Michael Spivak,

Calculus on Manifolds

A modern approach to classical theorems of advanced calculus

1965

Table of Contents

1. Function on Euclidean Space

1.1 Norm and Inner Product

Euclidean n-space R^n is defined as the set of all n-tuples $(x^1,\dots,x^n)$ of real numbers x^i (a "1-tuple of numbers" is just a number and $R^1 = R$, the set of all real numbers). An element of R^n is often called a **point** in R^n, and R^1, R^2, R^3 are often called the **line**, the **plane**, and **space**, respectively. If x denotes an element of R^n, then x is an n-tuple of numbers, the *i*th one of which is denoted x^i; thus we can write

$$x = (x^1,\dots,x^n).$$

A point in R^n is frequently also called a **vector** in R^n, because R^n, with $x + y = (x^1 + y^1,\dots,x^n + y^n)$ and $ax = (ax^1,\dots,ax^n)$, as operation, *is* a vector space (over the real numbers, of dimension n). In this vector space there is the notion of the length of a vector x, usually called the **norm** $|x|$ of x and defined by

$$|x| = \sqrt{(x^1)^2 + \cdots + (x^n)^2}.$$

If $n = 1$, then $|x|$ is the usual absolute value of x. The relation between the norm and the vector structure of R^n is very important.

1-1 Theorem. If $x, y \in R^n$ and $a \in R$, then

(1) $|x| \ge 0$, and $|x| = 0$ if and only if $x = 0$.

(2) $\left|\sum_{i=1}^{n} x^i y^i\right| \le |x| \cdot |y|$; equality holds if and only if x and y are linearly dependent.

(3) $|x + y| \le |x| + |y|$.

(4) $|ax| = |a| \cdot |x|$.

The quantity $\sum_{i=1}^{n} x^i y^i$ which appears in (2) is called the i**nner product** of x and y and denoted $\langle x, y\rangle$. The most important properties of the inner product are the following.

1-2 **Theorem.** If $x, x_1, x_2, y, y_1, y_2 \in R^n$, and $a \in R$, then

(1) $\langle x, y\rangle = \langle y, x\rangle$ (symmetry)

(2) $\langle ax, y\rangle = \langle x, ay\rangle = a\langle x, y\rangle$ (bilinearity)

$\langle x_1 + x_2, y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle$

$\langle x, y_1 + y_2\rangle = \langle x, y_1\rangle + \langle x, y_2\rangle$

(3) $\langle x, x\rangle \ge 0,$ and $\langle x, x\rangle = 0$ if and only if $x = 0$ (positive definiteness)

(4) $|x| = \sqrt{\langle x, x\rangle}.$

(5) $\langle x, y\rangle = \dfrac{|x + y|^2 - |x - y|^2}{4}$ (polarization identity)

We conclude this section with some important remarks about notation. The vector $(0,\dots,0)$ will usually be denoted simply 0. The **usual basis** of R^n is $e_1,\dots,e_n$, where $e_i = (0,\dots,1,\dots,0)$, with the 1 in the *i*th place. If $T : R^n \to R^m$ is a linear transformation, the matrix of T with respect to the usual bases of R^n and R^m is the $m \times n$ matrix $A = (a_{ij})$, where $T(e_i) = \sum_{j=1}^{m} a_{ji} e_j$ - the coefficient of $T(e_i)$ appear in the *i*th column of the

matrix. If $S : R^m \to R^p$ has the $p \times m$ matrix B, then $S \circ T$ has the $p \times n$ matrix BA [here $S \circ T(x) = S(T(x))$; most books on linear algebra denote $S \circ T$ simply ST]. To find $T(x)$ one computes the $m \times 1$ matrix

$$\begin{pmatrix} y^1 \\ \vdots \\ y^m \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}$$

then $T(x) = (y^1, \ldots, y^m)$. One notational convention greatly simplifies many formulas: if $x \in R^n$ and $y \in R^m$, then (x, y) denotes

$$(x^1, \ldots, x^n, y^1, \ldots, y^m) \in R^{n+m}.$$

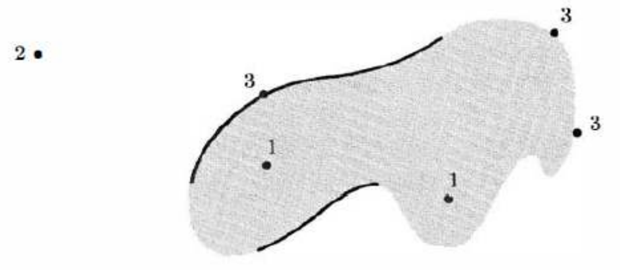
1.2 Subsets of Euclidean Space

The closed interval $[a,b]$ has a natural analogue in R^2. This is the **closed rectangle** $[a,b] \times [c,d]$, defined as the collection of all pairs (x, y) with $x \in [a,b]$ and $y \in [c,d]$. More generally, if $A \subset R^m$ and $B \subset R^n$, then $A \times B \subset R^{m+n}$ is defined as the set of all $(x, y) \in R^{m+n}$ with $x \in A$ and $y \in B$. In particular, $R^{m+n} = R^m \times R^n$. If $A \subset R^m$, $B \subset R^n$ and $C \subset R^p$, then $(A \times B) \times C = A \times (B \times C)$, and both of these are denoted simply $A \times B \times C$; this convention is extended to the product of any number of sets. The set $[a_1, b_1] \times \cdots \times [a_n, b_n] \subset R^n$ is called **closed rectangle** in R^n, while the set $(a_1, b_1) \times \cdots \times (a_n, b_n) \subset R^n$ is called **open rectangle.** More generally a set $U \subset R^n$ is called **open** if for each $x \in U$ there is an open rectangle A such that $x \in A \subset U$.

A subset C of R^n is **closed** if $R^n - C$ is open. For example, If C contains only finitely many points, then C is closed. The reader should supply the proof that a closed rectangle in R^n is indeed a closed set.

If $A \subset R^n$ and $x \in R^n$, then one of three possibilities must hold (Figure):

1. There is an open rectangle B such that $x \in B \subset A.$ (**interior**)
2. There is an open rectangle B such that $x \in B \subset R^n - A.$ (**exterior**)
3. If B is any open rectangle with $x \in B$, then B contains points of both A and $R^n - A.$ (**boundary**)



Those points satisfying (1) constitute the **interior** of A, those satisfying (2) the **exterior** of A, and those satisfying (3) the **boundary** of A. Problems 1-16 to 1-18 show that these terms may sometimes have unexpected meanings.

It is not hard to see that the interior of any set A is open, and the same is true for the exterior of A, which is, in fact, the interior of $R^n - A$. Thus (Problem 1-14) their union is open, and what remains, the boundary, must be closed.

A collection O of open sets is an **open cover** of A (or, briefly, **covers** A) if every point $x \in A$ is in some open set in the collection O. For example, if O is the collection of all open intervals $(a, a+1)$ for $a \in R$, then O is a cover of R. Clearly no finite number of the open sets in O will cover R or, for that matter, any unbounded subset of R. A similar situation can also occur for bounded sets. If O is the collection of all open intervals $(1/n, 1-1/n)$ for all integers $n > 1$, then O is an open cover of $(0,1)$, but again no finite collection of sets in O will cover $(0,1)$. Although this phenomenon may not appear particularly scandalous, sets for which this state of affairs cannot occur are of such importance that they have received a special designation: a set A is called

compact if every open cover O contains a finite subcollection of open sets which also covers A.

A set with only finitely many points is obviously compact and so is the infinite set A which contains 0 and the numbers $1/n$ for all integers n (reason: if O is a cover, then $O \in U$ for some open set U in O; there are only finitely many other points of A not in U, each requiring at most one more open set).

Recognizing compact sets is greatly simplified by the following results, of which only the first has any depth (i.e., uses any facts about the real numbers).

1-3 **Theorem (Heine-Borel)**. The closed interval $[a,b]$ is compact.

***

Theorem 13.1 (From Matsumoto) Let K be a bounded closed set in R^n and $\{U_\alpha\}_{\alpha \in A}$ be any open cover on K. Then you can choose limited number of open sets $U_{\alpha(1)}, U_{\alpha(2)}, \ldots, U_{\alpha(k)}$ from $\{U_\alpha\}_{\alpha \in A}$ which satisfy $K \subseteq U_{\alpha(1)} \cup U_{\alpha(2)} \cup \cdots \cup U_{\alpha(k)}$.

***

If $B \subset R^m$ is compact and $x \in B^n$, it is easy to see that $\{x\} \times B \subset R^{n+m}$ is compact. However, a much stronger assertion can be made,

1-4 **Theorem.** If B is compact and O is an open cover of $\{x\} \times B$, then there is an open set $U \subset R^n$ containing x such that $U \times B$ is covered by a finite number of sets in O.

1-5 **Corollary**. If $A \subset R^n$ and $B \subset R^m$ are compact, then $A \times B \subset R^{n+m}$ is compact.

1-6. **Corollary**. $A_1 \times \cdots \times A_k$ is compact if each A_i is. In particular, a closed rectangle in R^k is compact.

1-7 **Corollary**. A closed bounded subset of R^n is compact. (The converse is also true.)

1.3 Functions and Continuity

A **function** from R^n to R^m (sometimes called a (vector-valued) function of n variables) is a rule which associates to each point in R^n some point in R^m; the point function f associates to x is denoted $f(x)$. We write $f: R^n \to R^m$ (read "f takes R^n into R^m" or "f, taking R^n into R^m." depending on context) to indicate that $f(x) \in R^m$ is defined for $x \in R^n$. The notation $f: A \to R^m$ indicates that $f(x)$ is defined only for x in the set A, which is called **domain** of f. If $B \subset A$, we define $f(B)$ as the set of all $f(x)$ for $x \in B$, and if $C \subset R^m$ we define $f^{-1}(C) = \{x \in A : f(x) \in C\}$. The notation $f: A \to B$ indicates that $f(A) \subset B$.

A convenient representation of a function $f: R^2 \to R$ may be obtained by drawing a picture of its **graph**, the set of all 3-tuples of the form $(x, y, f(x,y))$, which is actually a figure in 3-space (see, e.g., Figures 2-1 and 2-2 of Chapter 2).

If $f, g: R^n \to R$, the functions $f+g$, $f-g$, $f \cdot g$, and f/g are defined precisely as in the one-variable case. If $f: A \to R^m$ and $g: B \to R^p$, where $B \subset R^m$, then the **composition** $g \circ f$ is defined by $g \circ f(x) = g(f(x))$; the domain of $g \circ f$ is $A \cap f^{-1}(B)$. If $f: A \to R^m$ is 1-1, that is, if $f(x) \neq f(y)$ when $x \neq y$, we define $f^{-1}: A \to R^n$ by the requirement that $f^{-1}(z)$ is the unique $x \in A$ with $f(x) = z$.

A function $f: A \to R^m$ determines m **component functions** $f^1, \ldots, f^m : A \to R$ by

$f(x)=(f^1(x),\dots,f^m(x))$. If conversely, m functions $g_1,\dots,g_m : A\to R$ are given, there is a unique function $f : A\to R^m$ such that $f^i = g_i$, namely $f(x)=(g_1(x),\dots,g_m(x))$. This function f will be denoted $(g_1,\dots,g_m)$, so that we always have $f=(f^1,\dots,f^m)$. If $\pi : R^n\to R^n$ is the identity function, $\pi(x)=x$, then $\pi^i(x)=x^i$; the function π^i is called the ith **projection function**.

The notation $\lim_{x\to a} f(x)=b$ means, as in the one-variable case, that we can get $f(x)$ as close to b as desired, by choosing x sufficiently close to, but not equal to, a. In mathematical terms this means that for every number $\varepsilon>0$ there is a number $\delta>0$ such that $|f(x)-b|<\varepsilon$ for all x in the domain of f which satisfy $0<|x-a|<\delta$. A function $f : A\to R^m$ is called **continuous** at $a\in A$ if $\lim_{x\to a} f(x)=f(a)$, and f is simply called continuous if it is continuous at each $a\in A$. One of the pleasant surprises about the concept of continuity is that it can be defined without using limits. It follows from the next theorem that $f : R^n\to R^m$ is continuous if and only if $f^{-1}(U)$ is open whenever $U\subset R^m$ is open; if the domain of f is not all of R^n, slightly more complicated condition is needed.

1-8 **Theorem**. If $A\subset R^n$, a function $f : A\to R^m$ is continuous if and only if for every open set $U\subset R^m$ there is some open set $V\subset R^n$ such that $f^{-1}(U)=V\cap A$.

***

(From Matsumoto: Let U be any open set of Y. Then a mapping $f : X\to Y$ is continuous if the inverse mapping $f^{-1}(U)$ is an open set of X.)

***

The following consequence of Theorem 1-8 is of great importance.

1-9 **Theorem**. If $f : A\to R^m$ is continuous, where $A\subset R^n$, and A is compact, then $f(A)\subset R^m$ is compact.

If $f : A\to R$ is bounded, the extent to which f fails to be continuous at $a\in A$ can be measured in a precise way. For $\delta>0$ let

$$M(a,f,\delta)=\sup\{f(x) : x\in A, |x-a|<\delta\},$$
$$m(a,f,\delta)=\inf\{f(x) : x\in A, |x-a|<\delta\}.$$

The **oscillation** $o(f,a)$ of f at a is defined by $o(f,a)=\lim_{\delta\to 0}[M(a,f,\delta)-m(a,f,\delta)]$. This limit always exists, since $M(a,f,\delta)-m(a,f,\delta)$ decreases as δ decreases. There are two important facts about $o(f,a)$.

1-10 **Theorem**. The bounded function f is continuous at a if and only if $o(f,a)=0$.

1-11 **Theorem**. Let $A\subset R^n$ be closed. If $f : A\to R$ is any bounded function, and $\varepsilon>0$, then $\{x\in A : o(f,x)\ge\varepsilon\}$ is closed.

2. Differentiation

2.1 Basic Definitions

Recall a function $f : R \to R$ is differentiable at $a \in R$ if there is a number $f'(a)$ such that

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = f'(a). \qquad (1)$$

This equation certainly makes no sense in the general case of a function $f : R^n \to R^m$, but can be reformulated in a way that does. If $\lambda : R \to R$ is the linear transformation defined by $\lambda(h) = f'(a) \cdot h$, then equation (1) is equivalent to

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \lambda(h)}{h} = 0. \qquad (2)$$

Equation (2) is often interpreted as saying that $\lambda + f(a)$ us a good approximation to f at a. Henceforth we focus our attention on the linear transformation λ and reformulate the definition of differentiability as follows.

A function $f : R \to R$ is differentiable at $a \in R$ if there is a linear transformation $\lambda : R \to R$ such that

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \lambda(h)}{h} = 0.$$

In this form the definition has a simple generalization to higher dimensions:

A function $f : R^n \to R^m$ is **differentiable** at $a \in R^n$ if there is a linear transformation $\lambda : R^n \to R^m$ such that

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0.$$

Note that $h \in R^n$ and $f(a+h) - f(a) - \lambda(h) \in R^m$, so the norm signs are essential. The linear transformation λ is denoted $Df(a)$ and called the **derivative** of f at a. This justification for the phrase "*the* linear transformation λ" is:

2-1 **Theorem**. If $f : R^n \to R^m$ is differentiable at $a \in R^n$ there is a **unique** linear transformation $\lambda : R^n \to R^m$ such that

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0.$$

Example Consider the function $f : R^2 \to R$ defined by $f(x, y) = \sin x$. Then $Df(a,b) = \lambda$ satisfies $\lambda(x, y) = (\cos a) \cdot x$. To prove this, note that

$$\lim_{(h,k) \to 0} \frac{|f(a+h, b+h) - f(a,b) - \lambda(h,k)|}{|(h,k)|} = \lim_{(h,k) \to 0} \frac{|\sin(a+h) - \sin a - (\cos a) \cdot h|}{|(h,k)|}$$

Since $\sin'(a) = \cos a$, we have

$$\lim_{h \to 0} \frac{|\sin(a+h) - \sin a - (\cos a) \cdot h|}{|h|} = 0.$$

Since $|(h,k)| \geq |h|$, it is also true that

$$\lim_{h \to 0} \frac{|\sin(a+h) - \sin a - (\cos a) \cdot h|}{|(h,k)|} = 0.$$

It is often convenient to consider the matrix of $Df(a) : R^n \to R^m$ with respect to the usual bases of R^n

and R^m. This $m \times n$ matrix is called the **Jacobian matrix** of f at a, and denoted $f'(a)$. If $f(x,y) = \sin x$, then $f'(a,b) = (\cos a, 0)$. If $f : R \to R$, then $f'(a)$ is a 1×1 matrix whose single entry is the number which is denoted $f'(a)$ in elementary calculus.

