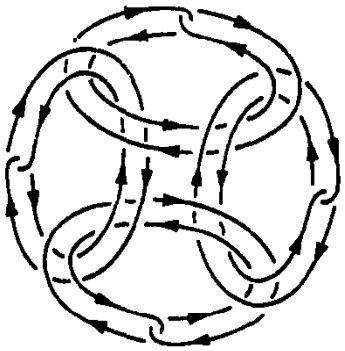


FIGURE 174.

The idea behind their construction is to determine how many pairwise nonhomologous cycles a given figure X can accommodate.

Problems

171. Show that every one-dimensional cycle on a sphere is homologous to zero.

172. Show that the cycle l_1 in Figure 105 is homologous to zero in the complement space of the set A_1 in Example 31 (and thus also in the complement space of Antoine's necklace $A^* \subset A_1$). This shows that a cycle homologous to zero need not be contractible.

173. Consider the figures in Figure 173 and Figure 174. Show that a cycle in either of these figures is homologous to zero in the exterior of the remainder of the relevant figure.

174. Show that if the linking number $m(z_1, z_2)$ is not equal to zero, then neither one of the cycles z_1 and z_2 is homologous to zero in the complement space of the other.

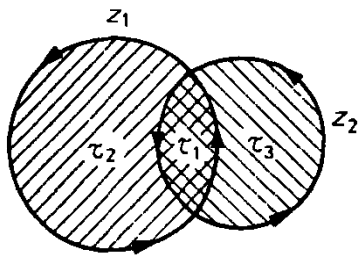


FIGURE 175.

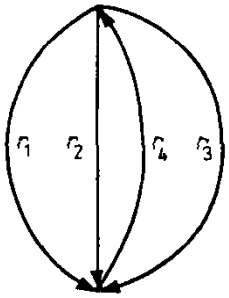


FIGURE 176.

In order to obtain the homology groups we must generalize the concept of a cycle and of the membrane spanning it. Each of the cycles z_1 and z_2 in Figure 175 is homologous to zero: z_1 is the boundary of the disk $\tau_1 + \tau_2$ and z_2 is the boundary of the disk $\tau_1 + \tau_3$. The sum $z_1 + z_2$ bounds the region $(\tau_1 + \tau_2) + (\tau_1 + \tau_3) = 2\tau_1 + \tau_2 + \tau_3$, with the cell τ_1 counted twice and each of the cells τ_2 and τ_3 counted once. This example shows that in order to find out whether a cycle $z_1 + z_2$ is homologous to zero we must assign suitable coefficients to its cells. Similarly, cycles may consist of cells with suitable coefficients. Thus the sum $r_1 + r_2 + r_3 + 3r_4$, related to Figure 176, is a cycle because at each vertex the number of incoming edges is equal to the number of outgoing edges.

Now we come to a theorem implied by the *duality principle of Alexander and Pontryagin.* The following is a simplified version of this principle: *Let P be a polyhedron embedded in three-space and let Q be its complement space. Let z_1 be a cycle in, say, P. The cycle z_1 is homologous to zero in P if and only if there is no cycle z_2 in Q that is linked with z_1 (i.e., if there is no cycle z_2 in Q such that* $\mathfrak{w}(z_1, z_2) \neq 0$).

Example 53 Figure 177 shows a curve P and a cycle z' in the complement space of P. The cycle z' is not linked with the 1-cycles of P, and so is homologous to zero in Q. Figure 177 also shows a 2-membrane $x' \subset Q$ with boundary z'.

Problems

175. Figure 178 shows four cycles m_1, m_2, m_3, m_4 on a pretzel surface. Give cycles z_1, z_2, z_3, z_4 in the complement space such that $\mathfrak{w}(m_i, z_j) = 1$ if $i = j$ and $\mathfrak{w}(m_i, z_j) = 0$ if $i \neq j$ $(i, j = 1, 2, 3, 4)$.

176. Show that for every knot $l \subset R^3$ there is a polyhedron $K \subset R^3$ that is

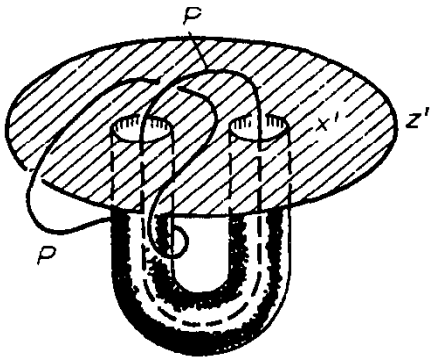


FIGURE 177.

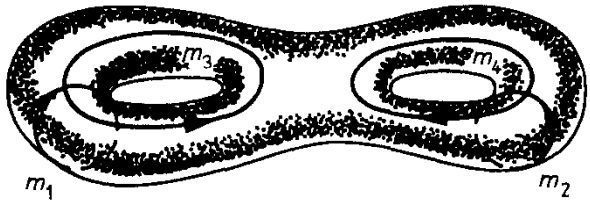


FIGURE 178.

homeomorphic to the lateral surface of a cylinder one of whose boundary curves is l and the other, l', is not linked with l (i.e., $\mathfrak{w}(l, l') = 0$).

177. For the figure P in Figure 132 give cycles m_1, m_2, m_3 in P and cycles z_1, z_2, z_3 in the complement space such that $\mathfrak{w}(m_i, z_j) = 1$ if $i = j$ and $\mathfrak{w}(m_i, z_j) = 0$ if $i \neq j$.

In the sequel we will consider not individual 1-cycles but *homology classes*, i.e., classes of pairwise homologous 1-cycles of a figure X. These classes form a group under addition, the one-dimensional *homology group* $H_1(X)$.

We now describe a procedure for the computation of one-dimensional homology groups, but note first that *two homotopic cycles* z_1 *and* z_2 (i.e., two cycles one of which can be obtained from the other by deformation) *are homologous*. Intuitively, this is so because the trace of z_1, when deformed into z_2, forms a membrane joining z_1 and z_2 (Figure 179). The converse is false. The cycles in Figure 180 are homologous but not homotopic: The holes in the surface make it impossible to move z_1 into z_2. It follows that for two cycles to be homologous it is sufficient, but not necessary, that they be homotopic.

It is easy to see that given a cell decomposition of a polyhedron, one can move any 1-cycle on the polyhedron by deformation into the 1-skeleton, i.e., into the graph made up of the vertices and edges of the cell decomposition (Figure 181). The folds that may come up in such a deformation can be straightened out. Hence any 1-cycle is homotopic (and therefore also homologous) to a cycle consisting of edges with suitable coefficients. Thus in order to compute the homology group $H_1(X)$ it suffices to consider the 1-cycles made up of edges with suitable integral coefficients. The membranes spanning the cycles can be regarded as 2-cells with suitable integral coefficients.

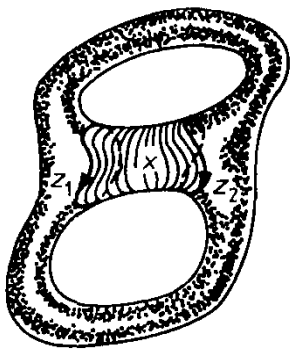


FIGURE 179.

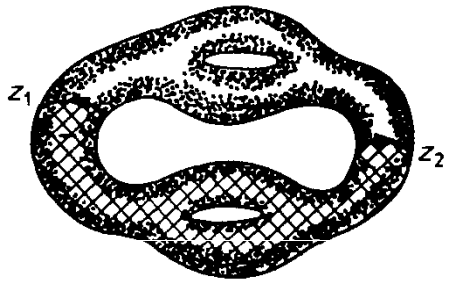


FIGURE 180.

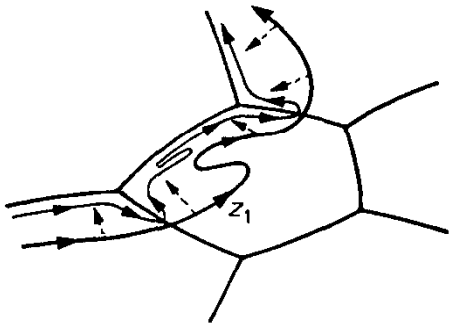


FIGURE 181.

More specifically: First, one must determine all 1-cycles (made up of edges). Second, one must compute the boundaries of the 2-cycles and determine which 1-cycles are homologous to one another. The first step is not difficult; all one has to do is make sure that the number of edges coming into a vertex is equal to the number of edges coming out of it, and this comes down to counting. As for the second step, we can already—in principle—carry it out. We go around the boundary in accordance with its orientation and write down the sum of the edges (and not their products, as was done in connection with edge paths) with appropriate signs. In other words, r appears in the boundary $\partial\tau$ of a cell τ with a coefficient equal to the sum of the powers with which it appeared in the edge path. For example, in the case of the cell in Figure 143, oriented counterclockwise, we have

$$\partial\tau_1 = a+b+c+d, \quad \partial\tau_2 = -d+f-h+k,$$
$$\partial\tau_3 = h+l, \quad \partial\tau_4 = -k-l.$$

Example 54 In Example 43 there is a cell decomposition of the 2-sphere P_0 with

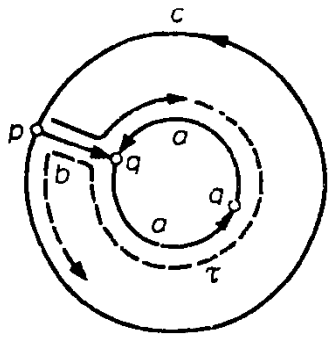


FIGURE 182.

just two cells. One of them is a 0-cell and the other a 2-cell. Since there are no 1-cells *the group* $H_1(P_0)$ *is trivial* (there are no nonzero 1-cycles).

Example 55 In Example 44 there is a cell decomposition of the projective plane with one 0-cell, one 1-cell r, and one 2-cell τ. Any 1-cycle has the form kr (for there are no edges other than r), and the cycle $2r$ is homologous to 0 (for $2r = \partial\tau$; see Figure 141). It follows that *the homology group* $H_1(N_0)$ *of the projective plane is the cyclic group of order* 2.

In Examples 54 and 55 we computed homology groups using particular cell decompositions of the polyhedra. Nonetheless, in each case we spoke of the homology group of the polyhedron, rather than of "the homology group with respect to the decomposition." The reason is that *the homology group of a polyhedron is completely determined by that polyhedron and does not depend on the choice of a decomposition.*

Problems

178. Figure 182 shows a cell decomposition of the Möbius strip (the two semicircles of the inner boundary must be glued together along the edge a). Show that $\partial\tau = c - 2a$. Deduce from this that the one-dimensional homology group of the Möbius strip is the free cyclic group.

179. Show that for the cell decomposition of the torus T in Example 45 we have $\partial\tau = 0$. Deduce from this that $H_1(T)$ is the free abelian group on two generators a and b.

180. Show that the cycle z in Figure 183 is homologous to $\pm 3a \pm 2b$ (the choice of sign depends on the orientation of the latitude circle a and the meridian circle b).

181. Show that the one-dimensional homology group of the pretzel surface P_2 is the free abelian group on four generators m_1, m_2, m_3, m_4 (cf. Figure 178).

182. Show that the one-dimensional homology group of the surface P_k is the free abelian group on $2k$ generators.

183. Compute the group $H_1(N_q)$. (*Answer.* $H_1(N_q)$ is an abelian group on q generators connected by the relation $2c_1 + 2c_2 + \cdots + 2c_q = 0$. This

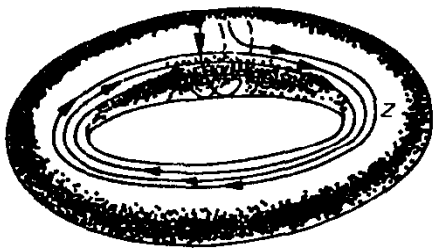


FIGURE 183.

