MATHEMATICAL SURVEYS · *Number 11*

FIXED POINTS AND TOPOLOGICAL DEGREE IN NONLINEAR ANALYSIS

BY
JANE CRONIN

1964
AMERICAN MATHEMATICAL SOCIETY
190 Hope Street, Providence, Rhode Island

PREFACE

Since this book is an introduction to the application of certain topological methods to nonlinear differential and integral equations, it is necessarily an incomplete account of each of these subjects. Only the topology which is needed will be introduced, and only those aspects of differential and integral equations which can be profitably studied by the use of the topological methods are discussed. The bibliography is intended to be representative, not complete.

I am greatly indebted to Professor Lamberto Cesari who read the entire manuscript and made a number of valuable suggestions and corrections concerning both the mathematics and the exposition in this book.

The preparation of this book was supported by the Air Force Office of Scientific Research. I wish to express my thanks to the Air Force Office of Scientific Research and also to the editorial staff of the American Mathematical Society for their work in preparing the manuscript for publication.

Jane Cronin.

FOREWORD

Our aim is to give a detailed description of fixed point theory and topological degree theory starting from elementary considerations and to explain how this theory is used in the study of nonlinear differential equations, ordinary and partial. The reader is expected to have a fair knowledge of advanced calculus, especially the point set theory of Euclidean n-space, but no further knowledge of topology is assumed. (Following this introduction is a list of definitions and notations from advanced calculus which will be used in the text.)

Why topological methods are used. It is easy to see that the solution $x(t)$ of the differential equation

$$\frac{d^2x}{dt^2} + x = 0,$$

such that $x(0) = 0$ and $dx(0)/dt = 1$, is $x = \sin t$. In finding this solution, we have obtained complete quantitative information, i.e., for any given value of t, we can find the corresponding value $x(t)$ by referring to a table of trigonometric functions. However, we have also obtained other useful information called qualitative or "in the large" information: we know that the values $x(t)$ are between -1 and $+1$ and that $x(t)$ has period 2π.

If all differential equations could be solved as easily, it would be unnecessary to introduce the distinction between quantitative and qualitative information concerning solutions. But this example is an exceptionally simple one. Most differential equations, especially nonlinear equations, must be studied with one technique to obtain quantitative information and by another to obtain qualitative information. If a method can be derived for finding the numerical values of the solution corresponding to given values of the independent variable, this method will usually not give us qualitative information. For example, to determine if the solution is periodic requires a different approach.

The topological methods we describe will give us qualitative information, e.g., information about the existence and stability of periodic solutions of ordinary differential equations and the existence of solutions of certain partial differential equations. However, they will gives no quantitative information, i.e., no information about how to compute any of the solution values. At best, we will get upper and lower bounds for solution values. The information that the topological methods give us is thus incomplete.

Nevertheless, topological methods are used because at present there are no other methods that yield as much qualitative information with so little effort. It is possible to envision a future in which topological methods will be supplanted by more efficient methods, but so far there is little encouragement for doing so. Such a venerable technique as the Poincaré-Bendixson Theorem has not been essentially improved in over sixty years although it is widely used. (The Poincaré-Bendixson Theorem is not one of the techniques which we will describe but as will be seen in Chapter II, it is intimately related to one of these techniques.)

The topological techniques to be developed. The two techniques to be developed, the fixed point theorem and local topological degree, are closely connected. The fixed point theorem has the advantage of being a comparatively elementary theorem (it can be proved without using any "topological machinery") which has many useful applications. The topological degree theory requires lengthier considerations for its development, but it has an important advantage over the fixed point theorem: it gives information about the number of distinct solutions, continuous families of solutions, and stability of solutions.

When to use topological techniques. We shall mostly be concerned with the question of *how* to apply topological techniques. The question of *when* to apply them is equally important. There can be no precise answer to the question, but we can formulate a rough rule. If we use an analytical method (like successive approximations), we establish existence and uniqueness of solution and a method (not necessarily practical) for computing the solution. If analytical means fail and especially if there seems to be no way to establish uniqueness, then we turn to the weaker question of establishing mere existence. For answering this weaker question, the topological methods (cruder and yielding less information than analytical methods) sometimes suffice. Thus topological methods should be regarded as a last resort or at least a later resort than analytical methods.