The definition of $Df(a)$ could be made if f were defined only in some open set containing a. Considering only functions defined on R^n streamlines the statement of theorems and produces no real loss of generality. It is convenient to define a function $f : R^n \to R^m$ to be differentiable on A if f is differentiable at a for each $a \in A$. If $f : A \to R^m$, then f is called differentiable if f can be extended to a differentiable function on some open set containing A.

*****

(From Matsumoto p.99.

Definition 9.II A mapping $(df)_p : T_p(M) \to T_q(N)$ is called the **differential** of $f : M \to N$ at point p.

Theorem 9.4 Let local coordinates around p and q be $(x_1, \ldots, x_m)$ and $(y_1, \ldots, y_n)$ and the function $y_j = f_j(x_1, \ldots, x_m)$. Then

$$(df)_p\left(\left(\frac{\partial}{dx_i}\right)_p\right) = \sum_{j=1}^{n} \frac{\partial f_j}{\partial x_i}(p)\left(\frac{\partial}{\partial y_j}\right)_q .$$

*****

2.2 Basic Theorems

2-2 **Theorem (Chain Rule)**. If $f : R^n \to R^m$ is differentiable at a, and $g : R^m \to R^p$ is differentiable at $f(a)$, then the composition $g \circ f : R^n \to R^p$ is differentiable at a, and

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a).$$

Remark. This equation can be written

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a).$$

If $m = n = p = 1$, we obtain the old chain rule.

2-3 **Theorem**

(1) If $f : R^n \to R^m$ is a **constant function** (that is, if for some $y \in R^m$ we have $f(x) = y$ for all $x \in R^n$), then

$$Df(a) = 0$$

(2) If $f : R^n \to R^m$ is a linear transformation, then

$$Df(a) = f.$$

(3) If $f : R^n \to R^m$, then f is differentiable at $a \in R^n$ if and only if each f^i is, and

$$Df(a) = \left(Df^1(a), \ldots, Df^m(a)\right)$$

Thus $f'(a)$ is the $m \times n$ matrix whose ith row is $(f^i)'(a)$.

(4) If $s : R^2 \to R$ is defined by $s(x,y) = x + y$, then

$$Ds(a,b) = s.$$

(5) If $p : R^2 \to R$ is defined by $p(x,y) = x \cdot y$, then

$$Dp(a,b)(x,y) = bx + ay.$$

Thus $p'(a,b) = (b,a).$

2-4 **Corollary**. If $f, g : R^n \to R$ are differentiable at a, then

$$D(f+g)(a) = Df(a) + Dg(a),$$
$$D(f \cdot g)(a) = g(a)Df(a) + f(a)Dg(a).$$

If, moreover, $g(a) \neq 0$, then

$$D(f/g)(a) = \frac{g(a)Df(a) - f(a)Dg(a)}{[g(a)]^2}.$$

We are now assured of the differentiability of those functions $f : R^n \to R^m$, whose component functions are obtained by addition, multiplication, division, and composition, from the functions π^i (which are linear transformations) and the functions which we can already differentiate by elementary calculus. Finding $Df(x)$ or $f'(x)$ may be a fairly formidable task.

Example Let $f : R^2 \to R$ be defined by $f(x,y) = \sin(xy^2)$. Since $f = \sin \circ (\pi^1 \cdot \left|\pi^2\right|^2)$, we have

$$\begin{aligned}
f'(a,b) &= \sin'(ab^2) \cdot \left[b^2 \left(\pi^1\right)'(a,b) + a\left(\left[\pi^2\right]^2\right)'(a,b) \right] \\
&= \sin'(ab^2) \cdot \left[b^2 \left(\pi^1\right)'(a,b) + 2ab\left(\pi^2\right)'(a,b) \right] \\
&= (\cos(ab^2)) \cdot \left[b^2(1,0) + 2ab(0,1) \right] \\
&= (b^2 \cos(ab^2), 2ab\cos(ab^2)).
\end{aligned}$$

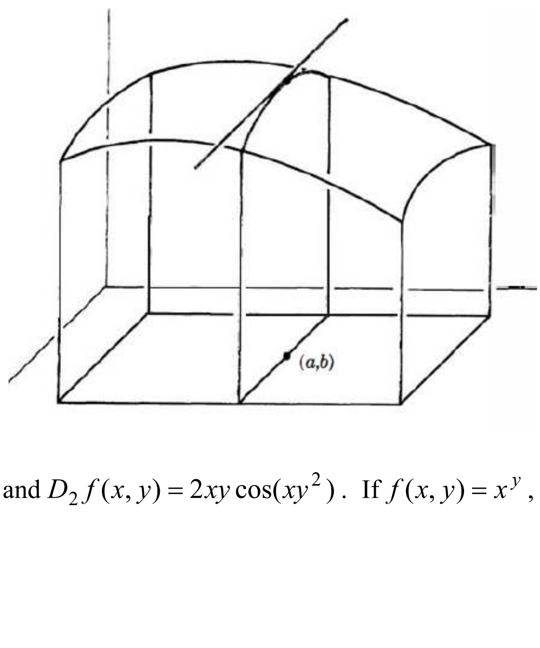
Fortunately, we will soon discover a much simpler method of computing f'.

2.3 Partial Derivatives

We begin the attack on the problem of finding derivatives "one variable at a time." If $f : R^n \to R$ and $a \in R^n$, the limit

$$\lim_{h \to 0} \frac{f(a^1, \ldots, a^i + h, \ldots, a^n) - f(a^1, \ldots, a^n)}{h}$$

if it exists, is denoted $D_i f(a)$, and called ith **partial derivative** of f at a. It is important to note that $D_i f(a)$ is the ordinary derivative of a certain function; in fact, if $g(x) = f(a^1, \ldots, x, \ldots, a^n)$, then $D_i f(a) = g'(a^i)$. This means that $D_i f(a)$ is the slope of the tangent line at $(a, f(a))$ to the curve obtained by intersecting the graph of f with the plane $x^i = a^i, j \neq i$ (Figure) It also means that computation of . $D_i f(a)$ is a problem we can already solve. If $f(x^1, \ldots, x^n)$ is given by some formula involving $x^1, \ldots, x^n$, then we find $D_i f(x^1, \ldots, x^n)$ by differentiating the function whose value at x^i is given by the formula when all x^j, for $j \neq i$, are thought of as constants.



Example If $f(x,y) = \sin(xy^2)$, then $D_1 f(x,y) = y^2 \cos(xy^2)$ and $D_2 f(x,y) = 2xy\cos(xy^2)$. If $f(x,y) = x^y$, then $D_1 f(x,y) = yx^{y-1}$ and $D_2 f(x,y) = x^y \log x$.

With a little practice (e.g., the problems at the end of this section) you should acquire as great a facility for computing $D_i f(a)$ as you already have for computing ordinary derivatives.

If $D_i f(x)$ exists for all $x \in R^n$, we obtain a function $D_i f : R^n \to R$. The *j*th partial derivative of this function at x, that is, $D_j(D_i f)(x)$, is often denoted $D_{i,j} f(x)$. Note that this notation reverses the order of i and j. As a matter of fact, the order is usually irrelevant, since most functions satisfy $D_{i,j} f = D_{j,i} f$ .There are various delicate theorems ensuring this equality: the following theorem is quite adequate. We state here but postpone the proof until later.

2-5 **Theorem**. If $D_{i,j} f$ and $D_{j,i} f$ are continuous in an open set containing a, then

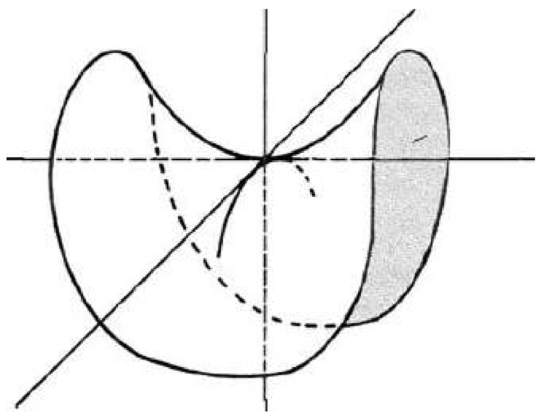
$$D_{i,j} f(a) = D_{j,i} f(a).$$

The function $D_{i,j} f$ is called a **second-order (mixed) partial derivative** of f. **Higher-order (mixed) partial derivatives** are defined in the obvious way. Clearly Theorem 2-5 can be used to prove the equality of higher-order mixed partial derivatives under appropriate conditions. The order of $i_1,\dots,i_k$ is completely immaterial in $D_{i1\dots,ik} f$ if f has continuous partial derivatives of all orders. A function with this property is called a C^∞ function. In later chapters it will frequently be convenient to restrict our attention to C^∞ functions.

Partial derivatives will be used in the next section to find derivatives. They also have another important use – finding maxima and minima of functions.

2-6 **Theorem**. Let $A \subset R^n$. If the maximum (or minimum) of $f : A \to R$ occurs at a point a in the interior of A and $D_i f(a)$ exists, then $D_i f(a) = 0$.

The reader is reminded that the converse of Theorem 2-6 is false even if $n = 1$ (if $f : R \to R$ is defined by $f(x) = x^3$, then $f'(0) = 0$, but 0 is not even a local maximum of minimum). If $n > 1$, the converse of Theorem 2-6 may fail to be true in a rather spectacular way. Suppose, for example, that $f : R^2 \to R$ is defined by $f(x, y) = x^2 - y^2$ (Figure). Then $D_1 f(0,0) = 0$ because g_1 has a minimum at 0, while $D_2 f(0,0) = 0$ because g_2 has a maximum at 0. Clearly $(0,0)$ is neither a relative maximum nor a relative minimum.



If Theorem 2-6 is used to find the maximum or minimum of f on A, the values of f at boundary points must be examined 'separately - a formidable task, since the boundary of A may be all A! Problem 2-27 indicates one way of doing this, and Problem 5-16 states a superior method which can often be used.

2.4 Derivatives

The reader who has compared Problems 2-10 and 2-17 has probably already guessed the following.

2-7 Theorem. If $f : R^n \to R^m$ is differentiable at a, then $D_j f^i(a)$ exists for $1 \le i \le m, 1 \le j \le n$ and $f'(a)$ is the $m \times n$ matrix $(D_j f^i(a))$.

There are several examples in the problems to show that the converse of Theorem 2-7 is false. It is true, however, if one hypothesis is added.

2-8 **Theorem**. If $f : R^n \to R^m$, then $Df(a)$ exists if all $D_j f^i(a)$ exist in an open set containing a and if each function $D_j f^i(a)$ is continuous at a.

(Such a function f is called **continuously differentiable** at a.)

Although the chain rule was used in the proof of Theorem 2-7, it could easily have been eliminated. With Theorem 2-8 to provide differentiable functions, and Theorem 2-7 to provide their derivatives, the chain rule may therefore seem almost superfluous. However, it has an extremely important corollary concerning partial derivatives.

2-9 **Theorem** Let $g_1, \ldots, g_m : R^n \to R$ be continuously differentiable at a, and let $f : R^m \to R$ be differentiable at $(g_1(a), \ldots, g_m(a))$. Define the function $F : R^n \to R$ by $F(x) = f(g_1(x), \ldots, g_m(a))$. Then

$$D_i F(a) = \sum_{j=1}^{m} D_j f(g_1(a), \ldots, g_m(a)) \cdot D_i g_j(a)\,.$$

Theorem 2-9 is often called the *chain rule*, but is weaker than Theorem 2-2 since g could be differentiable without g_i being continuously differentiable. Most computations requiring Theorem 2-9 are fairly straightforward. A slight subtlety is required for the function $F : R^2 \to R$ defined by

$$F(x,y) = f(g(x,y), h(x), k(y))$$

where $h, k : R \to R$. In order to apply Theorem 2-9 define $\bar{h}, \bar{k} : R^2 \to R$ by

$$\bar{h}(x,y) = h(x)\,, \qquad \bar{k}(x,y) = k(y)\,.$$

Then

$$D_1\bar{h}(x,y) = h'(x) \qquad D_2\bar{h}(x,y) = 0\,,$$
$$D_1\bar{k}(x,y) = 0 \qquad D_2\bar{k}(x,y) = k'(y)\,,$$

and we can write

$$F(x,y) = f(g(x,y), \bar{h}(x,y), \bar{k}(x,y))\,.$$

Letting $a = (g(x,y), h(x), k(y))$, we obtain

$$D_1 F(x,y) = D_1 f(a) \cdot D_2 g(x,y) + D_2 f(a) \cdot h'(x)\,,$$
$$D_2 F(x,y) = D_1 f(a) \cdot D_2 g(x,y) + D_3 f(a) \cdot k'(y)\,.$$

It should, of course, be unnecessary for you to actually write down the functions $\bar{h}$ and $\bar{k}$.

2.5 Inverse Functions

Suppose that $f : R \to R$ is continuously differentiable in an open set containing a and $f'(a) \neq 0$. If $f'(a) > 0$, there is an open interval V containing a such that $f'(a) > 0$ for $x \in V$, and a similar statement holds if $f'(a) < 0$. Thus f is increasing (decreasing) on V, and is therefore one-to-one with an inverse function f^{-1} defined on some open interval W containing $f(a)$. Moreover it is not hard to show that f^{-1} is differentiable, and for $y \in W$ that

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))}\,.$$

An analogous discussion in higher dimensions is much more involved, but the result (Theorem 2-11) is very important. We begin with a simple lemma.

2-10 **Lemma**. Let $A \subset R^n$ be a rectangle and let $f : A \to R^n$ be continuously differentiable. If there is a number M such that $\left|D_j f^i(x)\right| \le M$ for all x in the interior of A, then

$$\left|f(x) - f(y)\right| \le n^2 M\left|x - y\right|$$

for all $x, y \in A$.

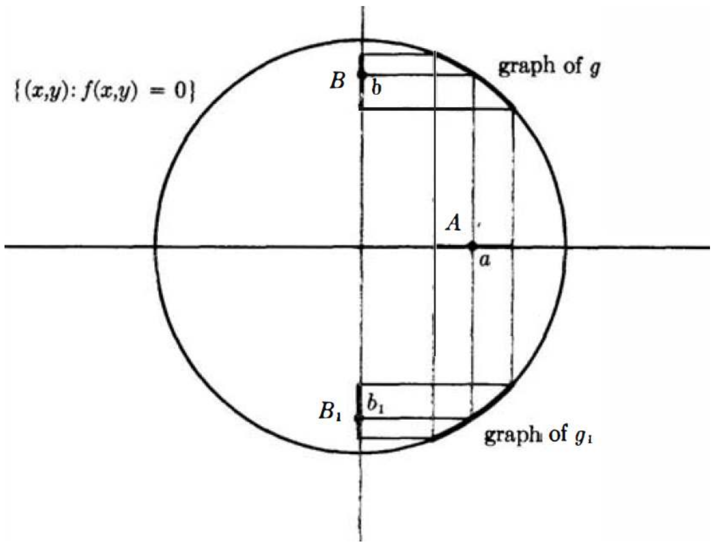
2-11 **Theorem (Inverse Function Theorem).** Suppose that $f : R^n \to R^n$ is continuously differentiable in an open set containing a, and $\det f'(a) \ne 0$. Then there is an open set V containing a and an open set W containing $f(a)$ such that $f : V \to W$ has a continuous inverse $f^{-1} : W \to V$ which is differentiable and for all $y \in W$ satisfies

$$\left(f^{-1}\right)'(y) = \left[f'\left(f^{-1}(y)\right)\right]^{-1}.$$

It should be noted that an inverse function f^{-1} may exist even if $\det f'(a) = 0$. For example, if $f : R \to R$ is defined by $f(x) = x^3$, then $\det f'(0) = 0$ but f has the inverse function $f^{-1}(x) = \sqrt[3]{x}$. One thing is certain however: if $\det f'(a) = 0$ then f^{-1} cannot be differentiable at $f(a)$. To prove this note that $f \circ f^{-1}(x) = x$. If f^{-1} were differentiable at $f(a)$, the chain rule would give $f'(a) \cdot (f^{-1})'\left(f(a)\right) = I$, and consequently $\det f'(a) \cdot \det(f^{-1})'(f(a)) = 1$, contradicting $\det f'(a) = 0$.

