SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Analysis of Stokes' Theorem on Differentiable Manifolds

تحليل نظرية ستوكس على متعدد الطيات التفاضلية

A Thesis Submitted in Partial Fulfillment of the Requirements for the M.Sc. Degree in Mathematics

Submitted by

NUSAIBA IBRAHIM MOHAMED IBRAHIM

Supervised by

DR. EMAD ELDEEN ABDALLAH ABDELRAHIM

February 2016

SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF GRADUATE STUDIES

Analysis of Stokes' Theorem on Differentiable Manifolds

تحليل نظرية ستوكس على متعدد الطيات التفاضلية

A Thesis Submitted in Partial Fulfillment of the Requirements for the M.Sc. Degree in Mathematics

Submitted by

NUSAIBA IBRAHIM MOHAMED IBRAHIM

B.Sc. (Hons), Sudan University of Science and Technology, 2012

Supervised by

DR. EMAD ELDEEN ABDALLAH ABDELRAHIM

February 2016

بسم الله الرحمن الرحيم

Sudan University of Science & Technology

College of Graduate Studies

جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا

كلية الدراسات العليا

Approval Page

Name of Candidate: Nusaiba Ibrahim Mohamed

Thesis title: Analysis of Stokes' Theorem on Differentiable manifold

Approved by:

1. External Examiner

Name: Mohammed Ali Bashir

Signature: Bashir Date: 22/2/2016

2. Internal Examiner

Name: Mohd. Arafa Ahmed Elkhatieb

Signature: MDHA.Elkhatieb Date: 22-2-2016

3. Supervisor

Name: Dr. Emad eldeen Abdalah Abdalrahim

Signature: Date: 22/2/2016

Sudan University of Science and Technology
College of Graduate Studies

Declaration

I, the signing here-under, declare that I'm the sole author of the (M.Sc.) thesis entitled ....

Analysis of Stokes' Theorem on Differentiable Manifolds

....

which is an original intellectual work. Willingly, I assign the copy-right of this work to the College of Graduate Studies (CGS), Sudan University of Science & Technology (SUST). Accordingly, SUST has all the rights to publish this work for scientific purposes.

Candidate's name: Nusaiba Ibrahim Mohamed Ibrahim ......

Candidate's signature: .... Date: 10/3/2016 ....

إقرار

أنا الموقع أدناه أقر بأننى المؤلف الوحيد لرسالة الماجستير المعنونة ....

تحليل نظرية ستوكس على متعدد الطيات التفاضلية ......

....

وهى منتج فكري أصيل . وباختياري أعطى حقوق طبع ونشر هذا العمل لكلية الدراسات العليا - جامعه السودان للعلوم والتكنولوجيا، عليه يحق للجامعه نشر هذا العمل للأغراض العلمية .

اسم الدارس : نسيبة إبراهيم محمد إبراهيم ....

توقيع الدارس : .... التاريخ : ١٠/٣/٢٠١٦ ....

بسم الله الرحمن الرحيم

Abstract

In this research, we deal with three forms of Stokes' theorem. The version known to Stokes' appears in the last chapter, along with its inseparable companions, Green's theorem and the Divergence theorem. We discuss how these three theorems can be derived from the modern Stokes theorem, which appears in chapter (4), with some applications on oriented manifolds with boundary. In addition to applications of Maxwell's field equations.

الخلاصة

في هذا البحث، نحن نتعامل مع ثلاثة أشكال لنظرية ستوكس. يظهر الإصدار المعروف لستوكس في الفصل الأخير، جنبا إلى جنب مع رفاقه الملازمة لها، نظرية جرين ونظرية التباعد. نناقش كيف يمكن الحصول على هذه النظريات الثلاث من نظرية ستوكس الحديثة التي تظهر في الفصل (٤)، مع بعض التطبيقات على متعدد الطيات الموجهة مع الحدود. بالإضافة إلى تطبيق النظرية في مجال ماكسويل.

Dedication

To my parents, husband, son, daughter, and rest family.

Acknowledgements

I thank almighty God for giving me the courage and the determination, as well as guidance in conducting this research study, despite all difficulties.

I offer my enduing gratitude to the faculty, staff and my fellow students at the SUST, who have inspired me to continue my work in this field. I take immense pleasure in thanking my supervisor Dr. Emad Eldeen Abdallah Abdelrahim for permitting me to carry out this project work. The supervision and support that he gave assisted me greatly in the progression and fulfillment of the project.

I would also like to express my sincere thanks to my beloved husband, parents, mother in law and uncle for their blessings, for their help and wishes for the successful completion of this project.

Table of Contents

Introduction

Today studying of mathematics are faced with an immense mountain of material. In addition to the traditional areas of mathematics as presented in the traditional manner, and these presentations do abound, there are the new and often enlightening ways of looking at these traditional areas. The area of differential geometry we discussed here, is one in which recent developments have effected great changes. This research centered about Stokes' theorem, and organised as follows:-

In chapter (1), we discuss the basic properties of functions on Euclidean space via our studying of norm and inner product, subsets of Euclidean space, functions of several variables and continuity. We consider the basic theorems of differentiation with some examples and applications.

In chapter (2), we treat the basic definitions of integration. We used the Fubini's theory to evaluate the integral of a function. We also discuss some applicable theorems with some examples.

In chapter (3), we discuss algebraic preliminaries required to fields and forms with basic properties. We study the exterior derivative, and its properties with some applications. We investigate the Poincaré lemma. In addition, we discuss the geometry and the fundamental theorem with some applications to deal with Stokes' theorem.

In chapter (4), we study the fields and forms on manifolds, this requires illustration of manifolds with boundary. We define the concept of oriented manifolds with boundary with some applications.

Chapter (1)

Euclidean Space and Differentiation

Section (1.1): Function on Euclidean Space

By a Euclidean space, we mean a vector space with a positive definite scalar product. Euclidean n-space $\mathbf{R}^n$ is defined as the set of all ordered n-tuples

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

where $(x^1, \dots, x^n)$ are real numbers. An element of $\mathbf{R}^n$ is often called a point, or vector, especially when $n > 1$, and $\mathbf{R}^1, \mathbf{R}^2, \mathbf{R}^3$, are often called the line, the plain, and space, respectively. If $y = (y^1, \dots, y^n)$ and if a is a real number, put

$$x + y = (x^1 + y^1, \dots, x^n + y^n),$$

$$ax = (ax^1, \dots, ax^n)$$

as operation, is a vector space. In this vector space there is the notion of the length of 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 space structure of $\mathbf{R}^n$ is very important. [7, 11]

Theorem (1.1.1.):

If $x, y \in \mathbf{R}^n$ and $\mathrm{a} \in \mathbf{R}$, then

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

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

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

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

Proof:

(1) $|x| = 0$ iff $|x|^2 = 0$ iff $\sum_{i=1}^{n} \left(x^i\right)^2 = 0$. If all $x^i = 0$, then $\sum_{i=1}^{n} \left(x^i\right)^2 = 0$. If some $x^i \neq 0$, then $\left(x^i\right)^2 > 0$, so $\sum_{j \neq i} \left(x^j\right)^2 + \left(x^i\right)^2 \geq \left(x^i\right)^2 > 0$.

(2) If x and y are linearly dependent, equality clearly holds.

If not, then $\lambda y - x \neq 0$ for all $\lambda \in \mathbf{R}$, so

$$0 < |\lambda y - x|^2 = \sum_{i=1}^{n} \left(\lambda y^i - x^i\right)^2$$

$$= \lambda^2 \sum_{i=1}^{n} \left(y^i\right)^2 - 2\lambda \sum_{i=1}^{n} x^i y^i + \sum_{i=1}^{n} \left(x^i\right)^2.$$

Therefore, the right side is a quadratic equation in λ with no real solution, and its discriminant must be negative. Thus

$$4\left(\sum_{i=1}^{n} x^i y^i\right)^2 - 4\sum_{i=1}^{n} \left(x^i\right)^2 \cdot \sum_{i=1}^{n} \left(y^i\right)^2 < 0.$$

(3) $|x+y|^2 = \sum_{i=1}^{n} \left(x^i + y^i\right)^2$

$$= \sum_{i=1}^{n} \left(x^i\right)^2 + \sum_{i=1}^{n} \left(y^i\right)^2 + 2\sum_{i=1}^{n} x^i y^i$$

$$\leq |x|^2 + |y|^2 + 2|x| \cdot |y| \qquad \text{by (2)}$$

$$= (|x| + |y|)^2.$$

(4) $|ax| = \sqrt{\sum_{i=1}^{n} (ax^i)^2} = \sqrt{a^2 \sum_{i=1}^{n} (x^i)^2} = |a| \cdot |x|$.

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

Theorem (1.1.2):

If x, x_1, x_2 and y, y_1, y_2 are vectors in $\mathbf{R}^n$ and $a \in \mathbf{R}$, 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) $\langle x, x\rangle \geq 0$, 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)

Proof:

(1) $\langle x, y\rangle = \sum_{i=1}^{n} x^i y^i = \sum_{i=1}^{n} y^i x^i = \langle y, x\rangle$.

(2) By (1) it suffices to prove

$$\langle ax, y\rangle = a\langle x, y\rangle,$$

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

These follow from the equations

$$\langle ax, y\rangle = \sum_{i=1}^{n} (ax^i) y^i = a \sum_{i=1}^{n} x^i y^i = a\langle x, y\rangle,$$

$$\langle x_1 + x_2, y\rangle = \sum_{i=1}^{n} ({x_1}^i + {x_2}^i) y^i = \sum_{i=1}^{n} {x_1}^i y^i + \sum_{i=1}^{n} {x_2}^i y^i$$

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

(3) $\langle x, x\rangle = \sum_{i=1}^{n} x^i x^i$. If some $x^i \neq 0$, then $\left(x^i\right)^2 > 0$, so

$\sum_{j\neq i} \left(x^j\right)^2 + \left(x^i\right)^2 \geq \left(x^i\right)^2 > 0$

If $x = 0$, then $\langle x, 0\rangle = \langle x, 0x\rangle = 0\langle x, x\rangle = \langle 0x, x\rangle = \langle 0, x\rangle$.

$\langle x, x\rangle = 0$, then by (1) it must be that $x = 0$.

(4) $|x|^2 = \sum_{i=1}^{n} \left(x^i\right)^2 = \sum_{i=1}^{n} x^i x^i = \langle x, x\rangle$ (Positivity).

$|ax| = \sqrt{\langle ax, ax\rangle} = \sqrt{a^2\langle x, x\rangle} = |a|\sqrt{\langle x, x\rangle}$ (Homogeneity).

$$\begin{aligned} |x + y|^2 &= \langle x + y, x + y\rangle \\ &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\ &\leq \langle x, x\rangle + |\langle x, y\rangle| + |\langle y, x\rangle| + \langle y, y\rangle \\ &\leq |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2 \end{aligned}$$ (Triangle inequality).

(5) By (4) $\dfrac{|x + y|^2 - |x - y|^2}{4}$

$$\begin{aligned} &= \tfrac{1}{4}[\langle x + y, x + y\rangle - (x - y, x - y)] \\ &= \tfrac{1}{4}[\langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle - (\langle x, x\rangle - 2\langle x, y\rangle + \langle y, y\rangle)] \\ &= \langle x, y\rangle. \end{aligned}$$

In the following, we will clarify some important remarks about notation. The vector $(0,\dots,0)$ will usually be denoted simply 0. The usual basis of $\mathbf{R}^n$ is $e_1,\dots,e_n$, where $e_i=(0,\dots,1,\dots,0)$, with the 1 in the ith place. If $T:\mathbf{R}^n\to\mathbf{R}^m$ is a linear transformation, the matrix of T with respect to the usual bases of $\mathbf{R}^n$ and $\mathbf{R}^m$ is the $m\times n$ matrix $A=(a_{ij})$, where $T(e_i)=\sum_{j=1}^{m}a_{ji}\,e_j$ –the coefficients of $T(e_i)$ appear in the ith column of the matrix. If $S:\mathbf{R}^m\to\mathbf{R}^p$ has the $p\times m$ matrix B, then $S\circ T$ has the $p\times n$ matrix BA. To find $T(x)$ one computes the $m\times 1$ matrix

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

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

The closed interval $[a,b]$ has a natural analogue in $\mathbf{R}^2$. This is the closed rectangle $[a,b]\times[c,d]$, defined as the collection of all pair (x,y) with $x\in[a,b]$ and $y\in[c,d]$. More generally, if $A\subset\mathbf{R}^m$ and $B\subset\mathbf{R}^n$, then $A\times B\subset\mathbf{R}^{m+n}$ is defined as the set of all $(x,y)\in\mathbf{R}^{m+n}$ with $x\in A$ and $y\in B$. In particular, $\mathbf{R}^{m+n}=\mathbf{R}^m\times\mathbf{R}^n$. If $A\subset\mathbf{R}^m$, $B\subset\mathbf{R}^n$, and $C\subset\mathbf{R}^p$, 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 $[a_1,b_1]\times\cdots\times[a_n,b_n]\subset\mathbf{R}^n$ is called a closed rectangle in $\mathbf{R}^n$, while the set $(a_1,b_1)\times\cdots\times(a_nb_n)\subset\mathbf{R}^n$ is called an open rectangle. More generally a set $U\subset\mathbf{R}^n$ is called open (Figure (1.1)) if for each $x\in U$ there is an open rectangle A such that $x\in A\subset U$.

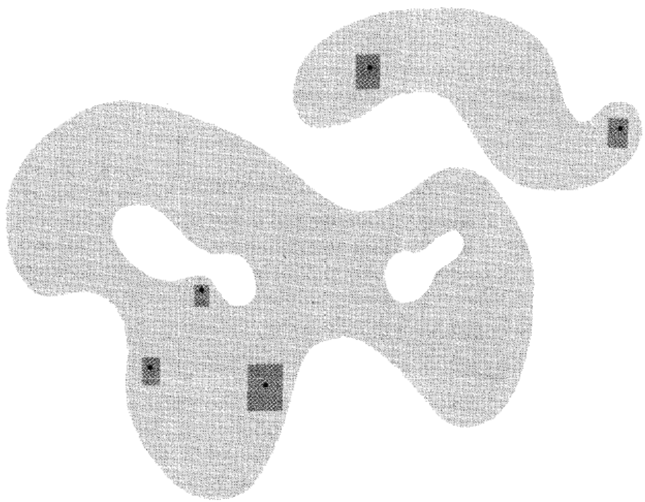


Figure (1.1)

A subset C of $\mathbf{R}^n$ is closed if $\mathbf{R}^n - C$ is open. For example, if C contains only finitely many points, then C is closed.

If $A \subset \mathbf{R}^n$ and $x \in \mathbf{R}^n$, then one of three possibilities must hold (Figure (1.2)):

1. There is an open rectangle B such that $x \in B \subset A$.
2. There is an open rectangle B such that $x \in B \subset \mathbf{R}^n - A$.
3. If B is any open rectangle with $x \in B$, then B contains points of both A and $\mathbf{R}^n - A$.

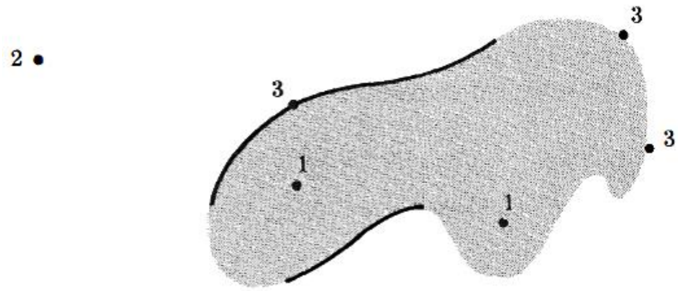


Figure (1.2)

Those points satisfying (1) constitute the interior of A, those satisfying (2) the exterior of A, and those satisfying (3) the boundary of A.

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 $\mathbf{R}^n - A$. Thus their union is open, and what remains, the boundary, must be closed.

A collection $\mathcal{O}$ of open sets is an open cover of A if every poin $x \in A$ is in some open set in the collection $\mathcal{O}$. For example, if $\mathcal{O}$ is the collection of all open intervals $(a, a+1)$ for $a \in \mathbf{R}$, then $\mathcal{O}$ is a cover of $\mathbf{R}$. Clearly no finite number of the open sets in $\mathcal{O}$ will cover $\mathbf{R}$ or, for that matter, any unbounded subset of $\mathbf{R}$. A similar situation can also occur for bounded sets. If $\mathcal{O}$ is the collection of all open intervals $(1/n, 1-1/n)$ for all integers $n > 1$, then $\mathcal{O}$ is an open 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 $\mathcal{O}$ 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.

Recognizing compact sets is greatly simplified by the following results, of which only the first has any depth.

Theorem (1.1.3): (Heine-Borel)

The closed interval $[a, b]$ is compact.

Proof:

if $\mathcal{O}$ is an open cover of $[a, b]$, let $A = \{x: a \leq x \leq b$ and $[a, x]$ is covered by some finite number of open sets in $\mathcal{O}\}$.

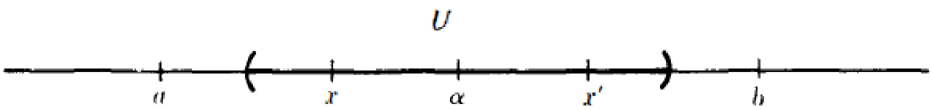


Figure (1.3)

Note that $a \in A$ and that A is clearly bounded above (by b). We would like to show that $b \in A$. This is done by proving two things about $\alpha =$ least upper bound of A; namely, (1) $\alpha \in A$ and (2) $b = \alpha$.

Since $\mathcal{O}$ is a vector, $\alpha \in U$ for some U in $\mathcal{O}$. Then all points in some interval to the left of α are also in U. Since α is the least upper bound of A, there is an x in this interval such that $x \in A$. Thus $[a, x]$ is covered by some finite number of open sets of $\mathcal{O}$, while $[x, \alpha]$ is covered by the single set U. Hence $[a, \alpha]$ is covered by finite number of open sets of $\mathcal{O}$, and $\alpha \in A$. This proves (1).

To prove that (2) is true, suppose instead that $\alpha < b$. Then there is a point x' between α and b such that $[\alpha, x'] \subset U$. Since $\alpha \in A$, the interval $[a, \alpha]$ is covered by finitely many open setes of $\mathcal{O}$, while $[\alpha, x']$ is covered by U. Hence $x' \in A$, contradicting the fact that α is an upper bound of A.

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

Theorem (1.1.4):

If B is compact and $\mathcal{O}$ is an open cover of $\{x\} \times B$, then there is an open set $U \subset \mathbf{R}^n$ containing x such that $U \times B$ is coverd by a finite number of sets in $\mathcal{O}$.

Proof:

Since $\{x\} \times B$ is compact, we can assume at the outset that $\mathcal{O}$ is finite, and we need only find the open set U such that $U \times B$ is covered by $\mathcal{O}$.

For each $y \in B$ the point (x, y) is in some open set W in $\mathcal{O}$.since W is open, we have $(x, y) \in U_y \times V_y \subset W$ for some open rectangle $U_y \times V_y$. The sets V_y cover the compact set B, so a finite number $V_{y_1}, \ldots, V_{y_k}$ also cover B. Let $U = U_{y_1} \cap \ldots \cap U_{y_k}$. then if $(x', y') \in U \times B$, we have $y' \in V_{y_i}$ for some i (Figure (1.4)), and certainly $x' \in U_{y_i}$. Hence $(x', y') \in U_{y_1} \times V_{y_1}$, which is contained in some W in $\mathcal{O}$.

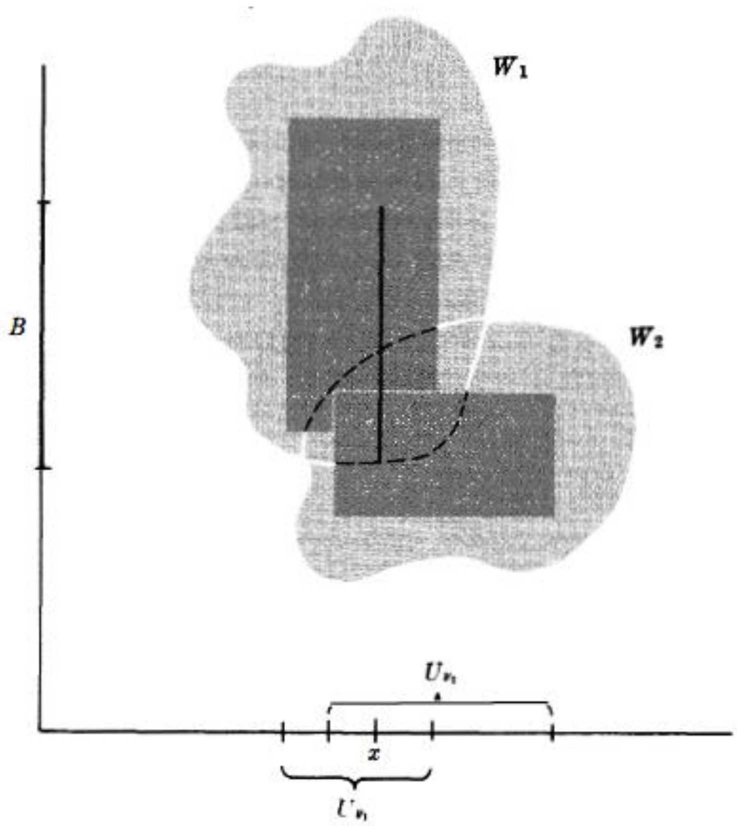


Figure (1.4)

Corollary (1.1.5):

If $A \subset \mathbf{R}^n$ and $B \subset \mathbf{R}^m$ are compact, then $A \times B \subset \mathbf{R}^{n+m}$ is compact.