Summary of contents. In Chapter I, a definition of the local topological degree in Euclidean n-space is given, the basic properties of topological degree are derived, and some methods for computing the degree are described. Also the Brouwer Theorem (the fixed point theorem in Euclidean n-space) is obtained. In Chapter II, the techniques described in Chapter I are applied to some problems in ordinary differential equations: existence and stability of periodic and almost periodic solutions. In Chapter III, the Euclidean n-space techniques developed in Chapter I are extended to spaces of arbitrary dimension. We obtain the Leray-Schauder degree and the Schauder and Banach fixed point theorems for mappings in Banach space. We obtain also a combination of analytical and topological techniques which can be used to study local problems in Banach space. Finally in Chapter IV, the theory developed in Chapter III is applied to integral equations, partial differential equations and to some further problems on periodic solutions of ordinary differential equations.

The applications in Chapters II and IV are treated in varying detail. Existence of periodic solutions of nonautonomous ordinary differential equations is treated in complete detail. Stability of periodic solutions is treated in full detail except for the basic stability theorem of Lyapunov which is stated without proof. The other topics in Chapter II are similarly treated: certain theorems, particularly those from other disciplines, are stated without proof. In Chapter IV, an elaborate apparatus of theorems from analysis must be used in applying the Leray-Schauder theory and the Schauder Fixed Point Theorem. Some of these theorems have lengthy and complicated proofs and we restrict ourselves to giving references for these proofs.

The numbering of definitions, theorems, etc., is done independently in each chapter. Unless otherwise stated, references to a numbered item means that item in the chapter in which the reference is made. E.g., if in Chapter II, reference is made to Theorem (3.8), that means Theorem (3.8) in Chapter II.

Some Terminology and Notation Used in This Text

▌ denotes end of proof.

nasc is abbreviation for necessary and sufficient condition.

Set notation

$\in$: element of
$\notin$: not an element of
$\bigcup$: union
$\bigcap$: intersection
$-$: difference
$\emptyset$: null set
$\subset$: contained in
A^c: complement of set A
$A \times B$: Cartesian product of sets A and B, i.e., the set of all ordered pairs (a, b) where $a \in A$, $b \in B$.

The set of elements having property P is denoted by:

$$[x \,/\, x \text{ has property } P].$$

If a, b are real numbers such that $a < b$, the following notation is used to indicate the various intervals:

$$[a, b] = [x \text{ real} \,/\, a \leqq x \leqq b],$$
$$[a, b) = [x \text{ real} \,/\, a \leqq x < b],$$
$$(a, b] = [x \text{ real} \,/\, a < x \leqq b],$$
$$(a, b) = [x \text{ real} \,/\, a < x < b].$$

$[a, b]$ is called a closed interval and (a, b) is called an open interval. R^n denotes real Euclidean n-space, i.e., the collection of n-tuples of real numbers

$(x_1, \cdots, x_n)$. The elements of R^n will be denoted by single letters, $p, q, \cdots$ when this is possible. If $p = (p_1, \cdots, p_n)$ and λ is a real number, then

$$\lambda p = (\lambda p_1, \cdots, \lambda p_n).$$

In particular,

$$(-1)p = (-p_1, \cdots, -p_n).$$

If $p = (p_1, \cdots, p_n)$ and $q = (q_1, \cdots, q_n)$, then

$$p + q = (p_1 + q_1, \cdots, p_n + q_n).$$

The *distance between points p and q*, denoted by $|p - q|$, is:

$$|p - q| = [(p_1 - q_1)^2 + \cdots + (p_n - q_n)^2]^{1/2}.$$

If $p \in R^n$, the neighborhoods of p are the sets:

$$N_\varepsilon(p) = [q \,/\, |p - q| < \varepsilon]$$

where ε is an arbitrary positive number. A set O in R^n is *open* if for each point $p \in O$, there is a neighborhood $N_\varepsilon(p)$ of p such that $N_\varepsilon(p) \subset O$. A set F in R^n is *closed* if F^c is open. A point p is a *limit point* (*cluster point, accumulation point*) of a set E in R^n if each neighborhood of p contains at least one point of E distinct from p. (Point p may or may not be in E.) If $D \subset R^n$, a *boundary point* of D is a point p such that each neighborhood of p contains a point of D and a point of D^c. (A boundary point of D may or may not be in D.) The collection of boundary points of D is denoted by D'. The set $D \cup D'$ (also denoted by $\bar{D}$) is called the *closure* of D. The set $D \cup D'$ is a closed set. If there is a neighborhood $N_\varepsilon(p)$ of a point p such that $N_\varepsilon(p)$ is contained in a set E, then p is an *interior point* of E.