2.6 Implicit Functions

Consider the function $f : R^2 \to R$ defined by $f(x, y) = x^2 + y^2 - 1$. If we choose (a, b) with $f(a, b) = 0$ and $a \ne 1, -1$, there are open interval A containing a and B containing b with the following property: if $x \in A$, there is a unique $y \in B$ with $f(x, y) = 0$. We can therefore define a function $g : A \to R$ by the condition $g(x) \in B$ and $f(x, g(x)) = 0$ (if $b > 0$, as indicated in Figure, then $g(x) = \sqrt{1 - x^2}$). For the function f we are considering there is another number b_1 such that $f(a, b_1) = 0$. There will also be an interval B_1 containing b_1 such that, when $x \in A$, we have $f(x, g_1(x)) = 0$ for a unique $g_1(x) \in B_1$ (here $g_1(x) = -\sqrt{1 - x^2}$). Both g and g_1 are differentiable. These functions are said to be defined **implicitly** by the equation $f(x, y) = 0$.



If we choose $a = 1$ or -1 it is impossible to find any such function g defined in an open interval containing a. We would like a simple criterion for deciding when, in general, such a function can be found. More

generally we may ask the following: If $f : R^n \times R \to R$ and $f(a^1,\dots,a^n,b) = 0$, when can we find, for each $(x^1,\dots,x^n)$ near $(a^1,\dots,a^n)$, a unique y near b such that $f(x^1,\dots,x^n,y) = 0$? Even more generally, we can ask about the possibility of solving m equations, depending upon parameters $x^1,\dots,x^n$, in m unknowns: If

$$f_i : R^n \times R^m \to R \qquad i = 1,\dots,m$$

and

$$f_i(a^1,\dots,a^n,b^1,\dots b^m) = 0 \qquad i = 1,\dots,m,$$

when can we find, for each $(x^1,\dots,x^n)$ near $(a^1,\dots,a^n)$ a unique $(y^1,\dots,y^m)$ near $(b^1,\dots,b^m)$ which satisfies $f_i(x^1,\dots,x^n,y^1,\dots y^m) = 0$? The answer is provided by the following Theorem.

2-12 **Theorem (Implicit Function Theorem).** Suppose $f : R^n \times R^m \to R^m$ is continuously differentiable in an open set containing (a,b) and $f(a,b) = 0$. Let M be the $m \times m$ matrix

$$\left(D_{n+j} f^i(a,b)\right) \qquad 1 \le i, j \le m.$$

If $\det M \ne 0$, there is an open set $A \subset R^n$ containing a and an open set $B \subset R^m$ containing b, with the following property: for each $x \in A$ there is a unique $g(x) \in B$ such that $f(x, g(x)) = 0$. The function g is differentiable.

Since the function g is known to be differentiable, it is easy to find its derivative. In fact, since $f^i(x, g(x)) = 0$, taking D_j of both sides gives

$$0 = D_j f^i(x, g(x)) + \sum_{\alpha=1}^{m} D_{n+\alpha} f^i(x, g(x)) \cdot D_j g^\alpha(x) \qquad i, j = 1,\dots m.$$

Since $\det M \ne 0$, these equations can be solved for $D_j g^\alpha(x)$. The answer will depend on the various $D_j f^i(x, g(x))$, and therefore on $g(x)$. This is unavoidable, since the function g is not unique. Reconsidering the function $f : R^2 \to R$ defined by $f(x,y) = x^2 + y^2 - 1$, we note that two possible functions satisfying $f(x, g(x)) = 0$ are $g(x) = \sqrt{1-x^2}$ and $g(x) = -\sqrt{1-x^2}$. Differentiating $f(x, g(x)) = 0$ gives

$$D_1 f(x, g(x)) + D_2 f(x, g(x)) \cdot g'(x) = 0,$$

or

$$2x + 2g(x) \cdot g'(x) = 0, \quad g'(x) = -x/g(x),$$

which is indeed the case for either $g(x) = \sqrt{1-x^2}$ or $g(x) = -\sqrt{1-x^2}$.

A generalization of the argument for Theorem 2-12 can be given, which will be important in Chapter 5.

2-13 **Theorem**. Let $f : R^n \to R^p$ be continuously differentiable in an open set containing a, where $p \le n$. If $f(a) = 0$ and the $p \times n$ matrix $\left(D_j f^i(a)\right)$ has *rank* p, then there is an open set $A \subset R^n$ containing a and a differentiable function $h : A \to R$ with differentiable inverse such that

$$f \circ h(x^1,\dots,x^n) = (x^{n-p+1},\dots,x^n).$$

2.7 Notation

The partial derivative $D_1 f(x,y,z)$ is denoted by

$$\frac{\partial f(x,y,z)}{\partial x} \quad \text{or} \quad \frac{\partial f}{\partial x} \quad \text{or} \quad \frac{\partial f}{\partial x}(x,y,z) \quad \text{or} \quad \frac{\partial}{\partial x} f(x,y,z)$$

or any other convenient similar symbols. This notation forces one to write

$$\frac{\partial f}{\partial x}(u,v,w)$$

for $D_1 f(u,v,w)$, although the symbol

$$\left.\frac{\partial f(x,y,z)}{\partial x}\right|_{(x,y,z)=(u,v,w)} \quad \text{or} \quad \frac{\partial f(x,y,z)}{\partial x}(u,v,w)$$

or something similar may be used (and must be used for an expression like $D_1 f(7,3,2)$. Similar notation is used for $D_2 f$ and $D_3 f$. Higher-order derivatives are denoted by symbols like

$$D_2 D_1 f(x,y,z) = \frac{\partial^2 f(x,y,z)}{\partial y \partial x}.$$

When $f: R \to R$, the symbol ∂ automatically reverts to d; thus

$$\frac{d \sin x}{dx}, \text{ not } \frac{\partial \sin x}{\partial x}.$$

The usual evaluation for $D_1(f \circ (g,h))$ runs as follows: If $f(u,v)$ is a function and $u = g(x,y)$ and $v = h(x,y)$, then

$$\frac{\partial f(g(x,y),h(x,y))}{\partial x} = \frac{\partial f(u,v)}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f(u,v)}{\partial v}\frac{\partial v}{\partial x}.$$

[The symbol $\partial u/\partial x$ means $\frac{\partial}{\partial x} g(x,y)$ and $\frac{\partial}{\partial u} f(u,v)$ means $D_1 f(u,v) = D_1 f(g(x,y),h(x,y))$] This equation is often written simply

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x}.$$

Note that f means something different on the two sides of the equation!

The notation df/dx, always a little too tempting, has inspired many (usually meaningless) definitions of dx and df separately, the sole purpose of which is to make the equation

$$df = \frac{df}{dx} \cdot dx$$

work out. If $f: R^2 \to R$ then df is *defined*, classically, as

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy$$

(whatever dx and dy mean).

Chapter 4 contains rigorous definitions which enable us to prove the above equations as theorems. It is a touchy question whether or not these modern definitions represent a real improvement over classical formalism; this the reader must decide for himself.

3. Integration

3.1 Basic Definitions

The definition of the integraI of a function $f : A \to R$, where $A \subset R^n$ is a closed rectangIe, is so similar to that of the ordinary integral that a rapid treatment will be given.

Recall that a partition P of a closed interval $[a,b]$ is a sequence $t_0,\dots,t_k$ where $a = t_0 \le t_1 \le \cdots \le t_k = b$. The partition P divides the interval $[a,b]$ into k subintervals $[t_{i-1},t_i]$. A **partition** of a rectangle $[a_1,b_1]\times\cdots\times[a_n,b_n]$ is a collection $P=(P_1,\dots,P_n)$, where each P_i is a partition of the interval $[a_i,b_i]$. Suppose, for example, that $P_1 = t_0,\dots,t_k$ is a partition of $[a_1,b_1]$ and $P_2 = s_0,\dots,s_l$ is a partition of $[a_2,b_2]$. Then the partition $P=(P_1,P_2)$ of $[a_1,b_1]\times[a_2,b_2]$ divides the closed rectangle $[a_1,b_1]\times[a_2,b_2]$ into $k\cdot l$ subrectangles, a typical one being $[t_{i-1},t_i]\times[s_{j-1},s_j]$. In generaI, if P_i divides $[a_i,b_i]$ into N_i subintervals, then $P=(P_1,\dots,P_n)$ divides $[a_1,b_1]\times\cdots\times[a_n,b_n]$ into $N = N_1\cdot\cdots\cdot N_n$ subrectangles. These subrectangles will be called **subrectangles of the partition** P.

Suppose now that A is a rectangle, $f : A \to R$ is a bounded function, and P is a partition of A. For each subrectangle S of the partition let

$$m_S(f) = \inf\{f(x) : x \in S\},$$
$$M_S(f) = \sup\{f(x) : x \in S\},$$

ane let $v(S)$ be the volume of S [the **volume** of a rectangle a rectangle $[a_1,b_1]\times\cdots\times[a_n,b_n]$, and also of $(a_1,b_1)\times\cdots\times(a_n,b_n)$, is defined as $(b_1-a_1)\cdots(b_n-a_n)$]. The **lower** and **upper sums** of f for P are defined by

$$L(f,P) = \sum_S m_S(f)\cdot v(S),$$
$$U(f,P) = \sum_S M_S(f)\cdot v(S),$$

Clearly $L(f,P) \le U(f,P)$, and an even stronger assertion is true.

3-1 **Lemma**. Suppose the partition P' refines P (that is, each subrectangle of P' is contained in a subrectangle of P). Then

$$L(f,P) \le L(f,P') \quad \text{and} \quad U(f,P') \le U(f,P).$$

3-2 **Corollary**. If P and P' are any two partitions, then $L(f,P') \le U(f,P)$.

It follows from CorolIary 3-2 that the least upper bound of all lower sums for f is less than or equal to the greatest lower bound of all upper sums for f. A function $f : A \to R$ is called **integrable** on the rectangle A if f is bounded and $\sup\{L(f,P)\} = \inf\{U(f,P)\}$. This common number is then denoted $\int_A f$, and called the **integral** of f over A. Often, the notation $\int_A f(x^1,\dots,x^n)dx^1\cdots dx^n$ is used. If $f : [a.b] \to R$, where $a \le b$, then $\int_a^b f = \int_{[a,b]} f$. A simple but useful criterion for integrability is provided by the following Theorem.

3-3 **Theorem**. A bounded function $f : A \to R$ is integrable if and only if for every $\varepsilon > 0$ there is a partition P of A such that $U(f,P) - L(f,P) < \varepsilon$.

In the following sections we will characterize the integrable functions and discover a method of

computing integrals. For the present we consider two functions, one integrable and one not.

Example 1. Let $f : A \to R$ be a constant function, $f(x) = c$. Then any partition P we have $m_S(f) = M_S(f) = c$, so that $L(f,P) = U(f,P) = \sum_S c \cdot v(S) = c \cdot v(A)$. Hence $\int_A f = c \cdot v(A)$.

Example 2. Let $f : [0,1] \times [0,1] \to R$ be defined by

$$f(x,y) = \begin{cases} 0\_if\_x\_is\_rational, \\ 1\_if\_x\_isirrational. \end{cases}$$

If P is a partition, then every subrectangle S will contain points (x,y) with x rational, and also points (x,y) with x irrational. Hence $m_S(f) = 0$ and $M_S(f) = 1$, so

$$L(f,P) = \sum_S 0 \cdot v(S) = 0$$

and

$$U(f,P) = \sum_S 1 \cdot v(S) = v([0,1] \times [0,1]) = 1.$$

Therefore f is not integrable.

3.2 Measure Zero and Content Zero

A subset A of R^n has (n-dimensional) **measure** 0 if for every $\varepsilon > 0$ there is a cover $\{U_1, U_2, U_3, \ldots\}$ of A by closed rectangles such that $\sum_{i=1}^{\infty} v(U_i) < \varepsilon$. It is obvious (but nevertheless useful to remember) that if A has measure 0 and $B \subset A$, then B has measure 0. The reader may verify that open rectangles may be used instead of closed rectangles in the definition of measure 0.

A set of only finitely many points clearly has measure 0. If A has infinitely many points which can be arranged in a sequence $a_1, a_2, a_3, \ldots$, then A also has measure 0, for if $\varepsilon > 0$, we can choose U_i to be a closed rectangle containing a_i with $v(U_i) < \varepsilon / 2^i$. Then $\sum_{i=1}^{\infty} v(U_i) < \sum_{i=1}^{\infty} \varepsilon / 2^i = \varepsilon$.

The set of all rational numbers between 0 and 1 is an important and rather surprising example of an infinite set whose members can be arranged in such a sequence.

3-4 **Theorem** If $A = A_1 \cup A_2 \cup A_3 \cup \cdots$ and each A_i has measure 0, then A has measure 0.

A subset A of R^n has (n-dimensional) **content** 0 if for every $\varepsilon > 0$ there is a finite cover $\{U_1, U_2, U_3, \ldots\}$ of A by closed rectangles such that $\sum_{i=1}^{n} v(U_i) < \varepsilon$. If A has content 0, then A clearly has measure 0. Again, open rectangles could be used instead of closed rectangles in the definition.

3-5 **Theorem**. If $a < b$, then $[a,b] \subset R$ does not have content 0. In fact, if $\{U_1, \ldots, U_n\}$ is a finite cover of $[a,b]$ by closed intervals, then $\sum_{i=1}^{n} v(U_i) \geq b - a$.

If $a < b$, it is also true that $[a,b]$ does not have measure 0. This follows from

3-6 **Theorem**. If A is compact and has measure 0, then A has content 0.

The conclusion of Theorem 3-6 is false if A is not compact. For example, let A be the set of rational numbers between 0 and 1; then A has measure 0. Suppose, however, that $\{[a_1,b_1],\dots,[a_n,b_n]\}$ covers A. Then A is contained in the closed set $[a_1,b_1]\cup\cdots\cup[a_n,b_n]$, and therefore $[0,1]\subset[a_1,b_1]\cup\cdots\cup[a_n,b_n]$. It follows from Theorem 3-5 that $\sum_{i=1}^{n}(b_i-a_i)\geq 1$ for any such cover, and consequently A does not have content 0.

3.3 Integrable Functions

3-7 **Lemma**. Let A be a closed rectangle and let $f:A\to R$ be a bounded function such that $o(f,x)<\varepsilon$ for all $x\in A$. Then there is a partition P of A with $U(f,P)-L(f,P)<\varepsilon\cdot v(A)$.

3-8 **Theorem**. Let A be a closed rectangle and $f:A\to R$ a bounded function. Let $B=\{x: f$ is not continuous at $x\}$. Then f is integrable if and only if B is a set of measure 0.

We have thus far dealt only with the integrals of functions over rectangles. Integrals over other sets are easily reduced to this type. If $C\subset R^n$, the **characteristic function** χ_C is defined

$$\chi_C(x)=\begin{cases}0\_\ x\notin C,\\ 1\_\ x\in C.\end{cases}$$

If $C\subset A$ for some closed rectangle A and $f:A\to R$ is bounded, then $\int_C f$ is defined as $\int_A f\cdot\chi_C$, provided $f\cdot\chi_C$ is integrable. This certainly occurs if f and χ_C are integrable.

3-9 **Theorem**. The function $\chi_C:A\to R$ is integrable if and only if the boundary of C has measure 0 (and hence content 0).

A bounded set C whose boundary has measure 0 is called **Jordan-measurable**. The integral $\int_C 1$ is called the (n-dimensional) **content** of C, or the (n-dimensional) **volume** of C. Naturally one-dimensional volume is often called **length**, and two-dimensional volume, **area**.

Problem 3-11 shows that even an open set C may not be Jordan-measurable, so that $\int_C f$ is not necessarily defined even if C is open and f is continuous. This unhappy state of affairs will be rectified soon.

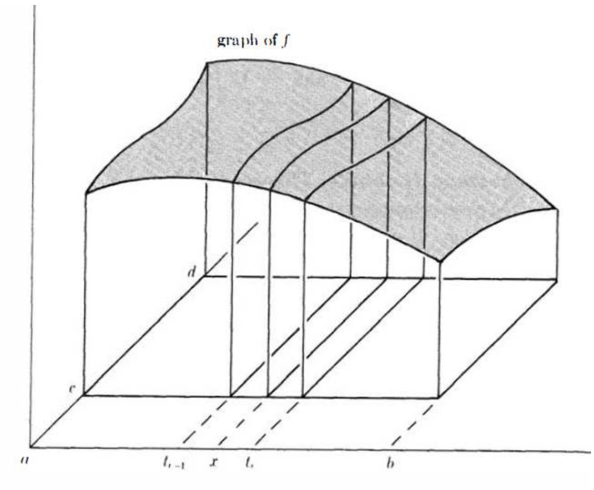
3.4 Fubini's Theorem

The problem of calculating integrals is solved, in some sense, by Theorem 3-10, which reduces the computation of integrals over a closed rectangle in $R^n, n>1$, to the computation of integrals over closed intervals in R. Of sufficient importance to deserve a special designation, this theorem is usually referred to as Fubini's theorem, although it is more or less a special case of a theorem proved by Fubini long after Theorem 3-10 was known.