Proof:

If $\mathcal{O}$ is an open cover of $A \times B$, then $\mathcal{O}$ covers $\{x\} \times B$ for each $x \in A$. By Theorem (1.1.4) there is an open set U_x containing x such that $U_x \times B$ is covered by finitely many sets in $\mathcal{O}$. Since A is compact, a finite number $U_{x_1}, \ldots, U_{x_n}$ of the U_x cover A. Since finitely many sets in $\mathcal{O}$ cover each $U_{x_i} \times B$, finitely many cover all of $A \times B$.

Corollary (1.1.6):

$A_1 \times \ldots \times A_k$ is compact if each A_i is. In particular, a closed rectangle in $\mathbf{R}^k$ is compact.

Corollary (1.1.7):

A closed bounded subset of $\mathbf{R}^n$ is compact.

Proof:

If $A \subset \mathbf{R}^n$ is closed and boundary, then $A \subset B$ for some closed rectangle B. If $\mathcal{O}$ is an open cover of A, then $\mathcal{O}$ together with $\mathbf{R}^n - A$ is an open cover of B. Hence a finite number $U_1, \ldots, U_n$ of sets in $\mathcal{O}$, together with $\mathbf{R}^n - A$ perhaps, cover B. Then $U_1, \ldots, U_n$ cover A.

A function from $\mathbf{R}^n$ to $\mathbf{R}^m$ is a rule which associates to each point in $\mathbf{R}^n$ some point in $\mathbf{R}^m$; the point a function f associates to x is denoted $f(x)$. We write $f: \mathbf{R}^n \to \mathbf{R}^m$ to indicate that $f(x) \in \mathbf{R}^m$ is defined for $x \in \mathbf{R}^n$. The notation $f: A \to \mathbf{R}^m$ indicates that $f(x)$ is defined only for x in the set A, which is called the domain of f. If $B \subset A$, we define $f(B)$ as the set of all $f(x)$ for $x \in B$, and if $C \subset \mathbf{R}^m$ 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 $f: \mathbf{R}^2 \to \mathbf{R}$ 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.

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

A function $f: A \to \mathbf{R}^m$ determines m component functions $f^1, \ldots, f^m: A \to \mathbf{R}$ by $f(x) = (f^1(x), \ldots, f^m(x))$. If conversely, m funcions $g_1, \ldots, g_m: A \to \mathbf{R}$ are given, there is unique function $f: A \to \mathbf{R}^m$ such that $f^i = g_i$, namely $f(x) = (g_1(x), \ldots, g_m(x))$. This function f will be denoted $(g_1, \ldots, g_m)$, so that we always have $f = (f^1, \ldots, f^m)$. If $\pi: \mathbf{R}^n \to \mathbf{R}^n$ 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 $\lim_{x \to \mathrm{a}} f(x) = b$ means, as in the one-variable case, that we can get $f(x)$ as close to b as desired, by chosing x sufficiently close to, but not equal to, a. In mathematical terms this means that for every number $\epsilon > 0$ there is the number $\delta > 0$ such that $|f(x) - b| < \epsilon$ for all x in the domain of f which satisfy $0 < |x - a| < \delta$. A function $f: A \to \mathbf{R}^m$ 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 $f: \mathbf{R}^n \to \mathbf{R}^m$ is continuous if and only if $f^{-1}(U)$ is open whenever $U \subset \mathbf{R}^m$ is open; if the domain of f is not all of $\mathbf{R}^n$, slightly more complicated condition is needed.

Theorem (1.1.8):

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

Proof:

Suppose f is continuous. If $a \in f^{-1}(U)$, the $f(a) \in U$. Since U is open, there is an open rectangle B with $F(a) \in B \subset U$. Since f is continuous at a, we can ensure that $f(x) \in B$, provided we choose x in some sufficiently small rectangle C containing a. Do this for each $a \in f^{-1}(U)$ and let V be the union of all such C. Clearly $f^{-1}(U) = V \cap A$.

Theorem (1.1.9):

If $f: A \to \mathbf{R}^m$ is continuous, where $A \subset \mathbf{R}^n$, and A is compact, the $f(A) \subset \mathbf{R}^m$ is compact.

Proof:

Let $\mathcal{O}$ be an open cover of $f(A)$. For each open set U in $\mathcal{O}$ there is an open set V_U such that $f^{-1}(U) = V_U \cap A$. The collection of all V_U is an open cover of A. Since A is compact, a finite number $V_{U_1}, \ldots, V_{U_n}$ cover A. Then $U_1, \ldots, U_n$ cover $f(A)$.

If $f: A \to \mathbf{R}$ 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

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

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

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 $o(f,a)$.

Theorem (1.1.10):

The bounded function f is continous at a if and only if $o(f,a) = 0$.

Proof:

Let f be continuous at a. For every number $\epsilon > 0$ we can choose a number $\delta > 0$ so that $|f(x) - f(a)| < \epsilon$ for all $x \in A$ with $|x-a| < \delta$; thus $M(a,f,\delta) - m(a,f,\delta) \le 2\epsilon$. Since this is true for every ϵ, we have $o(f,a) = 0$.

Theorem (1.1.11):

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

Proof:

Let $B = \{x \in A: o(f,x) \ge \epsilon\}$. We wish to show that $\mathbf{R}^n - B$ is open. If $x \in \mathbf{R}^n - B$, then either $x \notin A$ or else $x \in A$ and $o(f,x) < \epsilon$. In the first case, since A is closed, there is an open rectangle C containing x such that $C \subset \mathbf{R}^n - A \subset \mathbf{R}^n - B$. In the second case there is a $\delta > 0$ such that $M(x,f,\delta) - m(x,f,\delta) < \epsilon$. Let C be an open rectangle containing x such that $|x-y| < \delta$ for all $y \in C$. Then if $y \in C$ there is a δ_1 such that $|x-z| < \delta$ for all z satisfying $|z-y| < \delta_1$. Thus $M(y,f,\delta_1) - m(y,f,\delta_1) < \epsilon$, and consequently $o(y,f) < \epsilon$. Therefore $C \subset \mathbf{R}^n - B$. [13]

Section (1.2): The Concept of Differentiation

In order to arrive at a definition at a derivative of a function whose domain is $\mathbf{R}^n$, let us take another look at the familiar case $n = 1$, and let us see how to interpret the derivative in that case in a way which will naturally extend to $n > 1$. [11]

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

$$\lim_{h \to 0} \frac{(a+h) - f(a)}{h} = f'(a). \tag{1.1}$$

This equation certainly makes no sense in the general case of a function $f: \mathbf{R}^n \to \mathbf{R}^m$, but can be reformulated in a way that does. If $\lambda: \mathbf{R} \to \mathbf{R}$ is the linear transformation defined by $\lambda(h) = f'(a) \cdot h$, then equation (1.1) is equivalent to

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \lambda(h)}{h} = 0. \tag{1.2}$$

Equation (1.2) is often interpreted as saying that $\lambda + f(a)$ is 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 $f: \mathbf{R} \to \mathbf{R}$ is differentiable at $a \in \mathbf{R}$ if there is a linear transformation $\lambda: \mathbf{R} \to \mathbf{R}$ 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 $f: \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $a \in \mathbf{R}^n$ if there is a linear transformation $\lambda: \mathbf{R}^n \to \mathbf{R}^m$ such that

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

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

Theorem (1.2.1):

If $f: \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $a \in \boldsymbol{R}^n$ there is a unique linear transformation $\lambda: \mathbf{R}^n \to \mathbf{R}^m$ such that

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

Proof:

Suppose $\mu: \mathbf{R}^n \to \mathbf{R}^m$ satisfies

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

If $d(h) = f(a+h) - f(a)$, then

$$\lim_{h\to 0} \frac{|\lambda(h) - \mu(h)|}{|h|} = \lim_{h\to 0} \frac{|\lambda(h) - d(h) + d(h) - \mu(h)|}{|h|}$$

$$\leq \lim_{h\to 0} \frac{|\lambda(h) - d(h)|}{|h|} + \lim_{h\to 0} \frac{|d(h) - \mu(h)|}{|h|}$$

$$= 0.$$

If $x \in \mathbf{R}^n$, then $tx \to 0$ as $t \to 0$. Hence for $x \neq 0$ we have

$$0 = \lim_{t\to 0} \frac{|\lambda(tx) - \mu(tx)|}{|tx|} = \frac{|\lambda(x) - \mu(x)|}{|x|}.$$

Therefore $\lambda(x) = \mu(x)$.

We shall later discover a simple way of finding $Df(a)$. For the moment let us consider the function $f: \mathbf{R}^2 \to \mathbf{R}$ 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

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

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

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 $Df(a): \mathbf{R}^n \to \mathbf{R}^m$ with respect to the usual bases of $\mathbf{R}^n$ and $\mathbf{R}^m$. 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 $f: \mathbf{R} \to \mathbf{R}$, 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 $\mathbf{R}^n$ streamlines the statement of theorems and produces no real loss of generality. It is convenient to define a function $f: \mathbf{R}^n \to \mathbf{R}^m$ to be differentiable on A if f is differentiable at a for each $a \in A$. If $f: A \to \mathbf{R}^m$, then f is called differentiable if f can be extended to a differentiable function on some open set containing A. [13]

Theorem (1.2.2): (chain rule)

Let $A \subset \mathbf{R}^n$; let $B \subset \mathbf{R}^m$. Let $f: A \to \mathbf{R}^m$ and $g: B \to \mathbf{R}^p$, with $f(A) \subset B$. Suppose that f is differentiable at $a \in A$ and g is differentiable at $b = f(a)$. Then $h = g \circ f$ is differentiable at $x = a$ and we have

$$Dh(a) = Dg\big(f(a)\big) \cdot Df(a)$$

If f is differentiable on A and g on B, then this holds for every $a \in A$.

Proof:

According to the characterization above it is enough to show that the p-tuple $R_h(x,a)$ define by

$$h(x) - h(a) - Dg\big(f(a)\big) \cdot Df(a) \cdot (x-a) = \|x-a\| R_h(x,a)$$

approaches 0 as x approaches a. Using $y = f(x)$, $b = f(a)$, and the differentiability of f and g at a and b, we may write

$$h(x) - h(a) = g(y) - g(b) = Dg(b) \cdot (y-b) + \|y-b\| R_g(y,b),$$

and

$$y - b = f(x) - f(a) = Df(a) \cdot (x-a) + \|x-a\| R_f(x,a).$$

Then, replacing y by $f(x)$ and b by $f(a)$,

$$h(x) - h(a) = Dg\big(f(a)\big) \cdot Df(a) \cdot (x-a)$$
$$+ \|x-a\| \left\{ Dg\big(f(a)\big) \cdot R_f(x,a) + \frac{\|f(x)-f(a)\|}{\|x-a\|} R_g\big(f(x), f(a)\big) \right\}.$$

Using the continuity of f, which is an immediate consequence of differentiability, and the properties of $R_f(x, a)$ and $R_g(y, b)$, we see that $x \to a$ the expression in curly braces, which we may denote by $R_h(x, a)$. Goes to zero. [2]

Theorem (1.2.3):

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

$$Df(a) = 0.$$

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

$$Df(a) = f.$$

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

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

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

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

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

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

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

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

Proof:

(1) $\lim_{h \to 0} \dfrac{|f(a+h) - f(a) - 0|}{|h|} = \lim_{h \to 0} \dfrac{|y - y - 0|}{|h|} = 0.$

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

(3) If each f^i is differentiable at a and

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

then

$$f(a+h) - f(a) - \lambda(h)$$
$$= \left(f^1(a+h) - f^1(a) - Df^1(a)(h), \dots, f^m(a+h) + f^m(a) - Df^m(a)(h)\right).$$

Therefore

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} \le \lim_{h \to 0} \sum_{i=1}^{m} \frac{\left|f^i(a+h) - f^i(a) - Df^i(a)(h)\right|}{|h|}$$
$$= 0.$$

If, on the other hand, f is differentiable at a, then $f^i = \pi^i \circ f$ is differentiable at a by (2) and Theorem (1.2.2)

(4) follows from (2).

(5) Let $\lambda(x, y) = bx + ay$. Then

$$\lim_{(h,k)\to 0} \frac{|p(a+h, b+k) - p(a,b) - \lambda(h,k)|}{|(h,k)|} = \lim_{(h,k)\to 0} \frac{|hk|}{|(h,k)|}.$$

Now

$$|hk| \leq \begin{cases} |h|^2 & \text{if } |k| \leq |h|, \\ |k|^2 & \text{if } |h| \leq |k|. \end{cases}$$

Hence $|hk| \leq |h|^2 + |k|^2$. Therefore

$$\frac{|hk|}{|(h,k)|} \leq \frac{h^2 + k^2}{\sqrt{h^2 + k^2}} = \sqrt{h^2 + k^2},$$

$$\lim_{(h,k)\to 0} \frac{|hk|}{|(h,k)|} = 0.$$

Corollary (1.2.4):

If $f, g\colon \mathbf{R}^n \to \mathbf{R}$ 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

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

Proof:

(i) $f + g = s \circ (f, g)$, we have

$$\begin{aligned} D(f+g)(a) &= Ds\big(f(a), g(a)\big) \circ D(f,g)(a) \\ &= s \circ \big(Df(a), Dg(a)\big) \\ &= Df(a) + Dg(a). \end{aligned}$$

(ii) $f \cdot g = p \circ (f, g)$, so

$$\begin{aligned} D(f \cdot g)(a) &= Dp\big(f(a), g(a)\big) \circ D(f,g)(a) \\ &= Dp\big(f(a), g(a)\big)\big(Df(a), Dg(a)\big) \\ &= g(a)Df(a) + f(a)Dg(a). \end{aligned}$$

(iii) $f/g = q \circ (f,g)$

$$\begin{aligned} D(f/g)(a) &= Dq(f/g)\big(f(a),g(a)\big) \circ D(f,g)(a) \\ &= Dq\big(f(a),g(a)\big)\big(Df(a),Dg(a)\big) \\ &= \frac{g(a)Df(a) - f(a)Dg(a)}{[g(a)]^2}\,. \end{aligned}$$

We are now assured of the differentiability of those functions $f\colon \mathbf{R}^n \to \mathbf{R}^m$, whose component functions are obtained by addition, multiplication, division, and composition, from the functions π^i and the functions which we can already differentiate by elementary calculus. Finding $Df(x)$ or $f'(x)$, however, may be a fairly formidable task. For example, let $f\colon \mathbf{R}^2 \to \mathbf{R}$ be defined by $f(x,y) = \sin(xy^2)$. Since $f = \sin \circ\ (\pi^1 \cdot [\pi^2]^2)$, we have

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

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

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

if it exists, is denoted $D_i f(a)$, and called the 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 $D_i f(a) = g'\big(a^i\big)$. This means that $D_i f(a)$ is the slope of the tangent line at $\big(a, f(a)\big)$ to the curve obtained by intersecting the graph of f with the plane $x^j = a^j, j \neq i$ (Figure (1.6)). 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. For 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, instead, $f(x,y) = x^y$, then $D_1 f(x,y) \;=\; yx^{y-1}$ and $D_2 f(x,y) = x^y \log x$.

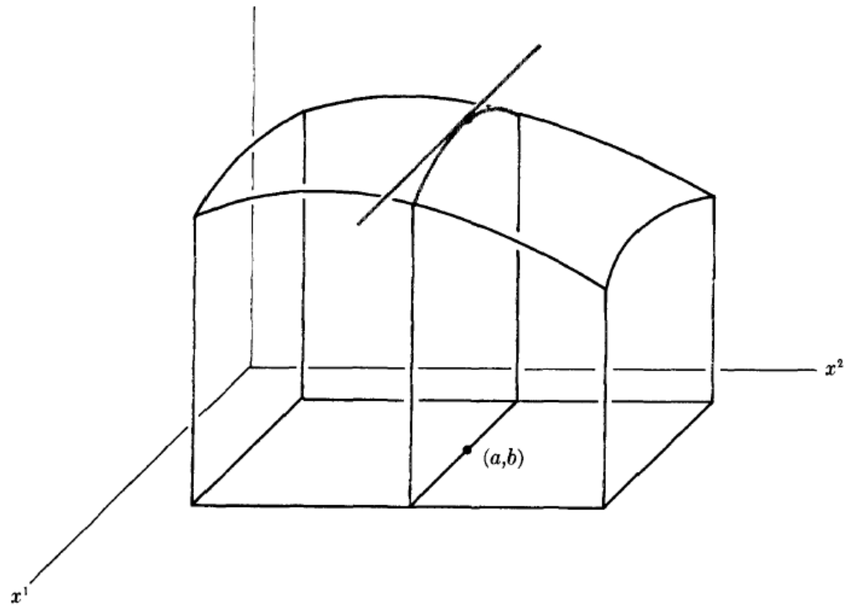


Figure (1.6)

To acquire as great a facility for computing $D_i f$ as we already have for computing ordinary derivatives solve problems by Spivak [13].

If $D_i f(x)$ exists for all $x \in \mathbf{R}^n$, we obtain a function $D_i f: \mathbf{R}^n \to \mathbf{R}$. The jth partial derivative of this function at x, that is, $D_j\big(D_i f(x)\big)$, 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.

Theorem (1.2.5):

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 (1.2.5) can be used to prove the equality of higher-order mixed partial derivatives under appropriate conditions. The order of $i_1, \ldots, i_k$ is completely immaterial in $D_{i_1, \ldots, i_k} f$ 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 section to find derivatives. They also have another important use-finding maxima and minima of functions.

Theorem (1.2.6):

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

Proof:

Let $g_i(x) = f(a^1, \dots, x, \dots, a^n)$. Clearly g_i has a maximum (or minimum) at a^i, and g_i is defined in an open interval containing a^i. Hence $0 = g_i{}'(a^i) = D_i f(a)$.

The converse of Theorem (1.2.6) is false even if $n = 1$ (if $f: \mathbf{R} \to \mathbf{R}$ is defined by $f(x) = x^3$, then $f'(0) = 0$, but 0 is not even a local maximum or minimum). If $n > 1$, the converse of Theorem (1.2.6) may fail to be true in a rather spectacular way. Suppose, for example, that $f: \mathbf{R}^2 \to \mathbf{R}$ is defined by $f(x, y) = x^2 - y^2$ (Figure (1.7)). Then $D_1 d(0,0) = 0$ 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.

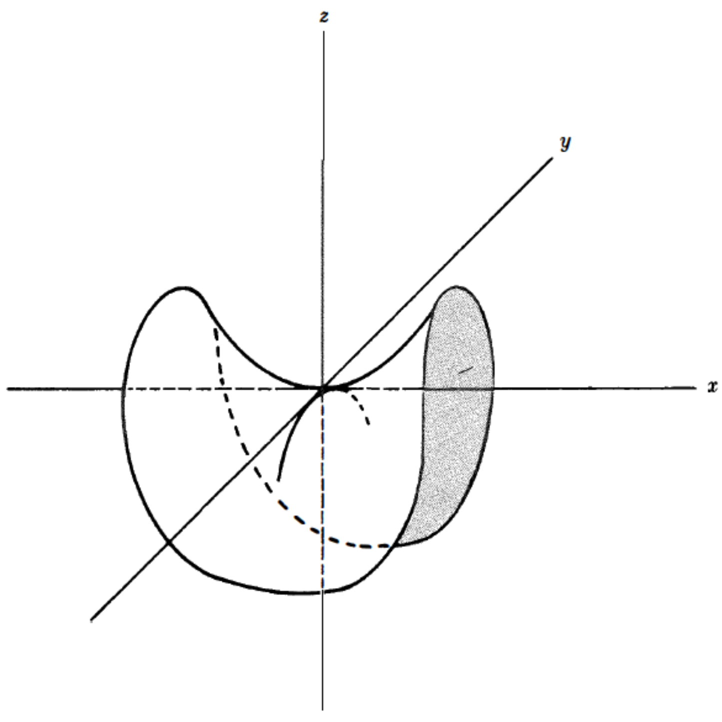


Figure (1.7)

If Theorem (1.2.6) 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 of A!