A *metric space* is a collection M of points $p, q, \cdots$ for which a function ρ from $M \times M$ into the non-negative real numbers is defined such that:

(1) $\rho(p, q) > 0$ if and only if $p \neq q$;
(2) $\rho(p, q) = \rho(q, p)$;
(3) $\rho(p, r) \leqq \rho(p, q) + \rho(q, r)$.

All the concepts defined for R^n, i.e., neighborhood, open set, closed set, etc., may be defined for a metric space by using the same definitions given for R^n only with $|p - q|$ replaced by $\rho(p, q)$.

A *separable metric* space is a metric space M such that there is a denumerable subset $[x_n]$ of M such that the closure of $[x_n]$ is M.

A *connected* set in a metric space M is a set S such that S is not the union of two disjoint nonempty sets A and B which are contained in disjoint open sets.

A *component* of an open set O in a metric space is a maximal connected subset of O.

A *locally connected* metric space is a metric space M such that if $p \in M$ and $N_\varepsilon(p)$ is a neighborhood of p then there is a connected neighborhood of p which is contained in $N_\varepsilon(p)$.

If A is a subset of a metric space M, the *diameter* of A is $\text{lub}_{a,b \in A}\, \rho(a, b)$.

If A, B are disjoint subsets of a metric space M, then the *distance between A and B* is

$$\underset{a \in A;\, b \in B}{\text{glb}} \quad \underset{(a,b) \in A \times B}{\rho(a, b)}$$

and is denoted by $d(A, B)$.

A function f from a metric space M_1 into a metric space M_2 is 1-1 if p, $q \in M_1$ and $p \neq q$ imply $f(p) \neq f(q)$.

Function f from metric space M_1 into metric space M_2 is *continuous* if for each open set O in M_2, the set $f^{-1}(O)$, where f^{-1} is the inverse of f, is also open.

If f is a 1-1 continuous function from M_1 into M_2 and if f^{-1} is continuous, then f is a *homeomorphism* from M_1 onto $f(M_1)$.

If f is a function from metric space M_1 into metric space M_2 and if A is a subset of M_1, then f/A denotes the function f regarded only on A, i.e., the function with domain A such that if $x \in A$, then the functional value is $f(x)$.

TABLE OF CONTENTS

CHAPTER I

Topological Techniques in Euclidean *n*-space

Introduction. The two topological concepts we use are the fixed point theorem and local topological degree (hereafter to be termed local degree or degree). The fixed point theorem is couched in simple terms and we will be able to state it with practically no introduction. But describing the local degree theory is lengthier. The description to be given (which is essentially the singular homology definition in Alexandroff-Hopf [1]) was chosen because it includes the "order" viewpoint and the "covering number" viewpoint and therefore seems to be the most suggestive of ways to compute the local degree. (For use in applications, we will have to make such computations.)

Nevertheless we describe briefly other methods of defining local degree. The earliest version is the definition in terms of the Kronecker integral, i.e., the degree is defined to be equal to a certain integral and the standard properties of the local degree are then proved. (See Alexandroff-Hopf [1, pp. 465–467] for discussion and references.) This definition holds only for differentiable mappings and does not give much basis for computing the degree. (Computing the integral itself is generally difficult.) Also we cannot obtain from this definition a theorem relating the degree and the number of solutions of a corresponding equation. Such a theorem is of considerable importance in applications. Another definition based entirely upon real analysis is given by Nagumo [1]. This definition is not long but would require some extension if it were to make a satisfactory basis for developing methods of computing the degree. A definition of degree in terms of cohomology is given by Rado and Reichelderfer [1]. From the point of view of a topologist, this is a more desirable definition than the one we give. It has, however, the disadvantages that it does not have as clear a geometric meaning as the Alexandroff-Hopf definition that we use (see Rado and Reichelderfer [1, p. 120, footnote 1]) and also it yields fewer suggestions for computing the degree. If the degree is defined only for mappings from the plane into itself, a much shorter definition can be given (see Alexandroff-Hopf [1, p. 464]). But it is important for later applications that our definition be applicable to mappings in Euclidean space of arbitrary finite dimension.