The idea behind the theorem is best illustrated (Figure) for a positive continuous function $f:[a.b]\times[c,d]\to R$. Let $t_0,\dots,t_n$ be a partition of $[a,b]$ and divide $[a.b]\times[c,d]$ into n strips by means of the line segments $\{t_i\}\times[c,d]$. If g_x is defined by $g_x(y)=f(x,y)$, then the area of the region under the graph of f and above $\{x\}\times[c,d]$ is

$$\int_c^d g_x = \int_c^d f(x,y)dy .$$

The volume of the region under the graph of f and above $[t_{i-1},t_i]\times[c,d]$ is therefore approximately equal to $(t_i - t_{i-1})\cdot \int_c^d f(x,y)dy$, for any $x\in[t_{i-1},t_i]$. Thus



$$\int_{[a,b]\times[c,d]} f = \sum_{i=1}^{n} \int_{[t_{i-1},t_i]\times[c,d]} f$$

is approximately $\sum_{i=1}^{n}(t_i - t_{i-1})\cdot\int_c^d f(x_i,y)dy$,

with x_i in $[t_{i-1},t_i]$. On the other hand, sums similar to these appear in the definition of $\int_a^b\left(\int_c^d f(x,y)dy\right)dx$.

Thus, if h is defined by $h(x)=\int_c^d g_x = \int_c^d f(x,y)dy$, it is reasonable to hope that h is integrable on $[a,b]$ and that

$$\int_{[a,b]\times[c,d]} f = \int_a^b h = \int_a^b\left(\int_c^d f(x,y)dy\right)dx .$$

This will indeed turn out to be true when f is continuous, but in the general case difficulties arise. Suppose, for example, that the set of discontinuities of f is $\{x_0\}\times[c,d]$ for some $x_0\in[a,b]$. Then f is integrable on $[a,b]\times[c,d]$ but $h(x_0)=\int_c^d f(x_0,y)dy$ may not even be defined. The statement of Fubini's theorem therefore looks a little strange, and will be followed by remarks about various special cases where simpler statements are possible.

We will need one bit of terminology. If $f: A\to R$ is a bounded function on a closed rectangle, then, whether or not f is integrable, the least upper bound of all upper sums, and the greatest lower bound of all upper sums, both exist. They are called he **lower** and **upper integrals** of f on A, and denoted

$$\boldsymbol{L}\int_A f \quad \text{and} \quad \boldsymbol{U}\int_A f ,$$

respectively.

3-10 **Theorem (Fumini's Theorem).** Let $A\subset R^n$ and $B\subset R^m$ be closed rectangles, and let $f: A\times B\to R$ be integrable. For $x\in A$ let $g_x: B\to R$ be defined by $g_x(y)=f(x,y)$ and let

$$\mathsf{L}(x)=\boldsymbol{L}\int_B g_x = \boldsymbol{L}\int_B f(x,y)dy,$$
$$\mathsf{U}(x)=\boldsymbol{U}\int_B g_x = \boldsymbol{U}\int_B f(x,y)dy.$$

Then L and U are integrable on A and

$$\int_{A\times B} f = \int_A \mathrm{L} = \int_A \left(\boldsymbol{L}\int_B f(x,y)dy \right) dx,$$

$$\int_{A\times B} f = \int_A \mathrm{U} = \int_A \left(\boldsymbol{U}\int_B f(x,y)dy \right) dx.$$

(The integrals on the right side are called **iterated integrals** for f .)

Remarks. 1. A similar proof shows that

$$\int_{A\times B} f = \int_B \left(\boldsymbol{L}\int_A f(x,y)dx \right) dy = \int_B \left(\boldsymbol{U}\int_A f(x,y)dx \right) dy.$$

These integrals are called *iterated integrals* for f in the reverse order from those of the theorem. As several problems show, the possibility of interchanging the orders of iterated integrals has many consequences.

2. In practice it is often the case that each g_x is integrable, so that $\int_{A\times B} f = \int_A \left(\int_B f(x,y)dy \right) dx$. This certainly occurs if f is continuous.

3. The worst irregularity commonly encountered is that g_x is not integrable for a finite number of $x \in A$. In this case $\mathrm{L}(x) = \int_B f(x,y)dy$ for all but these finitely many x . Since $\int_A \mathrm{L}$ remains unchanged if L is defined at a finite number of points, we can still write $\int_{A\times B} f = \int_A \left(\int_B f(x,y)dy \right) dx$, provided that $\int_B f(x,y)dy$ is defined arbitrarily, say as 0, when it does not exist.

4. There are cases when this will not work and Theorem 3-10 must be used as stated. Let $f : [0.1]\times[0,1] \to R$ be defined by

$$f(x,y) = \begin{cases} 1\_if\_x\_is\_irational \\ 1\_if\_x\_is\_rational\_and\_y\_is\_irrational \\ 1-1/q\_if\_x = p/q\_in\_lowest\_terms\_and\_y\_is\_rational. \end{cases}$$

Then f is integrable and $\int_{[0,1]\times[0,1]} f = 1$. Now $\int_0^1 f(x,y) = 1$ if x is irrational, and does not exist if x is rational.

Therefore h is not integrable if $h(x) = \int_0^1 f(x,y)$ is set equal to 0 when the integral does not exist.

5. If $A = [a_1, b_1]\times\cdots\times[a_n, b_n]$ and $f : A \to R$ is sufficiently nice, we can apply Fubini's theorem repeatedly to obtain

$$\int_A f = \int_{a_n}^{b_n} \left(\cdots \left(\int_{a_1}^{b_1} f(x^1,\dots,x^n)dx^1 \right) \cdots \right) dx^n .$$

6. If $C \subset A\times B$, Fubini's theorem can be used to evaluate $\int_C f$, since this is by definition $\int_{A\times B} \chi_C f$. Suppose, for example, that

$$C = [-1,1]\times[-1,1] - \{(x,y) : |(x,y)| < 1\}.$$

Then

$$\int_C f = \int_{-1}^{1}\left(\int_{-1}^{1} f(x,y)\cdot\chi_C(x,y)dy\right)dx$$

Now

$$\chi_C(x,y) = \begin{cases} 1\_if\_y > \sqrt{1-x^2}\_or\_y < -\sqrt{1-x^2}, \\ 0\_otherwise. \end{cases}$$

Therefore

$$\int_{-1}^{1} f(x,y)\cdot\chi_C(x,y)dy = \int_{-1}^{-\sqrt{1-x^2}} f(x,y)dy + \int_{\sqrt{1-x^2}}^{1} f(x,y)dy.$$

In general, if $C \subset A \times B$, then main difficulty in deriving expressions for $\int_C f$ will be determining $C \cap (\{x\} \times B)$ for $x \in A$. If $C \cap (A \times \{y\})$ for $y \in B$ is easier to determine, one should use the iterated integral

$$\int_C f = \int_B \left(\int_A f(x,y)\cdot\chi_C(x,y)dx\right)dy.$$

3-5 Partitions of Unity

In this section we introduce a tool of extreme importance in the theory of integration.

3-11 **Theorem**. Let $A \subset R^n$ and let O be and open cover of A. Then there is a collection Φ of C^∞ functions φ defined in an open set containing A, with the following properties:

(1) For each $x \in A$ we have $0 \le \varphi(x) \le 1$.

(2) For each $x \in A$ there is an open set V containing x such that all but finitely many $\varphi \in \Phi$ are 0 on V.

(3) For each $x \in A$ we have $\sum_{\varphi\in\Phi}\varphi(x) = 1$ (by (2) for each x this sum is finite in some open set containing x).

(4) For each $\varphi \in \Phi$ there is an open set U in O such that $\varphi = 0$ outside of some closed set containing in U.

(A collection Φ satisfying (1) to (3) is called a C^∞ **partition of unity** for A. If Φ also satisfies (4), it is said to be **subordinate** to the cover O. In this chapter we will only use continuity of the functions φ.)

An important consequence of condition (2) of the theorem should be noted. Let $C \subset A$ be compact. For each $x \in C$ there is an open set V_x containing x such that only finitely many $\varphi \in \Phi$ are not 0 on V_x. Since C is compact, finitely many such V_x cover C. Thus only finitely many $\varphi \in \Phi$ are not 0 on C.

One important application of partitions of unity will illustrate their main role – piecing together results obtained locally. An open cover O of an open set $A \subset R^n$ is **admissible** if each $U \in O$ is contained in A. If Φ is subordinate to O, $f : A \to R$ is bounded in some open set around each point of A, and $\{x : f$ is discontinuous at $x\}$ has measure 0, then each $\int_A \varphi\cdot|f|$ exists. We define f to be **integrable** (in the extended sense) if $\sum_{\varphi\in\Phi}\int_A \varphi\cdot|f|$ converges (the proof of Theorem 3-11 shows that the φ's may be arranged in a sequence). This

implies convergence of $\sum_{\varphi\in\Phi}\left|\int_A \varphi\cdot f\right|$, and hence absolute convergence of $\sum_{\varphi\in\Phi}\int_A \varphi\cdot f$, which we define to be $\int_A f$. These definitions do not depend on O or Φ .

3-12 **Theorem**.

(1) If Ψ is another partition of unity, subordinate to an admissible cover O' of A, then $\sum_{\psi\in\Psi}\int_A \psi\cdot|f|$ also converges, and

$$\sum_{\psi\in\Psi}\int_A \psi\cdot|f| = \sum_{\psi\in\Psi}\int_A \psi\cdot|f| .$$

(2) If A and f are bounded, then f is integrable in the extended sense.

(3) If A is Jordan-measurable and f is bounded, then this definition of $\int_A f$ agrees with the old one.

*****

From Matsumoto[1].

Theorem 14.1 Let M be a compact C^r -class manifold and let U,V be open sets in M and $M=U\cup V$. Then there are C^r -class functons on M , $f:M\to R$ and $g:M\to R$ and the followind three conditions hold:

(1) $0\le f\le 1$, $0\le g\le 1$

(2) $\sup p(f)\subset U$, $\sup p(g)\subset U$

(3) $f+g\equiv 1$ (i.e., for any $p\in M$, $f(p)+f(p)=1$

Theorem 14.4 Let M be a σ -compact C^r -class manifold and $\{U_\alpha\}_{\alpha\in A}$ is any open cover of M . Then there is a countable C^r -class function $f_j:M\to R$ $(j=1,2,3,\ldots)$ and the following conditons hold.

(1) $0\le f\le 1$

(2) $\{\mathrm{supp}(f_j)\}_{j=1}^{\infty}$ is a local cover of M , and also a refinement of $\{U_\alpha\}_{\alpha\in A}$,

(3) $\sum_{j=1}^{\infty} f_j\equiv 1$

*****

3.6 Change of Variable

If $g:[a,b]\to R$ is continuously differentiable and $f:R\to R$ is continuous, then, as is well known,

$$\int_{g(a)}^{g(b)} f=\int_a^b (f\circ g)\cdot g' .$$

[1] 松本幸夫「多様体の基礎」東京大学出版会，1988, pp186-191

The proof is very simple: if $F' = f$, then $(F \circ g)' = (f \circ g) \cdot g'$; thus the left side is $F(b(b)) - F(g(a))$, while the right side is $F \circ g(b) - F \circ g(a) = F(g(b)) - F(g(a))$.

We leave it to the reader to show that if g is one to one, then the above formula can be written

$$\int\limits_{g((a,b))} f = \int\limits_{(a,b)} f \circ g \cdot |g'| .$$

(Consider separately the cases where g is increasing and where g is decreasing.) The generalization of this formula to higher dimensions is by no means so trivial.

3-13 **Theorem.** Let $A \subset R^n$ be an open set and $g : A \to R^n$ a one-to-one, continuously differentiable function such that $\det g'(x) \neq 0$ for all $x \in A$. If $f : g(A) \to R$ is integrable, then

$$\int\limits_{g(A)} f = \int\limits_{A} (f \circ g) |\det g'| .$$

The condition $\det g'(x) \neq 0$ may be eliminated from the hypotheses of Theorem 3-13 by using the following theorem, which often plays ad unexpected role.

3-14. **Theorem (Sard's Theorem).** Let $g : A \to R^n$ be continuously differentiable. where $A \subset R^n$ is open, and let $B = \{x \in A : \det g'(x) = 0\}$. Then $g(B)$ has measure 0.

Theorem 3-14 is actually only the easy part of Sard's Theorem. The statement and proof of the deeper result will be found in Ref. 17, page 47.

"A mathematician is one to whom *that* is as obvious as that twice two makes four is to you. Liouville was a mathematician." Lord Kelvin

4. Integration on Chains

4.1 Algebraic Preliminaries

If V is a vector space (over R), we will denote the k-fold product $V \times \cdots \times V$ by V^k . A function $T: V^k \to R$ is called **multilinear** if for each i with $1 \le i \le k$ we have

$$T(v_1,\ldots,v_i+v_i',\ldots,v_k)=T(v_1,\ldots,v_i,\ldots,v_k)+T(v_1,\ldots,v_i',\ldots,v_k)$$
$$T(v_1,\ldots,av_i,\ldots,v_k)=aT(v_1,\ldots,v_i,\ldots,v_k).$$

A multilinear function $T: V^k \to R$ is called a **k-tensor** on V and the set of all k-tensors, denoted $T^k(V)$, becomes a vector space (over R) if for $S, T \in T^k(V)$ and $a \in R$ we define

$$(S+T)(v_1,\ldots,v_k)=S(v_1,\ldots,v_k)+T(v_1,\ldots,v_k),$$
$$T(v_1,\ldots,av_i,\ldots,v_k)=aT(v_1,\ldots,v_i,\ldots,v_k).$$

There is also an operation connecting the various spaces $T^k(V)$. If $S \in T^k(V)$ and $T \in T^l(V)$, we define the **tensor product** $S \otimes T \in T^{k+l}(V)$ by

$$S \otimes T(v_1,\ldots,v_k,v_{k+1},\ldots,v_{k+l})=S(v_1,\ldots,v_k)\cdot T(v_{k+1},\ldots,v_{k+l}).$$

Note that the order of the factors S and T is crucial here since $S \otimes T$ and $T \otimes S$ are far from equal. The following properties of $\otimes$ are left as easy exercises for the reader.

$$(S_1+S_2)\otimes T=S_1\otimes T+S_2\otimes T,$$
$$S\otimes(T_1+T_2)=S\otimes T_1+S\otimes T_2,$$
$$(aS)\otimes T=S\otimes(aT)=a(S\otimes T),$$
$$(S\otimes T)\otimes U=S\otimes(T\otimes U).$$

Both $(S\otimes T)\otimes U$ and $S\otimes(T\otimes U)$ are usually denoted simply $S\otimes T\otimes U$; higher-order products $T_1\otimes\cdots\otimes T_r$ are defined similarly.

The reader has probably already noticed that $T^1(V)$ is just the dual space V^* . The operation $\otimes$ allows us to express the other vector spaces $T^k(V)$ in terms of $T^1(V)$.

4-1 **Theorem**. Let $v_1,\ldots,v_n$ be a basis for V , and let $\varphi_1,\ldots,\varphi_n$ be the dual basis, $\varphi_i(v_j)=\delta_{ij}$. Then the set of all k-fold tensor products

$$\varphi_{i1}\otimes\cdots\otimes\varphi_{ik} \qquad 1\le i_1,\ldots,i_k\le n$$

is a bases for $T^k(V)$, which therefore has dimension n^k .

One important construction, familiar for the case of dual spaces, can also be made for tensors. If $f: V \to W$ is a linear transformation, a linear transformation $f^*: T^k(W)\to T^k(V)$ is defined by

$$f^*T(v_1,\ldots,v_k)=T(f(v_1),\ldots,f(v_k))$$

for $T\in T^k(W)$ and $v_1,\ldots,v_k\in V$. It is easy to verify that $f^*(S\otimes T)=f^*S\otimes f^*T$.

The reader is already familiar with certain tensors, aside from members of V^* . The first example is the inner product $\langle,\rangle\in T^2(R^n)$. On the grounds that any good mathematical commodity is worth generalizing, we define an **inner product** on V to be a 2-tensor T such that T is **symmetric**, that is $T(v,w)=T(w,v)$ for $v,w\in V$ and such that T is **positive-definite**, that is, $T(v,v)>0$ if $v\neq 0$. We distinguish $\langle,\rangle$ as the **usual inner product** on R^n .