Theorem (1.2.7):

If $f: \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at a, then $D_j f^i(a)$ exists for $1 < i < m$, $1 < j < n$ and $f'(a)$ is the $m \times n$ matrix $\left(D_j f^i(a)\right)$.

Proof:

Suppose first that $m = 1$, so that $f: \mathbf{R}^n \to \mathbf{R}$. Define $h: \mathbf{R} \to \mathbf{R}^n$ by $h(x) = (a^1, \dots, x, \dots, a^n)$, with x in the jth place. Then $D_j f(a) = (f \circ h)'(a^j)$. Hence, by Theorem (1.2.2),

$$(f \circ h)'(a^j) = f'(a) \cdot h'(a^j)$$

$$= f'(a) \cdot \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \longleftarrow j\text{th place.}$$

Since $(f \circ h)'(a^j)$ has the single entry $D_j f(a)$, this shows that $D_j f(a)$ exists and is the jth entry of the $1 \times n$ matrix $f'(a)$.

The theorem now follows for arbitrary m since, by Theorem (1.2.3), each f^i is differentiable and the ith row of $f'(a)$ is $\left(f^i\right)'(a)$.

There are several examples in the problems to show that the converse of Theorem (1.2.7) is false. It is true, however, if one hypothesis is added.

Theorem (1.2.8):

If $f: \mathbf{R}^n \to \mathbf{R}^m$, then $Df(a)$ exists if all $D_j f^i(x)$ exist in an open set containing a and if each function $D_j f^i$ is continuous at a.

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

Proof:

As in the proof of Theorem (1.2.7), it suffices to consider the case $m = 1$, so that $f: \mathbf{R}^n \to \mathbf{R}$. Then

$$\begin{aligned} f(a+h) - f(a) = {} & f(a^1 + h^1, a^2, \dots, a^n) - f(a^1, \dots, a^n) \\ & + f(a^1 + h^1, a^2 + h^2, a^3, \dots, a^n) - f(a^1 + h^1, a^2, \dots, a^n) + \dots \\ & + f(a^1 + h^1, \dots, a^n + h^n) - f(a^1 + h^1, \dots, a^{n-1} + h^{n-1}, a^n). \end{aligned}$$

Recall that $D_1 f$ is the derivative of the function g defined by $g(x) = f(x, a^2, \dots, a^n)$. Applying the mean-value theorem to g we obtain

$$f(a^1 + h^1,\ a^2, \dots,\ a^n) - f(a^1, \dots,\ a^n) = \ h^1 \cdot D_1 f(b_1,\ a^2, \dots,\ a^n)$$

for some b_1 between a^1 and $a^1 + h^1$. Similarly the ith term in the sum equals

$$h^i \cdot D_i f\big(a^1 + h^1, \dots, a^{i-1} + h^{i-1}, b_i, \dots, a^n\big) = h^i D_i f(c_i),$$

for some c_i. Then

$$\begin{aligned}
\lim_{h \to 0} \frac{\left|f(a+h) - f(a) - \sum_{i=1}^{n} D_i f(a) \cdot h^i\right|}{|h|} &= \lim_{h \to 0} \frac{\left|\sum_{i=1}^{n} [D_i f(c_i) - D_i f(a)] \cdot h^i\right|}{|h|} \\
&\le \lim_{h \to 0} \sum_{i=1}^{n} |D_i f(c_i) - D_i f(a)| \cdot \frac{|h^i|}{|h|} \\
&\le \lim_{h \to 0} \sum_{i=1}^{n} |D_i f(c_i) - D_i f(a)| \\
&= 0,
\end{aligned}$$

since $D_i f$ is continuous at a.

Although the chain rule was used in the proof of Theorem (1.2.7), it could easily have been eliminated. With Theorem (1.2.8) to provide differentiable functions, and Theorem (1.2.7) to provide their derivatives, the chain rule may therefore seem almost superfluous. However, it has an extremely important corollary concerning partial derivatives.

Theorem (1.2.9):

Let $g_1, \dots, g_m \colon \mathbf{R}^n \to \mathbf{R}$ be continuously differentiable at a, and let $f \colon \mathbf{R}^m \to \mathbf{R}$ be differentiable at $\big(g_1(a), \dots, g_m(a)\big)$. Define the function $F \colon \mathbf{R}^n \to \mathbf{R}$ by $F(x) = f\big(g_1(x), \dots, g_m(x)\big)$. Then

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

Proof:

The function F is just the composition $f \circ g$, where $g = (g_1, \dots, g_m)$. Since g_i is continuously differentiable at a, it follows from Theorem (1.2.8) that g is differentiable at a. Hence by Theorem (1.2.2),

$$F'(a) = f'(g(a)) \cdot g'(a)$$

$$= \left(D_1 f(g(a)), \ldots, D_m f(g(a))\right) \cdot \begin{pmatrix} D_1 g_1(a), & \cdots & , D_n g_1(a) \\ \vdots & & \vdots \\ D_1 g_m(a), & \cdots & , D_n g_m(a) \end{pmatrix}$$

But $D_i F(a)$ is the ith entry of the left side of this equation, while $\sum_{j=1}^{m} D_j f(g_1(a), \ldots, g_m(a)) \cdot D_i g_j(a)$ is the ith entry of the right side. Theorem (1.2.9) is often called the chain rule, but is weaker than Theorem (1.2.2) since g could be differentiable without g_i being continuously differentiable. Most computations requiring Theorem (1.2.9) are fairly straightforward. A slight subtlety is required for the function $F: \mathbf{R}^2 \to \mathbf{R}$ defined by

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

where $h, k: \mathbf{R} \to \mathbf{R}$. In order to apply Theorem (1.2.9) define $\bar{h}, \bar{k}: \mathbf{R}^2 \to \mathbf{R}$ 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

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

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

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

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

Suppose that $f: \mathbf{R} \to \mathbf{R}$ 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 $f'(x) > 0$ for $x \in V$, and a similar statement holds if $f'(a) < 0$. Thus f is increasing (or decreasing) on V, and is therefore $1-1$ 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 (1.2.11) is very important. We begin with a simple lemma.

Lemma (1.2.10):

Let $A \subset \mathbf{R}^n$ be a rectangle and let $f: A \to \mathbf{R}^n$ 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

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

for all $x, y \in A$.

Proof:

We have

$$f^i(y) - f^i(x) = \sum_{j=1}^{n}\left[f^i\left(y^1, \dots, y^j, x^{j+1}, \dots, x^n\right) - f^i\left(y^1, \dots, y^{j-1}, x^j, \dots, x^n\right)\right].$$

Applying the mean-value theorem, we obtain

$$f^i\left(y^1, \dots, y^j, x^{j+1}, \dots, x^n\right) - f^i\left(y^1, \dots, y^{j-1}, x^j, \dots, x^n\right) = \left(y^j - x^j\right) \cdot D_j f^i\left(z_{ij}\right)$$

for some z_{ij}. The expression on the right has absolute value less than or equal to $M \cdot \left|y^j - x^j\right|$. Thus

$$\left|f^i(y) - f^i(x)\right| \le \sum_{j=1}^{n}\left|y^j - x^j\right| \cdot M \le nM|y - x|$$

since each $\left|y^j - x^i\right| \le |y - x|$. Finally

$$|f(y) - f(x)| \le \sum_{i=1}^{n}\left|f^i(y) - f^i(x)\right| \le n^2 M \cdot |y - x|.$$

Theorem (1.2.11): (Inverse Function Theorem)

Suppose that $f: \mathbf{R}^n \to \mathbf{R}^n$ is continuously differentiable in an open set containing a, and $\det f'(a) \neq 0$. 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

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

Proof:

Let λ be the linear transformation $Df(a)$. Then λ is non-singular, since $\det f'(a) \neq 0$. Now $D(\lambda^{-1} \circ f)(a) = D(\lambda^{-1})(f(a)) \circ Df(a) = \lambda^{-1} \circ Df(a)$ is the identity linear transformation. If the theorem is true for $\lambda^{-1} \circ f$, it is clearly true for f. Therefore we may assume at the outset that λ is the identity. Thus whenever $f(a+h) = f(a)$, we have

$$\frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = \frac{|h|}{|h|} = 1.$$

But

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

This means that we cannot have $f(x) = f(a)$ for x arbitrarily close to, but unequal to, a. Therefore there is a closed rectangle U containing a in its interior such that

1. $f(x) \neq f(a)$ if $x \in U$ and $x \neq a$.

Since f is continuously differentiable in an open set containing a, we can also assume that

2. $\det f'(x) \neq 0$ for $x \in U$.
3. $\left|D_j f^i(x) - D_j f^i(a)\right| < 1/2n^2$ for all i, j, and $x \in U$.

Note that (3) and Lemma (1.2.10) applied to $g(x) = f(x) - x$ imply for $x_1, x_2 \in U$ that

$$|f(x_1) - x_1 - (f(x_2) - x_2)| \leq \tfrac{1}{2}|x_1 - x_2|.$$

Since

$$|x_1 - x_2| - |f(x_1) - f(x_2)| \leq |f(x_1) - x_1 - (f(x_2) - x_2)| \leq \tfrac{1}{2}|x_1 - x_2|,$$

we obtain

4. $|x_1 - x_2| \leq 2|f(x_1) - f(x_2)|$ for $x_{1,}x_2 \in U$.

Now f(boundary U) is a compact set which, by (1), does not contain $f(a)$ (Figure (1.8)). Therefore there is a number $d > 0$ such that $|f(a) - f(x)| \geq d$ for $x \in$ boundary U. Let $W = \{y: |y - f(a)| < d/2\}$. If $y \in W$ and $x \in$ boundary U, then

5. $|y - f(a)| < |y - f(x)|$.

We will show that for any $y \in W$ there is a unique x in interior U such that $f(x) = y$. To prove this consider the function $g: U \to \mathbf{R}$ defined by

$$g(x) = |y - f(x)|^2 = \sum_{i=1}^{n} \left(y^i - f^i(x)\right)^2.$$

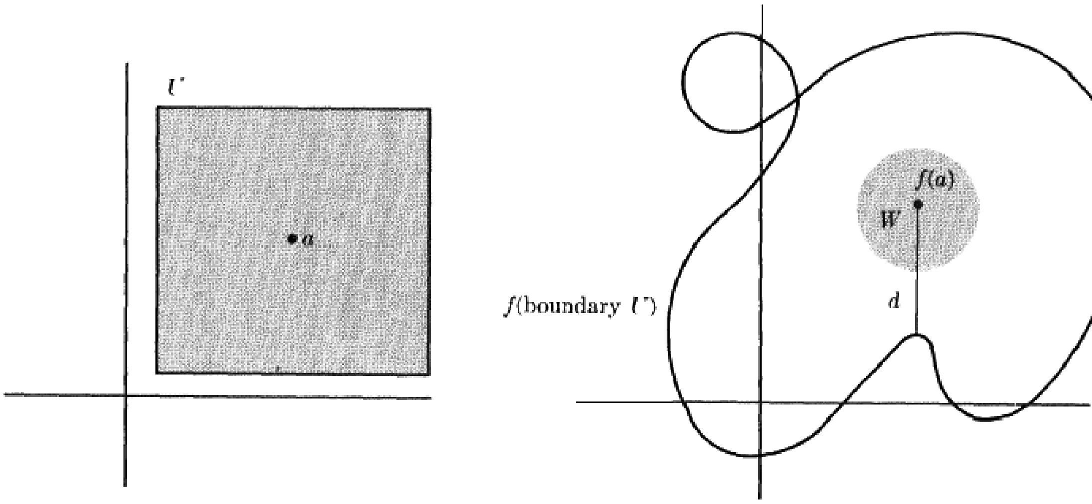


Figure (1.8)

This function is continuous and therefore has a minimum on U. If $x \in$ boundary U, then, by (5), we have $g(a) < g(x)$. Therefore the minimum of g does not occur on the boundary of U. By Theorem (1.2.6) there is a point $x \in$ interior U such that $D_j g(x) = 0$ for all j, that is

$$\sum_{i=1}^{n} 2\left(y^i - f^i(x)\right) \cdot D_j f^i(x) = 0 \qquad \text{for all } j.$$

By (2) the matrix $\left(D_j f^i(x)\right)$ has non-zero determinant. Therefore we must have $y^i - f^i(x) = 0$ for all i, that is $y = f(x)$. This proves the existence of x. Uniqueness follows immediately from (4).

If $V =$ (interior U) $\cap f^{-1}(W)$, we have shown that the function $f: V \to W$ has an inverse $f^{-1}: W \to V$. We can rewrite (4) as

6. $|f^{-1}(y_1) - f^{-1}(y_2)| \leq 2|y_1 - y_2|$ for $y_1, y_2 \in W$.

This shows that f^{-1} is continuous.

Only the proof that f^{-1} is differentiable remains. Let $\mu = Df(x)$. We will show that f^{-1} is differentiable at $y = f(x)$ with derivative μ^{-1}.

$$f(x_1) = f(x) + \mu(x_1 - x) + \varphi(x_1 - x),$$

where

$$\lim_{x_1 \to x} \frac{|\varphi(x_1 - x)|}{|x_1 - x|} = 0.$$

Therefore

$$\mu^{-1}\big(f(x_1) - f(x)\big) = x_1 - x + \mu^{-1}\big(\varphi(x_1 - x)\big).$$

Since every $y_1 \in W$ is of the form $f(x_1)$ for some $x_1 \in V$, this can be written

$$f^{-1}(y) = f^{-1}(y) + \mu^{-1}(y_1 - y) - \mu^{-1}\left(\varphi\big(f^{-1}(y_1) - f^{-1}(y)\big)\right),$$

and it therefore suffices to show that

$$\lim_{y_1 \to y} \frac{\left|\mu^{-1}\left(\varphi\big(f^{-1}(y_1) - f^{-1}(y)\big)\right)\right|}{|y_1 - y|} = 0.$$

Therefore it suffices to show that

$$\lim_{y_1 \to y} \frac{\left|\varphi\big(f^{-1}(y_1) - f^{-1}(y)\big)\right|}{|y_1 - y|} = 0.$$

Now

$$\frac{\left|\varphi\big(f^{-1}(y_1) - f^{-1}(y)\big)\right|}{|y_1 - y|} = \frac{\left|\varphi\big(f^{-1}(y_1) - f^{-1}(y)\big)\right|}{|f^{-1}(y_1) - f^{-1}(y)|} \cdot \frac{|f^{-1}(y_1) - f^{-1}(y)|}{|y_1 - y|}.$$

Since f^{-1} is continuous, $f^{-1}(y_1) \longrightarrow f^{-1}(y)$ as $y_1 - y$. Therefore the first factor approaches 0. Since, by (6), the second factor is less than 2, the product also approaches 0.

It should be noted that an inverse function f^{-1} may exist even if $\det f'(a) = 0$. For example, if $f: \mathbf{R} \to \mathbf{R}$ is defined by $f(x) = x^3$, then $f'(0) = 0$ 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 $f'(a) \cdot (f^{-1})'\big(f(a)\big) = I$, and consequently $\det f'(a) \cdot \det(f^{-1})'\big(f(a)\big) = 1$, contradicting $\det f'(a) = 0$.

Consider the function $f: \mathbf{R}^2 \to \mathbf{R}$ defined by $f(x, y) = x^2 + y^2 - 1$. If we choose (a, b) with $f(a, b) = 0$ and $a \neq 1, -1$, there are (Figure (1.9)) open intervals 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 $g: A \to \mathbf{R}$ by the condition $g(x) \in B$ and $f(x, g(x)) = 0$ (if $b > 0$, as indicated in Figure (1.9), 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 $f(x, y) = 0$. [13]

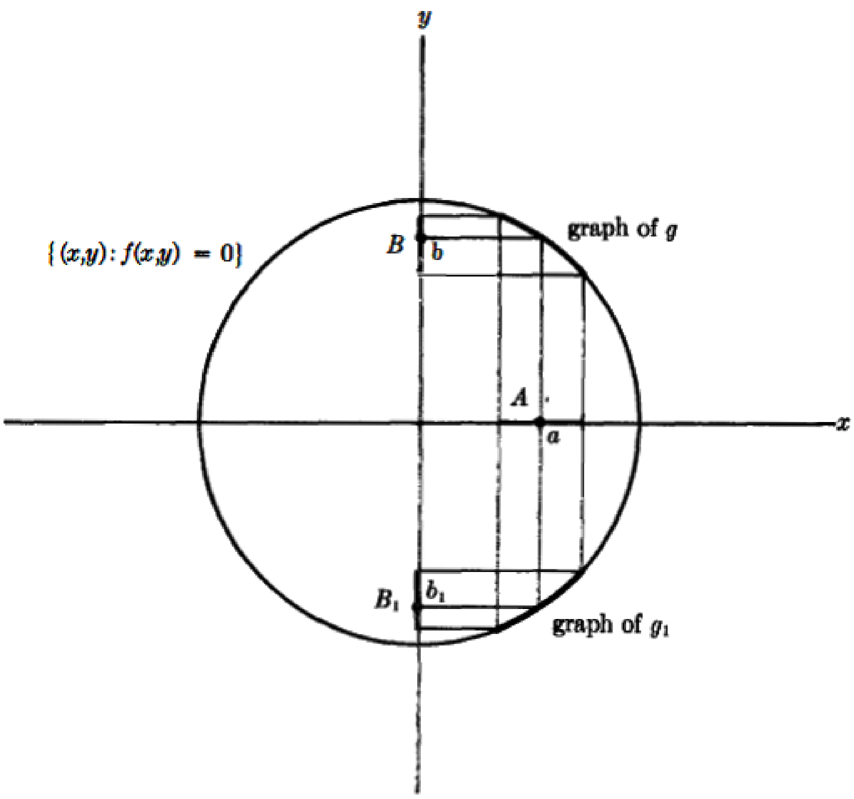


Figure (1.9)

Example (1.2.12):

Let $f: \mathbf{R}^2 \to \mathbf{R}$ be given by the equation

$$f(x, y) = x^2 - y^3.$$

Then (0,0) is a solution of the equation $f(x, y) = 0$. Because $\partial f / \partial y$ vanishes at (0,0), we do not expect to be able to solve this equation for y in terms of x near (0,0). But in fact, we can; and furthermore, the solution is unique! However, the function we obtain is not differentiable at $x = 0$. See Figure (1.10)

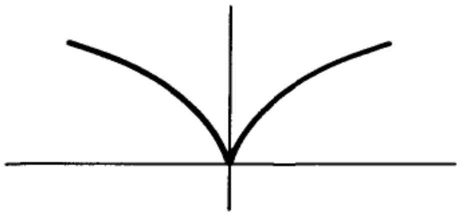


Figure (1.10)

Example (1.2.13):

Let $f: \mathbf{R}^2 \to \mathbf{R}$ be given by the equation

$$f(x,y) = y^2 - x^4.$$

Then $(0,0)$ is a solution of the equation $f(x,y) = 0$. Because $\partial f/\partial y$ vanishes at $(0,0)$, we do not expect to be able to solve for y in terms of x near $(0,0)$. In fact, however, we can do so, and we can do so in such a way that the resulting function is differentiable. However, the solution is not unique. The point $(1,2)$ also satisfies the equation $f(x,y) = 0$. Because $\partial f/\partial y$ is non-zero at $(1,2)$, one can solve this equation for y as a continuous function of x in a neighbourhood of $x = 1$. See Figure (1.11). One can in fact express y as a continuous function of x on a larger neighborhood than the one pictured, but if the neighborhood is large enough that it contains 0, then the solution is not unique on that larger neighborhood. [10]

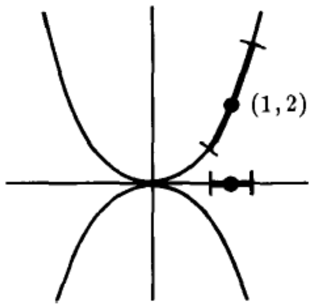


Figure (1.11)

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

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

and

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

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

Theorem (1.2.14): (Implicit Function Theorem)

Suppose $f: \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$ is continuously differentiable in an open set containing (a, b) and $f(a, b) = 0$. Let M be the $m \times m$ matrix

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

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

Proof:

Define $F: \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m$ by $(x, y) = \big(x, f(x, y)\big)$. Then $\det F'(a, b) = \det M \neq 0$. By Theorem (1.2.11) there is an open set $W \subset \mathbf{R}^n \times \mathbf{R}^m$ containing $F(a, b) = (a, 0)$ and an open set in $\mathbf{R}^n \times \mathbf{R}^m$ containing (a, b), which we may take to be of the form $A \times B$, such that $F: A \times B \longrightarrow W$ has a differentiable inverse $h: W \longrightarrow A \times B$. Clearly h is of the form $h(x, y) = \big(x, k(x, y)\big)$ for some differentiable function k (since F is of this form). Let $\pi: \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$ be defined by $\pi(x, y) = y$; then $\pi \circ F = f$. Therefore

$$f\big(x, k(x, y)\big) = f \circ h(x, y) = (\pi \circ F) \circ h(x, y) = \pi \circ (F \circ h)(x, y) = \pi(x, y) = y.$$

Thus $f\big(x, k\,(x, 0)\big) = 0$; in other words we can define $g(x) = k(x, 0)$.