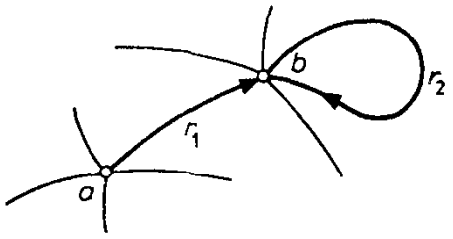


FIGURE 184.

group can be described as follows: $H_1(N_q)$ is the direct sum of a cyclic group of order 2 and the free abelian group on $q - 1$ generators).

184. Show that a closed surface Q is nonorientable if and only if the group $H_1(Q)$ contains an element of order 2. Show also that two closed surfaces are homeomorphic if and only if their one-dimensional homology groups are isomorphic.

185. Show that there is no cell decomposition of the torus with fewer than four cells.

We now discuss zero-dimensional homology. To obtain 0-cycles we must assign integral coefficients to the vertices of a cell decomposition. The boundary of an edge is equal to the difference of its endpoints. In Figure 184, $\partial r_1 = b - a$ and $\partial r_2 = 0$. Two 0-cycles are homologous if their difference is the boundary of a sum of 1-cells (with appropriate coefficients). We divide the 0-cycles into classes and put two cycles in the same class if they are homologous in the polyhedron X under consideration. The set of all classes forms a group under cycle addition. This group is the *zero-dimensional homology group* $H_0(X)$.

Problems

186. Show that if $r_1, r_2, \ldots, r_k$ is a simple chain of oriented edges leading from vertex a to vertex b, then $\partial(r_1 + r_2 + \cdots + r_k) = b - a$.

187. Show that if X is a connected polyhedron, then every 0-cycle in X is homologous to a point with an appropriate coefficient, i.e., $H_0(X)$ is a free cyclic group.

188. Show that the group $H_0(X)$ of a polyhedron with k components is a free abelian group on k generators.

The *two-dimensional homology group* $H_2(X)$ is defined in a similar way. One must consider the 2-cycles in X and the membranes spanning them.

Example 56 Let X be a *solid torus*, i.e., the set of points on a torus and in its interior. A cell decomposition of this polyhedron consists of cells o, a, b, τ on the torus, a 2-cell τ' that is a crosscut of the torus bounded by a meridian circle b, and a 3-cell v that is the interior of the torus cut by τ'. The boundaries of these cells satisfy the relations

$$\partial a = 0, \quad \partial b = 0, \quad \partial \tau = 0, \quad \partial \tau' = b, \quad \partial v = \tau.$$

The reason the 2-cell τ' does not appear in the expression for ∂v is that the 3-cell v borders on τ' from two sides with opposite orientations.

The 1-cycles of this cell decomposition have the form $ka + lb$ (k, l, integers) with b homologous to 0 (it bounds the membrane τ'). It follows that an arbitrary 1-cycle is homologous to ka, and so the group $H_1(X)$ is a free cyclic group. Since $\partial(m\tau + n\tau') = nb$, $m\tau + n\tau'$ is a 2-cycle (i.e., has just 0 as its boundary) if and only if $n = 0$. It follows that the 2-cycles have the form $m\tau$. But any such cycle is homologous to 0 (for $\partial v = \tau$, i.e., v is a "three-dimensional" membrane spanning the 2-cycle τ). Hence the group $H_2(X)$ is trivial.

Frequently, it is easier to determine the rank of $H_r(X)$ [the analogously defined group of r-cycles] than $H_r(X)$ itself. The rank of $H_r(X)$ is called the *r-dimensional Betti number* of the polyhedron X and is denoted by $p_r(X)$. The following is an alternative definition of the Betti number $p_r(X)$: We say that the r-cells $z_1, \ldots, z_n$ are homologically independent in X if there are no integers $k_1, \ldots, k_n$ not all 0 such that the cycle $k_1 z_1 + \cdots + k_n z_n$ is homologous to 0 in X. The r-dimensional Betti number $p_r(X)$ can be defined as the largest number of homologically independent r-cycles in X.

As an application of Betti numbers we state (without proof) a theorem for the computation of the Euler characteristic. Let X be a polyhedron represented as a cell decomposition. Let α_r, $r = 1, 2, \ldots$, be the number of r-cells of this decomposition. Then the *Euler characteristic of X*, i.e., the number $\chi(X) = \sum(-1)^r \alpha_r$, is given in terms of Betti numbers by the formula

$$\chi(X) = \sum(-1)^r p_r(X),$$

where the sum extends to the largest k for which there are k-cells in the decomposition.

Example 57 The *three-dimensional sphere* S^3 is defined as the boundary of a ball in 4-space R^4. Its equation in rectangular coordinates is

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1.$$

Consider a four-dimensional analogue of Figure 157. It is easy to show that the 3-sphere with a point removed is homeomorphic to Euclidean 3-space, i.e., to

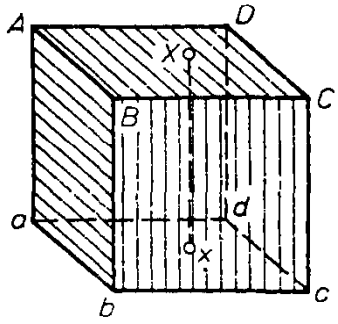


FIGURE 185.

an open 3-ball. Hence it is possible to represent the 3-sphere as a cell complex consisting of two cells: a 0-cell o and a 3-cell v. It follows (see Example 54) that *the homology groups $H_0(S^3)$ and $H_3(S^3)$ are free cyclic groups and that the homology groups of S^3 for other dimensions are trivial.* Hence $p_0(S^3) = p_3(S^3) = 1$ and $p_1(S^3) = p_2(S^3) = 0$.

Example 58 Gluing together the opposite sides of a square yields a torus. Similarly (Figure 185), gluing together the opposite sides of a cube yields a *three-dimensional torus* T^3 (which must not be confused with the solid torus considered in Example 56); in particular, points on the faces $ABCD$ and $abcd$, which are endpoints of segments parallel to Aa, are glued together. The vertices of the cube are all glued together and yield a 0-cell. All parallel edges are glued together and yield three 1-cells. Gluing together of opposite faces yields three 2-cells. There is also one 3-cell. All these yield a cell decomposition of the three-dimensional torus T^3. The boundary of each of these cells is 0. Hence *the homology group $H_3(T^3)$* of a 3-torus *is a free cyclic group and each of the groups $H_1(T^3)$ and $H_2(T^3)$ is a free abelian group on three generators.* It follows that $p_0(T^3) = p_3(T^3) = 1$ and $p_1(T^3) = p_2(T^3) = 3$.

Example 59 We will show that *we can obtain a three-dimensional sphere by gluing together two solid tori.* A torus (see Figure 5) divides space into two regions, an inner and an outer one. The inner region coincides with the interior of a solid torus. If we add a point to 3-space (this yields a 3-sphere), then the outer region also goes over into a solid torus, and this yields the division of a 3-sphere into two glued-together solid tori.

We clarify the claim in the first half of the previous sentence. Consider Figure 186. If we rotate this figure about the straight line l, then, from a topological viewpoint, each of the "lines of force" that begin at the "charge" A and end at the "charge" B (including the line $m \infty m'$, which is actually a single line (there is just one point at infinity!)) yields a disk that is uniquely determined by a point on the "charge" A. Thus the outer region determined by the torus (supplemented by a point at infinity) is indeed homeomorphic to a solid torus.

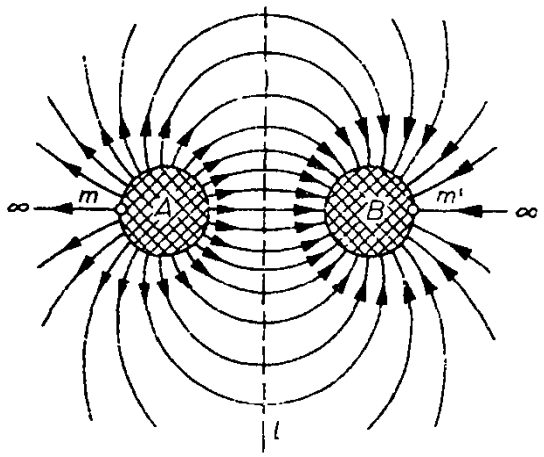


FIGURE 186.

Problems

189. Show that for a 3-ball X the groups $H_1(X)$, $H_2(X)$, and $H_3(X)$ are trivial.

190. A region in 3-space bounded by two concentric spheres is called a *spherical shell.* Let X be the polyhedron obtained by gluing together the diametrically opposite points on each of the boundary spheres. Show that the groups $H_0(X)$, $H_1(X)$, $H_2(X)$, and $H_3(X)$ are free cyclic groups.

191. Compute the homology groups of the polyhedron that is the union of the surface P_k and its interior.

192. Compute the homology groups of the three-dimensional figure (*three-dimensional projective space*) obtained from a 3-ball by gluing together the diametrically opposite points on its boundary.

193. Show that every closed surface can be embedded in three-dimensional projective space without self-intersections.

194. Let X be a polyhedron given in the form of a cell decomposition, and let α_r be the number of r-cells, $r = 0, 1, \ldots, n$, where n is the largest of the occurring dimensions. Show that if r is any one of the integers $0, 1, \ldots, n-1$, then

$$\sum_{k=1}^{r}(-1)^{r-k}\alpha_k \geq \sum_{k=1}^{n}(-1)^{r-k}p_k(X).$$

Hint. Consider the r-dimensional skeleton X^r of X (consisting of all cells of the decomposition of dimension $\leq r$) and prove the relations

$$p_0(X^r) = p_0(X),$$

$$p_1(X^r) = p_1(X), \ldots, p_{r-1}(X^r) = p_{r-1}(X),$$

$$p_r(X^r) \geq p_r(X).$$

We conclude this section by mentioning that when forming homology groups we need not limit the coefficients to integers. We can also compute modulo 2,

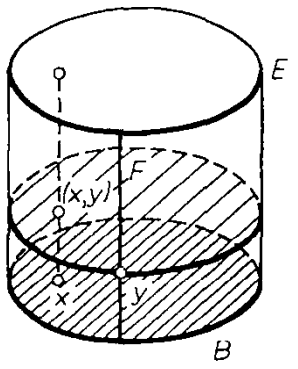


FIGURE 187.

modulo m, and, more generally, with elements of any abelian group G. The resulting homology groups are denoted by $H_r(X, Z_2)$, $H_r(X, Z_m)$, and $H_r(X, G)$ respectively. If we take as coefficients elements of Z_2, then we can regard all cells as unoriented. If we take as coefficients elements of a cyclic group of prime order p, then we obtain as homology groups $H_r(X, Z_p)$ direct sums of groups all of which are isomorphic to Z_p. The number of summands in such a direct sum is called the *r-dimensional Betti number* of the polyhedron X *modulo* p.

Problems

195. Show that for the projective plane N_1 the groups $H_0(N_1, Z_2)$, $H_1(N_1, Z_2)$, and $H_2(N_1, Z_2)$ have order 2.

196. Show that the surfaces P_k and N_{2k} have the same homology groups modulo 2 in all dimensions.

197. For three-dimensional projective space (see Problem 192) compute the homology groups modulo 2.

3.8. Topological Products

Example 60 Every point on the cylinder E (Figure 187) can be given as a pair of points (x, y), x on the lower base B and y on a generator F. The point x determines a segment parallel to F and y a disk parallel to B. The intersection of the two is the required point on the cylinder. In this way, the *cylinder* E can be viewed as *the set of all points* (x, y) *such that* x *varies over the set of points of the figure* B *(a disk) and* y *varies over the set of points of the figure* F *(a segment).*

Example 61 We consider two circles on the torus E (Figure 188), the *meridian* B and the *parallel* F. To determine a point on the torus it suffices to give a point $x \in B$ and a point $y \in F$. Then the required point is the intersection of the parallel through x and the meridian through y. In this way the *torus* E can be viewed as *the set of all pairs* (x, y) *such that* $x \in B$ *and* $y \in F$.