The following theorem shows that our generalization is not too general.

4-2 **Theorem**. If T is an inner product on V , here is a basis $v_1,\dots,v_n$ for V such that $T(v_i,v_j)=\delta_{ij}$. (Such a basis is called **orthonormal** with respect to T .) Consequently there is an isomorphism $f: R^n \to V$ such that $T(f(x),f(y))=\langle x,y\rangle$ for $x,y\in R^n$. In other words $f^*T=\langle,\rangle$.

Despite its importance, the inner product plays a far lesser role than another familiar, seemingly ubiquitous function, the tensor $\det \in T^n(R^n)$. In attempting to generalize this function, we recall that interchanging two rows of a matrix changes the sign of its determinant. This suggests the following definition. A k-tensor $\omega \in T^k(V)$ is called **alternating** if

$$\omega(v_1,\dots,v_i,\dots,v_j,\dots,v_k)=-\omega(v_1,\dots,v_j,\dots,v_i,\dots,v_k) \quad \text{for all } v_1,\dots,v_k\in V .$$

(In this equation v_i and v_j are interchanged and all other $v's$ are left fixed.) The set of all alternating k-tensors is clearly a subspace Λ^k of $T^k(V)$. Since it requires considerable work to produce the determinant, it is not surprising that alternation k-tensors are difficult to write down. There is, however, a uniform way of expressing all of them. Recall that the sign of a permutation σ , denoted $\operatorname{sgn}\sigma$, is +1 if σ is even and -1 if σ is odd. If $T\in T^k(V)$, we define $Alt(T)$ by

$$Alt(T)(v_1,\dots,v_k)=\frac{1}{k!}\sum \operatorname{sgn}\sigma\cdot T(v_{\sigma(1)},\dots,v_{\sigma(k)}),$$

where S_k is the set of all permutations of the number 1 to k .

4-3 **Theorem**.

(1) If $T\in T^k(V)$, then $Alt(T)\in \Lambda^k(V)$.

(2) If $\omega\in\Lambda^k(V)$, then $Alt(\omega)=\omega$.

(3) If $T\in T^k(V)$, then $Alt(Alt(T))=Alt(T)$.

To determine the dimensions of $\Lambda^k(V)$, we would like a theorem analogous to Theorem 4-1. Of course, if $\omega\in\Lambda^k(V)$ and $\eta\in\Lambda^l(V)$, then $\omega\otimes\eta$ is usually not in $\Lambda^{k+l}(V)$. We will therefore define a new product, the **wedge** product $\omega\wedge\eta\in\Lambda^{k+l}(V)$ by

$$\omega\wedge\eta=\frac{(k+l)!}{k!l!}Alt(\omega\otimes\eta).$$

(The reason for the strange coefficient will appear later.) The following properties of $\wedge$ are left as an exercise for the reader.

$$(\omega_1+\omega_2)\wedge\eta=\omega_1\wedge\eta+\omega_2\wedge\eta,$$
$$\omega\wedge(\eta_1+\eta_2)=\omega\wedge\eta_1+\omega\wedge\eta_2,$$
$$a\omega\wedge\eta=\omega\wedge a\eta=a(\omega\wedge\eta),$$
$$\omega\wedge\eta=(-1)^{kl}\eta\wedge\omega,$$
$$f^*(\omega\wedge\eta)=f^*(\omega)\wedge f^*(\eta).$$

The equation $(\omega\wedge\eta)\wedge\theta=\omega\wedge(\eta\wedge\theta)$ is true but requires more work.

4-4 **Theorem**

(1) If $S\in T^k(V)$ and $T\in T^k(V)$ and $Alt(S)=0$, then

$$Alt(S\otimes T)=Alt(T\otimes S)=0.$$

(2) $Alt(Alt(\omega\otimes\eta)\otimes\theta)=Alt(\omega\otimes\eta\otimes\theta)=Alt(\omega\otimes Alt(\eta\otimes\theta))$.

(3) If $\omega \in \Lambda^k(V), \eta \in \Lambda^l(V)$, and $\theta \in \Lambda^m(V)$, then

$$(\omega \wedge \eta) \wedge \theta = \omega \wedge (\eta \wedge \theta) = \frac{(k+l+m)!}{k!l!m!} Alt(\omega \otimes \eta \otimes \theta) .$$

Naturally $\omega \wedge (\eta \wedge \theta)$ and $(\omega \wedge \eta) \wedge \theta$ are both denoted simply $\omega \wedge \eta \wedge \theta$, and higher-order products $\omega_1 \wedge \cdots \wedge \omega_r$ are defined similarly. If $v_1, \ldots, v_n$ is a basis for V and $\varphi_1, \ldots, \varphi_n$ is the dual basis, a basis for $\Lambda^k(V)$ can now be constructed quite easily.

4-5 **Theorem**. The set of all

$$\varphi_{i1} \wedge \cdots \wedge \varphi_{ik} \qquad 1 \le i_1 < i_2 < \cdots < i_k \le n$$

is a basis for $\Lambda^k(V)$, which therefore has dimension

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} .$$

If V has dimension n, it follows from Theorem 4-5 that $\Lambda^n(V)$ has dimension 1. Thus all alternating n-tensors on V are multiples of any non-zero one. Since the determinant is an example of such a member of $\Lambda^n(R^n)$, it is not surprising to find it in the following theorem.

4-6 **Theorem**. Let $v_1, \ldots, v_n$ be a basis for V, and let $\omega \in \Lambda^k(V)$. If $w_i = \sum_{j=1}^n a_{ij} v_j$ are n vectors in V, then

$$\omega(w_1, \ldots, w_n) = \det(a_{ij}) \cdot \omega(v_1, \ldots, v_n) .$$

Theorem 4-6 shows that a non-zero $\omega \in \Lambda^k(V)$ splits the bases on V into two disjoint groups, those with $\omega(v_1, \ldots, v_n) < 0$; if $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are two bases and $A = (a_{ij})$ is defined by $w_i = \sum_{j=1}^n a_{ij} v_j$, then $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are in the same group if and only if $\det A > 0$. This criterion is independent of ω and can always be used to divide the bases of V into two disjoint groups. Either of these two groups is called an **orientation** for V. The orientation to which a basis $v_1, \ldots, v_n$ belongs is denoted $[v_1, \ldots, v_n]$ and the other orientation is denoted $-[v_1, \ldots, v_n]$. In R^n we define the **usual orientation** as $[e_1, \ldots, e_n]$.

The fact that $\dim \Lambda^n(R^n) = 1$ is probably not new to you, since det is often defined as the unique element $\omega \in \Lambda^n(R^n)$ such that $\omega(e_1, \ldots, e_n) = 1$. For a general vector space V there is no extra criterion of this sort to distinguish a particular $\omega \in \Lambda^n(V)$. Suppose, however, that an inner product T for V is given. If $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are two bases which are orthonormal with respect to T, and the matrix $A = (a_{ij})$ is defined by $w_i = \sum_{j=1}^n a_{ij} v_j$, then

$$\begin{aligned} \delta_{ij} = T(w_i, w_j) &= \sum_{k,l=1}^n a_{ik} a_{jl} T(v_k, v_l) \\ &= \sum_{k=1}^n a_{ik} a_{jk} \end{aligned}$$

In other words, if A^T denotes the transpose of the matrix V, then we have $A \cdot A^T = I$, so $\det A = \pm 1$. It follows from Theorem 4-6 that if $\omega \in \Lambda^n(V)$ satisfies $\omega(v_1, \ldots, v_n) = \pm 1$, then $\omega(w_1, \ldots, w_n) = \pm 1$. If an orientation μ for V has also been given, it follows that there is a unique $\omega \in \Lambda^n(V)$ such that whenever $v_1, \ldots, v_n$ is an orthonormal basis such that $[v_1, \ldots, v_n] = \mu$. This unique ω is called the **volume element** of V, determined by the

inner product T and orientation μ. Note that det is the volume element of R^n determined by the usual inner product and usual orientation, and that $|\det(v_1,\dots,v_n)|$ is the volume of the parallelepiped spanned by the line segments from 0 to each of $v_1,\dots,v_n$.

We conclude this section with a construction which we will restrict to R^n and φ is defined by

$$\varphi(w)=\begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix},$$

then $\varphi \in \Lambda^1(R^n)$; therefore there is a unique $z \in R^n$ such that

$$\langle w,z\rangle=\varphi(w)=\det\begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \\ w \end{pmatrix}$$

This z is denoted $v_1\times\cdots\times v_{n-1}$ and called the **cross product** of $v_1,\dots,v_{n-1}$. The following properties are immediate from the definition:

$$v_{\sigma(1)}\times\cdots\times v_{\sigma(n-1)}=\operatorname{sgn}\sigma\cdot v_1\times\cdots\times v_{n-1},$$

$$v_1\times\cdots\times av_i\times\cdots\times v_{n-1}=a\cdot(v_1\times\cdots\times v_{n-1}),$$

$$v_1\times\cdots\times(v_i+v_i')\times\cdots\times v_{n-1}=v_1\times\cdots\times v_i\times\cdots\times v_{n-1}+v_1\times\cdots\times v_i'\times\cdots\times v_{n-1}.$$

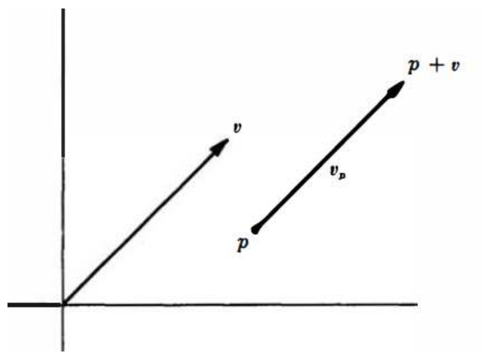
It is uncommon in mathematics to have a "product" that depends on more than two factors. In the case of two vectors $v,w\in R^3$, we obtain a more conventional looking product, $v\times w\in R^3$. For this reason it is sometimes maintained that the cross product ca be defined only in R^3.

4.2 Fields and Forms

If $p\in R^n$, the set of all pairs (p,v), for $v\in R^n$, is denoted $R^n{}_p$, and called the **tangent space** of R^n at p. This set is made into a vector space in the most obvious way, by defining

$$(p,v)+(p,w)=(p.v+w),$$

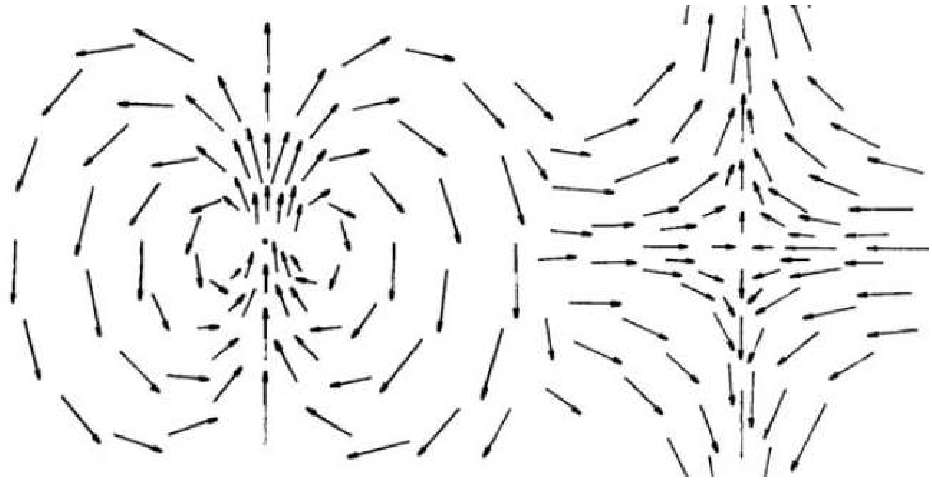
$$a\cdot(p,v)=(a,av).$$



A vector $v\in R^n$ is often pictured as an arrow from 0 to v; the vector $(p,v)R^n{}_p$ may be pictured (Figure) as an arrow with the same direction and length, but with initial point p. This arrow goes from p to the point $p+v$, and we therefore define $p+v$ to be the **end point** of (p,v). We will usually write (p,v) as v_p (read: the vector v at p).

The vector space $R^n{}_p$ is so closely allied to R^n that many of the structures on R^n have analogous on $R^n{}_p$. In particular the **usual inner product** $\langle,\rangle_p$ for $R^n{}_p$ is defined by $\langle v_p,w_p\rangle_p=\langle v,w\rangle$, and the **usual orientation** for $R^n{}_p$ is $\lfloor(e_1)_p,\dots,(e_n)_p\rfloor$.

Any operation which is possible in a vector space may be performed in each $R^n{}_p$, and most of this section is merely an elaboration of this theme. About the simplest operation in a vector space is the selection of a vector from it. If such a selection is made in each $R^n{}_p$, we obtain a **vector field** (Figure). To be precise, a vector



field is a function F such that $F(p) \in R^n{}_p$ for each $p \in R^n$. For each p there are numbers $F^1(p), \ldots, F^n(p)$ such that

$$F(p) = F^1(p) \cdot (e_1)_p + \cdots + F^n(p) \cdot (e_n)_p .$$

We thus obtain **n component function** $F^i : R^n \to R$. The vector field F is called continuous, differentiable, etc., if the functions F^i are. Similar definitions ca be made for a vector field defined only on an open subset of R^n. Operations on vectors yield operations on vector fields when applied at each point separately. For example, if F and G are vector fields and f is a function, we define

$$(F+G)(p) = F(p) + G(p),$$
$$\langle F, G \rangle(p) = \langle F(p), G(p) \rangle,$$
$$(f \cdot F)(p) = f(p)F(p).$$

If $F_1, \ldots, F_{n-1}$ are vector fields on R^n, then we can similarly define

$$(F_1 \times \cdots F_{n-1})(p) = F_1(p) \times \cdots \times F_{n-1}(p).$$

Certain other definitions are standard and useful. We define the **divergence**, $divF$ of F, as $\sum_{i=1}^n D_i F^i$. If we introduce the formal symbolism

$$\nabla = \sum_{i=1}^n D_i \cdot e_i$$

we can write, symbolically, $divF = \langle \nabla, F \rangle$. If $n = 3$ we write, in conformity with this symbolism,

$$(\nabla \times F)(p) = (D_2 F^3 - D_3 F^2)(e_1)_p + (D_3 F^1 - D_1 F^3)(e_2)_p + (D_1 F^2 - D_2 F^1)(e_3)_p .$$

The vector field $\nabla \times F$ is called $curlF$. The names "divergence" and "curl" are derived from physical considerations which are explained at the end of this book.

Many similar considerations may be applied to a function ω with $\omega(p) \in \Lambda^k(R^n{}_p)$; such a function is called a **k-form** on R^n, or simply a **differential form**. If $\varphi_1(p), \ldots, \varphi_n(p)$ is the dual basis to $(e_1)_p, \ldots, (e_n)_p$

then

$$\omega(p) = \sum_{i1<\cdots<ik} \omega_{i1,\dots,ik}(p) \cdot [\varphi_{i1}(p) \wedge \cdots \wedge \varphi_{ik}(p)]$$

for certain functions $\omega_{i1,\dots,ik}$; the form ω is called continuous, differentiable, etc., if these functions are. We shall usually assume tacitly that forms and vectors fields are differentiable, and "differentiable" will henceforth mean "C^∞"; this is a simplifying assumption that eliminates the need for counting how many times a function is differentiated in a proof. The sum $\omega + \eta$, product $f \cdot \omega$, and wedge product $\omega \wedge \eta$ are defined in the obvious way. A function f is considered to be a 0-form and $f \cdot \omega$ is also written $f \wedge \omega$.

If $f : R^n \to R$ is differentiable, then $Df(p) \in \Lambda^1(R^n)$. By a minor modification we therefore obtain a 1-form df , defined by

$$df(p)(v_p) = Df(p)(v) .$$

Let us consider in particular the 1-forms $d\pi^i$. It is customary to let x^i denote the function π^i . (On R^3 we often denote x^1, x^2, and x^3 by x, y, and z .) This standard notation has obvious disadvantages but it allows many classical results to be expressed by formulas of equally classical appearance. Since

$$dx^i(p)(v_p) = d\pi^i(p)(v_p) = D\pi^i(p)(v) = v \quad ,$$

we see that $dx^1(p), \dots, dx^n(p)$ is just the dual basis to $(e_1)_p, \dots, (e_n)_p$. Thus every k-form ω ca be written

$$\omega = \sum \omega_{i1,\dots,ik} dx^{i1} \wedge \cdots \wedge dx^{ik} .$$

The expression for df is of particular interest.