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

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

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

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

or

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

$$g'(x) = -x/g(x),$$

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

Theorem (1.2.15):

Let $f: \mathbf{R}^n \to \mathbf{R}^p$ be continuously differentiable in an open set containing a, where $p \leq n$. If $f(a) = 0$ and the $p \times n$ matrix $\big(D_j f^i(a)\big)$has rank p, then there is an open set $A \subset \mathbf{R}^n$ containing a and a differentiable function $h: A \to \mathbf{R}^n$ with differentiable inverse such that

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

Proof:

We can consider f as a function $f: \mathbf{R}^{n-p} \times \mathbf{R}^p \to \mathbf{R}^p$. If $\det M \neq 0$, then M is the $p \times p$ matrix $\big(D_{n-p+j} f^i(a)\big)$, $1 < i, j < p$, then we are precisely in the situation considered in the proof of Theorem (1.2.12), and as we showed in that proof, there is h such that $f \circ h(x^1, \ldots, x^n) = (x^{n-p+1}, \ldots, x^n)$.

In general, since $\big(D_j f^i(a)\big)$ has rank p, there will be it $j_1 < \cdots < j_p$ such that the matrix $\big(D_j f^i(a)\big)$ $1 \leq i \leq p, j = j_1, \ldots, j_p$ has non-zero determinant. If $g: \mathbf{R}^n \to \mathbf{R}^n$ permutes the x^j so that $g(x^1, \ldots, x^n) = \big(\ldots, x^{j_1}, \ldots, x^{j_p}\big)$, then $f \circ g$ is a function of the type already considered, so $\big((f \circ g) \circ k\big)(x^1, \ldots, x^n) = (x^{n-p+1}, \ldots, x^n)$ for some k. Let $h = g \circ k$. [13]

Chapter (2)

Measure, Content and Fubini's Theorem

Section (2.1): Measure Zero and Content Zero

In this section, we define the integral of real-valued function of several real variables, and derive its properties. The integral we study is called Riemann integral.

We begin by defining the volume of a rectangle. Let

$$A = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

be a rectangle in $\mathbf{R}^n$. Each of the intervals $[a_i, b_i]$ is called a component interval of A. The maximum of the numbers $b_1 - a_1, \ldots, b_n - a_n$ an is called the width of A. Their product

$$v(A) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n)$$

is called the volume of A .

In the case n= 1 , the volume and the width of the (1-dimensional) rectangle $[a, b]$ are the same, namely, the number $b - a$. This number is also called the length of $[a, b]$. [10]

Definition (2.1.1):

Let $[a, b]$ be a given interval. By a partition P of $[a, b]$ we mean a finite set of points $t_0, t_1, \ldots, t_k$, where

$$a = t_0 \leq t_1 \leq \cdot \cdot \cdot \leq t_{k-1} \leq t_k = b.$$

We write

$$\Delta t_i = t_i - t_{i-1} \qquad (i = 1, \ldots, k). \text{ [11]}$$

The intervals $[t_{i-1} - t_i]$, is called a subinterval determined by P, of the interval $[a, b]$. More generally, given a rectangle

$$A = [a_1, b_1] \times \cdots \times [a_n, b_n]$$

in $\mathbf{R}^n$, a partition P of A is an n-tuple $(P_1, \ldots, P_n)$ such that P_j; is a partition of $[a_j, b_j]$ for each j. If for each j, I_j is one of the subintervals determined by P_j of the interval $[a_j, b_j]$, then the rectangle

$$S = I_1 \times \cdot \cdot \cdot \times I_n$$

is called a subrectangle determined by P, of the rectangle A. The maximum width of these subrectangles is called the mesh of P.

Definition (2.1.2):

Let A be a rectangle in $\mathbf{R}^n$; let $f: A \to \mathbf{R}$; assume f is bounded. Let P be a partition of A. For each subrectangle S determined by P, let

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

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

We define the lower sum and the upper sum, respectively, of, determined by P, by the equations

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

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

where the summations extend over all subrectangles S determined by P. [10]

Clearly $L(f, P) \leq U(f, P)$, and an even stronger assertion (2.1.4) is true.

Lemma (2.1.3):

Suppose the partition P' refines P (that is, each subrectangle of P' is contained in a subrectangle of P). Then

$$L(f, P) \leq L(f, P') \qquad \text{and} \qquad U(f, P') \leq U(f, P).$$

Proof:

Each subrectangle S of P is divided into several subrectangles $S_1, \ldots, S_\alpha$ of P', so $v(S) = v(S_1) + \cdot \cdot \cdot + v(S_\alpha)$. Now $m_S(f) \leq m_{S_i}(f)$, since the values $f(x)$ for $x \in S$ include all values $f(x)$ for $x \in S_i$ (and possibly smaller ones). Thus

$$m_S(f) \cdot v(S) = m_S(f) \cdot v(S_1) + \cdot \cdot \cdot + m_S(f) \cdot v(S_\alpha)$$

$$\leq m_{S_1}(f) \cdot v(S_1) + \cdot \cdot \cdot + m_{S_\alpha}(f) \cdot v(S_\alpha).$$

The sum, for all S, of the terms on the left side is $L(f, P)$, while the sum of all the terms on the right side is $L(f, P')$. Hence $L(f, P) \leq L(f, P')$. The proof for upper sums is similar. [13]

Lemma (2.1.4):

Let A be a rectangle; let $f: A \to \mathbf{R}$ be a bounded function. If P and P' are any two partitions of A, then $L(f, P') \leq U(f, P)$.

Proof:

In the case where $P = P'$, the result is obvious: For any subrectangle S determined by P, we have $m_S(f) \leq M_S(f)$. Multiplying by $v(S)$ and summing gives the desired inequality.

In general, given partitions P and P' of A, let P'' be their common refinement. Using the preceding lemma, we conclude that

$$L(f, P') \leq L(f, P'') \leq U(f, P'') \leq U(f, P).$$

Now (finally) we define the integral.

Definition (2.1.5):

Let A be a rectangle; let $f: A \to \mathbf{R}$ be a bounded function. As P ranges over all partitions of A, define

$$L\int_A f = \sup_P\{L(f, P)\} \qquad \text{and} \qquad U\int_A f = \inf_P\{U(f, P)\}$$

These numbers are called the lower integral and upper integral, respectively, of f over A. They exist because the numbers $L(f, P)$ are bounded above by $U(f, P')$ where P' is any fixed partition of A; and the numbers $U(f, P)$ are bounded below by $L(f, P')$. If the upper and lower integrals of f over A are equal, we say f is integrable over A, and we define the integral of f over A to equal the common value of the upper and lower integrals. We denote the integral of f over A by either of the symbols

$$\int_A f \qquad \text{or} \qquad \int_{x \in A} f(x).$$

Example (2.1.6):

Let $f:[a,b] \to \mathbf{R}$ be a non-negative bounded function. If P is a partition of $I = [a,b]$, then $L(f,P)$ equals the total area of a bunch of rectangles inscribed in the region between the graph of f and the x-axis, and $U(f,P)$ equals the total area of a bunch of rectangles circumscribed about this region. See Figure (2.1)

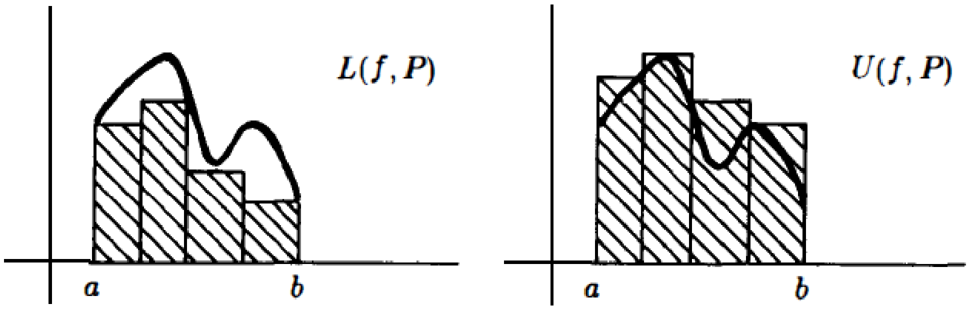


Figure (2.1)

The lower integral represents the "inner area" of this region, computed by approximating the region by inscribed rectangles, while the upper integral represents the "outer area," computed by approximating the region by circumscribed rectangles. If the "inner" and "outer" areas are equal, then f is integrable.

Similarly, if A is a rectangle in $\mathbf{R}^2$ and $f: A \to \mathbf{R}$ is non-negative and bounded, one can picture $L(f,P)$ as the total volume of a bunch of boxes inscribed in the region between the graph of f and the xy-plane, and $U(f,P)$ as the total volume of a bunch of boxes circumscribed about this region. See Figure (2.2).

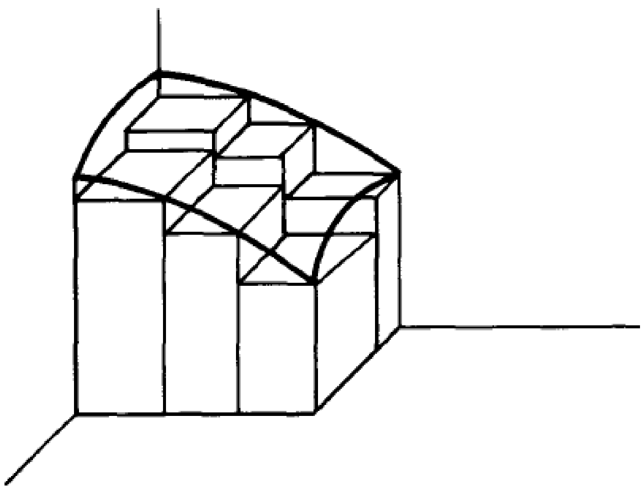


Figure (2.2)

Example (2.1.7):

Let $I = [0,1]$. Let $f: I \to \mathbf{R}$ be defined by setting $f(x) = 0$ if x is rational, and $f(x) = 1$ if x is irrational. We show that f is not integrable over I.

Let P be a partition of I. If S is any subinterval determined by P, then $m_S(f) = 0$ and $M_S(f) = 1$, since S contains both rational and irrational numbers. Then

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

and

$$U(f,P) = \sum_S 1 \cdot v(S) = 1.$$

Since P is arbitrary, it follows that the lower integral of f over I equals 0, and the upper integral equals 1. Thus f is not integrable over I.

A condition that is often useful for showing that a given function is integrable is the following:

Theorem (2.1.8): (The Riemann condition)

Let A be a rectangle; let $f: A \to \mathbf{R}$ be a bounded function. Then

$$L\int_A f \le U\int_A f\,;$$

equality holds if and only if given $\epsilon > 0$, there exists a corresponding partition P of A for which

$$U(f,P) - L(f,P) < \epsilon.$$

Proof:

Let P' be a fixed partition of A. It follows from the fact that $L(f,P) \le U(f,P')$ for every partition P of A, that

$$L\int_A f \le U(f,P').$$

Now we use the fact that P' is arbitrary to conclude that

$$L\int_A f \le U\int_A f.$$

Suppose now that the upper and lower integrals are equal. Choose a partition P so that $L(f,P)$ is within $\epsilon/2$ of the integral $\int_A f$, and a partition P' so that $U(f,P')$ is within $\epsilon/2$ of the integral $\int_A f$. Let P'' be their common refinement. Since

$$L(f,P') \leq L(f,P'') \leq \int\limits_A f \leq U(f,P'') \leq U(f,P),$$

the lower and upper sums for f determined by P'' are within ϵ of each other.

Conversely, suppose the upper and lower integrals are not equal. Let

$$\epsilon = U\int\limits_A f - L\int\limits_A f > 0.$$

Let P be any partition of A. Then

$$L(f,P) \leq L\int\limits_A f < U\int\limits_A f \leq U(f,P);$$

hence the upper and lower sums for f determined by P are at least ϵ apart. Thus the Riemann condition does not hold.

Here is an easy application of this theorem.

Theorem (2.1.9):

Every constant function $f(x) = c$ is integrable. Indeed, if A is a rectangle and if P is a partition of A, then

$$\int\limits_A c = c \cdot v(A) = c\sum_S v(S),$$

where the summation extends over all subrectangles determined by P.

Proof:

If S is a subrectangle determined by P, then $m_S(f) = c = M_S(f)$. It follows that

$$L(f,P) = c\sum_S v(S) = U(f,P),$$

so the Riemann condition holds trivially. Thus $\int_A c$ exists; since it lies between $L(f,P)$ and $U(f,P)$, it must equal $c\sum_S v(S)$.

This result holds for any partition. In particular, if P is the trivial partition whose only subrectangle is A itself,

$$\int_A c = c \cdot v(A).$$

Now we can derive a necessary and sufficient condition for the existence of the integral $\int_A f$. It involves the notion of a "set of measure 0." [10]

Definition (2.1.10):

We recall that a set has measure 0 in $\mathbf{R}^n$ if and only if, given ϵ, there exists a covering of the set by a sequence of rectangles $\{U_i\}$ such that

$$\sum v(U_i) < \epsilon. [7]$$

If this inequality holds, we often say that the total volume of the rectangles $U_1, U_2, \dots$ is less than ϵ.

We derive some properties of sets of measure 0.

Theorem (2.1.11):

(a) If $B \subset A$ and A has measure 0 in $\mathbf{R}^n$, then so does B.

(b) Let A be the union of the countable collection of sets $A_1, A_2, \dots$. If each A_i has measure 0 in $\mathbf{R}^n$, so does A.

(c) A set A has measure 0 in $\mathbf{R}^n$ if and only if for every $\epsilon > 0$, there is a countable covering of A by open rectangles $\mathrm{Int}\, U_1, \mathrm{Int}\, U_2, \dots$ such that

$$\sum_{i=1}^{\infty} v(U_i) < \epsilon\,.$$

(d) If U is a rectangle in $\mathbf{R}^n$, then $\mathrm{Bd}\, U$ has measure 0 in $\mathbf{R}^n$ but U does not.

Proof:

(a) is immediate. To prove (b), cover the set A_j by countably many rectangles

$$U_{1j},\ U_{2j}, U_{3j}, \dots$$

of total volume less than $\epsilon/2^j$. Do this for each j. Then the collection of rectangles $\{U_{ij}\}$ is countable, it covers A, and it has total volume less than

$$\sum_{j=1}^{\infty} \epsilon/2^j = \epsilon.$$

(c) If the open rectangles $\operatorname{Int} U_1, \operatorname{Int} U_2, \dots$ cover A, then so do the rectangles $U_1, U_2, \dots$. Thus the given condition implies that A has measure 0. Conversely, suppose A has measure 0. Cover A by rectangles $U'_1, U'_2, \dots$ of total volume less than $\epsilon/2$. For each i, choose a rectangle U_i such that

$$U'_i \subset \operatorname{Int} U_i \quad \text{and} \quad v(U_i) \leq 2v(U'_i).$$

(This we can do because $v(U)$ is a continuous function of the end points of the component intervals of U.) Then the open rectangles $\operatorname{Int} U_1, \operatorname{Int} U_2, \dots$ cover A, and $\sum v(U_i) < \epsilon$.

(d) Let

$$U = [a_1, b_1] \times \cdots \times [a_n, b_n].$$

The subset of U consisting of those points x of U for which $x_i = a_i$ is called one of the ith faces of U. The other ith face consists of those x for which $x_i = b_i$. Each face of U has measure 0 in $\mathbf{R}^n$; for instance, the face for which $x_i = a_i$ can be covered by the single rectangle

$$[a_1, b_1] \times \cdots \times [a_i, a_i + \delta] \times \cdots \times [a_n, b_n],$$

whose volume may be made as small as desired by taking δ small. Now $\operatorname{Bd} U$ is the union of the faces of U, which are finite in number. Therefore $\operatorname{Bd} U$ has measure 0 in $\mathbf{R}^n$.

Now we suppose U has measure 0 in $\mathbf{R}^n$, and derive a contradiction. Set $\epsilon = v(U)$. We can by (c) cover U by open rectangles $\operatorname{Int} U_1, \operatorname{Int} U_2, \dots$ with $\sum v(U_i) < \epsilon$. Because U is compact, we can cover U by finitely many of these open sets, say $\operatorname{Int} U_1, \dots, \operatorname{Int} U_n$. But

$$\sum_{i=1}^{n} v(U_i) < \epsilon. [10]$$

Definition (2.1.12):

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

Theorem (2.1.13):

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

Proof:

Clearly we can assume that each $U_i \subset [a, b]$. Let $a = t_0 < t_1 < \cdots < \ t_k = b$ be all endpoints of all U_i. Then each $v(U_i)$ is the sum of certain $t_j - t_{j-1}$. Moreover, each $\left[t_{j-1}, t_j\right]$ lies in at least one U_i, so $\sum_{i=1}^{n} v(U_i) \geq \sum_{j=1}^{k}\left(t_j - t_{j-1}\right) = b - a$.

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

Theorem (2.1.14):

If A is compact and has measure 0, then A has content 0.

Proof:

Let $\epsilon > 0$. Since A has measure 0, there is a cover $\{U_1, U_2, \ldots\}$ of A by open rectangles such that $\sum_{i=1}^{\infty} v(U_i) < \epsilon$. Since A is compact, a finite number $U_1, \ldots, U_n$ of the U_i also cover A and surely $\sum_{i=1}^{n} v(U_i) < \epsilon$.

The conclusion of Theorem (2.1.14) 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 $\{[a_1, b_1], \ldots, [a_n, b_n]\}$ 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 (2.1.14) that $\sum_{i=1}^{n}(b_i - a_i) \geq 1$ for any such cover, and consequently A does not have content 0.

Recall that $o(f, x)$ denotes the oscillation of f at x.

Lemma (2.1.15):

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

Proof:

For each $x \in A$ there is a closed rectangle U_x, containing x in its interior, such that $MU_x(f) - mU_x(f) < \epsilon$. Since A is compact, a finite number $U_{x_1}, \ldots, U_{x_n}$ of the sets U_x cover A. Let P be a partition for A such that each subrectangle S of P is contained in some U_{x_i}. Then $M_S(f) - m_S(f) < \epsilon$ for each subrectangle S of P, so that $U(f,P) - L(f,P) = \sum_S [M_S(f) - m_S(f)] \cdot v(S) < \epsilon \cdot v(A)$.

Theorem (2.1.16):

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

Proof:

Suppose first that B has measure 0. Let $\epsilon > 0$ and let $B_\epsilon = \{x: o(f,x) \geq \epsilon\}$. Then $B_\epsilon \subset B$, so that B_ϵ has measure 0. Since (Theorem (1.1.10) B_ϵ is compact, B_ϵ has content 0. Thus there is a finite collection $U_1, \ldots, U_n$ of closed rectangles, whose interiors cover B_ϵ, such that $\sum_{i=1}^{n} v(U_i) < \epsilon$. Let P be a partition of A such that every subrectangle S of P is in one of two groups (see Figure (2.3))

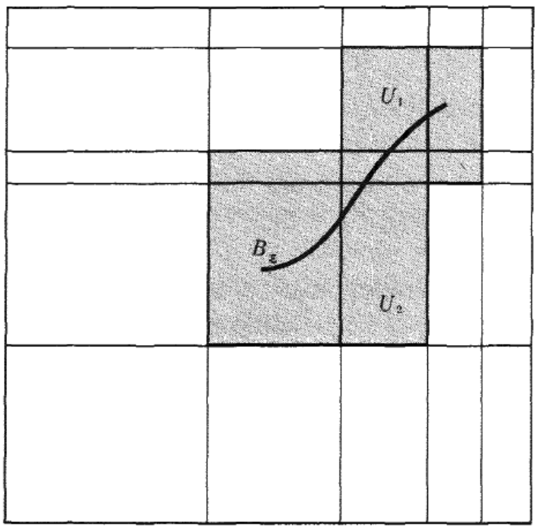


Figure (2.3): The shaded rectangles are in $\mathcal{S}_1$.

(1) $\mathcal{S}_1$, which consists of subrectangles S, such that $S \subset U_i$ for some i.

(2) $\mathcal{S}_2$, which consists of sub rectangles S with $S \cap B_\epsilon = \emptyset$.