In the two previous examples we obtained the *topological product* of two figures B and F— the cylinder as the topological product of a disk and a segment and the

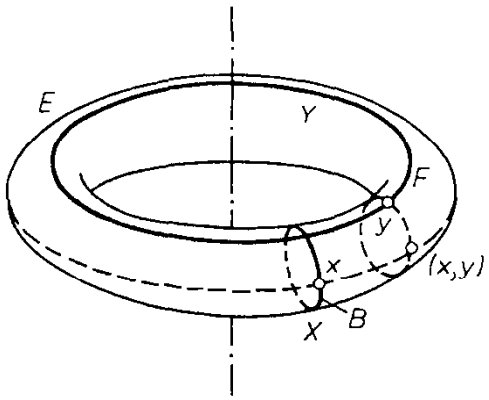


FIGURE 188.

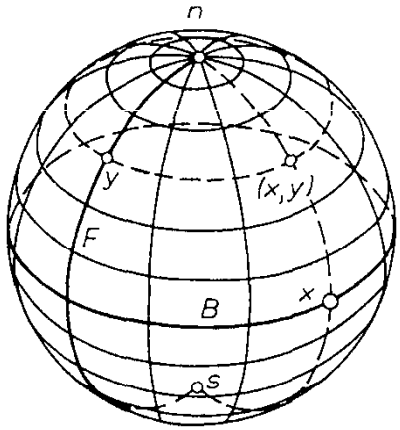


FIGURE 189.

torus as the topological product of two circles. Quite generally, we call a figure E the *topological product of figures B and F if E can be represented as the set of all pairs (x, y) such that $x \in B$ and $y \in F$*. Note that so far, we have described the set of points of E but have not said anything about its topology. The latter can be described intuitively as follows: We say that the points (x_1, y_1) and (x_2, y_2) in E are "close" if x_1 and x_2 in B are "close" and y_1 and y_2 in F are "close." Here it is essential that to each point of E there corresponds one pair (x, y) and the points of E corresponding to different pairs are different.

Example 62 Consider the *equator* B and the *zero meridian* F on a sphere (Figure 189). To determine a point on the sphere it suffices to give its geographic coordinates, i.e., points $x \in B$ and $y \in F$. The required point is the intersection of the circles determined by x and y respectively. But this does not mean that the sphere is the topological product of the equator and the zero meridian. Indeed, if x and x' are different points on the equator and n is the north pole on the zero meridian, then the same point n on the sphere corresponds to the different pairs (x, n) and (x', n).

Problems

198. Show that an annulus is the topological product of a segment and a circle.

199. Show that a solid torus is the topological product of a disk and a circle.

200. Show that the polyhedron X in Problem 190 is the topological product of a sphere and a circle.

201. Show that the 3-torus T^3 is the topological product of a 2-torus and a circle. This means that T^3 is the topological product of three circles.

We will investigate *homological properties of topological products*, but, for the sake of simplicity, will limit ourselves to considering Betti numbers rather than homology groups. If in Example 61 we view the meridian and parallel as 1-cycles, then their topological product (the torus) is a 2-cycle. Quite generally, let z be an r-cycle in a polyhedron B and let z' be an r'-cycle in a polyhedron F. The product of these cycles is an $(r + r')$-cycle in the polyhedron E that is the topological product of B and F. In this way, i.e., by multiplication of cycles in B and F, we can obtain a system of homologically independent cycles in the polyhedron E. To this end we choose first a maximal system of homologically independent 0-cycles in B and a maximal system of homologically independent r-cycles in F. By multiplying these cycles we obtain $p_0(B)p_r(F)$ r-cycles in E. Next we choose a maximal system of homologically independent 1-cycles in B and a maximal system of homologically independent $(r-1)$-cycles in F. By multiplying these cycles we obtain $p_1(B)p_{r-1}(F)$ r-cycles in E. Then we do the same with 2-cycles in B and $(r-2)$-cycles in F, etc. The set of cycles obtained in this way is a maximal set of homologically independent r-cycles in the polyhedron E. It follows that for the *topological product E of two polyhedra B and F we have*

$$p_r(E) = p_0(B)p_r(F) + p_1(B)p_{r-1}(F) + p_2(B)p_{r-2}(F) + \cdots + p_r(B)p_0(F).$$

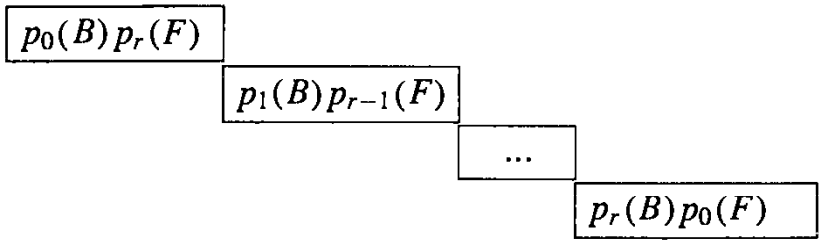
We can clarify this formula graphically as follows. Set up a table and enter the number $p_i(B)p_j(F)$ in the box in the jth column and ith row:

...	...	...	...	...	...
$p_j(F)$	$p_0(B)p_j(F)$	$p_1(B)p_j(F)$	...	$p_i(B)p_j(F)$	...
...	...	...	...	...	...
$p_2(F)$	$p_0(B)p_2(F)$	$p_1(B)p_2(F)$	...	$p_i(B)p_2(F)$	...
$p_1(F)$	$p_0(B)p_1(F)$	$p_1(B)p_1(F)$	...	$p_i(B)p_1(F)$	...
$p_0(F)$	$p_0(B)p_0(F)$	$p_1(B)p_0(F)$	...	$p_i(B)p_0(F)$	...

$p_0(B)$	$p_1(B)$	...	$p_i(B)$	...

Then the sum on the rth diagonal of this table yields the r-dimensional Betti

number of the polyhedron E:



Problems

202. Set up a table like the one above for the topological product of a sphere and a circle and use the above procedure to compute the Betti numbers of the polyhedron in Problem 190 (see also Problem 200).

203. Show that if a polyhedron E is the topological product of two polyhedra B and F, then $\chi(E) = \chi(B) \cdot \chi(F)$.

204. Compute the Betti numbers of an n-torus (i.e., the topological product of n circles).

205. Show that neither the 3-sphere nor three-dimensional projective space is homeomorphic to the topological product of a circle and a surface.

3.9. Fiber Bundles

In Example 60 we denoted by p the projection of the cylinder E on its base B. For every $x \in B$ the preimage $p^{-1}(x)$ is a segment parallel to F. We call such segments *fibers*. Above each x in the base figure B there is ("grows") a fiber, and the whole cylinder is their union (much like a bundle of fibers).

If E is the topological product of B and F, then the projection p that associates with every point $(x, y) \in E$ the point $x \in B$ is a mapping of E on the base B such that the preimage $p^{-1}(x)$ of every $x \in B$ (the fiber that grows over x) is homeomorphic to F. This can be easily checked against Example 61 as well as against Problems 198 to 201.

Consider the projection p of a helix E on a circle B (Figure 145). Each preimage $p^{-1}(x)$ (the fiber growing over x) is homeomorphic to the set F of points $\ldots, -4\pi, -2\pi, 0, 2\pi, 4\pi, 6\pi, \ldots$ on the number line. This example is different from the one in Figure 190 below, which shows the topological product of this fiber and the circle B, consisting of infinitely many disjoint circles. However (we are back to the case of the spiral), $p^{-1}(U)$ of a neighborhood U (see Figure 146) splits into disjoint sheets, i.e., $p^{-1}(U)$ is the topological product of a neighborhood U with a fiber. This means that E is a topological product locally (in the neighborhood of each point $x \in B$) but not globally. In such cases we speak in topology of *locally trivial fiber bundles*. Every covering is a locally trivial fiber bundle, where the fiber F of the bundle consists of isolated points. In Example 49 the fiber consists of two points, the covering E is an orientable surface, while the base B is nonorientable.

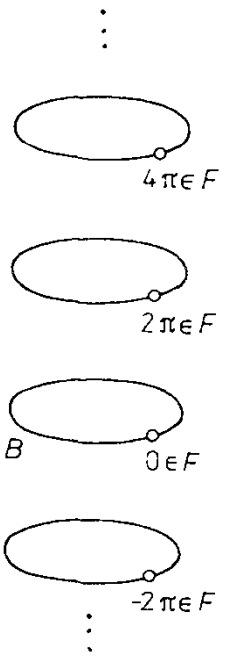


FIGURE 190.

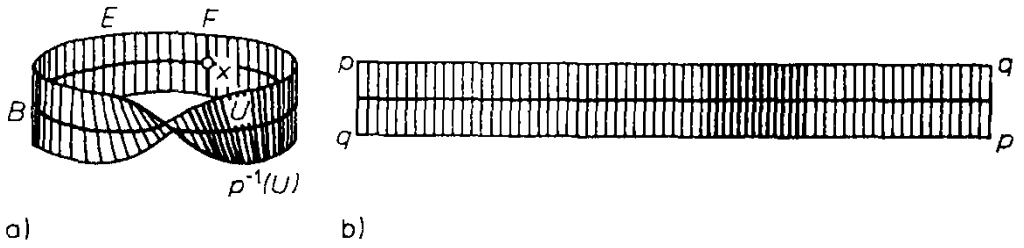


FIGURE 191.

Example 63 Denote by B the midline of a Möbius strip E. At each point $x \in B$ there is a segment stretching across the strip from edge to edge. We call it the fiber that grows over x (these cross segments come from the cross segments of the rectangular strip that is made into a Möbius strip by gluing). If we map each cross segment on the corresponding x, then we obtain a projection p: $E \to B$, where $p^{-1}(x)$ is the fiber over x. This fiber bundle is locally trivial. Indeed, if we choose on the circle B an arc U, then its preimage $p^{-1}(U)$ is the topological product of U and the fiber F (Figure 191). But the Möbius strip is not the topological product of the circle B and the fiber F (cf. Problem 198).

Example 64 Another example of a locally trivial fiber bundle is the *normed tangent bundle* of a surface. Let B be an orientable surface. Let E be the set of all unit vectors tangent to B. Let p: $E \to B$ be the mapping that associates with every tangent vector $z \in E$ its point of origin $x \in B$. The fiber $p^{-1}(x)$ over $x \in B$ (which consists of all unit vectors tangent to the surface at x) is homeomorphic to a circle. The mapping p: $E \to B$ is a locally trivial fiber bundle. Indeed, every small neighborhood U of a point $x \in B$ can be viewed as a small piece of the plane, and therefore every vector z tangent to B at a point $x \in U$ can be given as a

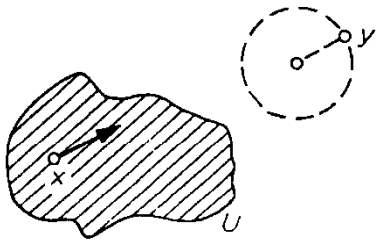


FIGURE 192.

pair (x, y), where y is a point of the unit circle (Figure 192). Hence $p^{-1}(U)$ can be represented as the topological product of U and the circle.