1. **The fixed point theorem.** The fixed point theorem says that if f is a continuous mapping of a solid sphere into itself, then f takes at least one point into itself, i.e., f leaves at least one point fixed. In the precise statement of the theorem which follows, we include a slightly wider class of sets than spheres.

DEFINITION. A *topological mapping* g of a set $E \subset R^n$ into R^n is a 1-1 continuous mapping such that g^{-1} is also continuous.

NOTATION. Let σ^n denote the solid unit sphere in R^n, i.e.,

$$\sigma^n = [(x_1, \cdots, x_n) \,/\, x_1^2 + \cdots + x_n^2 \leqq 1].$$

(1.1) BROUWER FIXED POINT THEOREM. *Let $B^n = g(\sigma^n)$ where g is a topological mapping. Let f be a continuous mapping of B^n into itself. Then there is an element x of B^n such that $x = f(x)$, i.e., the mapping f has a fixed point.*

This theorem is intuitively reasonable in the sense that any mapping f that one considers clearly does have such a fixed point. However, this observation is far from a proof of the theorem.

We will postpone proving the theorem until we have defined the local degree. The comparatively sophisticated degree theory will make possible a very short proof of the fixed point theorem. There are elementary proofs of the theorem, i.e., proofs which require few facts about topology (see Alexandroff-Hopf [**1**, p. 376 ff.]). As might be expected, an elementary proof is somewhat longer.

The fixed point theorem illustrates some typical traits of qualitative techniques. In terms of analysis, the theorem tells us that under certain circumstances the equation in R^n,

$$x - f(x) = 0, \tag{1.2}$$

has at least one solution. The theorem has two disadvantages: first it gives no information about how to find (i.e., compute) a solution of (1.2); secondly, it gives no information about how many solutions (1.2) has beyond the statement that (1.2) has at least one solution. Equation (1.2) may have just one solution or it may have an infinite set of solutions. For example: if $B^n = \sigma^n$ and f is the identity mapping, then every point $x \in \sigma^n$ is a solution of $x - f(x) = 0$; if $B^n = \sigma^2$ and f is a rotation of π radians, the only solution of (1.2) is (0, 0).

The first disadvantage, that no method for computing the solution is given, is inherent in the qualitative approach. A qualitative method usually establishes only the existence of a solution. When a qualitative method is used, we must expect to regard the computation of the solution as a separate problem. The second disadvantage, that no indication of the number of solutions is given, will be remedied when we have developed the local degree. At the cost of developing some topological "machinery," we will obtain estimates on the number of solutions.

2. **The order of a point relative to a cycle.** From the viewpoint of the analyst who wishes to apply degree theory, the local degree is a kind of estimate of the number of points mapped into a given point by a given mapping. If f is a continuous mapping from Euclidean n-space R^n into R^n, then the degree of f at point $p \in R^n$ is to be an estimate of the number of

points mapped by f into point p (these points are called p-points). We will require that this estimate remain unchanged or invariant if the mapping f or the point p is varied slightly. (As will be seen later, this condition is of crucial importance in applications of degree in analysis.) This important requirement of invariance unfortunately excludes the possibility of making the definition of degree the simplest possible one, i.e., defining the degree of f at point p as equal to the number of p-points of f. For suppose f is the mapping of R^1 into R^1 defined by

$$f: x \to x'$$

where $x' = x^2$. Then, if the degree were simply the number of p-points, the degree of f at 0 would be one. However, if our mapping f were changed to

$$f_\varepsilon: x \to x''$$

where $x'' = x^2 + \varepsilon$ and ε is a small positive number, then the degree of f at 0 would be zero no matter how small ε were chosen. If ε were a negative number, no matter how close to zero, the degree of f_ε at 0 would be two. Thus our requirement of invariance of the degree under small changes of f could not be satisfied.