Let $|f(x)| < M$ for $x \in A$. Then $M_S(f) - m_s(f) < 2M$ for every S. Therefore

$$\sum_{S \in \mathcal{S}_1} [M_S(f) - m_S(f)] \cdot v(S) < 2M \sum_{i=1}^{n} v(U_i) < 2M\epsilon.$$

Now, if $S \in \mathcal{S}_2$, then $o(f,x) < \epsilon$ for $x \in S$. Lemma (2.1.15) implies that there is a refinement P' of P such that

$$\sum_{S' \subset S} [M_{S'}(f) - m_{S'}(f)] \cdot v(S') < \epsilon \cdot v(S)$$

for $S \in \mathcal{S}_2$. Then

$$\begin{aligned}
&U(f, P') - L(f, P') \\
&\qquad = \sum_{S' \subset S \in \mathcal{S}_1} [M_{S'}(f) - m_{S'}(f)] \cdot v(S') + \sum_{S' \subset S \in \mathcal{S}_2} [M_{S'}(f) - m_{S'}(f)] \cdot v(S') \\
&\qquad < 2M\epsilon + \sum_{S \in \mathcal{S}_2} \epsilon \cdot v(S) \\
&\qquad \leq 2M\epsilon + \epsilon \cdot v(A).
\end{aligned}$$

Since M and $v(A)$ are fixed, this shows that we can find a partition P' with $U(f, P') - L(f, P')$ as small as desired. Thus f is integrable.

Suppose, conversely, that f is integrable. Since $= B_1 \cup B_{\frac{1}{2}} \cup B_{\frac{1}{3}} \cup \cdots$, it suffices (Theorem (2.1.11)(b)) to prove that each $B_{1/n}$ has measure 0. In fact we will show that each $B_{1/n}$ has content 0 (since $B_{1/n}$ is compact, this is actually equivalent).

If $\epsilon > 0$, let P be a partition of A such that $U(f, P) - L(f, P) < \epsilon/n$. Let $\mathcal{S}$ be the collection of subrectangles S of P which intersect $B_{1/n}$. Then $\mathcal{S}$ is a cover of $B_{1/n}$. Now if $S \in \mathcal{S}$, then $M_S(f) - m_S(f) \geq 1/n$. Thus

$$\begin{aligned}
\frac{1}{n} \sum_{S \in \mathcal{S}} v(S) &\leq \sum_{S \in \mathcal{S}} [M_S(f) - m_S(f)] \cdot v(S) \\
&\leq \sum_{S} [M_S(f) - m_S(f)] \cdot v(S) \\
&< \frac{\epsilon}{n},
\end{aligned}$$

and consequently $\sum_{S \in \mathcal{S}} v(S) < \epsilon$.

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

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

If $C \subset A$ for some closed rectangle A and $f: A \to \mathbf{R}$ 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.

Theorem (2.1.17):

The function $\chi_C: A \to \mathbf{R}$ is integrable if and only if the boundary of C has measure 0 (and hence content 0).

Proof:

If x is in the interior of C, then there is an open rectangle U with $x \in U \subset C$. Thus $\chi_C = 1$ on U and χ_C is clearly continuous at x. Similarly, if x is in the exterior of C, there is an open rectangle U with $x \in U \subset \mathbf{R}^n - C$. Hence $\chi_C = 0$ on U and χ_C is continuous at x. Finally, if x is in the boundary of C, then for every open rectangle U containing x, there is $y_1 \in U \cap C$, so that $\chi_C(y_1) = 1$ and there is $y_2 \in U \cap (\mathbf{R}^n - C)$, so that $\chi_C(y_2) = 0$. Hence χ_C is not continuous at x. Thus $\{x: \chi_C \text{ is not continuous } at\ x\} = \text{boundary } C$, and the result follows from Theorem (2.1.16).

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. [13]

Section (2.2): Fubini’s Theory and Change of Variables

In this section we can evaluate the integral of a function, such that $f: U \to \mathbf{R}$ is integrable, in some sense, by Fubini’s theorem.

Even in the case of a function $f: [a, b] \to \mathbf{R}$ of a single variable, the problem is not easy. One tool is provided by the fundamental theorem of calculus, which is applicable when f is continuous. This theorem is familiar to us from single-variable analysis. For reference, we state it here: [10]

Theorem (2.2.1):

Let $f \in \mathbf{R}$ is continuous on $[a, b]$. For $a \leq x \leq b$, put

$$F(x) = \int_a^x f(t)\, dt.$$

Then F is continuous on $[a, b]$; furthermore, if f is continuous at a point x_0 of $[a, b]$, then F is differentiable at x_0, and

$$F'(x_0) = f(x_0).$$

Theorem (2.2.2): (Fundamental theorem of calculus)

If $f \in \mathbf{R}$ on $[a, b]$ and if there is a differentiable function F on $[a, b]$ such that $F' = f$, then

$$\int_a^b f(x)\, dx = F(b) - F(a). [11]$$

The conclusions of this theorem are summarized in the two equations

$$D \int_a^x f = f(x) \qquad \text{and} \qquad \int_a^x DF = F(x) - F(a).$$

In each case, the integrand is required to be continuous on the interval in question.

Theorem (2.2.2) tells us we can calculate the integral of a continuous function f if we can find an antiderivative of f, that is, a function g such that $F' = f$. Theorem (2.2.1) tells us that such an antiderivative always exists, since F is such an antiderivative. The problem, of course, is to find such an antiderivative in practice. That is what the so-called "Technique of Integration," as studied in calculus, is about.

The same difficulties of evaluating the integral occur with n-dimensional integrals. One way of approaching the problem is to attempt to reduce the computation of an n-dimensional integral to the presumably simpler problem of computing a sequence of lower-dimensional integrals. One might even be able to reduce the problem to computing a sequence of one-dimensional integrals, to which, if the integrand is continuous, one could apply the fundamental theorem of calculus.

This is the approach used in calculus to compute a double integral. To integrate the continuous function $f(x, y)$ over the rectangle $U = [a, b] \times [c, d]$, for example, one integrates f first with respect to y, holding x fixed, and then integrates the resulting function with respect to x. In doing so, one is using the formula

$$\int_U f = \int_{x=a}^{x=b} \int_{y=c}^{y=d} f\,(x,y)$$

or its reverse. These formulas are not usually proved in calculus. In fact, it is seldom mentioned that a proof is needed; they are taken as "obvious." We shall prove them, and their appropriate n-dimensional versions, in this section.

These formulas hold when f is continuous. But when f is integrable but not continuous, difficulties can arise concerning the existence of the various integrals involved. For instance, the integral

$$\int_{y=c}^{y=d} f\,(x,y)$$

may not exist for all x even though $\int_U f$ exists, for the function f can behave badly along a single vertical line without that behavior affecting the existence of the double integral.

One could avoid the problem by simply assuming that all the integrals involved exist. What we shall do instead is to replace the inner integral in the statement of the formula by the corresponding lower integral, which we know exists. When we do this, a correct general theorem results; it includes as a special case the case where all the integrals exist.

Theorem (2.2.3): (Fubini's theorem)
Let $U = A \times B$, where A is a rectangle in $\mathbf{R}^n$ and B is a rectangle in $\mathbf{R}^m$. Let $f: U \to \mathbf{R}$ be a bounded function; write f in the form $f(x,y)$ for $x \in A$ and $y \in B$. For each $x \in A$, consider the lower and upper integrals

$$L\int_{y\in B} f(x,y) \quad \text{and} \quad U\int_{y\in B} f(x,y).$$

If f is integrable over U, then these two functions of x are integrable over A, and

$$\int_U f = \int_{x\in A} L\int_{y\in B} f(x,y) = \int_{x\in A} U\int_{y\in B} f(x,y).$$

Proof:

For purposes of this proof, define

$$\mathcal{L}(x) = L\int_{y\in B} f(x,y) \qquad \text{and} \qquad \mathcal{U}(x) = U\int_{y\in B} f(x,y)$$

for $x \in A$. Assuming $\int_U f$ exists, we show that $\mathcal{L}$ and $\mathcal{U}$ are integrable over A, and that their integrals equal $\int_U f$.

Let P be a partition of U. Then P consists of a partition P_A of A, and a partition P_B of B. We write $P = (P_A, P_B)$. If S_A is the general sub rectangle of A determined by P_A, and if S_B is the general subrectangle of B determined by P_B, then $S_A \times S_B$ is the general subrectangle of U determined by P.

We begin by comparing the lower and upper sums for f with the lower and upper sums for $\mathcal{L}$ and $\mathcal{U}$.

We first show that

$$L(f,P) \leq L(\mathcal{L}, P_A);$$

that is, the lower sum for f is no larger than the lower sum for the lower integral, $\mathcal{L}$.

Consider the general subrectangle $S_A \times S_B$ determined by P. Let x_0 be a point of S_A. Now

$$m_{S_A\times S_B}(f) \leq f(x_0, y)$$

for all $y \in S_B$; hence

$$m_{S_A\times S_B}(f) \leq m_{S_B}\big(f(x_0,y)\big).$$

See Figure (2.4). Holding x_0 and S_A fixed, multiply by $v(S_B)$ and sum over all subrectangles S_B. One obtains the inequalities

$$\sum_{S_B} m_{S_A\times S_B}(f)v(S_B) \leq L(f(x_0,y), P_B) \leq L\int_{y\in B} f(x_0,y) = \mathcal{L}(x_0).$$

This result holds for each $x_0 \in S_A$. We conclude that

$$\sum_{S_B} m_{S_A\times S_B}(f)v(S_B) \leq m_{S_A}(\mathcal{L}).$$

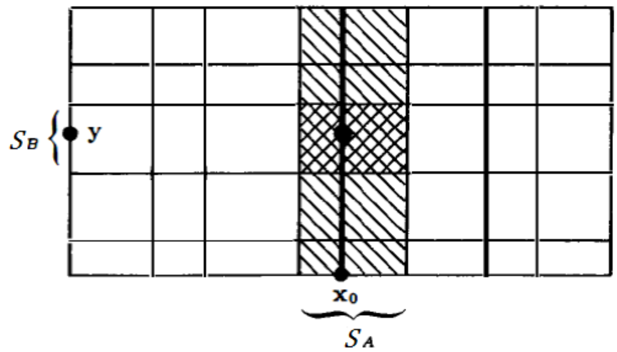


Figure (2.4)

Now multiply through by $v(S_A)$ and sum. Since $v(S_A)v(S_B) = v(S_A \times S_B)$, one obtains the desired inequality

$$L(f,P) \leq L(\mathcal{L}, P_A).$$

Secondly, an entirely similar proof shows that

$$U(f,P) \geq U(\mathcal{U}, P_A);$$

that is, the upper sum for f is no smaller than the upper sum for the upper integral, $\mathcal{U}$.

Thirdly, we summarize the relations that hold among the upper and lower sums of, $\mathcal{L}$, and $\mathcal{U}$ in the following diagram:

$$\begin{array}{ccc} & \leq U(\mathcal{L}, P_A) \leq & \\ L(f,P) \leq L(\mathcal{L}, P_A) & & U(\mathcal{U}, P_A) \leq U(f,P). \\ & \leq L(\mathcal{U}, P_A) \leq & \end{array}$$

The first and last inequalities in this diagram come from the first and second Steps. Of the remaining inequalities, the two on the upper left and lower right follow from the fact that $L(h,P) \leq U(h,P)$ for any h and P. The ones on the lower left and upper right follow from the fact that $\mathcal{L}(x) \leq \mathcal{U}(x)$ for all x. This diagram contains all the information we shall need.

Fourthly, we prove the theorem. Because f is integrable over U, we can, given $\epsilon > 0$, choose a partition $P = (P_A, P_B)$ of U so that the numbers at the extreme ends of the diagram in third Step are within ϵ of each other. Then the upper and lower sums for $\mathcal{L}$ are within ϵ of each other, and so are the upper and lower sums for $\mathcal{U}$. It follows that both $\mathcal{L}$ and $\mathcal{U}$ are integrable over A.

Now we note that by definition the integral $\int_A \mathcal{L}$ lies between the upper and lower sums of $\mathcal{L}$. Similarly, the integral $\int_A \mathcal{U}$ lies between the upper and lower sums for $\mathcal{U}$. Hence all three numbers

$$\int_A \mathcal{L} \quad \text{and} \quad \int_A \mathcal{U} \quad \text{and} \quad \int_U f$$

lie between the numbers at the extreme ends of the diagram. Because ϵ is arbitrary, we must have

$$\int_A \mathcal{L} = \int_A \mathcal{U} = \int_U f.$$

This theorem expresses $\int_U f$ as an iterated integral. To compute $\int_U f$, one first computes the lower integral (or upper integral) of f with respect to y, and then one integrates the resulting function with respect to x. There is nothing special about the order of integration; a similar proof shows that one can compute $\int_U f$ by first taking the lower integral (or upper integral) of f with respect to x, and then integrating this function with respect to y. [10]

Remarks:

1. A similar proof shows that

$$\int_U f = \int_{y\in B} L \int_{x\in A} f(x,y) = \int_{y\in B} U \int_{x\in A} f(x,y).$$

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 $f(x,y)$ is integrable, so that $\int_U f = \int_A \int_B f(x,y)$. This certainly occurs if f is continuous.
3. The worst irregularity commonly encountered is that $f(x,y)$ is not integrable for a finite number of $x \in A$. In this case. $\mathcal{L}(x) = \int_B f(x,y)$ for all but these finitely many x. Since $\int_A \mathcal{L}$ remains unchanged if $\mathcal{L}$ is redefined at a finite number of points, we can still write $\int_U f = \int_A \int_B f(x,y)$, provided that $\int_B f(x,y)$ is defined arbitrarily, say as 0, when it does not exist.

4. There are cases when this will not work and Theorem (2.2.3) must be used as stated. Let $f:[0,1]\times[0,1]\to\mathbf{R}$ be defined by

$$f(x,y)=\begin{cases}1 & \text{if } x \text{ is irrational,}\\ 1 & \text{if } x \text{ is rational and y is irrational,}\\ 1-1/q & \text{if } x=p/q \text{ in lowest terms and y is rational.}\end{cases}$$

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 $h(x)=\int_0^1 f(x,y)\,dy$ 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 $f:A\to\mathbf{R}$ is sufficiently nice, we can apply Fubini's theorem repeatedly to obtain

$$\int_A f=\int_{a^n}^{b^n}\left(\cdots\left(\int_{a^1}^{b^1} f(x^1,\ldots,x^n)dx^1\right)\cdots\right)dx^n.$$

6. If $C\subset U$, Fubini's theorem can be used to evaluate $\int_C f$, since this is by definition $\int_U \chi_C f$. 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

$$\chi_C(x,y)=\begin{cases}1 & \text{if } y>\sqrt{1-x^2}\ or\ y<-\sqrt{1-x^2},\\ 0 & \text{otherwise.}\end{cases}$$

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 $C\subset A\times B$, the main difficulty in deriving 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

$$\int_C f = \int_B \left(\int_A f(x,y) \cdot \chi_C(x,y)\, dx \right) dy. [13]$$

Now we introduce a tool of extreme importance in the theory of integration. To define the integral in the general case we use a partition of unity subordinate to the cover $\mathcal{O}$ of A, i.e. a family of differentiable functions on A, φ such that follows: [5]

Theorem (2.2.4):

Let $A \subset \mathbf{R}^n$ and let $\mathcal{O}$ be an 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 $\mathcal{O}$ such that $\varphi = 0$ outside of some closed set contained 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 $\mathcal{O}$. In this chapter we will only use continuity of the functions φ.)

Proof:

$Case\ 1$: $A\ is\ compact$

Then a finite number $U_1, \dots, U_n$ of open sets in $\mathcal{O}$ cover A. It clearly suffices to construct a partition of unity subordinate to the cover $\{U_1, \dots, U_n\}$. We will first find compact sets $D_i \subset U_i$ whose interiors cover A. The sets D_i are constructed inductively as follows. Suppose that $D_1, \dots, D_k$ have been chosen so that {interior $D_1, \dots$, interior $D_k, U_{k+1}, \dots, U_n$} covers A. Let

$$C_{k+1} = A - (\operatorname{int} D_1 \cup \cdots \cup \operatorname{int} D_k \cup U_{k+2} \cup \cdots \cup U_n).$$

Then $C_{k+1} \subset U_{k+1}$ is compact. Hence we can find a compact set D_{k+1} such that

$$C_{k+1} \subset \text{interior } D_{k+1} \quad \text{and} \quad D_{k+1} \subset U_{k+1}.$$

Having constructed the sets $D_1, \dots, D_n$, let ψ_i be a nonnegative C^∞ function which is positive on D_i and 0 outside of some closed set contained in U_i. Since $\{D_1, \dots, D_n\}$ covers A, we have $\psi_1(x) + \cdots + \psi_n(x) > 0$ for all x in some open set U containing A. On U we can define

$$\varphi_i(x) = \frac{\psi_i(x)}{\psi_1(x) + \cdots + \psi_n(x)}.$$

If $f: U \to [0,1]$ is a C^∞ function which is 1 on A and 0 outside of some closed set in U, then $\Phi = \{f \cdot \varphi_1, \dots, f \cdot \varphi_n\}$ is the desired partition of unity.

$Case\ 2.\ A = A_1 \cup A_2 \cup A_3 \cup \cdots, where\ each\ A_i is\ compact\ and A_i \subset interior\ A_{i+1}.$

For each i let $\mathcal{O}_i$ consist of all $U \cap (\text{interior } A_{i+1} - A_{i-2})$ for U in $\mathcal{O}$. Then $\mathcal{O}_i$ is an open cover of the compact set $B_i = A_i - \text{interior } A_{i-1}$. By case 1 there is a partition of unity Φ_i for B_i, subordinate to $\mathcal{O}_i$. For each $x \in A$ the sum

$$\sigma(x) = \sum_{\varphi \in \Phi_i,\ \forall i} \varphi(x)$$

is a finite sum in some open set containing x, since if $x \in A_i$ we have $\varphi(x) = 0$ for $\varphi \in \Phi_j$ with $j \geq i + 2$. For each φ in each Φ_i, define $\varphi'(x) = \varphi(x)/\sigma(x)$. The collection of all φ' is the desired partition of unity.

$Case\ 3$: $A\ is\ open$

Let $A_i = \{x \in A: |x| \leq i\ and\ distance\ from\ x\ to\ boundary\ A \geq 1/i\}$, and apply case 2.

$Case\ 4$: $A\ is\ arbitrary$

Let B be the union of all U in $\mathcal{O}$. By case 3 there is a partition of unity for B; this is also a partition of unity for A.

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 $\mathcal{O}$ of an open set $A \subset \mathbf{R}^n$ is admissible if each $U \in \mathcal{O}$ is contained in A. If Φ is subordinate to $\mathcal{O}$, $f: A \to \mathbf{R}$ is bounded in some open set around each point of A, and $\{x: f \text{ is discontinuous at } x\}$ has measure 0, then each $\int_A \varphi \cdot |f|$ exists. We define f to be integrable if $\sum_{\varphi\in\Phi} \int_A \varphi \cdot |f|$ converges. 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 $\int_A f \cdot$ These definitions do not depend on $\mathcal{O}$ or Φ.

Theorem (2.2.5):

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

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

(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.

Proof:

(1) Since $\varphi \cdot f = 0$ except on some compact set C, and there are only finitely many ψ, which are non-zero on C, we can write

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

This result, applied to $|f|$, shows the convergence of $\sum_{\varphi\in\Phi} \sum_{\psi\in\Psi} \int_A \psi \cdot \varphi \cdot |f|$, and hence of $\sum_{\varphi\in\Phi} \sum_{\psi\in\Psi}\left|\int_A \psi \cdot \varphi \cdot f\right|$. This absolute convergence justifies interchanging the order of summation in the above equation; the resulting double sum clearly equals $\sum_{\psi\in\Psi} \int_A \psi \cdot f$. Finally, this result applied to $|f|$ proves convergence of $\sum_{\psi\in\Psi} \int_A \psi \cdot |f|$.

(2) If A is contained in the closed rectangle B and $|f(x)| \le M$ for $x \in A$, and $F \subset \Phi$ is finite, then

$$\sum_{\varphi\in F} \int_A \varphi \cdot |f| \le \sum_{\varphi\in F} M \int_A \varphi = M \int_A \sum_{\varphi\in F} \varphi \le M v(B),$$

since $\sum_{\varphi\in F} \varphi \le 1$ on A.