In particular, let E be the normed tangent bundle of the sphere S^2 (i.e., the set of all unit vectors tangent to this sphere). We can represent E as a cell decomposition with four cells. Specifically, let $x_0 \in S^2$ and let F_0 be the fiber over x_0. We choose a point $\tau^0 \in F_0$ and denote the rest of F_0, a 1-cell, by τ^1. Let v be a vector field on S^2 with just one singularity (with index +2) at x_0. The field v can be viewed as a 2-cell in E. This cell is projected on S^2 minus the point x_0. It has just one point in common with every fiber other than F_0. If we remove the cells τ^0, τ^1, and $\tau^2 = v$, then we obtain a set τ^3 homeomorphic to an open 3-ball. Thus E can be represented as the cell decomposition $\{\tau^0, \tau^1, \tau^2, \tau^3\}$. The boundary of the cell $v = \tau^2$ is the fiber F_0 gone over twice, i.e., $\partial\tau^2 = 2\tau^1$. The boundaries of the other cells are empty: $\partial\tau^3 = 0$, $\partial\tau^1 = 0$, $\partial\tau^0 = 0$. It follows easily that the Betti numbers of E are $p_0(E) = p_3(E) = 1$ and $p_1(E) = p_2(E) = 0$. If we consider homology modulo 2, then $\partial\tau^2 = 0$. Hence the Betti numbers modulo 2 are $p_0(E) = p_1(E) = p_2(E) = p_3(E) = 1$.

Problems

206. Show that the Klein bottle can be represented as a locally trivial fiber bundle with circles as both base and fiber.

207. Show that the normed tangent bundle of a 2-torus T is homeomorphic to a 3-torus.

208. Show that if a locally trivial fiber bundle has a 2-sphere as base and a circle as fiber then the space E of this fiber bundle can be obtained by gluing together the boundaries of two solid tori.

209. Show that if a locally trivial fiber bundle has a circle as base and a segment as fiber, then the space E of this fiber bundle is homeomorphic to an annulus or a Möbius strip.

The following important theorem on homologies of fiber bundles is due to the French mathematician Jean Leray. We state it in a simplified form.

Let p: $E \to B$ be a fiber bundle whose base is a connected polyhedron with trivial fundamental group and whose fiber F is an arbitrary polyhedron. As in the previous section, we set up a table for the topological product of B and F. In this table we indicate "knight's moves" by arrows (Figure 193). Each arrow is labeled

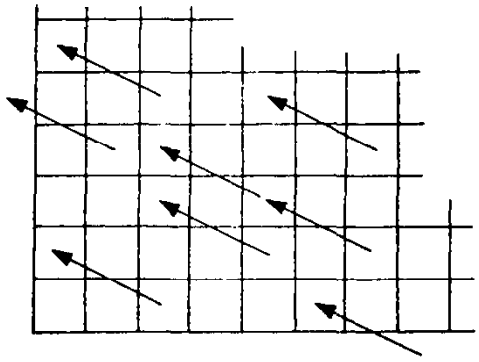


FIGURE 193.

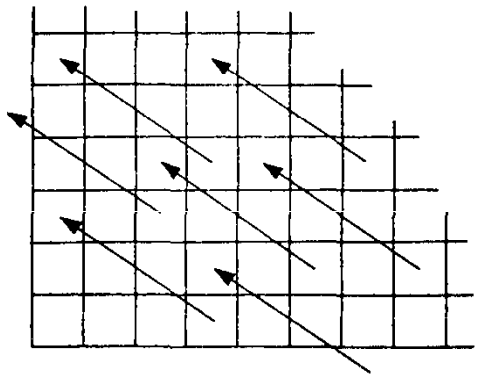


FIGURE 194.

with a nonnegative integer and observe the following conditions: 1. The number in a cell is not less than the sum of the respective numbers at the two arrows that enter and leave the cell. 2. If either the beginning or end of an arrow goes beyond the bounds of the table, then we label it with 0. We call the result *table* E_2.

Now we set up a new table. With each cell we associate as the new number the difference between the old number and the sum of the respective numbers at the two arrows leading in and out. Then we mark the "elongated knight's moves" (Figure 194) and write next to them nonnegative integers satisfying conditions 1 and 2 above. In this way we obtain *table* E_3.

In the same way we obtain *table* E_4 from *table* E_3, etc. In *table* E_n the arrows move n cells to the left and $n-1$ cells upward.

Regardless of the choice of cell decomposition, the numbers in the tables eventually stop changing; they stabilize: The arrows keep getting longer and finally go beyond the bounds of the table. We mark the table in which the numbers stop changing E_∞ (it has no arrows that run completely in the interior). The theorem of Leray is the assertion: *It is possible to choose the numbers to be attached to the arrows so that the sum of the numbers on the rth diagonal of E yields the Betti number of dimension r for the space E.* A similar result holds for the Betti numbers with respect to a prime modulus p.

Example 65 Let $p\colon E \to B$ be a fiber bundle with base S^2 and fiber S^1. Then $p_0(B) = p_2(B) = 1$, $p_0(E) = p_1(E) = 1$, and the remaining Betti numbers

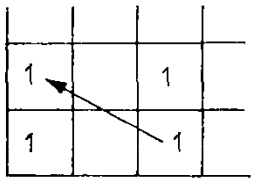


FIGURE 195.

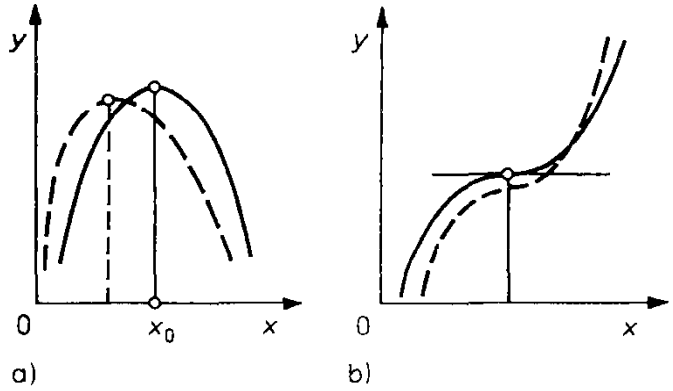


FIGURE 196.

for base and fiber are 0. Hence the table E_2 looks like the table in Figure 195 (the numbers in the cells other than the four containing 1 are 0 and the arrows other than the indicated one are likewise labeled 0). The number at the indicated arrow is either 0 or 1. The table E_3 is already the table E_∞ (all arrows go beyond the edge of the table). By Leray's theorem, our space E has the Betti numbers $p_0(E) = p_1(E) = p_2(E) = 1$ (if the number at the indicated arrow is 0) or the Betti numbers $p_0(E) = p_3(E) = 1$, $p_1(E) = p_2(E) = 0$ (if the number at the indicated arrow is 1). The first possibility is realized for the topological product of S^2 and S^1 (cf. Problem 200). The second is realized for the tangent bundle of S^2 (see Example 64).

3.10. Morse Theory

A necessary condition for a differentiable function to have a (local) minimum or maximum at an interior point x_0 of its domain of definition is that its tangent be horizontal at x_0. But this condition is not sufficient. At a point of inflection with a horizontal tangent the function has neither a minimum nor a maximum.

Minimum and maximum points are stable with respect to small perturbations of their graphs (Figure 196a). This is not the case for inflection points (with horizontal tangent): Such a point can disappear as a result of a small perturbation (i.e., there is no nearby point where the tangent is horizontal; Figure 196b).

There is an analogous necessary condition for functions of two variables x and y (defined in a region of the plane): *For a function $f(x, y)$ to have a local maximum at a given interior point (x_0, y_0) of its domain of definition it is necessary that this*

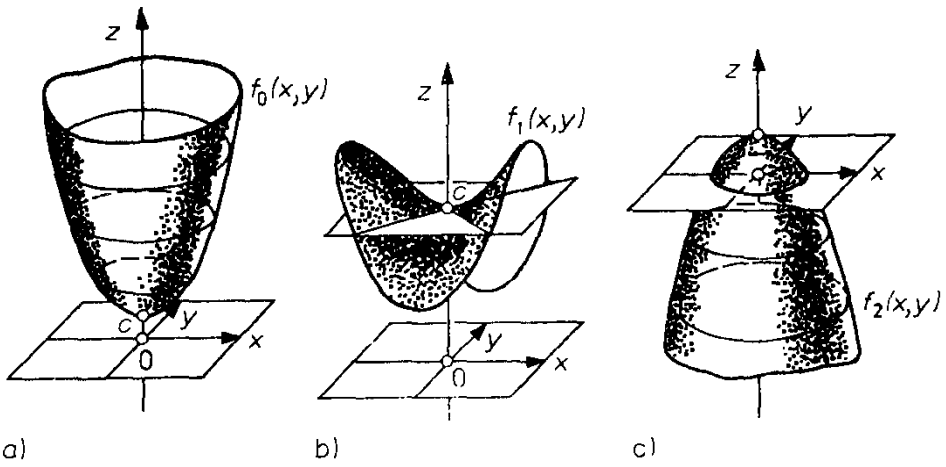


FIGURE 197.

point be stationary, i.e., that the graph of the function have a horizontal tangent plane at (x_0, y_0).

Example 66 Figure 197 shows the graphs of the functions

$$\begin{aligned} f_0(x, y) &= c + x^2 + y^2, \\ f_1(x, y) &= c + x^2 - y^2, \\ f_2(x, y) &= c - x^2 - y^2. \end{aligned} \tag{21}$$

In all three cases the point (0, 0) is stationary. It is a minimum point for f_0, a maximum point for f_2, and neither a minimum nor a maximum for f_1, for which it is a so-called *saddle point*. All these points are stable under small perturbations of the graph of the function. There are even more complicated stationary points. Thus the function $f(x, y) = x^3 - 3xy^2$ has at the origin a saddle point of order three (three depressions and three elevations (rather than two as in Figure 197b)). It is easy to see that this stationary point is likewise stable under small perturbations of the graph.

We can also consider functions given not on a plane but on surfaces that are topologically equivalent to the plane in a neighborhood of each of their points.

Example 67 Let p be a point on a torus T and let $f(p)$ denote the height of p above a horizontal plane Π. If the torus is located as shown in Figure 198, then this function has a maximum point a, a minimum point d, and two saddle points b and c. Let C_0 be the number of minimum points, C_2 the number of maximum points, and C_1 the number of saddle points. In this example we have $C_0 = 1$, $C_1 = 2$, $C_2 = 1$. Hence $C_0 - C_1 + C_2 = 0$.

Example 68 Consider the function $f(p)$ defined in the previous example on a sphere rather than on a torus. In this case we have two stationary points: a minimum point (the south pole) and a maximum point (the north pole). In this case we have $C_0 = 1$, $C_1 = 0$, $C_2 = 1$. Hence $C_0 - C_1 + C_2 = 2$.

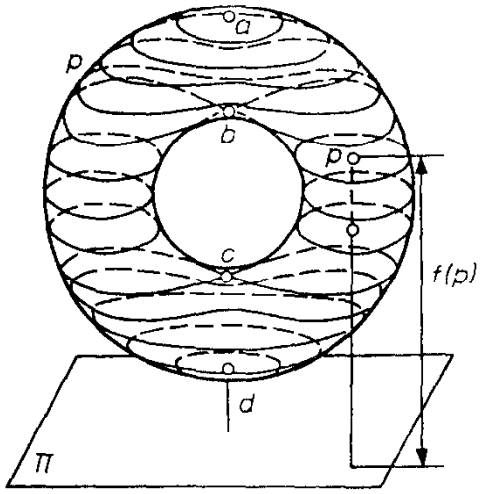


FIGURE 198.