To remedy this, we count the points mapped by f into p in a special way. Each p-point is counted with a plus or minus sign depending on whether the mapping f maps the points near the p-point so that directions are preserved or reversed. For example, if f is the mapping from R^1 into R^1:

$$f: x \to x^2 - \varepsilon$$

where ε is a positive number, the mapping f takes the interval $[0, \delta]$ where $\delta^2 > \varepsilon$ onto the interval $[-\varepsilon, \delta^2 - \varepsilon]$ without changing directions on $[0, \delta]$. On the other hand, the interval $[-\delta, 0]$ which is also mapped onto $[-\varepsilon, \delta^2 - \varepsilon]$ is "flipped over" in the process of being mapped. Roughly speaking, its direction is reversed. The 0-point $\sqrt{\varepsilon}$ is counted with a plus sign and the 0-point $-\sqrt{\varepsilon}$ is counted with a negative sign; hence we say that the degree of f at 0 is zero. With this definition, the degree is a crude estimate of the number of p-points—crude in that if the degree is nonzero, then there exists at least one p-point but if the degree is zero then there may be p-points (as in the example described) or there may be none at all.

The preceding is clearly far from a precise definition. We have not even said exactly what is meant by "flipping over" or reversing directions, much less given any indication of how this is done in an n-space R^n where $n > 1$.

Our first purpose is to show that the rough description of degree given above can be made into a precise definition for continuous mappings in Euclidean n-spaces. Then we define exactly what is to be meant by changing the mapping continuously, and prove that the degree is invariant under such changes. This fairly lengthy procedure will occupy the next five sections of Chapter I.

In order to define the degree of a mapping at a point, we will need some "combinatorial" concepts. The purpose of introducing these concepts is to make possible the definition of the order of a point relative to a cycle. This notion of order can be regarded as the simplest version of the local degree.

CELLS, CHAINS, AND CYCLES.

DEFINITION. A *convex* set $E \subset R^n$ is a set with the property: if $a, b \in E$ then all points $\lambda a + \mu b$ where $0 \leqq \lambda \leqq 1$, $0 \leqq \mu \leqq 1$, and $\lambda + \mu = 1$, are contained in E.

REMARK. It follows from the definition that if $\{C_\nu\}$ is a collection (finite or infinite) of convex sets, then $\bigcap_\nu C_\nu$ is convex.

DEFINITION. Let $a_0, a_1, \cdots, a_q$ be a finite set of distinct points in R^n. The *convex hull* of $a_0, a_1, \cdots, a_q$ is the convex set which is the intersection of all the convex sets which contain $a_0, a_1, \cdots, a_q$. We denote the convex hull by $\overline{a_0 a_1 \cdots a_q}$.

DEFINITION. Let U be an open set in R^n. (In particular U may be R^n itself.) Let $a_0, a_1, \cdots, a_q$ be a set of $(q + 1)$ distinct points in U such that $\overline{a_0 a_1 \cdots a_q} \subset U$. The set $\overline{a_0 a_1 \cdots a_q}$ is a *q-cell in U* or a *cell of order q in U*. The q-cell will sometimes be denoted by $\bar{x}^q$. (Also, if the superscript q is not needed, it will be omitted.) The points $a_0, a_1, \cdots, a_q$ are the *vertices* of $\bar{x}^q$.

DEFINITION. If $a_{k_1}, \cdots, a_{k_m}$ is a subset of $a_0, a_1, \cdots, a_q$ such that $\overline{a_{k_1} \cdots a_{k_m}}$ is a subset of the boundary of $\overline{a_0 a_1 \cdots a_q}$ then $\overline{a_{k_1} \cdots a_{k_m}}$ is a *side* of $\overline{a_0 a_1 \cdots a_q}$.