4-7 **Theorem**. If $f : R^n \to R$ is differentiable, then

$$df = D_1 f \cdot dx^1 + \cdots + D_n f \cdot dx^n .$$

In classical notation,

$$df = \frac{\partial f}{\partial x^1} dx^1 + \cdots + \frac{\partial f}{\partial x^n} dx^n .$$

If we consider now a differentiable function $f : R^n \to R^m$ we have a linear transformation $Df(p) : R^n \to R^m$. Another minor modification therefore produces a linear transformation defined by

$$f_*(v_p) = (Df(p)(v))_{f(p)} .$$

This linear transformation induces a linear transformation $f^* \colon \Lambda^k(R^m{}_{f(p)}) \to \Lambda^k(R^n{}_p)$. If ω is a k-form on R^m we can therefore define a k-form $f^*\omega$ on by $(f^*\omega)(p) = f^*(\omega(f(p)))$. Recall this means that if $v_1, \dots, v_k \in R^n{}_p$, then we have $f^*\omega(p)(v_1, \dots, v_k) = \omega(f(p))(f_*(v_1), \dots, f_*(v_k))$. As an antidote to the abstractness of these definitions we present a theorem, summarizing the important properties of f^* , which allows explicit calculations of $f^*\omega$.

4-8 **Theorem**. If $f : R^n \to R^m$ is differentiable, then

(1) $f^*(dx^i) = \sum_{j=1}^{n} D_j f^i \cdot dx^j = \sum_{j=1}^{n} \frac{\partial f^i}{\partial x^j} dx^j$.

(2) $f^*(\omega_1 + \omega_2) = f^*(\omega_1) + f^*(\omega_2)$.

(3) $f^*(g \cdot \omega) = (g \circ f) \cdot f^*\omega$.

(4) $f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta$.

By repeatedly applying Theorem 4-8 we have, for example,

$$f^*(Pdx^1 \wedge dx^2 + Pdx^2 \wedge dx^3) = (P \circ f)\left[f^*(dx^1) \wedge f^*(dx^2)\right] + (Q \circ f)\left[f^*(dx^2) \wedge f^*(dx^3)\right].$$

The expression obtained by expanding out each $f^*(dx^i)$ is quite complicated. (It is helpful to reminder, however, that we have $dx^i \wedge dx^i = (-1) dx^i \wedge dx^i = 0$.) In one special case it will be worth our while to make an explicit evaluation.

4-9 **Theorem**. If $f : R^n \to R^n$ is differentiable, then

$$f^*(hdx^1 \wedge \cdots \wedge dx^n) = (h \circ f)(\det f') dx^1 \wedge \cdots \wedge dx^n .$$

An important construction associated with forms is a generalization of the operator d which changes 0-forms into 1-forms. If

$$\omega = \sum_{i1 < \cdots < ik} \omega_{i1,\ldots,ik} dx^{i1} \wedge \cdots \wedge dx^{ik} ,$$

we define a (k+1)-form $d\omega$, the **differential** of ω , by

$$\begin{aligned} d\omega &= \sum_{i1 < \cdots < ik} d\omega_{i1,\ldots,ik} \wedge dx^{i1} \wedge \cdots \wedge dx^{ik} \\ &= \sum_{i1 < \cdots < ik} \sum_{\alpha=1}^{n} D_\alpha(\omega_{i1,\ldots,ik}) \cdot dx^\alpha \wedge dx^{i1} \wedge \cdots \wedge dx^{ik} \end{aligned} .$$

4-10 **Theorem**

(1) $d(\omega + \eta) = d\omega + d\eta$.

(2) If ω is a k-form and η is an l-form, then $d(\omega \wedge \eta) = d\omega + (-1)^k \omega \wedge \eta$.

(3) $d(d\omega) = 0$. Briefly, $d^2 = 0$.

(4) If ω is a k-form on R^m and $f : R^n \to R^m$ is differentiable, then $f^*(d\omega) = d(f^*\omega)$.

A form ω is called **closed** if $d\omega = 0$ and **exact** if $\omega = d\eta$, for some η . Theorem 4-10 shows that every exact form is closed, ant it is natural to ask whether, conversely, every closed form is exact. If ω is the 1-form $Pdx + Qdy$ on R^2 , then

$$\begin{aligned} d\omega &= (D_1 Pdx + D_2 Pdy) \wedge dx + (D_1 Qdx + D_2 Qdy) \wedge dy \\ &= (D_1 Q - D_2 P) dx \wedge dy \end{aligned} .$$

Thus, if $d\omega = 0$, then $D_1 Q = D_2 P$. Problems 2-21 and 3-34 show that there is a 0-form f such that $\omega = df = D_1 f dx + D_2 f dy$. If ω is defined only on a subset of R^2 , however, such a function may not exist. The classical example is the form

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

defined on $R^2 - 0$. This form is usually denoted $d\theta$ (where θ is defined in Problem 3-41), since (Problem 4-21) it equals $d\theta$ on the set $\{(x, y) : x < 0, or\_x \geq 0\_and\_y \neq 0\}$, where θ is defined. Note, however, that θ

cannot be defined continuously on all of $R^2 - 0$. If $\omega = df$ for some function $f : R^2 - 0 \to R$, then $D_1 f = D_1 \theta$ and $D_2 f = D_2 \theta$, so $f = \theta + cons\tan t$, showing that such an f cannot exist.

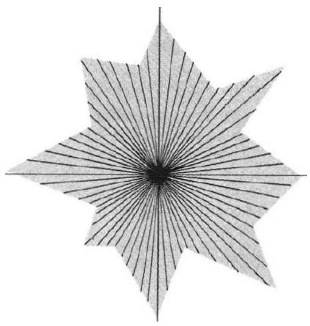
Suppose that $\omega = \sum_{i=1}^{n} \omega_i dx^i$ is a 1-form on R^n and ω happens to equal $df = \sum_{i=1}^{n} D_i f \cdot dx^1$. We can clearly assume that $f(0) = 0$. As in Problem 2-35, we have

$$\begin{aligned} f(x) &= \int_0^1 \frac{d}{dt} f(tx)dt \\ &= \int_0^1 \sum_{i=1}^{n} D_i f(tx) \cdot x^i dt . \\ &= \int_0^1 \sum_{i=1}^{n} \omega_i(tx) \cdot x^i dt \end{aligned}$$

This suggests that in order to find f, given ω, we consider the function $I\omega$, defined by

$$I\omega(x) = \int_0^1 \sum_{i=1}^{n} \omega_i(tx) \cdot x^i dt .$$

Note that the definition of $I\omega$ makes sense if ω is defined only on an open set $A \subset R^n$ with the property that whenever $x \in A$, the linesegment from 0 to x is contained in A; such an open set is called **star-shaped** with respect to 0 (Figure). A somewhat involved calculation shows that (on a star-shaped open set) we have $\omega = d(I\omega)$ provided that ω satisfies the necessary condition $d\omega = 0$. The calculation, as well as the definition of $I\omega$, may be generalized considerably:



4-11 **Theorem (Poincaré Lemma).** If $A \subset R^n$ is an open set star-shaped with respect to 0, then every closed form on A is exact.

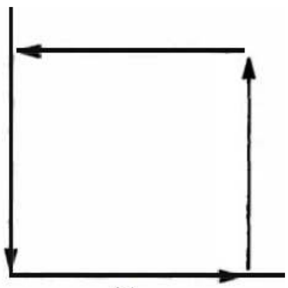
4.3 Geometric Preliminaries

A **singular n-cube** in $A \subset R$ is a continuous function $c : [0,1]^n \to A$ (here $[0,1]^n$ denotes the n-fold product $[0,1] \times \cdots \times [0,1]$). we let R^0 and $[0,1]^0$ both denote $\{0\}$. A singular 0-cube in A is then a function $f : \{0\} \to A$ or, what amounts to the same thing, a point in A. A singular 1-cube is often called a **curve**. A particularly simple, but particularly important example of a singular n-cube in R^n is the **standard n-cube** $I^n : [0,1]^n \to R^n$ denoted by $I^n = x$ for $x \in [0,1]^n$.

We shall need to consider formal sums of singular n-cubes in A multiplied by integers, that is, expressions like

$$2c_1 + 3c_2 - 4c_3 ,$$

where c_1, c_2, c_3 are singular n-cubes in A. Such a finite sum of singular n-cubes with integer coefficients is called an **n-chain** in A. In particular a singular n-cube c is also considered as an n-chain $1 \cdot c$. It is clear how n-chains can be



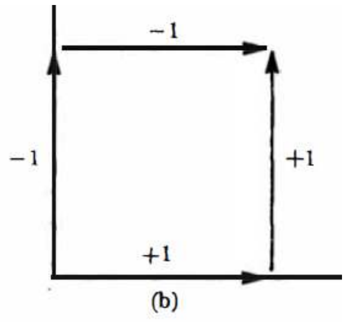
(a)

added, and multiplied by integers. For example

$$2(c_1+3c_4)+(-2)(c_1+c_3+c_2)=-2c_2-2c_3+6c_4 .$$

(A rigorous exposition of this formalism is presented in Problem 4-22.)

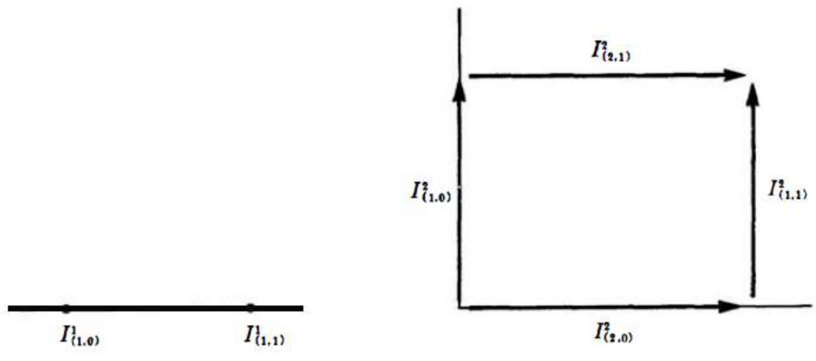
For each singular n-chain c in A we shall define an (n-1)-chain in A called the **boundary** of c and denoted ∂c. The boundary of I^2, for example, might be defined as the sum of four singular 1-cubes arranged counterclockwise around the boundary of $[0,1]^2$, as indicated in Figure(a). It is actually much more convenient to define ∂I^2 as the sum, with the indicated coefficients, of the four singular 1-cubes shown in Figure (b). The precise definition of ∂I^n requires some preliminary notions. For each i with $1 \le i \le n$ we define two singular (n-1)-cubes $I^n{}_{(i,0)}$ and $I^n{}_{(i,1)}$ as follows. If $x \in [0,1]^{n-1}$, then



(b)

$$\begin{aligned} I^n{}_{(i,0)}(x) &= I^n(x^1,\ldots,x^{i-1},0,x^i,\ldots,x^{n-1}) \\ &= (x^1,\ldots,x^{i-1},0,x^i,\ldots,x^{n-1}), \\ I^n{}_{(i,1)}(x) &= I^n(x^1,\ldots,x^{i-1},1,x^i,\ldots,x^{n-1}) \\ &= (x^1,\ldots,x^{i-1},1,x^i,\ldots,x^{n-1}) \end{aligned}$$

We call $I^n{}_{(i,0)}$ the $(i,0)$ **-face** of I^n and $I^n{}_{(i,1)}$ the $(i,1)$ **-face** (Figure).



(a) (b)

We then define [**boundary chain** ∂I^n as]

$$\partial I^n = \sum_{i=1}^{n} \sum_{\alpha=0,1} (-1)^{i+\alpha} I^n{}_{(i,\alpha)} .$$

For a general singular n-cube $c:[0,1]^n \to A$ we first define the (i,α) -face,

$$c_{(i,\alpha)} = c \circ (I^n{}_{(i,\alpha)})$$

and then define

$$\partial c = \sum_{i=1}^{n} \sum_{\alpha=0,1} (-1)^{i+\alpha} c_{(i,\alpha)} .$$

Finally we define the **boundary of an n-chain** $\sum a_i c_i$ by

$$\partial\left(\sum a_i c_i\right) = \sum a_i \partial(c_i) .$$

Although these few definitions suffice for all applications in this book, we include here the one standard property of ∂.

4-12 **Theorem**. If c is an n-chain in A, then $\partial(\partial c)=0$. Briefly, $\partial^2=0$.

It is natural to ask whether Theorem 4-12 has a converse: If $\partial c=0$, is there a chain d in A such that $c=\partial d$? The answer depends on A and is generally "no." For example, define $c:[0,1]\to R^2-0$ by $c(t)=(\sin 2\pi nt,\cos 2\pi nt)$, where n is a non-zero integer, Then $c(1)=c(0)$, so $\partial c=0$. But (Problem 4-26) there is no 2-chain c' in R^2-0, with $\partial c'=c$.

*****

Examples 1. Standard p-cube[2]

$R^p\subset V$, $I^p:[0,1]\ \to R^p, I^p(t):=t, t\in[0,1]^p$

Example 2. **Closed disk**

$c_R^2(t^1,t^2):=(t^1R\cos(2\pi t^2, t^1R\sin(2\pi t^2),\ (t^1,t^2)\in[0,1]^2$

Example 3. Singular 2-cube

$c_{f,g}(t^1,t^2):=\left(l(t^1), g(l(t^1))+\left[f(l(t^1))-g(l(t^1))\right]t^2\right),\ (t^1,t^2)\in[0,1]^2$

Example 4 .**Modified ring area**

$\rho(r,t):=r_1(t)+(r_2(t)-r_1(t))r,\ (r,t)\in[0,1]^2$

$c(r,t):=(\rho(r,t)\cos(2\pi t), \rho(r,t)\sin(2\pi t))$

Example 5. Singular 3-cube

$c_R^3:=(t^1R\cos(2\pi t^2)\sin(2\pi t^3), t^1R\sin(2\pi t^2)\sin(t^3), t^1R\cos(\pi t^3)),\ t=(t^1,t^2,t^3)=\in[0,1]^3$

Example 6. Faces for p=1

$I^2_{(1,0)}=(0,t),\ \ I^2_{(1,1)}=(1,t),\ \ I^2_{(2,0)}=(t,0),\ \ I^2_{(2,1)}=(t,1)$, $t\in R$

Example 7. Boundary chain

$(c_R^2)_{(1,0)}(t)=(0,0)$, $(c_R^2)_{(1,1)}(t)=(R\cos(2\pi t,)R\sin(2\pi t))$, $(c_R^2)_{(2,0)}(t)=(tR,0)$, $(c_R^2)_{(2,1)}(t)=(tR,0)$, $t\in[0,1]$

$\partial c_R^2=C_R-C_0$, $C_R(t):=(R\cos 2\pi t, R\sin 2\pi t)$, $C_0(t):=(0,0)$, $t\in[0,1]$

*****

4.4 The Fundamental Theorem of Calculus

The fact that $d^2=0$ and $\partial^2=0$, not to mention the typographical similarity of d and ∂, suggests some connection between chains and forms. This connection is established by integrating forms over chains. Henceforth only differentiable singular n-cubes will be considered.

If ω is a k-form on $[0,1]^k$, then $\omega=fdx^1\wedge\cdots\wedge dx^k$ for a unique function f. We define

$$\int_{[0,1]^k}\omega=\int_{[0,1]^k}f.$$

We could also write this as

---

[2] 朝井朝雄「現代ベクトル解析の原理と応用」共立出版 2006, pp-275-281.

$$\int_{[0,1]^k} f dx^1 \wedge \cdots \wedge dx^k = \int_{[0,1]^k} f(x^1,\ldots,x^k) dx^1 \cdots dx^k ,$$

one of the reasons for introducing the functions x^i.