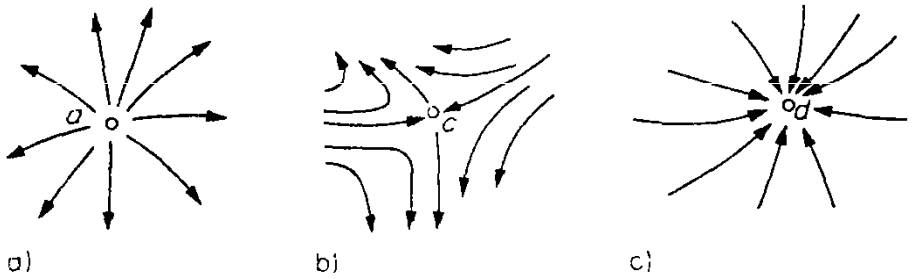


FIGURE 199.

These examples lead to the formulation of a theorem due to the American mathematician Morse. We assign to a minimum point the index 0, to a saddle point the index 1, and to a maximum point the index 2. We can now formulate the first half of Morse's theorem (for the case of a surface): *Consider a function on a surface Q all of whose stationary points are nondegenerate [i.e., at each stationary point the matrix $(\partial^2 f/\partial x_i \partial x_j)$ is invertible]. If C_0 is the number of stationary points of index 0 (i.e., the number of minimum points), C_1 the number of stationary points of index 1 (i.e., the number of saddle points), and C_2 the number of stationary points of index 2 (i.e., the number of maximum points), then*

$$C_0 - C_1 + C_2 = \chi(Q). \tag{22}$$

We argue as follows. The function f determines on the surface Q level lines (i.e., lines on which its value remains fixed). One can investigate on Q the lines of steepest descent, i.e., lines on which the function decreases most rapidly. These lines are orthogonal to the level lines. The vectors that point in the direction of the lines of steepest descent form a vector field on Q. This vector field has no singularities at the nonstationary points of the function. [Thus a singular point of the vector field is a stationary point of the function.] Figure 199 shows the vector field near a minimum point (a), near a saddle point (b), and near a maximum point (c). It is easy to see that if j is the index of a singular point of the vector field and k is its index as a stationary point, then $j = (-1)^k$ (see Figure 89). Thus the vector

field under consideration has C_0 singular points with index $+1$ (minimum points), C_1 singular points with index -1 (saddle points), and C_2 singular points with index $+1$ (maximum points). Now Poincaré's theorem on vector fields (Section 2.6) implies the validity of (22) for closed orientable surfaces.

Problems

210. Show that equation (22) holds also for nonorientable surfaces.

211. On the surface P_k there is a function all of whose stationary points are nondegenerate. Show that the number of its stationary points is at least $2k + 2$.

Now we formulate the second half of Morse's theorem: *If on a surface Q there is a function all of whose stationary points are nondegenerate, then*

$$C_0 \geq p_0(Q), \qquad C_1 - C_0 \geq p_1(Q) - p_0(Q). \tag{23}$$

We assume that the values that the function takes on at its stationary points $a_1, \ldots, a_q$ are distinct and that the stationary points are numbered so that $f(a_1) > \cdots > f(a_q)$.

Near a_1 (the point with the largest local maximum, Figure 200a) the level lines are closed and circle a_1. We cut the surface along such a level line and obtain a two-dimensional cap τ_1 and a remainder Q_1 (Figure 200b).

We simplify Q_1. To this end we draw a level line slightly above a_2. Now we remove from Q_1 the part F_1 above this level line (Figure 200c). We denote the remaining part of the surface by Q_1'. Since there are no stationary points between a_1 and a_2, F_1 consists of "parallel" lines of steepest descent, and one can "pull down" F_1 to Q_1' along these lines. This procedure (like any homotopy) takes a cycle to a homologous cycle. This being so, the homology of Q_1 is the same as that of Q_1'.

Now let l be a level line that is slightly below a_2. We denote by Q_2 the part of the surface below this line. Suppose a_2 is a saddle point. Except for a neighborhood of a_2 we can pull down Q_1' to Q_2 along lines of steepest descent (see Figure 200d). Now we can contract the remaining strip to a 1-cell τ_2 glued to Q_2 (Figure 200e). This deformation does not change the homology. It follows that the initial surface Q has the same homology as the figure obtained from Q_2 by gluing to it the 1-cell τ_2 (corresponding to the saddle point) and then the 2-cell τ_1 (corresponding to the maximum).

Next we pull Q_2 down to a part Q_2' lying below a level line that is somewhat higher than a_3. If a_3 is a minimum point, then the part F' of the surface that remains near a_3 (a disk) has the same homology as a point, i.e., as a 0-cell τ_3 (Figure 200f). Let Q_3^* be the figure obtained from Q_3 by removing F'. Then we see that the initial surface Q has the same homology as the figure obtained from Q_3^* by adding to it the 0-cell τ_3, the 1-cell τ_2, and the 2-cell τ_1.

Finally we come down to the last minimum point a_q (Figure 200g). If we proceed in reverse order then we see that we obtain a figure with the same homology as

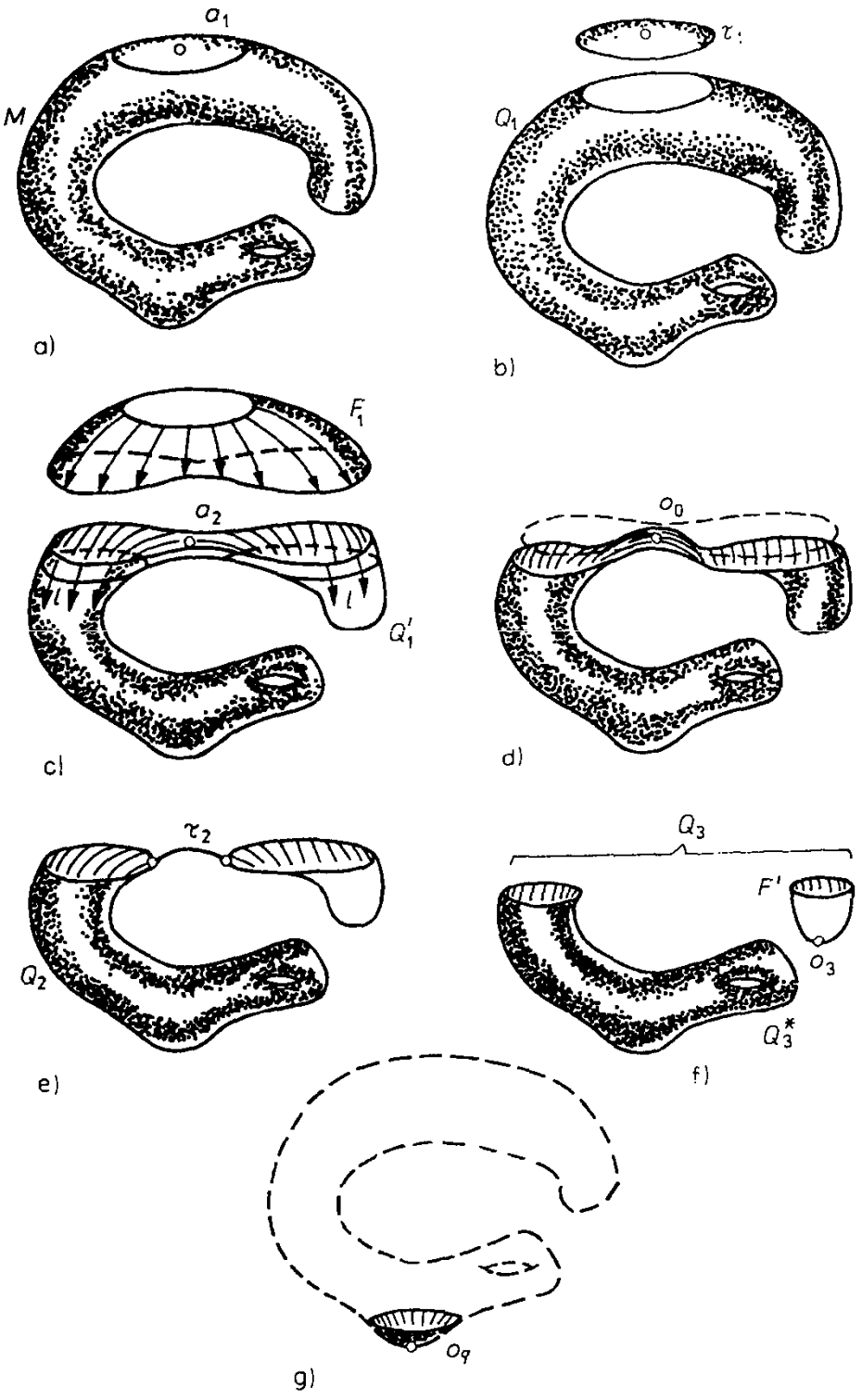


FIGURE 200.

Q by successive addition of cells, namely, of a 0-cell for each minimum point, of a 1-cell for each saddle point, and of a 2-cell for each maximum point. In other words, *Q has the same homology as a polyhedron with C_0 0-cells, C_1 1-cells, and C_2 2-cells.* This implies the validity of the inequality (23) (see Problem 194).

Problems

212. Show that formulas (22) and (23) remain valid if we take the Betti numbers modulo a prime p.

213. Show that $C_r \geq p_r(Q), r = 0, 1, 2$.

We conclude with the observation that functions can be considered on *"multidimensional surfaces"* (manifolds). An n-dimensional "surface" is a figure each of whose points has a neighborhood homeomorphic to an n-dimensional open ball. (In Problems 190 and 192 and in Examples 57 and 58 we considered three-dimensional "surfaces.") On an n-dimensional "surface" Q there are $n + 1$ types of nondegenerate stationary points (on the right-hand side of formulas like those in (21) there can appear $0, 1, \ldots, n$ minus signs). Morse's theorem (and the arguments we used to prove it) remains in force here. For example, the inequality (23) goes over into the inequality

$$\sum_{k=0}^{r}(-1)^{r-k}C_k \geq \sum_{k=0}^{r}(-1)^{r-k}p_k(Q), \quad r = 0, 1, \ldots, n.$$

A

Topological Objects in Nematic Liquid Crystals, V.P. Mineev

Many mathematical concepts, and even whole theories, exist for many years without being applied outside mathematics. For example, it took centuries to clarify the concept of a complex number and to use these numbers extensively in physics and technology. A more recent example is topology. In the last decade [this was written close to twenty years ago (tr.)] a number of problems arose in unrelated areas of physics whose adequate formulation and subsequent solution involved the language of topology. At the same time, this brought about significant advances in the affected areas of physics.

One striking illustration of this point comes from the biophysics of polymers, which studies huge protein molecules and nucleic acids. If we investigate the spatial disposition of such a molecule, then we encounter restrictions of a topological nature. From a purely mathematical standpoint, a long, closed molecule is a closed curve. We know that such curves form knots. Different knots cannot be deformed into one another without cutting the curve and subsequently gluing it together. In the case of polymers, adherence to the condition that curves must not be cut is secured by the fact that cutting a polymer chain requires breaking the chemical bond at an appropriate point of the chain, which requires considerable amounts of energy. This means that, at low enough temperatures, the probability of a break is small, and the molecules of the polymer can exist indefinitely in a particular knot configuration. The important question of what fraction of the molecules of a given length has a particular knot configuration is treated on the basis of insights from algebraic topology concerning topologically different types of knots.