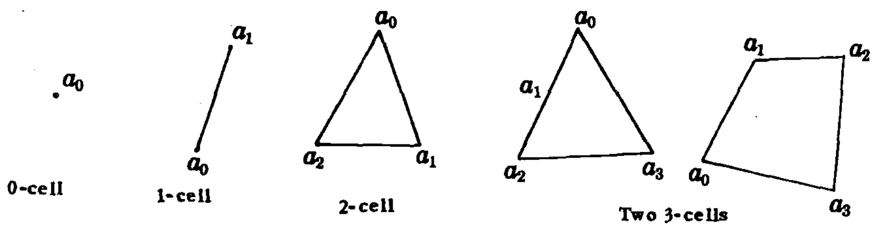


FIGURE 1. SOME CELLS IN R^2

DEFINITION. Consider the collection of all orderings of the vertices of a cell $\bar{x}^q$. Two orderings are *equivalent* if one can be obtained from the other by an even permutation of the vertices.

It is easy to show that this is a genuine equivalence relation, i.e., the relation is symmetric, reflexive, and transitive. Also just two equivalence classes of orderings are obtained.

DEFINITION. The two equivalence classes of orderings are called the *orientations* of $\bar{x}^q$. An *oriented cell in an open set U* is a cell $\bar{x}^q$ in U with one of the orientations specified. (More precisely, an oriented cell is a set with two elements: the cell $\bar{x}^q$ and one of the orientations.) Corresponding to each cell $\bar{x}^q$, there are two oriented cells which are denoted by $+x^q$ (or just x^q) and $-x^q$. These are called the ***positively oriented cell*** and the ***negatively oriented cell*** or more briefly the positive cell and the negative cell. If $a_0, a_1, \cdots, a_q$ are the vertices of $\bar{x}^q$, and if $a_0a_1 \cdots a_q$ is an ordering in the positive orientation, the oriented cell x^q is sometimes denoted by $(a_0a_1 \cdots a_q)$.

Note that the terms "positively oriented" and "negatively oriented" are assigned in an entirely arbitrary way. There is no real reason for calling a particular orientation the positive one.

If we try to apply the above definition of orientation to a 0-cell (which consists of just one point), then only one orientation is obtained for the cell. Consequently it would seem purposeless to introduce an orientation for a 0-cell. However, if a definition of orientation is not made for the 0-cell, then in any reference to orientation, we would have to treat the 0-cell as a special case. Consequently we make the following definition.

DEFINITION. An ***oriented*** ***0-cell*** is a 0-cell with the only ***possible orientation*** specified. This orientation is called the positive orientation.

The geometric notion of a cell and the algebraic notion of "counting" the cell a certain number of times are combined in the concept of chain in which we associate with each one of a finite set of cells a "coefficient," i.e., a positive or negative integer, and in this association it is required that "multiplying" a cell by a negative integer $-n$ is "the same" as changing the orientation of the cell and "multiplying" the cell with reversed orientation by $+n$. We make this idea precise with the following definition.

DEFINITION. Let C^q be the collection of oriented q-cells in U where q is fixed. A *q-chain on U* is a function c^q with domain C^q and range a subset of the integers and with the properties:

(1) if $q > 0$,

$$c^q(-x^q) = -c^q(x^q)$$

for all $x^q \in C^q$;

(2) $c^q(x^q) \neq 0$ for at most a finite number of elements x^q of C^q.

(If it is not needed, the superscript q in c^q will be omitted. Also the phrase "on U" will often be omitted.)

NOTATION. The chain which has the value $+n$ on an oriented cell x^q, the value $-n$ on $-x^q$, and is zero elsewhere will be denoted by nx^q or if $n = 1$, the chain may be simply denoted by x^q. In general, the q-chain c will be denoted by $\sum_{i=1}^{m} t^i x_i$ where x_i is the oriented q-cell such that $c(x_i) = t^i$ and $x_1, \cdots, x_m, -x_1, \cdots, -x_m$ are the oriented q-cells for which the functional value is nonzero.

DEFINITION. If c_1, c_2 are the q-chains $\sum_i t^i x_i$, $\sum_j u^j x_j$, respectively, the *sum of* c_1 *and* c_2, denoted by $c_1 + c_2$, is the q-chain $\sum_k (t^k + u^k) x_k$ where the x_k are those oriented q-cells such that $t^k + u^k = c_1(x_k) + c_2(x_k) \neq 0$. If a is an integer and c is the q-chain $\sum t^i x_i$, then ac is defined to be the q-chain $\sum a t^i x_i$. In particular if $a = -1$, then ac is denoted by $-c$. If c is the q-chain for which all the functional values are zero, then c is called the null q-chain and is denoted by θ_q.