If ω is a k-form on A and c is a singular k-cube in A, we define

$$\int_c \omega = \int_{[0,1]^k} c^* \omega .$$

Note, in particular, that

$$\begin{aligned}\int_{I^k} f dx^1 \wedge \cdots \wedge dx^k &= \int_{[0,1]^k} (I^k)^* (f dx^1 \wedge \cdots \wedge dx^k) \\ &= \int_{[0,1]^k} f(x^1,\ldots,x^k) dx^1 \cdots dx^k .\end{aligned}$$

A special definition must be made for $k = 0$. A 0-form ω is a function; if $c : \{0\} \to A$ is a singular 0-cube in A we define

$$\int_c \omega = \omega(c(0)) .$$

The integral of ω over a k-chain $c = \sum a_i c_i$ is defined by

$$\int_c \omega = \sum a_i \int_{c_i} \omega .$$

The integral of a 1-form over a 1-chain is often called a **line integral**. if $Odx + Qdy$ is a 1-form on R^2 and $c : [0,1] \to R^2$ is a singular 1-cube (a curve), then one can (but we will not) prove that

$$\int_c Pdx + Qdy = \lim \sum_{i=1}^n \left[c^1(t_i) - c^1(t_{i-1})\right] \cdot P(c(t^i)) + \left[c^2(t_i) - c^2(t_{i-1})\right] \cdot Q(c(t^i))$$

where $t_0,\ldots,t_n$ is a partition of $[0,1]$, the choice of t^i in $[t_{i-1}, t_i]$ is arbitrary, and the limit is taken over all partitions as the maximum of $|t_i - t_{i-1}|$ goes to 0. The right side is often taken as a definition of $\int_c Pdx + Qdy$.

This is a natural definition to make, since these sums are very much like sums appearing in the definition of ordinary integrals. However such an expression is almost impossible to work with and is quickly equated with an integral equivalent to $\int_{[0,1]} c^*(Pdx + Qdy)$. Analogous definitions for surface integrals, that is, integrals of 2-forms over singular 2-cubes, are even more complicated and difficult to use. This is one reason why we have avoided such an approach. The other reason is that the definition given here is the one that makes sense in the more general situations considered in Chapter 5.

The relationship between forms, chains, d, and ∂ is summed up in the neatest possible way by Stokes' theorem, sometimes called the fundamental theorem of calculus in higher dimensions (if $k = 1$ and $c = I^1$, it really is the fundamental theorem of calculus).

4-13 **Theorem (Stokes' Theorem).** If ω is a (k-1)-form on an open set $A \subset R^n$ and c is a k-chain in A, then

$$\int_c d\omega = \int_{\partial\omega} \omega .$$

Stokes’ theorem shares three important attributes with many fully evolved major theorems:

1. It is trivial.
2. It is trivial because the terms appearing in it have been properly defined.
3. It has significant consequences.

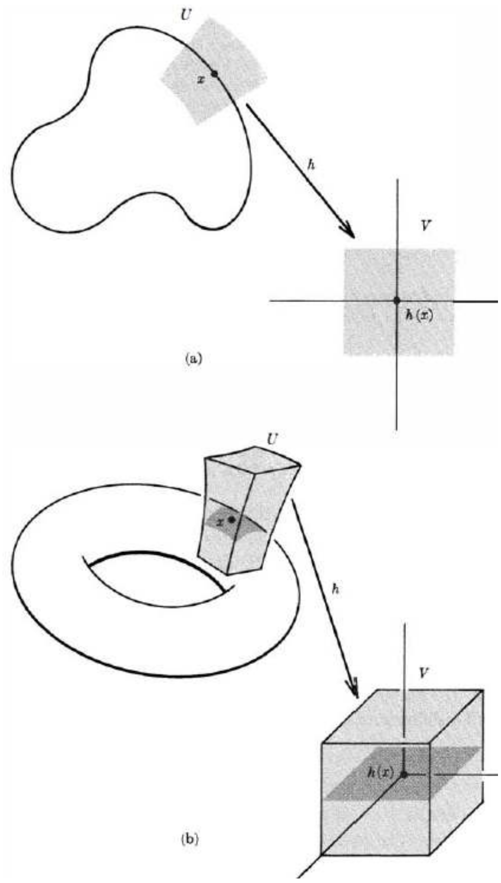
Since this entire chapter was little more than a series of definitions which made the statement and proof of Stokes' theorem possible, the reader should be willing to grant the first two of these attributes to Stokes' theorem. The rest of the book is devoted to justifying the third.

5. Integration on Manifolds

5.1 Manifolds

I f U and V are open sets in R^n, a differentiable function $h:U \to V$ with a differentiable inverse $h^{-1}:V \to U$ will be called a **diffeomorphism**. ("Differentiable" henceforth means "C^∞".)

A subset M of R^n is called a **k-dimensional manifold** (in R^n) if for every point $x \in M$ the following condition is satisfied :



(M) There is an open set U containing x, an open set $V \subset R^n$, and a diffeomorphism $h:U \to V$ such that

$$h(U \cap M) = V \cap \left(R^k \times \{0\}\right) = \left\{y \in V : y^{k+1} = \cdots = y^n = 0\right\}.$$

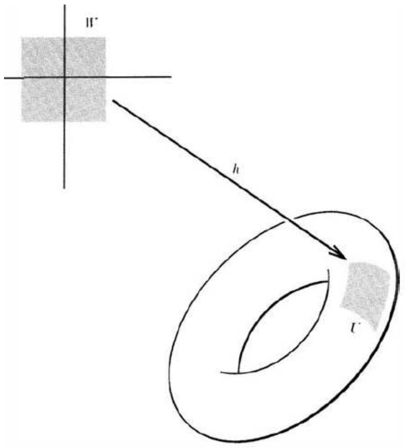
In other words, $U \cap M$ is, "up to diffeomorphism," simply $R^k \times \{0\}$ (see Figure). The two extreme cases of our definition should be noted: a point in R^n is a 0-dimensional manifold, and an open subset of R^n is an n-dimensional manifold.

One common example of an n-dimensional manifold is the **n-sphere** S^n, as defined as $\left\{x \in R^{n+1} : |x| = 1\right\}$. We leave it as an exercise for the reader to prove that condition (M) is satisfied. If you are unwilling to trouble yourself with the details, you may instead use the following theorem, which provides many examples of manifolds (note that $S^m = g^{-1}(0)$, where $g: R^{n+1} \to R$ is defined by $g(x) = |x|^2 - 1$).

5-1 **Theorem**. Let $A \subset R^n$ be open and let $g: A \to R^p$ be a differentiable function such that $g'(x)$ has *rank p* whenever $g(x) = 0$. Then $g^{-1}(0)$ is an (n-p)-dimensional manifold in R^n.

There is an alternative characterization of manifolds which is very important.

5-2 **Theorem**. A subset M of R^n is a k-dimensional manifold if and only if for each point $x \in M$ the following "coordinate condition" is satisfied:



(C) There is an open set U containing x, an open set $W \subset R^k$, and a 1-1 differentiable function $f: W \to R^n$ such that

(1) $f(W) = M \cap U$,

(2) $f'(y)$ has *rank k* for each $y \in W$,

(3) $f^{-1}: f(W) \to W$ is continuous.

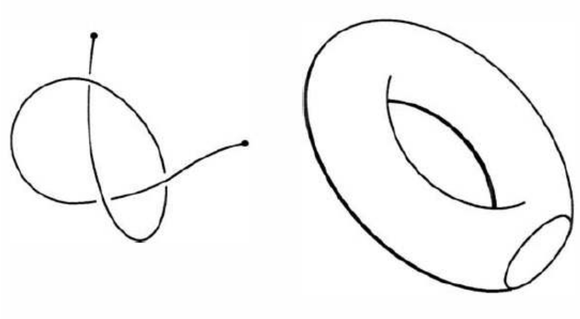
(Such a function f is called a **coordinate system** around x. See Figure.)

One consequence of the proof of Theorem 5-2 should be noted. If $f_1: W_1 \to R^n$ and $f_2: W_2 \to R^n$ are two coordinate systems, then

$$f_2^{-1} \circ f_1 : f_1^{-1}(f_2(W_2)) \to R^k$$

is differentiable with non-singular Jacobian. In fact, $f_2^{-1}(y)$ consists of the first k components of $h(y)$.

The **half-space** $H^k \subset R^k$ is defined as $\{x \in R^k : x^k \geq 0\}$. A subset M of R^n is a **k-dimensional manifold-with-boundary** (Figure) if for every point $x \in M$ either condition (M) or the following condition is satisfied:



(a) (b)

(M') There is an open set U containing x, an open set $V \subset R^n$, and a diffeomorphism $h : U \to V$ such that

$$h(U \cap M) = V \cap (H^k \times \{0\}) = \{y \in V : y^k \geq 0\_and\_y^{k+1} = \cdots = y^n = 0\}$$

and $h(x)$ has kth component=0.

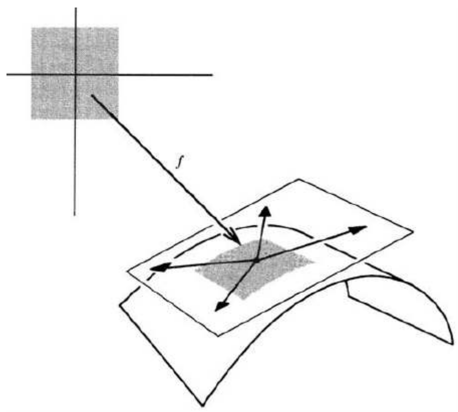
It is important to note that condition (M) and (M') cannot both hold for the same x. In fact, if $h_1 : U_1 \to V_1$ and $h_2 : U_2 \to V_2$ satisfied (M) and (M'), respectively, then $h_2 \circ h_1^{-1}$ would be a differentiable map that takes an open set in R^k, containing $h(x)$, into a subset of H^k which is not open in R^k. Since $\det(h_2 \circ h_1^{-1})' \neq 0$, this contradicts Problem 2-36. The set of all points $x \in M$ for which condition M' is satisfied is called the **boundary** of M and denoted ∂M. This must not be confused with the boundary of a set, as defined in Chapter1 (see Problems 5-3 and5-8).

5.2 Fields and Forms on Manifolds

Let M be a k-dimensional manifold in R^n and let $f : W \to R^n$ be a coordinate system around $x = f(a)$. since $f'(a)$ has $rank\ k$, the linear transformation $f_* : R^k{}_a \to R^n{}_x$ is one-to-one, and $f_*(R^k{}_a)$ is a k-dimensional subspace of $R^n{}_x$. If $g : V \to R^n$ is another coordinate system, with $x = g(b)$, then

$g_*(R^k{}_b) = f_*(f^{-1} \circ g)_*(R^k{}_b) = f_*(R^k{}_a)$.

Thus the k-dimensional subspace $f_*(R^k{}_a)$ does not depend on the coordinate system f. This subspace is denoted M_x, and is called the **tangent space** of M at x (see Figure). In later sections we will use the fact that there is a natural inner product T_x on M_x, introduced by that on $R^n{}_x$: if $v, w \in M_x$ define $T_x(v, w) = \langle v, w \rangle_x$



Suppose that A is an open set containing M, and F is a differentiable vector field on A such that $F(x) \in M_x$ for each $x \in M$. If $f : Q \to R^n$ is a coordinate system, there is a unique (differentiable) vector field G on W such that $f_*(G(a)) = F(f(a))$ for each $a \in W$. We can also consider a function F which merely assigns a vector $F(x) \in M_x$ for each $x \in M$; such a function is called a **vector field on** M. There is still a unique vector field G on W such that $f_*(G(a)) = F(f(a))$ for $a \in W$; we define F to be differentiable if G is differentiable. Note that our definition does not depend on the coordinate system chosen: if $g : V \to R^n$ and $g_*(H(b)) = F(g(b))$ for all $b \in V$, then the component function of $H(b)$ must be equal the

component function of $G(f^{-1}(g(b)))$, so H is differentiable if G is.

Precisely the same considerations hold for forms. A function ω which assigns $\omega(x) \in \Lambda^p(M_x)$ for each $x \in M$ is called a **p-form on** M. If $f : W \to R^n$ is a coordinate system, then $f^*\omega$ is a p-form on W; we define ω to be differentiable if $f^*\omega$ is. A p-form ω on M ca be written as

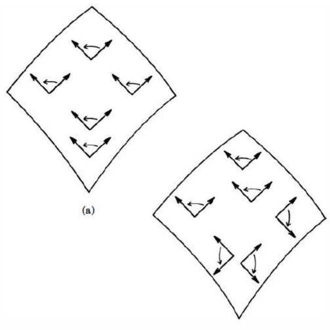
$$\omega = \sum_{i1<\cdots<ip} \omega_{i1,\ldots,ip} dx^{i1} \wedge \cdots dx^{ip} .$$

Here the functions $\omega_{i1,\ldots,ip}$ are defined only on M. The definitions of $d\omega$ given previously would make no sense here, since $D_j(\omega_{i1,\ldots,ip})$ has no meaning. Nevertheless, there is a reasonable way of defining $d\omega$.

5-3 **Theorem**. There is a unique (p+1)-form $d\omega$ on M such that for every coordinate system $f : W \to R^n$ we have

$$f^*(d\omega) = d(f^*\omega) .$$

It is often necessary to choose an orientation μ_x for each tangent space M_x of a manifold M. Such choices are called **consistent** (Figure) provided that for every coordinate system $f : W \to R^n$ and $a, b \in W$ the relation



(a)

(b)

$$\left[f_*((e_1)_a), \ldots, f_*((e_k)_a)\right] = \mu_{f(a)}$$

holds if and only if

$$\left[f_*((e_1)_b), \ldots, f_*((e_k)_b)\right] = \mu_{f(b)} .$$

Suppose orientation μ_x have been chosen consistently. If $f : W \to R^n$ is a coordinate system such that

$$\left[f_*((e_1)_a), \ldots, f_*((e_k)_a)\right] = \mu_{f(a)}$$

for one, and hence for every $a \in W$, then f is called **orientation-preserving**. If f is *not* orientation-preserving and $T : R^k \to R^k$ is a linear transformation with $\det T = -1$, then $f \circ T$ is orientation-preserving. Therefore there is an orientation-preserving coordinate system around each point. If f and g are orientation-preserving and $x = f(a) = g(b)$, then the relation

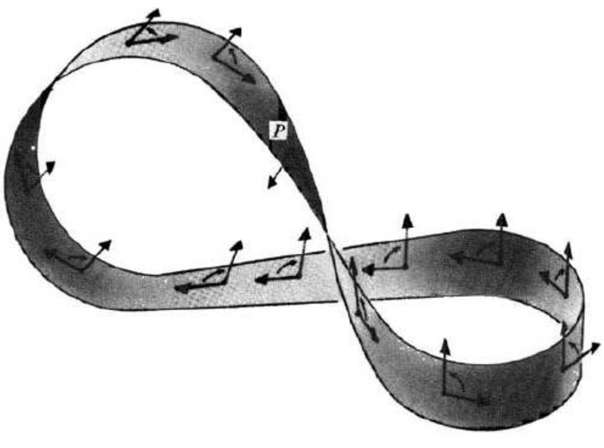
$$\left[f_*((e_1)_a), \ldots, f_*((e_k)_a)\right] = \mu_x = \left[g_*((e_1)_b), \ldots, g_*((e_k)_b)\right]$$

implies that

$$\left[(g^{-1} \circ f)_*((e_1)_a), \ldots, (g^{-1} \circ f)_*((e_k)_a)\right] = \left[(e_1)_b, \ldots, (e_k)_b\right],$$

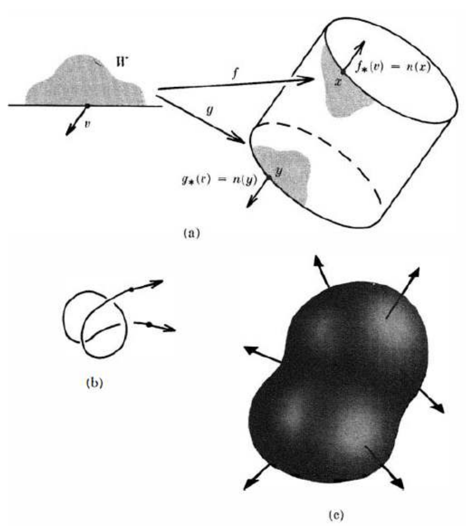
so that $\det(g^{-1} \circ f)' > 0$, an important fact to remember.

A manifold for which orientations μ_x can be chosen consistently is called **orientable**, and a particular choice of the μ_x is called an **orientation** μ of M. A manifold together with an orientation μ is called an **oriented** manifold. The classical example of a non-orientable manifold is the Möbius strip. A model can be made by gluing together the ends of a strip of paper which has been given a half twist (Figure).



Our definitions of vector fields, forms, and orientations can be made for manifold-with-boundary and $x \in \partial M$, then $(\partial M)_x$ is a (k-1)-dimensional

subspace of the k-dimensional vector space M_x. Thus there are exactly two unit vectors in M_x which are perpendicular to $(\partial M)_x$; they can be distinguished as follows (Figure). If $f : W \to R^n$ is a coordinate system with $W \subset H^k$ and $f(0) = x$, then only one of these unit vectors is $f_*(v_0)$ for some v_0 with $v^k < 0$. This unit vector is called the **outward unit normal** $n(x)$; it is not hard to check that this definition does not depend on the coordinate system f.