In the biophysics of polymers the long molecules themselves form the topological objects—the knots. In other areas of physics we encounter objects with less direct topological pecularities. For example, in field theory we encounter particles described by vector fields with topological characteristics. In solid state physics the stability of certain defects in well-ordered materials, such as ordinary and liquid crystals, superconductors, plasma, and ferromagnets, turned out to be of a topological nature. In the present essay we will talk about the simplest materials that exhibit defects whose stability is of a topological nature. These are the nematic liquid crystals, often simply called nematics. The necessary mathematical concepts, such as index of a vector field, fundamental group, degree of a mapping, etc., are merely sketched in this esssay. The precise definitions and explanations are found in the body of the book.

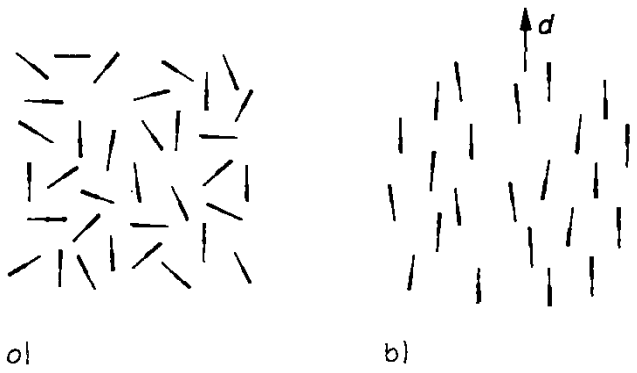


FIGURE 1.

A.1. Nematics

Nematic liquid crystals are long molecules whose interactions tend to arrange the molecules in parallel. At high temperatures this tendency is blocked by motion due to heat, and the material behaves like an ordinary liquid (Figure 1a). Below a certain critical temperature (of a few tens of degrees) there arises in the liquid a definite preferred direction of orientation of the axes of the molecules. But the distribution of the centers of gravity of the molecules of a nematic liquid is as chaotic as in an ordinary liquid. The slight departures of the axes from parallelism are due to heat oscillations. Mathematically, the direction of the dominant orientation is described by a unit vector **d** called the *director*. The special name of this vector reflects the fact that while the ends of the long molecules are different, their positions are unordered (Figure 1b). The nematic states with oppositely directed vectors (**d** and **-d**) are not physically distinguishable. In other words, the vectors must be regarded as short marks that determine directions rather than as oriented arrows.

Conditioned as it is by the container walls and by the external (e.g., magnetic) field, the nematic state is always inhomogeneous. This means that the direction of the director **d** changes gradually from point to point. The distribution of **d** in space is called the *vector field* of the vector **d**. (We call the reader's attention to the fact that, since the director is not oriented, we must not, strictly speaking, use formally the terminology of vector fields and must modify statements involving vector fields accordingly.)

A.2. Disclination in the Nematic

Because of its strong diffusion of light, a nematic liquid crystal has the appearance of an opaque, nontransparent liquid. If we look at it under a microscope, then we can see long, thin threads drifting in the liquid. Nematic liquid crystals owe their name to these threads ($\nu\eta\mu\alpha$ is Greek for thread). Already at the beginning of our century scientists conjectured what is now an established fact: the threads are not foreign disseminations but reflections of peculiarities of the disposition of the molecules.

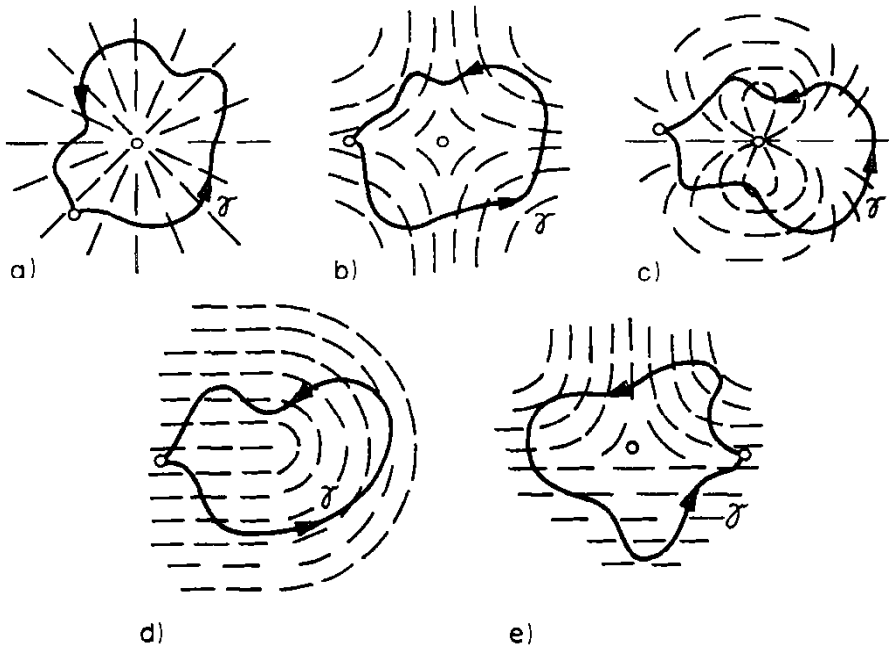


FIGURE 2.

The direction field of the director **d** may include curves on which the direction of **d** is undetermined (discontinuous). Such a distribution of **d** is easiest to represent in the case of a plane vector field, i.e., a vector field in space all of whose vectors are parallel to a plane (see Figure 2, in which the field of the director is indicated by dashed lines). We know that the singularities of a plane vector field are characterized by their indices, where the index of a singular point is defined as the number v of complete revolutions of **d** in the positive direction when one goes around it on a closed curve γ. For example, the indices of the singular points in Figures 2a, 2b, and 2c are 1, -1, and 2 respectively. However, we recall that states that differ only in the sign of **d** are indistinguishable. Therefore, there can be singular points such that if one goes around them on a closed contour γ, then the vector **d** executes a half-integral number of rotations. Thus the index of the singular point in Figure 2d is 1/2 and the index of the singular point in Figure 2e is $-1/2$. The singular points in Figure 2 are traces on the plane of the image of singular curves in the direction field of **d**. [This remark must be kept in mind when we look at many of the drawings beginning with Figure 2 (translator).] If we look at the singular curve in Figure 2a not from above but from the side, then the distribution of **d** looks like the representation in Figure 3a. Following the suggestion of the English physicist Frank, the breakline in the direction field of the director is called a *disclination.*

Since the mutual interaction of the molecules tends to align them in parallel, singular curves in the distribution of **d** are undesirable from the point of view of energy. It follows that in a nematic liquid there are bound to arise deformations of the distribution of **d** that tend to eliminate the singularities and to shift the distribution to a homogeneous state with the lowest energy level. This is illustrated by the deformations shown in Figures 3a, b, c, which eliminate the disclination

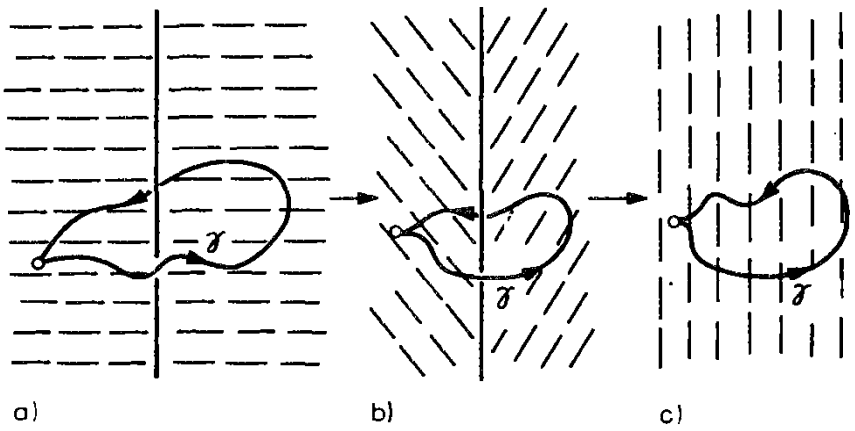


FIGURE 3.

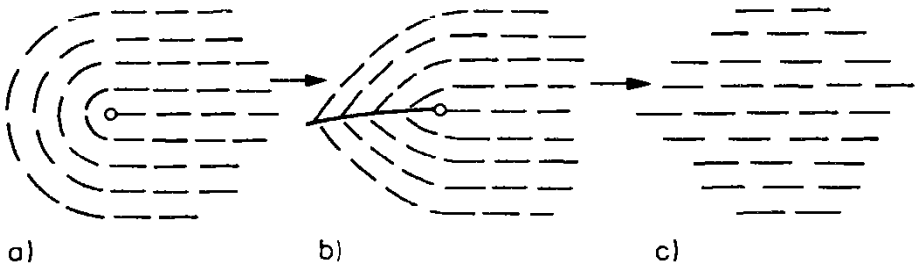


FIGURE 4.

shown in Figure 3a and shift the distribution to a homogeneous state without singularities. This deformation of the field **d**, reminiscent of the closing of an umbrella, has been called "escape to the third dimension." The name reflects the change in the direction of **d** from horizontal in Figure 2a to vertical in Figure 3c. Thus the disclination shown in Figures 2a and 3a is unstable with respect to escape to the third dimension. It is natural to ask whether other disclinations are stable. How should a test be formulated so that one can tell a stable disclination from an unstable one?

Of course any disclination can be eliminated by creating a break in the director field, as shown in the sequence of Figures 4a, b, c. We see, however, that in the vicinity of the break the molecules are not parallel as in the nematic but form angles as in an ordinary liquid. Thus forming such a break is equivalent to dissolving the nematic order in the half-plane that rests on the singular curve. This calls for a great expenditure of energy. In other words, this process has a huge energy barrier. It follows that when investigating stability of disclinations we must limit ourselves from the outset to continuous deformations of the field **d**. If this is done, then topology can be of use.

A.3. Disclination and Topology

In a region filled with a nematic liquid the vectors have a certain distribution, i.e., with each point **r** in the region there is associated a vector **d**, and this determines a vector field **d**(**r**). We translate the vectors **d** in the region so that they all begin at the same point (Figures 5a and 5b). Then their tips lie on a unit sphere. Thus the vector field **d**(**r**) effects a mapping of the vectors in our region on a unit sphere.

The sphere with the tips of the **d** vectors is different from an ordinary sphere. After all, **d** is the director vector, and the states **d** and **-d** are physically indistinguishable, i.e., the diametrically opposite points on the sphere are equivalent, or—as we say in topology—the diametrically opposite points on the sphere are glued together. Such a sphere is called a *projective plane* and is denoted by RP^2. The result of gluing together the diametrically opposite points on a sphere cannot be represented in space but such a representation is not needed. It suffices to note that **d** and **-d** represent the same point. Thus the vector field **d**(**r**) yields a mapping of the points r of coordinate space on the points of the projective plane RP^2. We will now investigate the relation between RP^2 and the stability of disclinations in the nematic.

Let L be a disclination curve in the vector field **d**(**r**), i.e., a curve on which the vector field **d**(**r**) has a discontinuity (Figure 6a). We draw around it a closed contour γ. Each point **r** on this contour has an image **d**(**r**) on the surface RP^2 and

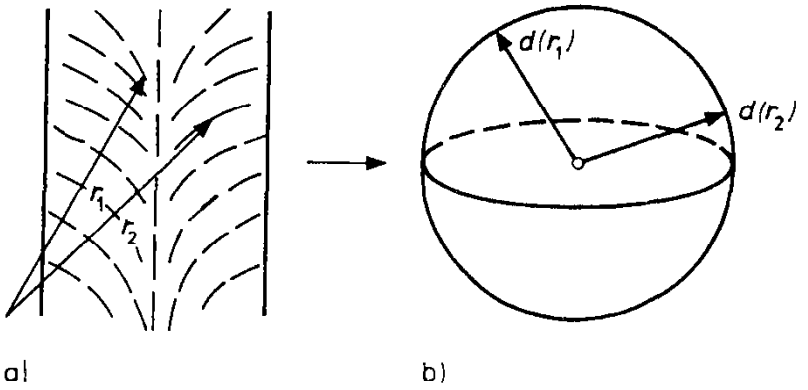


FIGURE 5.