DEFINITION. If $x^q = (a_0 a_1 \cdots a_q)$ is the q-chain which has the value 1 on x^q and -1 on $-x^q$ and is zero elsewhere, the *boundary of* x^q, denoted by $b(x^q)$, is the $(q-1)$-chain

$$b(x_q) = \sum_{i=0}^{q} (-1)^i x_i^{q-1}$$

where $x_i^{q-1} = (a_0 a_1 \cdots \hat{a}_i \cdots a_q)$, i.e., the vertices of $\bar{x}^q$ with the same ordering as in x^q and with the vertex a_i omitted.

To make this definition valid, we must show that it is independent of the particular ordering $a_0 a_1 \cdots a_q$ in the orientation. Suppose $a_0 a_1 \cdots a_q$ and $a_{i_0} a_{i_1} \cdots a_{i_q}$ are orderings in the same orientation so that

$$(a_{i_0} a_{i_1} \cdots a_{i_q}) = (a_0 a_1 \cdots a_q).$$

If $a_{i_j} = a_k$, it is sufficient to show that

$$(-1)^k (a_0 a_1 \cdots \hat{a}_k \cdots a_q) = (-1)^{i_j} (a_{i_0} a_{i_1} \cdots \hat{a}_{i_j} \cdots a_{q_j}).$$

But

$$(a_0 a_1 \cdots a_k \cdots a_q) = (-1)^k (a_k a_0 a_1 \cdots \hat{a}_k \cdots a_q)$$

and

$$(a_{i_0} a_{i_1} \cdots a_{i_j} \cdots a_q) = (-1)^{i_j} (a_{i_j} a_{i_0} a_{i_1} \cdots \hat{a}_{i_j} \cdots a_q)$$

or

$$(-1)^k (a_k a_0 a_1 \cdots \hat{a}_k \cdots a_q) = (-1)^{i_j} (a_{i_j} a_{i_0} a_{i_1} \cdots \hat{a}_{i_j} \cdots a_q).$$

Since $a_k = a_{i_j}$, we have:

$$(-1)^k (a_0 a_1 \cdots \hat{a}_k \cdots a_q) = (-1)^{i_j} (a_{i_0} a_{i_1} \cdots \hat{a}_{i_j} \cdots a_q).$$

For examples of the boundary of a chain, see Figure 2.

REMARK. A definition of boundary for a 0-chain can be introduced but since we do not need it, we omit it.

DEFINITION. If c is the q-chain $\sum_i t^i x_i$ where $q \geqq 1$, the boundary of c, denoted by $b(c)$, is:

$$b(c) = \sum_i t^i b(x_i).$$

REMARK. From these definitions, it follows that:

$$b(c_1 + c_2) = b(c_1) + b(c_2)$$
$$b(-c) = -b(c)$$
$$b(\theta_q) = \theta_{q-1}.$$

DEFINITION. The q-chain c is a *q-cycle* if $b(c) = \theta_{q-1}$.

NOTATION. A q-cycle will be denoted by z^q.

THEOREM. *If $q \geqq 2$, then $b(x^q)$ is a $(q-1)$-cycle.*

PROOF. Let

$$x^q = (a_0 a_1 \cdots a_q),$$
$$x_i^{q-1} = (a_0 a_1 \cdots \hat{a}_i \cdots a_q),$$
$$x_j^{q-1} = (a_0 a_1 \cdots \hat{a}_j \cdots a_q),$$
$$x_{ij}^{q-2} = (a_0 a_1 \cdots \hat{a}_i \cdots \hat{a}_j \cdots a_q).$$