Suppose that μ is an orientation of a k-dimensional manifold-with-boundary M. If $x \in \partial M$, choose $v_1, \ldots, v_{k-1} \in (\partial M)_x$ so that $[n(x), v_1, \ldots, v_{k-1}] = \mu_x$. If it is also true that $[n(x), w_1, \ldots, w_{k-1}] = \mu_x$, then both $[v_1, \ldots, v_{k-1}]$ and $[w_1, \ldots, w_{k-1}]$ are the same orientation for $(\partial M)_x$. This orientation is denoted $(\partial\mu)_x$. It is easy to see that the orientation $(\partial\mu)_x$, for $x \in \partial M$, are consistent on ∂M. Thus if M is orientable, ∂M is also orientable, and an orientation μ for M determines an orientation $\partial\mu$ for ∂M, called the **induced orientation**. If we apply these definitions to H^k with the usual orientation, we find that the induced orientation on $R^{k-1} = \{x \in H^k : x^k = 0\}$ is $(-1)^k$ times the usual orientation. The reason for such a choice will become clear in the next section.

If M is an *oriented* (n-1)-dimensional manifold in R^n, a substitute for outward unit normal vectors ca be defined, even though M is not necessarily the boundary of an n-dimensional manifold. If $[v_1, \ldots, v_{n-1}] = \mu_x$, we choose $n(x)$ in $R^n{}_x$ so that $n(x)$ is a unit vector perpendicular to M_x and $[n(x), v_1, \ldots, v_{n-1}]$ is the usual orientation of $R^n{}_x$. We still call $n(x)$ the outward unit normal to M 8denoted by μ). The vectors $n(x)$ vary continuously on M, in an obvious sense. Conversely, if a continuous family of unit normal vectors $n(x)$ is defined on all of M, then we can determine an orientation of M. This shows that such a continuous choice of normal vectors is impossible on the Möbius strip. In the paper model of the Möbius strip the two sides of the paper (which has thickness) may be thought of as the end points of the unit normal vectors in both directions. The impossibility of choosing normal vectors continuously is reflected by the famous property of the paper model. The paper model is one-sided (if you start to paint it on one side you end up painting it all over); in other words, choosing $n(z)$ arbitrarily at one point, and then by the continuity requirement at other points, eventually forces the opposite choice for $n(x)$ at the initial point.

5.3 Stokes' Theorem on Manifold

If ω is a p-form on a k-dimensional manifold-with-boundary M and c is a singular p-cube in M, we define

$$\int_c \omega = \int_{[0,1]^p} c^*\omega$$

precisely as before; integrals over p-chains are also defined as before. In the case $p = k$ it may happen that there is an open set $W \supset [0,1]^k$ and a coordinate system $f * W \to R^n$ such that $c(x) = f(x)$ for $x \in [0,1]^k$; a k-cube in M will always be understood to be of this type. If M is oriented, the singular k-cube c is called **orientation-preserving**

if f is.

5-4 **Theorem**. If $c_1, c_2 : [0,1]^k \to M$ are two orientation preserving singular k-cubes in the oriented k-dimensional manifold M and ω is a k-form on M such that $\omega = 0$ outside of $c_1([0,1]^k) \cap c_2([0,1]^k)$, then

$$\int_{c_1} \omega = \int_{c_2} \omega .$$

The last equation in this proof should help explain why we have had to be so careful about orientations.

Let ω be a k-form on an oriented k-dimensional manifold M. If there is an orientation-preserving singular k-cube c in M such that $\omega = 0$ outside of $c([0,1]^k)$, we define

$$\int_M \omega = \int_c \omega .$$

Theorem 5-4 shows $\int_M \omega$ doe not depend on the choice of c. Suppose now that ω is an arbitrary k-form on M. There is an open cover O of M such that for each $U \in O$ there is an orientation-preserving singular k-cube c with $U \subset c([0,1]^k)$. Let Φ be a partition of unity for M subordinate to this cover. We define

$$\int_M \omega = \sum_{\varphi \in \Phi} \int_M \varphi \cdot \omega$$

provided the sum converges as described in the discussion preceding Theorem 3-12 (this is certainly true if M is compact). An argument similar to that in Theorem 3-12 shows that $\int_M \omega$ does not depend on the cover O or on Φ.

All our definitions could have been given for a k-dimensional manifold-with-boundary M with orientation μ. Let ∂M have the induced orientation $\partial\mu$. Let c be an orientation-preserving k-cube in M such that $c_{(k,0)}$ lies in ∂M and is the only face which has any interior points in ∂M. As the remarks after the definition of $\partial\mu$ show, $c_{(k,0)}$ is orientation-preserving if k is even, but not if k is odd. Thus, if ω is a (k - l)-form on M which is 0 outside of $c([0,1]^k)$, we have

$$\int_{c_{(k,0)}} \omega = (-1)^k \int_{\partial M} \omega .$$

On the other hand, $c_{(k,0)}$ appears with coefficient $(-1)^k$ in ∂c.

Therefore

$$\int_{\partial c} \omega = \int_{(-1)^k c_{(k,0)}} \omega = (-1)^k \int_{c_{(k,0)}} \omega = \int_{\partial M} \omega .$$

Our choice of $\partial\mu$ was made to eliminate any minus signs in this equation, and in the following theorem.

5-5 **Theorem (Stokes' Theorem).** If M is a compact oriented k-dimensional manifold-with-boundary and ω is a (k-1)-form on M, then

$$\int_M \omega = \int_{\partial M} \omega .$$

(Here ∂M is given the induced orientation.)

5.3 The Volume Element

Let M be a k-dimensional manifold (or manifold-with-boundary) in R^n, with an orientation μ. If $x \in M$, then μ_x and the inner product T_x we defined previously determine a volume element $\omega(x) \in \Lambda^k(M_x)$. We therefore obtain a nowhere-zero k-form ω on M, which is called the **volume element** on M (determined by μ) and denoted dV, even though it is not generally the differential of a (k -1)-form. The **volume** of M is defined as $\int_M dV$, provided this integral exists, which is certainly the case if M is compact. "Volume" is usually called **length** or **surface area** for one- and two-dimensional manifolds, and dV is denoted ds (the "element of length") or dA [or dS] (the "element of [surface] area").

A concrete case of interest to us is the volume element of an oriented surface (two-dimensional manifold) M in R^3. Let $n(x)$ be the unit outward normal at $x \in M$. If $\omega \in \Lambda^2(M_x)$ is defined by

$$\omega(v,w) = \det\begin{pmatrix} v \\ w \\ n(x) \end{pmatrix},$$

then $\omega(v,w) = 1$ if v and w are an orthonormal basis of M_x with $[v,w] = \mu_x$. Thus $dA = \omega$. On the other hand, $\omega' v,w) = \langle v \times w, n(x) \rangle$ by definition of $v \times w$. Thus we have

$$dA(v,w) = \langle v \times w, n(x) \rangle.$$

Since $v \times w$ is a multiple of $n(x)$ for $v, w \in M_x$, we conclude that

$$dA(v,w) = |v \times w|$$

if $[v,w] = \mu_x$. If we wish to compute the area of M, we must evaluate $\int_{[0,1]^2} c^*(dA)$ for orientation-preserving singular 2-cubes c. Define

$$E(a) = [D_1c^1(a)]^2 + [D_1c^2(a)]^2 + [D_1c^3(a)]^2,$$

$$F(a) = D_1c^1(a) \cdot D_2c^1(a) + D_1c^2(a) \cdot D_2c^2(a) + D_1c^3(a) \cdot D_2c^3(a),$$

$$G(a) = [D_2c^1(a)]^2 + [D_2c^2(a)]^2 + [D_2c^3(a)]^2.$$

Then

$$\begin{aligned} c^*(dA)((e_1)_a,(e_2)_a) &= dA(c_*((e_1)_a,(e_2)_a)) \\ &= |(D_1c^1(a), D_1c^2(a), D_1c^3(a)) \times (D_2c^1(a), D_2c^2(a), D_2c^3(a))| \\ &= \sqrt{E(a)G(a) - F(a)^2} \end{aligned}$$

by Problem 4-9. Thus

$$\int_{[0,1]^2} c^*(dA) = \int_{[0,1]^2} \sqrt{EG - F^2}.$$

Calculating surface area is clearly a foolhardy enterprise; fortunately one seldom needs to know the area of a surface. Moreover, there is a simple expression for dA which suffices for theoretical considerations.

5-6 **Theorem**. Let M be an oriented two-dimensional manifold (or manifold-with-boundary) in R^3 and let n be

the unit outward normal. Then

(1) $dA = n^1 dy \wedge dz + n^2 dz \wedge dx + n^3 dx \wedge dy$.

Moreover, on M we have

(2) $n^1 dA = dy \wedge dz$.

(3) $n^2 dA = dz \wedge dx$.

(4) $n^3 dA = dx \wedge dy$.

A word of caution: if $\omega \in \Lambda^2(R^3{}_a)$ is defined by

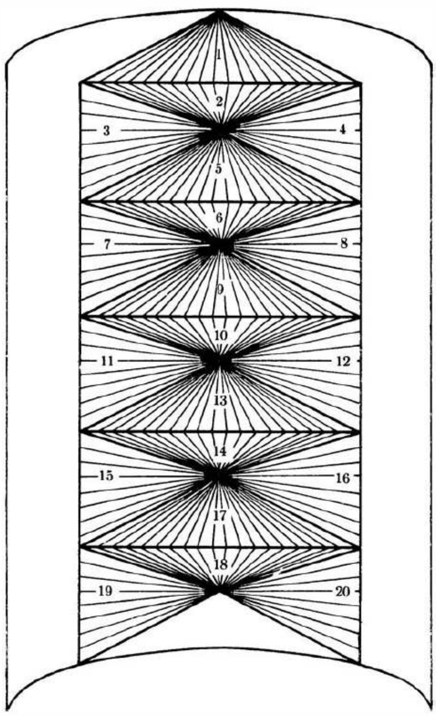
$$\omega = n^1(a) \cdot dy(a) \wedge dz(a) + n^2(a) \cdot dz(a) \wedge dx(a) + n^3(a) \cdot dx(a) \wedge dy(a),$$

it is *not* true, for example, that

$$n^1(a) \cdot \omega = dy(a) \wedge dz(a).$$

The two sides give the same result only when applied to $v, w \in M_a$.

A few remarks should be made to justify the definition of length and surface area we have given. If $c : [0,1] \to R^n$ is differentiable and $c([0,1])$ is a one-dimensional manifold-with-boundary, it can be shown, but the proof is messy, that the length of $c([0,1])$ is indeed the least upper bound of the lengths of inscribed broken lines. If $c : [0,1]^2 \to R^n$, one naturally hopes that the area of $c([0,1]^2)$ will be the least upper bound of the areas of surfaces made up of triangles whose vertices lie in $c([0,1]^2)$. Amazingly enough, such a least upper bound is usually nonexistent - one can find inscribed polygonal surfaces arbitrarily close to $c([0,1]^2)$ with arbitrarily large area! This is indicated for a cylinder in Figure. Many definitions of surface area have been proposed, disagreeing with each other, but all agreeing with our definition for differentiable surfaces. For a discussion of these difficult questions the reader is referred to References [3] or [15].



5.4 Classical Theorems

We have now prepared all the rnachinery necessary to state and prove the classical "Stokes' type" of theorems. We will indulge in a little bit of self-explanatory classical notation.

5-7 **Theorem (Green's Theorem).** Let $M \subset R^2$ be a compact two-dimensional manifold-with-boundary. Suppose that $\alpha, \beta : M \to R$ are differentiable. Then

$$\begin{aligned}\int_{\partial M} \alpha dx + \beta dy &= \int_M (D_1\beta - D_2\alpha) dx \wedge dy \\ &= \iint_M \left(\frac{\partial \beta}{\partial x} - \frac{\partial \alpha}{\partial y} \right) dxdy\end{aligned}.$$

(Here M is given the usual orientation, and ∂M the induced orientation, also known as the counterclockwise

orientation.)

5-8 **Theorem (Divergence Theorem).** Let $M \subset R^3$ be a compact three-dimensional manifold-with-boundary and n the unit outward normal on ∂M . Let F be a differentiable vector field on M . Then

$$\int_M divF dV = \int_{\partial M} \langle F, n \rangle dA .$$

This equation is also written in terms of three differentiable functions $\alpha, \beta, \gamma : M \to R$:

$$\iiint_M \left(\frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial x} + \frac{\partial \alpha}{\partial x} \right) dV = \iint_{\partial M} (n^1 \alpha + n^2 \beta + n^3 \gamma) dS .$$

5-9 **Theorem (Stokes' Theorem).** Let $M \subset R^3$ be a compact oriented two-dimensional manifold-with-boundary and n the unit outward normal on M determined by the orientation of M . Let ∂M have the induced orientation. Let T be the vector field on ∂M with $ds(T) = 1$ and let F be a differentiable vector field in an open set containing M . Then

$$\int_M \langle (\nabla \times F), n \rangle dA = \int_{\partial M} \langle F, T \rangle ds .$$

This equation is sometimes written

$$\iint_{\partial M} \alpha dx + \beta dy + \gamma dz = \iint_M \left[n^1 \left(\frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z} \right) + n^2 \left(\frac{\partial \alpha}{\partial z} - \frac{\partial \gamma}{\partial x} \right) + n^3 \left(\frac{\partial \beta}{\partial x} - \frac{\partial \alpha}{\partial y} \right) \right] dS .$$

Theorems 5-8 and 5-9 are the basis for the names $divF$ and $curlF$. If $F(x)$ is the velocity vector of a fluid at x (at some time) then $\int_{\partial M} \langle F, n \rangle dA$ is the amount of fluid "diverging" from M . Consequently the condition $divF = 0$ expresses the fact that the fluid is incompressible. If M is a disc, then $\int_{\partial M} \langle F, n \rangle ds$ measures the amount that the fluid curls around the center of the disc. If this is zero for all discs, then $\nabla \times F = 0$, and the fluid is called *irrotational*.

These interpretations of $divF$ and $curlF$ are due to Maxwell [13]. Maxwell actually worked with the negative of $divF$, which he accordingly called the *convergence*. For $\nabla \times F$ Maxwell proposed "with great diffidence" the terminology *rotation of* F ; this unfortunate term suggested the abbreviation $rotF$ which one occasionally still sees.

The classical theorems of this section are usually stated in somewhat greater generality than they are here. For example, Green's Theorem is true for a square, and the Divergence Theorem is true for a cube. These two particular facts can be proved by approximating the square or cube by manifolds-with-boundary. A thorough generalization of the theorems of this section requires the concept of manifolds-with-corners; these are subsets of R^n which are, up to diffeomorphism, locally a portion of R^k which is bounded by pieces of (k -1)-planes. The ambitious reader will find it a challenging exercise to define manifolds-with-corners rigorously and to investigate how the results of this entire chapter may be generalized.

Bibliography

l. Ahlfors, *Complex Analysis*, McGraw-Hill, New York, 1953.

2. Auslander and MacKenzie, *Introduction to Differentiable Manifolds*, McGraw-Hill, New York, 1963.

3. Cesari, *Surface Area*, Princeton University Press, Princeton, New Jersey, 1956.

4. Courant, *Differential and Integral Calculus*, Volume II, Interscience, New York, 1937.

5. Dieudonné, *Foundations of Modem Analysis*, Academic Press, New York, 1960.

6. Fort, *Topology of 3-Manifolds*, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

7. Gauss, *Zur mathematischen Theorie der electrodynamischen Wirkungen*, (4) (Nachlass) Werke V, 605.

8. Helgason, *Differential Geometry and Symmetric Spaces*, Academic Press, New York, 1962.

9. Hilton and Wylie, *Homology Theory*, Cambridge University Press, New York, 1960.

10. Hu, *Homotopy Theory*, Academic Press, New York, 1959.

11. Kelley, *General Topology*, Van Nostrand, Princeton, New Jersey, 1955.

12. Kobayashi and Nomizu, *Foundations of Differential Geometry*, Interscience, New York, 1963.

13. Maxwell, *Electricity and Magnetism*, Dover, New York, 1954.

14. Natanson, *Theory of Functions of a Real Variable*, Frederick Ungar, New York, 1955.

15. Radó, *Length and Area*, Volume XXX, American Mathematical Society, Colloquium Publications, New York, 1948.

16. de Rham, *Variétés Differentiables*, Hermann, Paris, 1955.

17. Sternberg, *Lectures on Differential Geometry*, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.