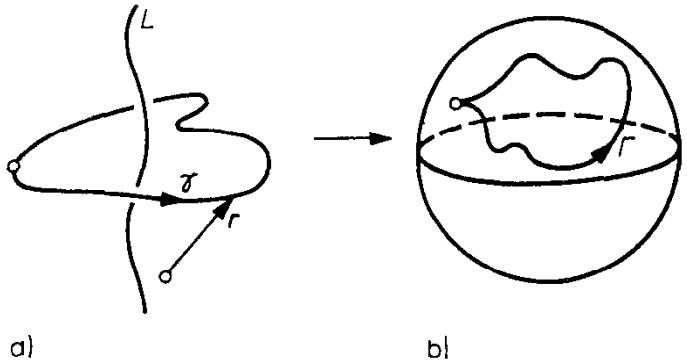


FIGURE 6.

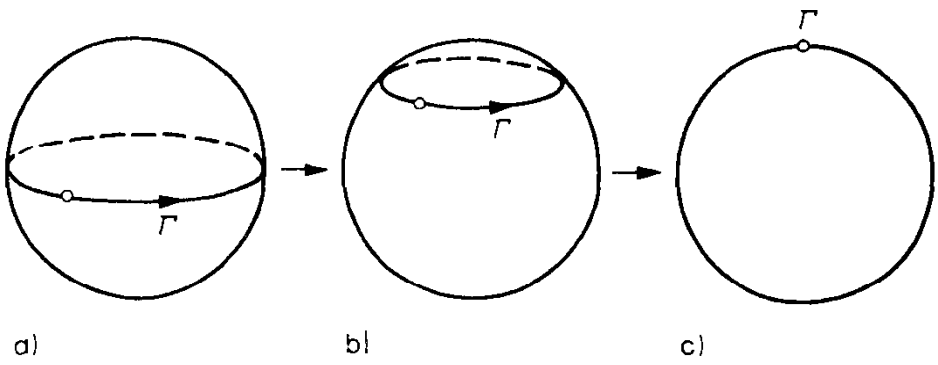


FIGURE 7.

the image of the closed contour γ itself is a closed contour Γ on RP^2 (Figure 6b). Clearly, to each continuous deformation of the field $\mathbf{d}(\mathbf{r})$ near γ there corresponds a continuous deformation of the contour Γ on RP^2. For example, the deformation of the distribution $\mathbf{d}(\mathbf{r})$ that we called an escape to the third dimension results in the contraction of Γ on RP^2 to a point (Figures 7a, b, c).

Quite generally, if an (unstable) disclination can be eliminated, then the corresponding contour Γ on RP^2 can be contracted to a point. We will denote the class of these contours and the corresponding disclinations by Γ_0. It is easy to see that in the case of plane fields this class includes all disclinations with integral index v of the vector field $\mathbf{d}(\mathbf{r})$ (Figures 2a, b, c). All contours in Γ_0, and thus the fields of disclinations in this class, can be continuously deformed into one another.

On the other hand, there are in the nematic disclinations (for example, those in Figures 2d to 2e) such that the images of the contours γ that circle them are contours of type $\Gamma_{1/2}$ that link diametrically opposite points on the sphere (Figure 8). We know that such points are equivalent, i.e., the contours $\Gamma_{1/2}$ are closed! Unlike a contour of type Γ_0, a contour of type $\Gamma_{1/2}$ cannot be contracted to a point on RP^2. In this sense, contours of type $\Gamma_{1/2}$ resemble contours on an annulus that circle the hole at its center. We can easily imagine this situation if we deform a sphere and glue together just two diametrically opposite points A, the beginning and end of a contour $\Gamma_{1/2}$.

While contours of type $\Gamma_{1/2}$ cannot be contracted to a point, they can be deformed into one another. The disclinations in Figures 2d and 2e are stable. While it is impossible to make the corresponding distributions of the field $\mathbf{d}(\mathbf{r})$ into homogeneous distributions of $\mathbf{d}(\mathbf{r})$ by continuous deformations, we can easily map them on one another. The reader should try to deform the field in Figure 2d into the field in Figure 2e. A preliminary step for this is to deform into one another the corresponding contours $\Gamma_{1/2}$ shown in Figure 8.

We emphasize that the existence of closed contours $\Gamma_{1/2}$ that are not contractible to a point, and thus of topologically stable disclinations, is solely the consequence of the equivalence of diametrically opposite points $\mathbf{d}$ and $-\mathbf{d}$. On an ordinary sphere it is possible to contract any closed contour to a point. Hence, if $\mathbf{d}$ and $-\mathbf{d}$ were different, there would be no stable singular curves in materials such as, for

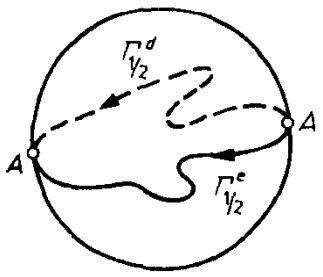


FIGURE 8.

example, isotropic ferromagnets. A sphere and RP^2 are locally equivalent, but they have different global topological properties.

Thus from a topological standpoint there are two types of disclination curves in nematic liquid crystals. On the projective plane they are characterized, respectively, by contours of type Γ_0 that are contractible to a point and by closed contours of type $\Gamma_{1/2}$ that are not contractible to a point. Stable curves of type $\Gamma_{1/2}$ cannot terminate in the interior of the region occupied by the nematic liquid. They are either closed or they extend to the boundary of the liquid. We give an indirect proof of this assertion. Suppose that a curve of type $\Gamma_{1/2}$ terminated in the interior of the liquid. Then the contour γ could be removed from around it and contracted to a point. But then the image of γ could also be contracted to a point on the projective plane. But this is impossible for contours of type $\Gamma_{1/2}$. On the other hand, there are no topological obstructions to unstable curves terminating in the interior of the liquid. However, it is energetically advantageous for segments of singular curves to shorten their length and they either vanish or, if they are attached to the surface at one end, they contract and turn into a singular point on the surface.

Disclination curves interact. If they attract one another, then they can merge into one. What is the result of such a merger? Do we get out of two stable curves another stable curve, or is the resulting curve unstable so that it vanishes, i.e., the initial curves are annihilated?

Topology answers this question. The image of the contour $\gamma = \gamma_1 + \gamma_2$, which goes around two curves of type $\Gamma_{1/2}$ at one time, is a contour of type $\Gamma_{1/2}$ traversed twice; we can speak of the product $\Gamma_{1/2} \cdot \Gamma_{1/2}$. Figure 9b shows that such a contour is equivalent to a contour Γ_0 and can therefore be contracted to a point. This means that when two stable disclinations combine into one, the result is an unstable disclination of type Γ_0, i.e., the two are annihilated. To see that this is so we need only deform the distribution $\mathbf{d}(\mathbf{r})$ in Figure 9a into the distribution in Figure 2a. On the other hand, the result of the merger of a stable and an unstable disclination is always a stable disclination; we can write briefly $\Gamma_{1/2} \cdot \Gamma_0 = \Gamma_0 \cdot \Gamma_{1/2} = \Gamma_{1/2}$.

The multiplication rule just formulated states that the set of classes of contours on RP^2 consisting of two elements Γ_0 and $\Gamma_{1/2}$ forms a group $\pi_1(RP^2)$, the *fundamental group* of the projective plane. The multiplication of the elements in this group can be replaced by the addition of the indices of the contours subject to

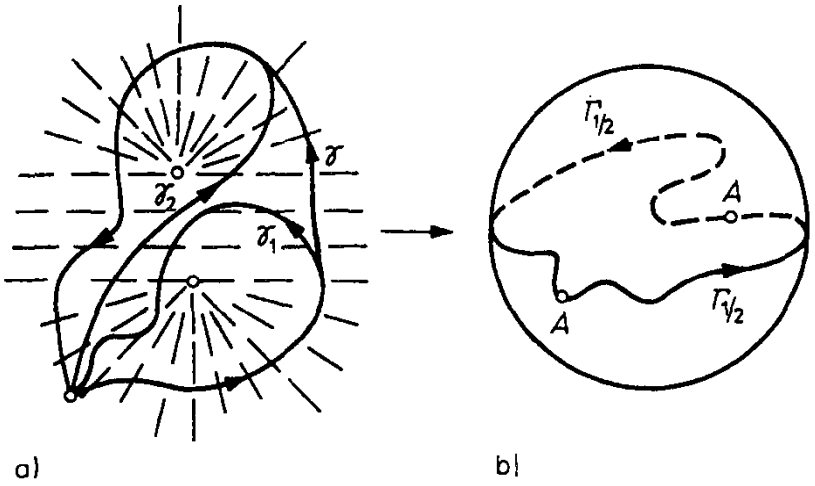


FIGURE 9.

the condition that all integers are equivalent to 0:

$$\Gamma_{1/2} \cdot \Gamma_{1/2} = \Gamma_0 \Rightarrow \frac{1}{2} + \frac{1}{2} = 0;$$

$$\Gamma_{1/2} \cdot \Gamma_0 = \Gamma_{1/2} \Rightarrow \frac{1}{2} + 0 = \frac{1}{2}.$$

A.4. Singular Points

In the region occupied by a nematic liquid there can be, in addition to singular curves of the vector field $\mathbf{d}(\mathbf{r})$, singular points—discontinuities, of the field $\mathbf{d}(\mathbf{r})$. The simplest example of this type is that of a point at which the direction $\mathbf{d}$ of the vector field coincides with the direction of the radius vector $\mathbf{d}(\mathbf{r}) = \mathbf{r}/|\mathbf{r}|$. The vectors issuing from such a singular point resemble the spines of a hedgehog rolled up in a ball, hence the name *hedgehog* for such a singular point. Is a hedgehog stable? In other words, can one use a continuous deformation of the field $\mathbf{d}(\mathbf{r})$ to remove this singular point and to turn the field $\mathbf{d}(\mathbf{r})$ into a homogeneous field? To answer this question we surround the singular point by a sphere σ. The image of σ on RP^2 is all of RP^2 traversed once. In this way, the field $\mathbf{d}(\mathbf{r})$ effects about the hedgehog a mapping of degree 1 of the sphere σ on RP^2. We deform continuously the field $\mathbf{d}(\mathbf{r})$. Then the image of σ, which can be thought of as a closed membrane (i.e., a membrane without boundary) spanning RP^2, will also be deformed and will form folds. However, while spanning RP^2, this membrane is not contractible to a point. But this means that the singular point of the field $\mathbf{d}(\mathbf{r})$ cannot be eliminated. The degree of the mapping is a topological invariant.