It is sufficient to show that the value of the chain $b[b(x^q)]$ on x_{ij}^{q-2} is zero. First the chain $b(x^q)$ has the value $(-1)^m$ on x_m^{q-1} for $m = i, j$. If $i < j$, then $b(x_i^{q-1})$ has value $(-1)^{j-1}$ on x_{ij}^{q-2} and $b(x_j^{q-1})$ has value $(-1)^i$ on x_{ij}^{q-2}. Thus if $i < j$, the value of $b[b(x^q)]$ on x_{ij}^{q-2} is:

$$[(-1)^i(-1)^{j-1} + (-1)^j(-1)^i] = 0. \blacksquare$$

DEFINITION. Let U, V be open sets in R^n. The q-cycle z^q on U is a *bounding cycle in V* if there is a $(q+1)$-chain c^{q+1} on V such that:

$$b(c^{q+1}) = z^q.$$

(2.1) THEOREM. *If z^q is a cycle on U, an open set in R^n, and if $q \geqq 1$, then z^q is a bounding cycle in R^n.*

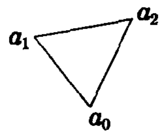
PROOF. Let $z^q = \sum_{i=1}^m t^i x_i^q$. Let a be a point in R^n which is not a vertex of any of the cells $x_1^q, \cdots, x_m^q$. If $x_i^q = (a_0 a_1 \cdots a_q)$, let $(a x_i^q)$ denote the oriented cell $(a a_0 a_1 \cdots a_q)$ and let $a z^q$ denote the chain

$$\sum_{i=1}^m t^i (a x_i^q).$$

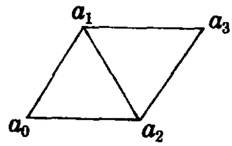
In particular, if $z^q = \theta_q$, then $a z^q$ denotes θ_{q+1}.

If $b(x_i^q)$ is $\sum_{j=0}^n (-1)^j x_j^{q-1}$, let $ab(x_i^q)$ denote the chain $\sum_{j=0}^n (-1)^j (a x_j^{q-1})$. Then

$$\begin{aligned} b[(az^q)] &= \sum_{i=1}^m t^i b(a x_i^q) \\ &= \sum_{i=1}^m t^i [x_i^q - ab(x_i^q)] \\ &= \sum_{i=1}^m t^i x_i^q - ab(z^q) \\ &= \sum_{i=1}^m t^i x_i^q = z^q. \blacksquare \end{aligned}$$

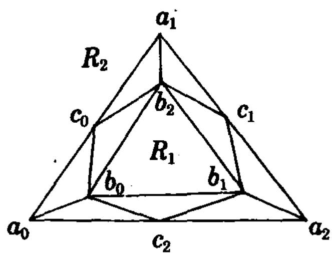


(i) $b(a_0a_1a_2) = (a_0a_1) - (a_0a_2) + (a_1a_2).$



(ii) $$b[(a_0a_1a_2) + (a_1a_3a_2)] = (a_0a_1) - (a_0a_2) + (a_1a_2) - (a_1a_2) + (a_1a_3) + (a_3a_2)$$
$$= (a_0a_1) - (a_0a_2) + (a_1a_3) + (a_3a_2).$$

Note that (a_1a_2) appears in the boundary of each of the 2-cells and "cancels out."



(iii) $$b[(a_0c_2b_0) + (b_0c_2b_1) + (b_1c_1a_2) + (b_1a_2c_1) + (b_1c_1b_2) + (b_2c_1a_1) + (b_2a_1c_0) + (b_2c_0b_0) + (b_0c_0a_0)]$$
$$= (a_0c_2) + (c_2a_2) + (a_2c_1) + (c_1a_1) + (a_1c_0) + (c_0a_0) - (b_0b_1) - (b_1b_2) - (b_2b_0).$$

Thus if U is the open set bounded by circles R_1 and R_2, then

$$(a_0c_2) + (c_2a_2) + (a_2c_1) + (c_1a_1) + (a_1c_0) + (c_0a_0) \sim [(b_0b_1) + (b_1b_2) + (b_2b_0)]$$

in U. But neither of these 1-chains is homologous to θ_1 in U.

FIGURE 2. SOME BOUNDARIES

DEFINITION. Suppose z_1^q, z_2^q are cycles in U. Then z_1^q is *homologous in* U to z_2^q if there exists a chain c^{q+1} in U such that

$$b(c^{q+1}) = z_1^q - z_2^q.$$

In particular if there exists c^{q+1} such that

$$b(c^{q+1}) = z^q,$$

then z^q is homologous to θ_q. We write $z_1^q \sim z_2^q$ in U to indicate that z_1^q is homologous in U to z_2^q.