(3) If $\epsilon > 0$ there is a compact Jordan-measurable $C \subset A$ such that $\int_{A-C} 1 < \epsilon$. There are only finitely many $\varphi \in \Phi$ which are non-zero on C. If $F \subset \Phi$ is any finite collection which includes these, and $\int_A f$ has its old meaning, then

$$\left|\int_A f - \sum_{\varphi \in F} \int_A \varphi \cdot f\right| \leq \int_A \left|f - \sum_{\varphi \in F} \varphi \cdot f\right| \leq M \int_A \left(1 - \sum_{\varphi \in F} \varphi\right)$$

$$= M \int_A \sum_{\varphi \in \Phi - F} \varphi \leq M \int_{A-C} 1 \leq M\epsilon. \text{ [12]}$$

Example (2.2.6):

Let $f: \mathbf{R} \to \mathbf{R}$ be defined by the equation

$$f(x) = \begin{cases} (1 + \cos x)/2 & \text{for } -\pi \leq x \leq \pi, \\ 0 & \text{otherwise.} \end{cases}$$

Then f is of class C^1. For each integer $m \geq 0$, set $\varphi_{2m+1}(x) = f(x - m\pi)$. For each integer $m \geq 1$, set $\varphi_{2m}(x) = f(x + m\pi)$. Then the collection $\{\varphi_i\}$ forms a partition of unity on $\mathbf{R}$. The support U_i of φ_i is a closed interval of the form $[k\pi, (k+2)\pi]$, which is compact, and each point of $\mathbf{R}$ has a neighbourhood that intersects at most three of the sets U_i. We leave it to us to check that $\sum \varphi_i(x) = 1$. Thus $\{\varphi_i\}$ is a partition of unity on $\mathbf{R}$. See Figure (2.5).

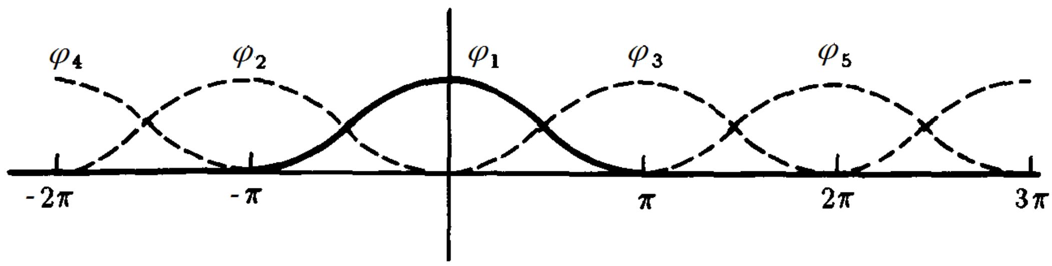


Figure (2.5)

Now we discuss the general change of variables theorem. We begin by reviewing the version of it used in calculus and we will prove it.

Recall the common convention that if f is integrable over $[a, b]$, then one defines

$$\int_b^a f = -\int_a^b f.$$

Theorem (2.2.7): (Substitution rule)

Let $I = [a, b]$. Let $g: I \to \mathbf{R}$ be a function of class C^1, with $g'(x) \neq 0$ for $x \in (a, b)$. Then the set $g(I)$ is a closed interval J with end points $g(a)$ and $g(b)$. If $f: J \to \mathbf{R}$ is continuous, then

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

or equivalently,

$$\int_{J} f = \int_{I} (f \circ g) \cdot |g'|.$$

Proof:

Continuity of g' and the intermediate-value theorem imply that either $g'(x) > 0$ or $g'(x) < 0$ on all of (a, b). Hence g is either strictly increasing or strictly decreasing on I, by the mean-value theorem, so that g is $1 - 1$. In the case where $g' > 0$, we have $g(a) < g(b)$; in the case where $g' < 0$, we have $g(a) > g(b)$. In either case, let $J = [c, d]$ denote the interval with end points $g(a)$ and $g(b)$. See Figure (2.6). The intermediate-value theorem implies that g carries $I \; onto \; J$. Then the composite function $f\big(g(x)\big)$ is defined for all x in $[a, b]$, so the theorem at least makes sense.

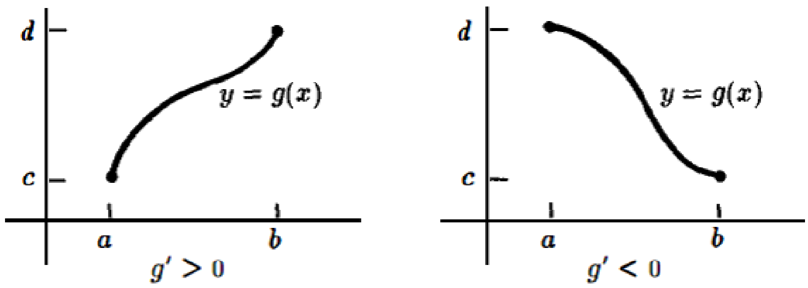


Figure (2.6)

Define

$$F(y) = \int_{c}^{y} f$$

for y in $[c, d]$. Because f is continuous, the fundamental theorem of calculus implies that $F'(y) = f(y)$. Consider the composite function $h(x) = F\big(g(x)\big)$; we differentiate it by the chain rule. We have

$$h'(x) = F'(g(x)) \cdot g'(x) = f(g(x)) \cdot g'(x).$$

Because the latter function is continuous, we can apply the fundamental theorem of calculus to integrate it. We have

$$\begin{aligned}\int_{x=a}^{x=b} f(g(x)) \cdot g'(x) &= h(b) - h(a) \\ &= F(g(b)) - F(g(a)) \\ &= \int_{c}^{g(b)} f - \int_{c}^{g(a)} f.\end{aligned}$$

Now c equals either $g(a)$ or $g(b)$. In either case, this equation can be written in the form

$$\int_a^b (f \circ g) \cdot g' = \int_{g(a)}^{g(b)} f. \tag{2.1}$$

This is the first of our desired formulas.

Now in the case where $g' > 0$, we have $J = [g(a), g(b)]$. Since $|g'| = g'$ in this case, equation (2.1) can be written in the form

$$\int_I (f \circ g) \cdot |g'| = \int_J f. \tag{2.2}$$

In the case where $g' < 0$, we have $J = [g(b), g(a)]$. Since $|g'| = -g'$ in this case, equation (2.1) can again be written in the form (2.2).

Example (2.2.8):

Consider the integral

$$\int_{x=0}^{x=1} (2x^2 + 1)^{10}(4x).$$

Set $f(y) = y^{10}$ and $g(x) = 2x^2 + 1$. Then $g'(x) = 4x$, which is positive for $0 < x < 1$. See Figure (2.7). The substitution rule implies that

$$\int_{x=0}^{x=1} (2x^2 + 1)^{10}(4x) = \int_{x=0}^{x=1} f(g(x)) \cdot g'(x) = \int_{y=1}^{y=3} f(y) = \int_{y=1}^{y=3} y^{10}.$$

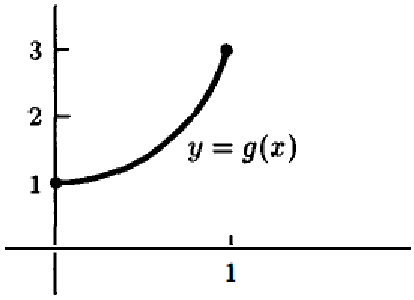


Figure (2.7)

Definition (2.2.9):

Let A be open in $\mathbf{R}^n$. Let $g: A \to \mathbf{R}^n$ be a $1-1$, continuously differentiable function such that $\det g'(x) \neq 0$ for $x \in A$. Then g is called a change of variables in $\mathbf{R}^n$. [10]

Theorem (2.2.10):

Let $A \subset \mathbf{R}^n$ be an open set and $g: A \to \mathbf{R}^n$ a $1-1$, continuously differentiable function such that $\det g'(x) \neq 0$ for all $x \in A$. If $f: g(A) \to \mathbf{R}$ is integrable, then

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

Proof:

We begin with some important reductions.

1. Suppose there is an admissible cover $\mathcal{O}$ for A such that for each $U \in \mathcal{O}$ and any integrable f we have

$$\int_{g(U)} f = \int_{U} (f \circ g)|\det g'|.$$

Then the theorem is true for all of A. (Since g is automatically $1-1$ in an open set around each point, it is not surprising that this is the only part of the proof using the fact that g is $1-1$ on all of A.)

Proof of (1): The collection of all $g(U)$ is an open cover of $g(A)$. Let Φ be a partition of unity subordinate to this cover. If $\varphi = 0$ outside of $g(U)$, then, since g is $1-1$, we have $(\varphi \cdot f) \circ g = 0$ outside of U. Therefore the equation

$$\int_{g(U)} \varphi \cdot f = \int_{U} [(\varphi \cdot f) \circ g]|\det g'|.$$

can be written

$$\int_{g(A)} \varphi \cdot f = \int_A [(\varphi \cdot f) \circ g]|\det g'|.$$

Hence

$$\int_{g(A)} f = \sum_{\varphi \in \Phi} \int_{g(A)} \varphi \cdot f = \sum_{\varphi \in \Phi} \int_A [(\varphi \cdot f) \circ g]|\det g'|$$

$$= \sum_{\varphi \in \Phi} \int_A [(\varphi \circ g)(f \circ g)]|\det g'|$$

$$= \int_A (f \circ g)|\det g'|.$$

Remark: The theorem also follows from the assumption that

$$\int_V f = \int_{g^{-1}(V)} (f \circ g)|\det g'|$$

for V in some admissible cover of $g(A)$. This follows from (1) applied to g^{-1}.

2. It suffices to prove the theorem for the function $f = 1$.

$Proof\ of$ (2): If the theorem holds for $f = 1$, it holds for constant functions. Let V be a rectangle in $g(A)$ and P a partition of V. For each subrectangle S of P let f_S be the constant function $m_S(f)$. Then

$$L(f,P) = \sum_S m_S(f) \cdot v(S) = \sum_S \int_{\operatorname{int} S} f_S$$

$$= \sum_S \int_{g^{-1}(\operatorname{int} S)} (f_S \circ g)|\det g'| \leq \sum_S \int_{g^{-1}(\operatorname{int} S)} (f \circ g)|\det g'|$$

$$\leq \int_{g^{-1}(V)} (f \circ g)|\det g'|.$$

Since $\int_V f$ is the least upper bound of all $L(f,P)$, this proves that $\int_V f \leq \int_{g^{-1}(V)} (f \circ g)|\det g'|$. A similar argument, letting $f_S = M_S(f)$, shows that $\int_V f \geq \int_{g^{-1}(V)} (f \circ g)\,|\det g'|$. The result now follows from the above Remark.

3. If the theorem is true for $g: A \to \mathbf{R}^n$ and for $h: B \to \mathbf{R}^n$, where $g(A) \subset B$, then it is true for $h \circ g: A \to \mathbf{R}^n$.

$Proof\ of$ (3):

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

$$= \int\limits_{A} [(f \circ h) \circ g] \cdot [|\det h'| \circ g] \cdot |\det g'|$$

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

4. The theorem is true if g is a linear transformation.

$Proof\ of\ (4)$: By (1) and (2) it suffices to show for any open rectangle U that

$$\int\limits_{g(U)} 1 = \int\limits_{U} |\det g'|.$$

Observations (3) and (4) together show that we may assume for any particular $a \in A$ that $g'(a)$ is the identity matrix: in fact, if T is the linear transformation $Dg(a)$, then $(T^{-1} \circ g)'(a) = I$; since the theorem is true for T, if it is true for $T^{-1} \circ g$ it will be true for g.

We are now prepared to give the proof, which proceeds by induction on n. The remarks before the statement of the theorem, together with (1) and (2), prove the case $n = 1$.

Assuming the theorem in dimension $n - 1$, we prove it in dimension n. For each $a \in A$ we need only find an open set U with $a \in U \subset A$ for which the theorem is true. Moreover we may assume that $g'(a) = I$.

Define $h: A \to \mathbf{R}^n$ by $h(x) = (g^1(x), \dots, g^{n-1}(x), x^n)$. Then $h'(a) = I$. Hence in some open U' with $a \in U' \subset A$, the function h is $1 - 1$ and $\det h'(x) \neq 0$. We can thus define $k: h(U') \to \mathbf{R}^n$ by $k(x) = \left(x^1, \dots, x^{n-1}, g^n\left(h^{-1}(x)\right)\right)$ and $g = k \circ h$. We have thus expressed g as the composition of two maps, each of which changes fewer than n coordinates (Figure (2.8)).

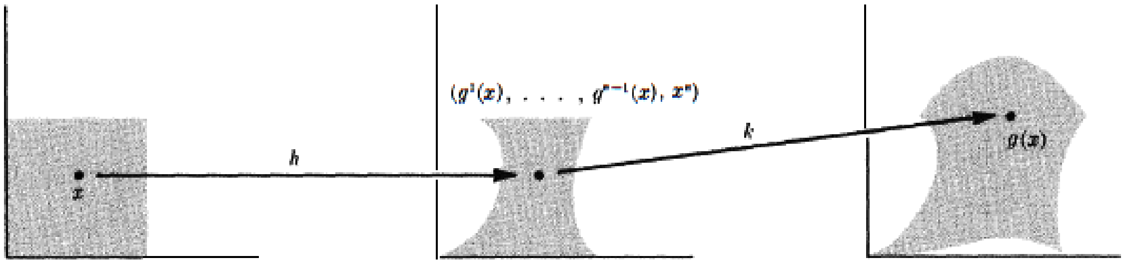


Figure (2.8)

We must attend to a few details to ensure that k is a function of the proper sort. Since

$$(g^n \circ h^{-1})'\big(h(a)\big) = (g^n)'(a) \cdot [h'(a)]^{-1} = (g^n)'(a),$$

we have $D_n(g^n \circ h^{-1})\big(h(a)\big) = D_n g^n(a) = 1$, so that $k'\big(h(a)\big) = I$. Thus in some open set V with $h(a) \in V \subset h\,(U')$, the function k is $1-1$ and $\det k'(x) \neq 0$. Letting $U = k^{-1}(V)$ we now have $g = k \circ h$, where $h\colon U \to \mathbf{R}^n$ and $k\colon V \to \mathbf{R}^n$ and $h(U) \subset V$. By (3) it suffices to prove the theorem for h and k. We give the proof for h; the proof for k is similar.

Let $W \subset U$ be a rectangle of the form $D \times [a_n, b_n]$, where D is a rectangle in $\mathbf{R}^{n-1}$. By Fubini's theorem

$$\int\limits_{h(W)} 1 = \int\limits_{[a_n,b_n]} \left(\int\limits_{h(D\times|x^n|)} 1\, dx^1 \ldots dx^{n-1} \right) dx^n.$$

Let $h_{x^n}\colon D \to \mathbf{R}^{n-1}$ be defined by

$h_{x^n}(x^1,\ldots,x^{n-1}) = \big(g^1(x^1,\ldots,x^n),\ldots,g^{n-1}(x^1,\ldots,x^n)\big)$. Then each h_{x^n} is clearly $1-1$ and

$$\det(h_{x^n})'\big((x^1,\ldots,x^{n-1})\big) = \det h'(x^1,\ldots,x^n) \neq 0.$$

Moreover

$$\int\limits_{h(D\times|x^n|)} 1\, dx^1 \ldots dx^{n-1} = J \int\limits_{h_{x^n}(D)} 1\, dx^1 \ldots dx^{n-1}.$$

Applying the theorem in the ease $n-1$ therefore gives

$$\int\limits_{h(W)} 1 \;=\; \int\limits_{[a_n,b_n]} \left(\int\limits_{h_{x^n}(D)} 1\, dx^1 \ldots dx^{n-1} \right) dx^n$$

$$= \int\limits_{[a_n,b_n]} \left(\int\limits_{D} |\det(h_{x^n})'(x^1,\ldots,x^{n-1})| dx^1 \ldots\ dx^{n-1} \right) dx^n$$

$$= \int\limits_{[a_n,b_n]} \left(\int\limits_{D} |\det h'(x^1,\ldots,x^n)| dx^1 \ldots\ dx^{n-1} \right) dx^n$$

$$= \int\limits_{W} |\det h'|.$$

The condition $\det g'(x) \neq 0$ may be eliminated from the hypotheses of Theorem (2.2.10) by using the following theorem, which often plays an unexpected role. [13]

Example (2.2.11):

Let $f(x,y) = e^{-x^2-y^2}$. We wish to evaluate $\int_{\mathbf{R}^2} f$. Let $U = (0,\infty) \times (0,2\pi)$ and $V = \mathbf{R}^2 - \{(x,y): x \geq 0 \text{ and } y = 0\}$. Note that $\{(x,y): x \geq 0 \text{ and } y = 0\}$ is a null set, so that

$$\int\limits_{V} f = \int\limits_{\mathbf{R}^2} f,$$

assuming f is integrable, which we have not yet shown. It is easily seen that the function $g: U \to V$ defined by $g(r,\theta) = (r\cos(\theta), r\sin(\theta))$ is a C^1 diffeomorphism of V onto U and $\det|g'(r,\theta)| = r.$ By the change-of-variables theorem and Fubini's theorem, again relying on the as-yet-unproven integrability of f, we have

$$\int\limits_{V} f = \int\limits_{U} (f \circ g)|\det(g')| = \int\limits_{0}^{2\pi} \int\limits_{0}^{\infty} r\, e^{-r^2}\, dr\, d\theta.$$

For each natural number n let

$$\begin{cases} r\, e^{-r^2} & \text{if } 0 < r < n \text{ and } 0 < \theta < 2\pi. \\ 0 & \text{otherwise.} \end{cases}$$

Each h_n is integrable and has integral $\pi(1 - e^{-n^2})$. By the monotone convergence theorem we see that $(f \circ g)|\det(g')|$ is integrable and that we must have

$$\int\limits_{0}^{2\pi} \int\limits_{0}^{\infty} r\, e^{-r^2}\, dr\, d\theta = \pi.$$

Using the change-of-variables theorem with g^{-1} as our diffeomorphism, we see that f must be integrable over V, and hence over $\mathbf{R}^2$. We conclude that

$$\int_V f = \pi. \text{ [9]}$$

Theorem (2.2.12): (Sard's Theorem)

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

Proof:

Let $U \subset A$ be a closed rectangle such that all sides of U have length l, say. Let $\epsilon > 0$. If N is sufficiently large and U is divided into N^n rectangles, with sides of length l/N, then for each of these rectangles S, if $x \in S$ we have

$$|Dg(x)(y-x) - g(y) - g(x)| < \epsilon|x-y| \leq \epsilon\sqrt{n}\,(l/N)$$

for all $y \in S$. If S intersects B we can choose $x \in S \cap B$; since $\det g'(x) = 0$, the set $\{Dg(x)(y-x): y \in S\}$ lies in an $(n-1)$-dimensional subspace V of $\mathbf{R}^n$. Therefore the set $\{g(y) - g(x): y \in S\}$ lies within $\epsilon\sqrt{n}\,(l/N)$ of V, so that $\{g(y): y \in S\}$ lies within $\epsilon\sqrt{n}\,(l/N)$ of the $(n-1)$-plane $V + g(x)$. On the other hand, by Lemma (1.2.10) there is a number M such that

$$|g(x) - g(y)| < M|\,x-y| \leq M\sqrt{n}\,(l/N).$$

Thus, if S intersects B, the set $\{g(y): y \in S\}$ is contained in a cylinder whose height is $< 2\epsilon\sqrt{n}\,(l/N)$ and whose base is an $(n-1)$-dimensional sphere of radius $< M\sqrt{n}\,(l/N)$. This cylinder has volume $< C(l/N)^n\epsilon$ for some constant C. There are at most N^n such rectangles S, so $g(U \cap B)$ lies in a set of volume $< C(l/N)^n \cdot \epsilon \cdot N^n = Cl^n \cdot \epsilon$. Since this is true for all $\epsilon > 0$, the set $g(U \cap B)$ has measure 0. Since we can cover all of A with a sequence of such rectangles U, the desired result follows from Theorem (2.1.11)(b). [13]

Chapter (3)

Multi-Linear Algebra

Section (3.1): Fields and Forms

We turn our attention to fields and forms, the discussion of which requires algebraic preliminaries.

Definition (3.1.1):

Let V be a vector space. Let $V^k = V \times \cdots \times V$ denote the set of all k-tuples $(v_1, \ldots, v_k)$ of vectors of V. A function $T: V^k \to \mathbf{R}$ is called multilinear if for each i with $1 < i < k$ we have

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

Definition (3.1.2):

A multilinear function $T: V^k \to \mathbf{R}$ is called a k-tensor on V and the set of all k-tensors, denoted $\mathcal{J}^k(V)$. [8, 10]

Example (3.1.3):

(1) The space of 1-tensors $\mathcal{J}^1(V^*)$ is equal to V^*, the dual space of V, that is, the space of real-valued linear functions on V.

(2) The usual inner product on $\mathbf{R}^n$ is an example of a 2-tensor.

(3) The determinant is an n-tensor on $\mathbf{R}^n$. [5]

Theorem (3.1.4):

The set of all k-tensors on V constitutes a vector space if we define

$$(S + T)(v_1, \ldots, v_k) = S(v_1, \ldots, v_k) + T(v_1, \ldots, v_k),$$

$$(aS)(v_1, \ldots, v_k) = a \cdot S(v_1, \ldots, v_k).$$

Now we introduce a product operation into the set of all tensors on V. The product of a k-tensor and an l-tensor will be a $k + l$ tensor.