Quite generally, to investigate the stability of a singular point of the field $\mathbf{d}(\mathbf{r})$ one must surround it by a sphere σ and investigate its image on RP^2. To each stable singular point there corresponds a membrane (the image of σ on RP^2). To unstable singular points there correspond closed membranes contractible to a point on RP^2. The only way to remove a stable singular point or a stable singular

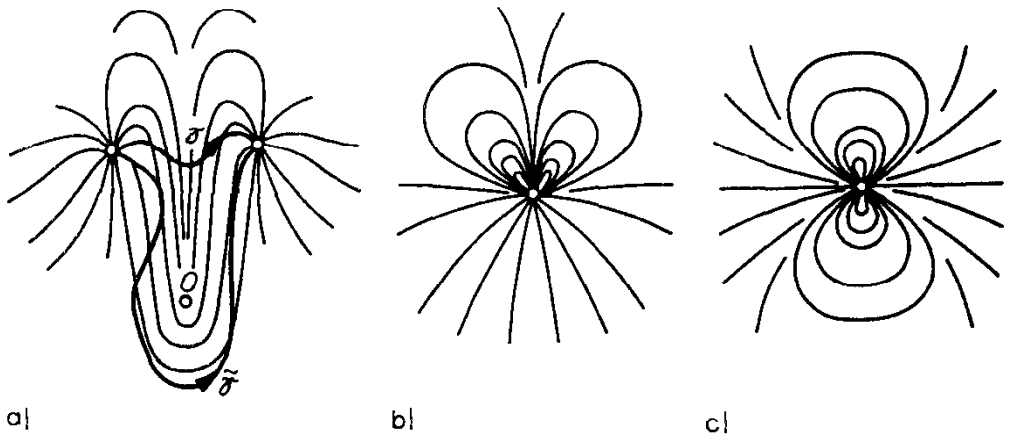


FIGURE 10.

curve is to create breaks in the field $\mathbf{d}(\mathbf{r})$. This corresponds to overcoming a huge energy barrier. We considered such a process for a singular curve (Figure 4). In distinction to the case of a singular curve, removal of a singular point requires the creation of a break in the field $\mathbf{d}(\mathbf{r})$ on a curve that issues from the field.

The degree of a mapping is an integral index. There arises the question of the difference between two hedgehogs whose degrees, say $N = 1$ and $N = -1$, have the same absolute value. The difference must lie in the directions of the spines, i.e., in the different orientations of the membranes that are the images of the sphere σ. In the first case the spines point outward and in the second case inward. But we know that opposite directions on $\mathbf{d}(\mathbf{r})$ are indistinguishable. Hence hedgehogs with $N = 1$ and $N = -1$ are the same singular point with $|N| = 1$. On the other hand, if we merge two hedgehogs, then it would seem that this must result in the addition of their indices. Hence if we merge two hedgehogs with indices $|N_1| = 1$ and $|N_2| = 1$, then we should presumably obtain a hedgehog of index 2 or a hedgehog of index 0, i.e., a removable (unstable) singular point. Such behavior seems unlikely. But what *does* happen in this case?

The actual process of merger of the hedgehogs yields a result that depends on the choice of path leading to the merger. This is due to the nontrivial nature of the fundamental group $\pi_1(RP^2)$ of a nematic liquid crystal. In topology we speak of the *influence* of π_1. For example, we can carry out the merger of two singular points with $|N_1| = 1$ and $|N_2| = 1$ along two paths γ and $\tilde{\gamma}$ on different sides of a curve of stable disclination. In Figure 10a the field lines of $\mathbf{d}(\mathbf{r})$ are indicated by thin lines. The disclination curve is perpendicular to the page and is marked with the letter O. It is easy to see that the merger along γ leads to a point singularity with $|N| = 2$ shown in Figure 10b. We note that the field distribution in Figure 10b has a plane of symmetry perpendicular to the page but no axis of symmetry. The merger along $\tilde{\gamma}$ leads to a point singularity with $N = 0$ shown in Figure 10c. In this case, the field distribution $\mathbf{d}(\mathbf{r})$ has a horizontal axis of symmetry on the plane of the drawing. Thus the existence of a disclination in the field $\mathbf{d}(\mathbf{r})$ leads to different results when hedgehogs are merged. By this we mean that differences

arise only in the presence of topologically stable disclinations that correspond to the nontrivial elements of the fundamental group.

A.5. What Else Is There?

We have seen that the topological properties of closed contours and membranes on the projective plane enable us to analyze a number of questions connected with stability and merging of disclinations and singular points in nematic liquid crystals. In addition to helping us classify singular points of a field $\mathbf{d}(\mathbf{r})$, topology also enables us to classify stable nonsingular field configurations, such as boundaries of regions and solitons, that arise in $\mathbf{d}(\mathbf{r})$ under the influence of electric and magnetic fields.

The nematic is not unique. There is a rather large class of ordered materials, such as ordinary and liquid crystals of all kinds, ferro and antiferromagnetics, superconductors, and superliquid liquids, whose investigation is aided by topological methods.

We used the director $\mathbf{d}$ to describe topologically the singularities in nematic liquids, and the projective plane RP^2 to classify them. In other ordered materials we realize other types of fields, such as vector fields, matrix fields, and other regions D of variation of the order-describing parameters. In general, the fundamental group is not commutative. Of the materials found in nature the only ones with a noncommutative fundamental group $\pi_1(D)$ are the biaxial liquid crystals (the nematic liquid crystals investigated above were uniaxial). The noncommutativity of $\pi_1(D)$ has beautiful, not yet experimentally verified, consequences.

The most impressive application of topology began in 1972, in connection with the discovery of the superfluid phase of the light helium isotope ^{3}He. It turned out that the properties of the superfluidity of this phase are largely dictated by topology.

Bibliography

Alexandrov, P. S. *Combinatorial topology*. Vols. 1, 2, and 3. Mineola, NY: Dover Publications, 1998.

Bollobás, B. *Modern graph theory*. New York: Springer-Verlag, 1998.

Courant, R., and H. Robbins. *What is mathematics*? 2d ed. New York: Oxford University Press, 1996.

Diestel, R. *Graph theory*. 2d ed. New York: Springer-Verlag, 2000.

Fritsch, R., and G. Fritsch. *The four-color theorem*. New York: Springer-Verlag, 1998.

Gamelin, T. W., and R. E. Greene. *Introduction to topology*. Mineola, NY: Dover Publications, 1999.

Gardner, M. *My best mathematical and logic puzzles*. New York: Dover Publications, 1994.

Harary, F. *Graph theory*, Reading, PA: Addsion-Wesley, 1969.

Henle, M. *A combinatorial introduction to topology*. Mineola, NY: Dover Publications, 1994.

Hilbert, D., and S. Cohn-Vossen. *Geometry and the imagination*. New York: Chelsea Publishing, 1952.

Jänich, K. *Topology*. New York: Springer-Verlag, 1984.

Kinsey, L. C. *Topology of surfaces*. New York: Springer-Verlag, 1993.

Massey, W. S. *Algebraic topology*. New York: Springer-Verlag, 1991.

Milnor, J. W. *Topology from the differentiable viewpoint*. Princeton, NJ: Princeton University Press, 1997.

Pontryagin, L. S. *Foundations of combinatorial topology*. Mineola, NY: Dover Publications, 1999.

Prasolov, V. V., and A. B. Sossinsky. *Knots, links, braids and 3-manifolds*. Providence, RI: American Mathematical Society, 1997.

Rademacher, H., and O. Toeplitz. *The enjoyment of math*. Princeton, NJ: Princeton University Press, 1994.

Reidemeister, K. *Knot theory*. Moscow, ID: BCS Associates, 1983.

Seifert, H., and W. Threlfall. *Seifert and Threlfall: A textbook of topology*. New York and London: Academic Press, 1980.

Spanier, E. *Algebraic topology*. New York: Springer-Verlag, 1994.

Stillwell, J. *Classical topology and combinatorial group theory*. New York: Springer-Verlag, 1980.

Index

Alexander, J. W., 68, 106
Antione, L., 68
Antoine's necklace, 68
Appel, K., 61

Betti number, 111–114, 116–119
bouquet, 91
Brouwer's theorem, 60
B_K^1, 60

cap, 42
cell, 87
cell decomposition, 87–91, 107–110, 120
chain, 16
chord, 42–43
chromatic number, 62
circuit, 12
col (Q), 62
complete graph, 10
complex number, 97, 127
complex sphere, 97
connected graph, 10–11
continuity, 1, 3
contractible, 68
core, 42
country, 62
covering, 91–95
covering path, 93
current, 14
cut point, 8
cycle, 16
$\chi(G)$, 12
$\chi(Q)$, 39

decomposition into polygons, 38
degree, 9
degree of a mapping, 95–96
dimension, 25
direction field, 55
disclination, 128, 132
disconnected graph, 11
discontinuity, 131
domain, 3
double knot, 70–71
duality principle, 106
$\partial\tau$, 108

edge, 9
edge path, 89
erasable edge, 13
Euclid, 22
Euclidian geometry, 94
Euler, vii, 9–12
Euler characteristic, 12, 39
Eulerian, 9–10
Euler's theorem, 31
exterior, 20, 67

fiber, 117
fiber bundle, 117
finite graph, 9
forest, 13
four color problem, 60
Frank, C. F., 129
Frankl, F., 73
free cyclic group, 85
fundamental group, 83–84
fundamental region, 93
fundamental theorem of algebra 95–96

granny knot, 71
graph, 1, 9–10
group, 83
Guthrie, F., 61
$G(L)$, 100

Haken, W., 61
half-strip, 43
handle, 33
Heawood, P. J., 61–63
Heawood formula, 62
Heawood's theorem, 62
hedgehog, 55, 134
hedgehog theorem, 55
homeomorphism, 4
homologous, 104
homologous to zero, 104
homology group, 104–105
homotopic paths, 31
hyperbolic geometry, 94
hyperbolic plane, 94–95
$H_0(X)$, 110–113
$H_1(X)$, 111–113
$H_2(X)$, 111–113
$H_r(X)$, 111
$H_r(X, G)$, 114

index, 55
interior, 20, 67
intersection index, 15–19
inverse mapping, 4–5
isotopy, 7
isotopy invariant, 77

Jordan curve theorem, 20
    space analogue of, 67
$J(a, b)$, 16
$J(G, Q)$, 66

Kirchhoff vertex theorem, 14
Klein bottle, 49
knot, 70
knot group, 99–100
Königsberg bridge problem, 10
Kuratowski, C., 15

Leray, J., 119
Leray's theorem, 119–121
link, 73
linking number, 76
Listing, J., 33
loop, 9

manifold, 126
mapping, 3
member, 16
Menger, K., 25, 27
meridian, 7
Möbius, A. F., 33
Möbius strip, 34
monomial, 89
Morse, H. M., 123
Morse's theorem, 123–124

nematic, 128
nonorientable surface, 36
normal projection, 71
north pole, 97–98
null homotopic, 84
N_q, 52

one-sided surface, 34–35
one-to-one and onto mapping
    (bijective mapping), 4
order of a group, 97
Ore, O., 61
orientable surface, 36
oriented boundary, 19

parallel, 7
path, 28
Peano curve, 28
Poincaré, H., vii, 55–56, 85
Poincaré model, 94
Poincaré's theorem, 56–60
point at infinity, 51
polyhedron, 88
Pontryagin, L. S., 106
projective plane, 51–52
projective space, 113

P_k, 38
$\pi(X)$, 83

ramification, 33
representation, 97
Riemann sphere, 97
RP^2, 131

saddle point, 55
sailor knot, 71
sheet, 91
Sierpiński, W., 26
Sierpiński carpet, 26–28
simple arc, 23
simple closed curve, 23
simply connected, 85
singular point, 55
solid torus, 111
south pole, 97
spanning tree, 13
sphere, 36, 69, 97, 111
spherical shell, 113
stationary point, 122–123
Stemple, J., 61
strip, 42
support line, 2
surface, 33
surface with boundary, 33

tangent bundle, 118
three-sheeted covering, 91
topological invariant, 6, 8
topological product, 114
topological property, 6
topology, 6
torus, 5
trail, 11
tree, 12
trefoil knot, 70–71
two-dimensional cycle, 66
two-dimensional homology group, 111
two-sheeted covering, 91
two-sided surface, 36

universal covering, 93
Uryson, P. S., vii, 25–26

vertex, 9

Wada, 24
winding number, 86
$w(X, Y)$, 76

Youngs, J. W. T., 63

zero-dimensional figure, 26
zero-dimensional homology group, 110