Definition (3.1.5):

Let S be a k-tensor on V and let T be an l-tensor on V. We define a $k+l$ tensor $S \otimes T$ on V by the equation

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

We list some of the properties of this product operation:

Theorem (3.1.6):

Let S, T, U be tensors on V. Then the following properties hold:

(1) $S \otimes (T \otimes U) = (S \otimes T) \otimes U.$ (Associativity)

(2) $(aS) \otimes T = a(S \otimes T) = S \otimes (aT).$ (Homogeneity)

(3) Suppose S and T have the same order, Then: (Distributivity)

$$(S + T) \otimes U = S \otimes U + T \otimes U,$$

$$U \otimes (S + T) = U \otimes S + U \otimes T.$$

Proof:

Associativity is proved, for instance, by noting that (if S, T, U have orders k, l, m, respectively).

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

The value of $(S \otimes T) \otimes U$ on the given tuple is the same.

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 defied similarly. [10]

Theorem (3.1.7):

If $\{\varphi_1, \dots, \varphi_n\}$ is a basis for $\mathcal{J}^1(V^*) = V^*$,

then the set $\{\varphi_{i_1} \otimes \cdots \otimes \varphi_{i_k} : 1 \leq i_1, \dots, i_k \leq n\}$ is a basis of $\mathcal{J}^k(V^*)$, and therefore $\dim \mathcal{J}^k(V^*) = n^k$.

Proof:

We will first show that the elements of this set are linearly independent. If

$$\varphi = \sum_{i_1, \dots, i_k}^{n} a_{i_1, \dots, i_k} \cdot \varphi_{i_1} \otimes \cdots \otimes \varphi_{i_k} = 0,$$

then, taking the basis $\{v_1, \ldots, v_n\}$ of V dual to $\{\varphi_1, \ldots, \varphi_n\}$, meaning that $\varphi_i(v_j) = \delta_{ij}$, we have $\varphi(v_{j_1}, \ldots, v_{j_k}) = a_{j_1,\ldots,j_k} = 0$ for every $1 \leq j_1, \ldots, j_k \leq n$.

To show that $\{\varphi_{i_1} \otimes \cdots \otimes \varphi_{i_k} : 1 \leq i_1, \ldots, i_k \leq n\}$ span $\mathcal{J}^k(V^*)$, we take any element $\varphi \in \mathcal{J}^k(V^*)$ and consider the k-tensor T defined by

$$T = \sum_{i_1,\ldots,i_k}^{n} T(v_{i_1}, \ldots, v_{i_k}) \cdot \varphi_{i_1} \otimes \cdots \otimes \varphi_{i_k}.$$

Clearly, $T(v_{i_1}, \ldots, v_{i_k}) = \varphi(v_{i_1}, \ldots, v_{i_k})$ for every $1 \leq i_1, \ldots, i_k \leq n$, and so, by linearity, $T = \varphi$. [5]

Example (3.1.8):

Consider the case $V = \mathbf{R}^n$. Let $e_1, \ldots, e_n$ be the usual basis for $\mathbf{R}^n$; let $\varphi_1, \ldots, \varphi_n$ be the dual basis for $\mathcal{J}^1(V)$. Then if x has components $x_1, \ldots, x_n$, we have

$$\varphi_i(x) = \varphi(x_1\, e_1 + \cdots + x_n\, e_n) = x_i.$$

Thus $\varphi_i : \mathbf{R}^n \to \mathbf{R}$ equals projection onto the ith coordinate.

More generally, given $I = (i_1, \ldots, i_k)$, the elementary tensor φ_I satisfies the equation

$$\varphi_I(x_1, \ldots, x_k) = \varphi_{i_1}(x_1) \cdots \varphi_{i_k}(x_k).$$

Let us write $X = [x_1 \cdots x_k]$, and let x_{ij} denote the entry of X in row i and column j. Then x_j is the vector having components $x_{1j}, \ldots, x_{nj}$. In this notation,

$$\varphi_I(x_1, \ldots, x_k) = x_{i_1 1}\, x_{i_2 2} \ldots x_{i_k k}.$$

Thus φ_I is just a monomial in the components of the vectors $x_1, \ldots, x_k$; and the general k-tensor on $\mathbf{R}^n$ is a linear combination of such monomials.

It follows that the general 1-tensor on $\mathbf{R}^n$ is a function of the form

$$f(x) = d_1 x_1 + \cdots + d_n x_n,$$

for some scalars d_i. The general 2-tensor on $\mathbf{R}^n$ has the form

$$g(x, y) = \sum_{i,j=1}^{n} d_{ij}\, x_i y_j,$$

for some scalars d_{ij}. And so on.

Finally, we examine how tensors behave with respect to linear transformation of the underlying vector spaces.

Definition (3.1.9):
Let $f: V \to W$ be a linear transformation. We define the dual transformation

$$f^{*}: \mathcal{J}^{k}(W) \to \mathcal{J}^{k}(V),$$

for $T \in \mathcal{J}^{k}(V)$ and if $v_1, \ldots, v_k \in V$, then

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

The transformation f^{*} is the composite of the transformation $f \times \cdot \cdot \cdot \times f$ and the transformation T, as indicated in the following diagram:

$$\begin{tikzcd} V^k \arrow[rr, "f \times \cdots \times f"] \arrow[drr, "f^{*}T"'] & & W^k \arrow[d, "T"] \\ & & \mathbf{R} \end{tikzcd}$$

Figure (3.1)

It is immediate from the definition that $f^{*}T$ is multilinear, since f is linear and T is multilinear. It is also true that f^{*} itself is linear, as a map of tensors, as we now show.

Theorem (3.1.10):
Let $f: V \to W$ be a linear transformation; let

$$f^{*}: \mathcal{J}^{k}(W) \to \mathcal{J}^{k}(V)$$

be the dual transformation. Then:

(1) f^{*} is linear.
(2) $f^{*}\,(T \otimes S) = f^{*}T \otimes f^{*}S$.
(3) If $U: W \to X$ is a linear transformation, then $(U \circ f)^{*}\,T = f^{*}(U^{*}T)$.

Proof:
The proofs are straightforward. One verifies (1), for instance, as follows:

$$\big(f^*(aT+bS)\big)(v_1,\ldots,v_k) = (aT+bS)\,\big(f(v_1),\ldots,f(v_k)\big)$$

$$= aT\big(f(v_1),\ldots,f(v_k)\big) + bS\big(f(v_1),\ldots,f(v_k)\big)$$

$$= af^*T(v_1,\ldots,v_k) + bf^*S(v_1,\ldots,v_k),$$

whence $f^*(aT+bS) = af^*T + bf^*S$.

The following diagrams illustrate property (3): [10]

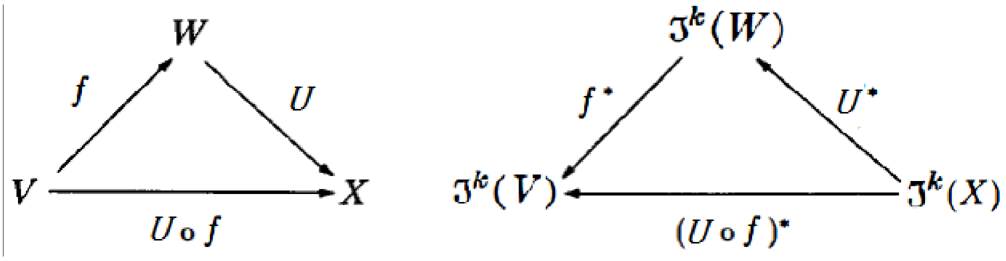


Figure (3.2)

The first example is the inner product $\langle , \rangle \in \mathcal{J}^2(\mathbf{R}^n)$. 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 $\mathbf{R}^n$. The following theorem shows that our generalization is not too general.

Theorem (3.1.11):

If T is an inner product on V, there is a basis $v_1,\ldots,v_n$ 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 $f: \mathbf{R}^n \to V$ such that $T\big(f(x),f(y)\big) = \langle x,y\rangle$ for $x,y \in \mathbf{R}^n$. In other words $f^*T = \langle , \rangle$.

Proof:

Let $w_1,\ldots,w_n$, be any basis for V. Define

$$w_1' = w_1,$$

$$w_2' = w_2 - \frac{T(w_1',w_2)}{T(w_1',w_1{}')}\cdot w_1'\,,$$

$$w_3' = w_3 - \frac{T(w_1', w_3)}{T(w_1', w_1')} \cdot w_1' - \frac{T(w_2', w_3)}{T(w_2', w_2')} \cdot w_2' , \qquad \text{etc.}$$

It is easy to check that $T(w_i', w_j') = 0$ if $i \neq j$ and $w_i' \neq 0$ so that $T(w_i', w_i') > 0$. Now define $v_i = w_i' / \sqrt{T(w_i', w_i')}$. The isomorphism f may be defined by $f(e_i) = v_i$. [13]

We continue to assume that V is a finite-dimensional real vector space. A covariant k-tensor T on V is said to be alternating (or antisymmetric or skew-symmetric) if it changes sign whenever two of its arguments are interchanged. This means that for all vectors $v_1, \dots, v_k \in V$ and every pair of distinct indices i, j it satisfies

$$T(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = -T(v_1, \dots, v_j, \dots, v_i, \dots, v_k). \text{ [8]}$$

The space of all alternating k-tensors is a vector subspace $\Lambda^k(V)$ of $\mathcal{J}^k(V)$. Note that, for any alternating k-tensor T, we have $T(v_1, \dots, v_k) = 0$ if $v_i = v_j$ for some $i \neq j$. [5]

We can easily verify that $T = \operatorname{Alt} T$ if and only if T is alternating. [6]

Example (3.1.12):

(1) All 1-tensors are trivially alternating, that is, $\Lambda^1(V) = \mathcal{J}^1(V) = V$.

(2) The determinant is an alternating n-tensor on $\mathbf{R}^n$. [5]

Example (3.1.13):

If $T \in \mathcal{J}^2(V)$,

$$(\operatorname{Alt} T)(x, y) = \tfrac{1}{2}\big(T(x, y) - T(y, x)\big). \text{ [6]}$$

Consider now S_k, the group of all possible permutations of $\{1, \dots, k\}$. If $\sigma \in S_k$, we set $\sigma(v_1, \dots, v_k) = (v_{\sigma(1)}, \dots, v_{\sigma(k)})$. Given a k-tensor $T \in \mathcal{J}^k(V)$ we can define a new alternating k-tensor, called $\operatorname{Alt}(T)$, in the following way:

$$\operatorname{Alt}(T) = \frac{1}{k!} \sum_{\sigma \in S_k} (\operatorname{sgn} \sigma)\, (T \circ \sigma),$$

where $\operatorname{sgn} \sigma$ is $+1$ or -1 according to whether σ is an even or an odd permutation.

Example (3.1.14):

If $T \in \mathcal{J}^3(V)$,

$\mathrm{Alt}(T)(v_1, v_2, v_3) = \frac{1}{6}(T\,(v_1, v_2, v_3) + T(v_3, v_1, v_2) + T\big(v_2, v_{3,} v_1\big)$

$-T(v_1, v_3, v_2) - T(v_2, v_1, v_3) - \ T\,(v_3, v_2, v_1))$. [5]

Theorem (3.1.15):

(1) If $T \in \mathcal{J}^k(V)$, then $\mathrm{Alt}(T) \in \Lambda^k(V)$.

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

(3) If $T \in \mathcal{J}^k(V)$, then $\mathrm{Alt}\big(\mathrm{Alt}(T)\big) = \mathrm{Alt}(T)$.

Proof:

(1) Let (i, j) be the permutation that interchanges i and j and leaves all other numbers fixed. If $\sigma \in S_k$, let $\sigma' = \sigma \cdot (i, j)$. Then

$$\begin{aligned}
&\mathrm{Alt}(T)\big(v_1, \dots, v_j, \dots, v_i, \dots, v_k\big) \\
&\quad = \frac{1}{k!}\sum_{\sigma \in S_k} \mathrm{sgn}\ \sigma \cdot T\big(v_{\sigma(1)}, \dots, v_{\sigma(j)}, \dots, v_{\sigma(i)}, \dots, v_{\sigma(k)}\big) \\
&\quad = \frac{1}{k!}\sum_{\sigma \in S_k} \mathrm{sgn}\ \sigma \cdot T\big(v_{\sigma\prime(1)}, \dots, v_{\sigma\prime(i)}, \dots, v_{\sigma\prime(j)}, \dots, v_{\sigma\prime(k)}\big) \\
&\quad = \frac{1}{k!}\sum_{\sigma\prime \in S_k} -\mathrm{sgn}\ \sigma' \cdot T\big(v_{\sigma'(1)}, \dots, v_{\sigma'(k)}\big) \\
&\quad = -\mathrm{Alt}(T)(v_1, \dots, v_k).
\end{aligned}$$

(2) If $\omega \in \Lambda^k(V)$, and $\sigma \in (i, j)$, then $\omega\big(v_{\sigma(1)}, \dots, v_{\sigma(k)}\big) = \mathrm{sgn}\,\sigma \cdot \omega(v_1, \dots, v_k)$. Since every σ is a product of permutations of the form (i, j), this equation holds of all σ. Therefore

$$\begin{aligned}
\mathrm{Alt}(\omega)(v_1, \dots, v_k) &= \frac{1}{k!}\sum_{\sigma \in S_k} \mathrm{sgn}\ \sigma \cdot \omega\big(v_{\sigma(1)}, \dots, v_{\sigma(k)}\big) \\
&= \frac{1}{k!}\sum_{\sigma \in S_k} \mathrm{sgn}\ \sigma \cdot \mathrm{sgn}\,\sigma \cdot \omega(v_1, \dots, v_k) \\
&= \omega(v_1, \dots, v_k).
\end{aligned}$$

(3) follows immediately from (1) and (2). [12,13]

We will now define the wedge product between alternating tensors: if $\omega \in \Lambda^k(V)$ and $\eta \in \Lambda^l(V)$, then $\omega \wedge \eta \in \Lambda^{k+l}(V)$ is given by

$$\omega \wedge \eta = \frac{(k+l)!}{k!\,l!}\,\mathrm{Alt}(\omega \otimes \eta). \text{ [5]}$$

The funny coefficient is not essential, but it makes some things work out more nicely, as we shall soon see. It is clear that

(1) $\wedge$ is bilinear:

$$(\omega_1 + \omega_2) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta$$

$$\omega \wedge (\eta_1 + \eta_2) = \omega \wedge \eta_1 + \omega \wedge \eta_2$$

$$a\omega \wedge \eta = \omega \wedge a\eta = a(\omega \wedge \eta)$$

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

Moreover, it is easy to see that

(3) $\wedge$ is "anti-commutative": $\omega \wedge \eta = (-1)^{kl}\eta \wedge \omega$.
In particular, if k is odd then

$$\omega \wedge \omega = 0. \text{ [12]}$$

Example (3.1.16):
If $\omega, \eta \in \Lambda^1(V) = V$, then

$$\omega \wedge \eta = 2\ \mathrm{Alt}(\omega \otimes \eta) = \omega \otimes \eta - \eta \otimes \omega. \text{ [5]}$$

To prove associativity we need the following proposition.

Theorem (3.1.17):
(1) Let $S \in \mathcal{J}^k(V)$ and $T \in \mathcal{J}^l(V)$. If $\mathrm{Alt}(S) = 0$, then

$$\mathrm{Alt}(S \otimes T) = \mathrm{Alt}(T \otimes S) = 0.$$

(2) $\mathrm{Alt}(\mathrm{Alt}(\omega \otimes \eta) \otimes \theta) = \mathrm{Alt}(\omega \otimes \eta \otimes \theta) = \mathrm{Alt}\big(\omega \otimes \mathrm{Alt}(\eta \otimes \theta)\big)$.

Proof:
(1) Let us consider
$(k+l)!\,\mathrm{Alt}(S \otimes T)(v_1, \dots, v_{k+l})$

$$= \sum_{\sigma \in S_{k+l}} (\operatorname{sgn} \sigma) \cdot S\big(v_{\sigma(1)}, \dots, v_{\sigma(k)}\big) \cdot T\big(v_{\sigma(k+1)}, \dots, v_{\sigma(k+l)}\big).$$

Taking the subgroup G of S_{k+l} formed by the permutations of $\{1, \dots, k+l\}$ that leave $k+1, \dots, k+l$ fixed. We have

$$\sum_{\sigma\in G} \operatorname{sgn}\sigma \cdot S\big(v_{\sigma(1)},\dots,v_{\sigma(k)}\big)\cdot T\big(v_{\sigma(k+1)},\dots,v_{\sigma(k+l)}\big)$$
$$= \left(\sum_{\sigma\in G} (\operatorname{sgn}\sigma)\cdot S\big(v_{\sigma(1)},\dots,v_{\sigma(k)}\big)\right)\cdot T(v_{k+1},\dots,v_{k+l})$$
$$= k!\ (\operatorname{Alt}(S)\otimes T)(v_1,\dots,v_{k+l}) = 0.$$

Then, since G decomposes S_{k+l} into disjoint right cosets $G\cdot\tilde{\sigma} = \{\sigma\tilde{\sigma} : \sigma\in G\}$, and for each coset

$$\sum_{\sigma\in G\cdot\tilde{\sigma}} \operatorname{sgn}\sigma\cdot(S\otimes T)\big(v_{\sigma(1)},\dots,v_{\sigma(k+l)}\big)$$
$$= \operatorname{sgn}\tilde{\sigma}\cdot\sum_{\sigma'\in G}\operatorname{sgn}\sigma\cdot(S\otimes T)\left(v_{\sigma(\tilde{\sigma}(1))},\dots,v_{\sigma(\tilde{\sigma}(k+l))}\right)$$
$$= \operatorname{sgn}\tilde{\sigma}\cdot k!\,(\operatorname{Alt}(S)\otimes T)\big(v_{\tilde{\sigma}(1)},\dots,v_{\tilde{\sigma}(k+l)}\big) = 0$$

we have that $\operatorname{Alt}(S\otimes T) = 0$. Similarly, we prove that $\operatorname{Alt}(T\otimes S) = 0$.

(2) By linearity of the operator Alt and the fact that $\operatorname{Alt}\circ\operatorname{Alt} = \operatorname{Alt}$, we have

$$\operatorname{Alt}(\operatorname{Alt}(\eta\otimes\theta) - \eta\otimes\theta) = 0.$$

Hence, by (1),

$$0 = \operatorname{Alt}\big(\omega\otimes(\operatorname{Alt}(\eta\otimes\theta) - \eta\otimes\theta)\big)$$
$$= \operatorname{Alt}\big(\omega\otimes\operatorname{Alt}(\eta\otimes\theta)\big) - \operatorname{Alt}(\omega\otimes\eta\otimes\theta),$$

Using these properties we can show the following.

Theorem (3.1.18):

$$(\omega\wedge\eta)\wedge\theta = \omega\wedge(\eta\wedge\theta).$$

Proof:

By Theorem (3.1.17), for $\omega\in\Lambda^k(V)$, $\eta\in\Lambda^l(V)$, and $\theta\in\Lambda^m(V)$, we have

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

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

and

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

$$= \frac{(k+l+m)!}{k!\,m!\,l!}\ \mathrm{Alt}(\omega \otimes \eta \otimes \theta). [5]$$

Theorem (3.1.19):

The set

$$\{\phi_{i_1} \wedge \cdots \wedge \phi_{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$$

is a basis for $\Lambda^k(V)$, and

$$\dim \Lambda^k(V) = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Proof:

If $\omega \in \Lambda^k(V) \subset \mathcal{J}^k(V)$, by theorem (3.1.7)

$$\omega = \sum_{i_1,\dots,i_k} a_{i_1,\dots,i_k} \phi_{i_1} \otimes \cdots \otimes \phi_{i_k}.$$

And, since ω is alternating,

$$\omega = \mathrm{Alt}(\omega) = \sum_{i_1,\dots,i_k} a_{i_1,\dots,i_k} \mathrm{Alt}\left(\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}\right).$$

We can show by induction that $\mathrm{Alt}\left(\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}\right) = \frac{1}{k!} \phi_{i_1} \wedge \phi_{i_2} \wedge \cdots \wedge \phi_{i_k}$. Indeed, for $k = 1$, the result is trivially true, and, assuming it is true for k basis tensors, we have, by Theorem (3.1.17), that

$$\mathrm{Alt}\left(\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}\right) = \mathrm{Alt}\left(\mathrm{Alt}\left(\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}\right) \otimes \phi_{i_{k+1}}\right)$$

$$= \frac{k!}{(k+1)!} \mathrm{Alt}\left(\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}\right) \wedge \phi_{i_{k+1}}$$

$$= \frac{k!}{(k+1)!} \phi_{i_1} \wedge \phi_{i_2} \wedge \cdots \wedge \phi_{i_k}.$$

Hence,

$$\omega = \frac{1}{k!} \sum_{i_1,\dots,i_k} a_{i_1,\dots,i_k} \phi_{i_1} \wedge \phi_{i_2} \wedge \cdots \wedge \phi_{i_k}.$$

However, the tensors $\phi_{i_1} \wedge \cdots \wedge \phi_{i_k}$ are not linearly independent. Indeed, due to anticommutativity, if two sequences $(i_1, \dots, i_k)$ and $(j_1, \dots, j_k)$ differ only in their orderings, then $\phi_{i_1} \wedge \cdots \wedge \phi_{i_k} = \pm \phi_{j_1} \wedge \cdots \wedge \phi_{j_k}$. In addition, if any two of the indices are equal, then $\phi_{i_1} \wedge \cdots \wedge \phi_{i_k} = 0$. Hence, we can avoid repeating terms by considering only increasing index sequences:

$$\omega = \sum_{i_1,\ldots,i_k} b_{i_1,\ldots,i_k} \phi_{i_1} \otimes \cdots \otimes \phi_{i_k}$$

and so the set $\{\phi_{i_1} \wedge \cdots \wedge \phi_{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$ spans $\Lambda^k(V)$. Moreover, the elements of this set are linearly independent. Indeed, if

$$0 = \omega = \sum_{i_1,\ldots,i_k} b_{i_1,\ldots,i_k} \phi_{i_1} \otimes \cdots \otimes \phi_{i_k},$$

then, taking a basis $\{v_1, \ldots, v_n\}$ and an increasing index sequence $(j_1, \ldots, j_k)$, we have

$$\begin{aligned} 0 = \phi(v_{j_1}, \ldots, v_{j_k}) &= k! \sum_{i_1,\ldots,i_k} b_{i_1,\ldots,i_k} \mathrm{Alt}(\phi_{i_1} \otimes \cdots \otimes \phi_{i_k})(v_{j_1}, \ldots, v_{j_k}) \\ &= \sum_{i_1,\ldots,i_k} b_{i_1,\ldots,i_k} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \cdot \phi_{i_1}(v_{j\sigma(1)}) \ldots \phi_{i_k}(v_{j\sigma(k)}). \end{aligned}$$

Since $(i_1, \ldots, i_k)$ and $(j_1, \ldots, j_k)$ are both increasing, the only term of the second sum that may be different from zero is the one for which $\sigma = \mathrm{id}$. Consequently,

$$0 = \phi(v_{j_1}, \ldots, v_{j_k}) = b_{i_1,\ldots,i_k}.$$

The following result is clear from the anticommutativity shown in Example (3.1.16). [5]

Notice that Theorem (3.1.19) implies that every $\omega \in \Lambda^k(\mathbf{R}^n)$ is a linear combination of the functions

$$(v_1, \ldots, v_n) \mapsto \text{determinant of a } k \times k \text{ minor of } \begin{pmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_k \end{pmatrix}. \text{ [12]}$$

Remark (3.1.20):

(1) Another consequence of Theorem (3.1.19) is that $\dim \Lambda^n(V) = 1$. Hence, if $V = \mathbf{R}^n$, any alternating n-tensor in $\mathbf{R}^n$ is a multiple of the determinant.

(2) It is also clear that $\Lambda^k(V) = 0$ if $k > n$. Moreover, the set $\Lambda^0(V)$ is defined to be equal to $\mathbf{R}$ (identified with the set of constant functions on V). [5]

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 $\Lambda^n(\mathbf{R}^n)$, it is not surprising to find it in the following theorem.

Theorem (3.1.21):
Let $v_1,\ldots,v_n$ be a basis for V, and let $\omega \in \Lambda^n(V)$, and let

$$\omega = \sum_{j=1}^{n} a_{ji} v_j \qquad i = 1,\ldots,n.$$

Then

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

Proof:
Define $\eta \in \mathcal{J}^n(\mathbf{R}^n)$ by

$$\eta\left((a_{11},\ldots,a_{n1}),\ldots,(a_{n1},\ldots,a_{nn})\right) = \omega\left(\sum_{j=1}^{n} a_{j1} v_j\,,\ldots,\sum_{j=1}^{n} a_{jn} v_j\right).$$

Then clearly $\eta \in \Lambda^n(\mathbf{R}^n)$, so $\eta = \lambda.\det$ for some $\lambda \in \mathbf{R}$, and

$$\lambda = \eta(e_1,\ldots,e_n) = \omega(v_1,\ldots,v_n).$$

The following properties of k-forms are obvious from the corresponding properties for $\Lambda^n(V)$:

$$(\omega_1 + \omega_2) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta$$

$$\omega \wedge (\eta_1 + \eta_2) = \omega \wedge \eta_1 + \omega \wedge \eta_2$$

$$a\omega \wedge \eta = \omega \wedge a\eta = a(\omega \wedge \eta)$$

$$\omega \wedge \eta = (-1)^{kl} \eta \wedge \omega$$

$$f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta.$$ [12, 13]

Theorem (3.1.21) shows that a non-zero $\omega \in \Lambda^n(V)$ splits the bases of V into two disjoint groups, those with $\omega(v_1,\ldots,v_n) > 0$ and those for which $\omega(v_1,\ldots,v_n) < 0$; if $v_1,\ldots,v_n$ and $w_1,\ldots,w_n$ are two bases and $A = \left(a_{ij}\right)$ is define by $w_i = \sum a_{ij} v_j$, then $v_1,\ldots,v_n$ and $w_1,\ldots,w_n$ are in the same group 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 $v_1, \dots, v_n$ belongs is denoted $[v_1, \dots, v_n]$ and the other orientation is denoted $-[v_1, \dots, v_n]$. In $\mathbf{R}^n$ we define the usual orientation as $[e_1, \dots, e_n]$.

The fact that $\dim \Lambda^n(\mathbf{R}^n) = 1$ is probably not new to us, since det is often defined as the unique element $\omega \in \Lambda^n(\mathbf{R}^n)$ such that $\omega(e_1, \dots, e_n) = 1$. 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 $v_1, \dots, v_n$ and $w_1, \dots, w_n$ 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 $A^{\mathbf{T}}$ denotes the transpose of the matrix A, then we have $A \cdot A^{\mathbf{T}} = I$, so $\det A = \pm 1$. It follows from Theorem (3.1.21) that if $\omega \in \Lambda^n(V)$ satisfies $\omega(v_1, \dots, v_n) = \pm 1$, then $\omega(w_1, \dots, w_n) = \pm 1$. If an orientation μ for V has also been given, it follows that there is a unique $\omega \in \Lambda^n(V)$ such that $\omega(v_1, \dots, v_n) = 1$ whenever $v_1, \dots, v_n$ is an orthonormal basis such that $[v_1, \dots, v_n] = \mu$. 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 $\mathbf{R}^n$ 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 $v_1, \dots, v_n$.

Next is a construction which we will restrict to $\mathbf{R}^n$. If $v_1, \dots, v_{n-1} \in \mathbf{R}^n$ and φ is defined by

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

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

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

This z is denoted $v_1 \times \cdots \times v_{n-1}$ and called the cross product of $v_1, \ldots, v_{n-1}$. 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 $v, w \in \mathbf{R}^3$, we obtain a more conventional looking product, $v \times w \in \mathbf{R}^3$. For this reason it is sometimes maintained that the cross product can be defined only in $\mathbf{R}^3$.

Now we can study tangent space and vectors fields, and we will be concerned Poincaré Lemma.

Definition (3.1.22):

Given $p \in \mathbf{R}^n$, we define a tangent space to $\mathbf{R}^n$ at p to a pairs (p, v), where $v \in \mathbf{R}^n$, and is denoted $\mathbf{R}^n{}_p$. This set is made into a vector space in the most obvious way, by defining [10]

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

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

A vector $v \in \mathbf{R}^n$ is often pictured as an arrow from 0 to v; the vector $(p, v) \in \mathbf{R}^n{}_p$ may be pictured (Figure (3.3)) 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.

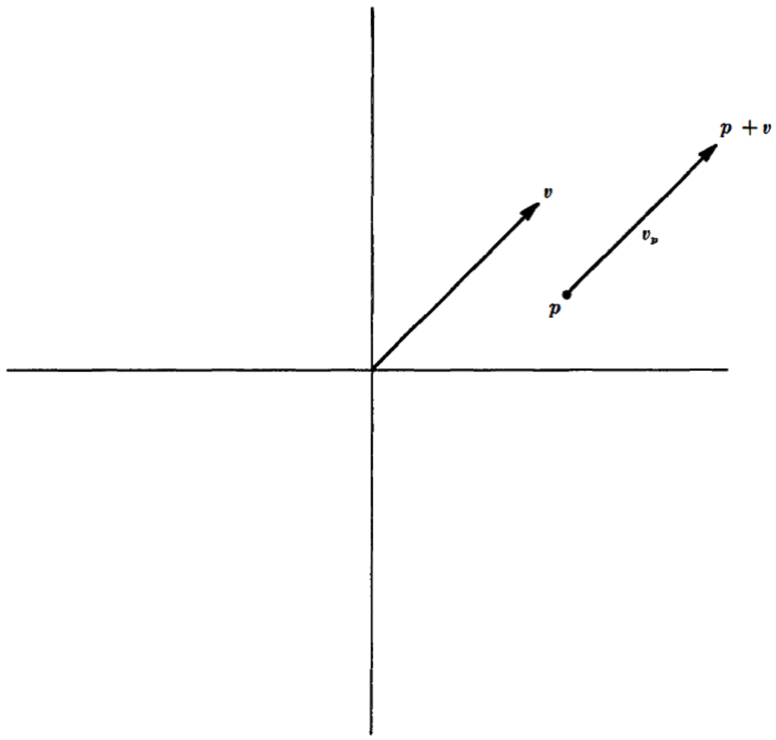


Figure (3.3)

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

Any operation which is possible in a vector space may be performed in each $\mathbf{R}^n{}_p$, 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 $\mathbf{R}^n{}_p$, we obtain a vector field (Figure (3.4)). To be precise, a vector field is a function F such that $F(p) \in \mathbf{R}^n{}_p$ for each $p \in \mathbf{R}^n$. For each p there are numbers $F^1(p), \dots, F^n(p)$ such that

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

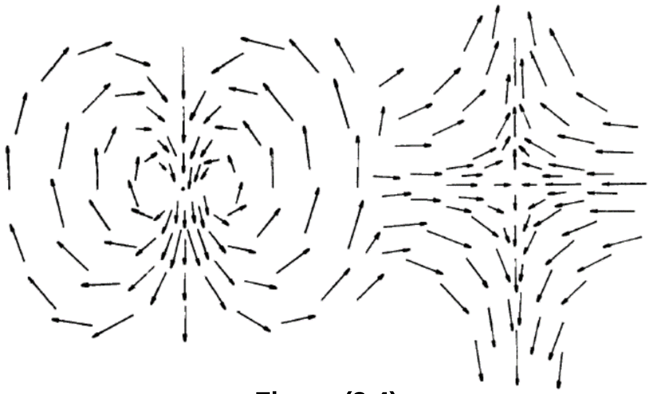


Figure (3.4)

We thus obtain n component functions $F^i\colon \mathbf{R}^n \to \mathbf{R}$. The vector field F is called continuous, differentiable, etc., if the functions F^i are. Similar definitions can be made for a vector field defined only on an open subset of $\mathbf{R}^n$. 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 $\mathbf{R}^n$, then we can similarly define

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

Certain other definitions are standard and useful. We define the divergence, $\operatorname{div} F$ 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, $\operatorname{div} F = \langle \nabla, F\rangle$. If $n = 3$ we write, in conformity with this symbolism,

$$(\nabla \times F)(p) = (D_2F^3 - D_3F^2)(e_1)_p + (D_3F^1 - D_1F^3)(e_2)_p + (D_1F^2 - D_2F^1)(e_3)_p.$$

The vector field $\nabla \times F$ is called $\operatorname{curl} F$. The names “divergence” and “curl” are derived from physical considerations which are explained at the end of this research.

Many similar considerations may be applied to a function ω with $\omega(p) \in \Lambda^k(\mathbf{R}^n{}_p)$; such a function is called a k-form on $\mathbf{R}^n$, 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

$$\omega(p) = \sum_{i_1 < \ldots < i_k} \omega_{i_1,\ldots,i_k}(p) \cdot \left[\varphi_{i_1}(p) \wedge \cdots \wedge \varphi_{i_k}(p)\right]$$

for certain functions $\omega_{i_1,\ldots,i_k}$; the form ω is called continuous, differentiable, etc., if these functions are. We shall usually assume tacitly that forms and vector 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 $f: \mathbf{R}^n \to \mathbf{R}$ is differentiable, then $Df(p) \in \Lambda^1(\mathbf{R}^n)$. 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. This standard notation has obvious disadvantages but it allows many classical results to be expressed by formulas of equally classical appearance. Since $dx^i(p)(v_p) = d\pi^i(p)(v_p) = D\pi^i(p)(v) = v^i$ we see that $dx^1(p), \ldots, dx^n(p)$ is just the dual basis to $(e_1)_p, \ldots, (e_n)_p$. Thus every k-form ω can be written

$$\omega = \sum_{i_1 < \ldots < i_k} \omega_{i_1, \ldots, i_k} \, dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

The expression for df is of particular interest.

Theorem (3.1.23):
If $f: \mathbf{R}^n \to \mathbf{R}$ is differentiable, then

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

In classical notation,

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

Proof:

$$df(p)(v_p) = Df(p)(v) = \sum_{i=1}^n v^i \cdot D_i f(p) = \sum_{i=1}^n dx^i(p)(v_p) \cdot D_i f(p).$$

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

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

This linear transformation induces a linear transformation $f_*: \Lambda^k\left({\mathbf{R}^m}_{f(p)}\right) \to \Lambda^k({\mathbf{R}^n}_p)$. If ω is a k-form on $\mathbf{R}^m$ we can therefore define a k-form $f^*\omega$ on $\mathbf{R}^n$ by $(f^*\omega)(p) = f^*\left(\omega\big(f(p)\big)\right)$.

Recall this means that if $v_1, \dots, v_k \in {\mathbf{R}^n}_p$, then we have $f^*\omega(p)(v_1, \dots, v_k) = \omega\big(f(p)\big)\big(f_*(v_1), \dots, f_*(v_k)\big)$. As an antidote to the abstractness of these definitions we present a theorem, summarizing the important properties of f^*, which allows explicit calculations of $f^*\omega$.

Theorem (3.1.24):

If $f: \mathbf{R}^n \to \mathbf{R}^m$ is differentiable, then

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

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

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

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

Proof:

(1)
$$\begin{aligned} f^*\left(dx^i\right)(p)\left(v_p\right) &= dx^i\big(f(p)\big)\left(f_* v_p\right) \\ &= dx^i\big(f(p)\big)\left(\sum_{j=1}^n v^j \cdot D_j f^1(p), \dots, \sum_{j=1}^n v^j \cdot D_j f^m(p)\right)_{f(p)} \\ &= \sum_{j=1}^n v^j \cdot D_j f^i(p) \\ &= \sum_{j=1}^n D_j f^i(p) \cdot dx^j(p)\left(v_p\right). \end{aligned}$$

(2) We notice that given $p \in \mathbf{R}^n$ and $v_1, \dots, v_k \in {\mathbf{R}^n}_p$. Then

$$\begin{aligned} f^*(\omega_1 + \omega_2)(p)(v_1, \dots, v_k) &= (\omega_1 + \omega_2)\big(f(p)\big)\left(df_p(v_1), \dots, df_p(v_k)\right) \\ &= (f^*\omega_1)(p)(v_1, \dots, v_k) + (f^*\omega_2)(p)(v_1, \dots, v_k) \\ &= (f^*\omega_1 + f^*\omega_2)(p)(v_1, \dots, v_k). \end{aligned}$$

(3) We have

$$\begin{aligned} f^*(g \cdot \omega)_p(v_1, \dots, v_k) &= (g \cdot \omega)_{f(p)}\big((df)_p v_1, \dots, (df)_p v_k\big) \\ &= g\big(f(p)\big)\omega_{f(p)}\left((df)_p v_1, \dots, (df)_p v_k\right) \\ &= (g \circ f)(p)(f^*\omega)_p(v_1, \dots, v_k). \end{aligned}$$

(4) By setting $(y_1, \ldots, y_m) = \big(f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)\big) \in \mathbf{R}^m$, $(x_1, \ldots, x_n) \in \mathbf{R}^n$, $\omega = \sum_{i_1 < \cdots < i_k} a_{i_1,\ldots,i_k}\, dy^{i_1} \wedge \cdots \wedge dy^{i_k}$, $\eta = \sum_{j_1 < \cdots < j_l} b_{j_1,\ldots,j_l}\, dy^{j_1} \wedge \cdots \wedge dy^{j_l}$

$$f^*(\omega \wedge \eta) = f^*\left(\sum_{i_1 < \cdots < i_k} \sum_{j_1 < \cdots < j_l} \big(a_{i_1,\ldots,i_k} b_{j_1,\ldots,j_l}\big)\big(dy^{i_1} \wedge \cdots \wedge dy^{i_k}\big) \wedge \big(dy^{j_1} \wedge \cdots \wedge dy^{j_l}\big)\right)$$

$$= \sum_{i_1 < \cdots < i_k} \sum_{j_1 < \cdots < j_l} a_{i_1,\ldots,i_k}(f_1, \ldots, f_m) b_{j_1,\ldots,j_l}(f_1, \ldots, f_m)\big(df^{i_1} \wedge \cdots \wedge df^{i_k}\big) \wedge \big(df^{j_1} \wedge \cdots \wedge df^{j_l}\big)$$

$$= \sum_{i_1 < \cdots < i_k} a_{i_1,\ldots,i_k}(f_1, \ldots, f_m)\big(df^{i_1} \wedge \cdots \wedge df^{i_k}\big) \sum_{j_1 < \cdots < j_l} b_{j_1,\ldots,j_l}(f_1, \ldots, f_m) \wedge \big(df^{j_1} \wedge \cdots \wedge df^{j_l}\big)$$

$= f^*\omega \wedge f^*\eta$. [3, 5]

By repeatedly applying Theorem (3.1.24) we have, for example,

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

The expression obtained by expanding out each $f^*\big(dx^i\big)$ is quite complicated. In one special case it will be worth our while to make an explicit evaluation.

Theorem (3.1.25):

If $f: \mathbf{R}^n \to \mathbf{R}^n$ is differentiable, then

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

Proof:

Since

$$f^*(h\, dx^1 \wedge \cdots \wedge dx^n) = (h \circ f) f^*(dx^1 \wedge \cdots \wedge dx^n),$$

it suffices to show that

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

Let $p \in \mathbf{R}^n$ and let $A = (a_{ij})$ be the matrix of $f'(p)$. Here, and whenever convenient and not confusing, we shall omit "p" in $dx^1 \wedge \cdots \wedge dx^n(p)$, etc. Then

$$\begin{aligned}
f^*(dx^1 \wedge \cdots \wedge dx^n)(e_1, \ldots, e_n) & \\
&= dx^1 \wedge \cdots \wedge dx^n(f_* e_1, \ldots, f_* e_n) \\
&= dx^1 \wedge \cdots \wedge \left(\sum_{i=1}^{n} a_{i1} e_i, \ldots, \sum_{i=1}^{n} a_{in} e_i\right) \\
&= \det(a_{ij}) \cdot dx^1 \wedge \cdots \wedge dx^n(e_1, \ldots, e_n),
\end{aligned}$$

By Theorem (3.1.21).

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

$$\omega = \sum_{i_1 < \cdots < i_k} \omega_{i_1,\ldots,i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k},$$

we define a $(k+1)$-form called exterior derivative of ω, by

$$\begin{aligned}
d\omega &= \sum_{i_1 < \cdots < i_k} d\omega_{i_1,\ldots,i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k} \\
&= \sum_{i_1 < \cdots < i_k} \sum_{\alpha=1}^{n} D_\alpha\left(\omega_{i_1,\ldots,i_k}\right) \cdot dx^\alpha \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}. \quad [13]
\end{aligned}$$

Example (3.1.26):

Consider the form $\omega = \frac{-y}{x^2+y^2}\, dx + \frac{x}{x^2+y^2}\, dy$ defined on $\mathbf{R}^2 - 0$. Then,

$$\begin{aligned}
d\omega &= d\left(\frac{-y}{x^2+y^2}\right) \wedge dx + d\left(\frac{x}{x^2+y^2}\right) \wedge dy \\
&= \frac{y^2 - x^2}{(x^2+y^2)^2}\, dy \wedge dx + \frac{y^2 - x^2}{(x^2+y^2)^2}\, dx \wedge dy = 0.
\end{aligned}$$

The exterior derivative satisfies the following properties: [5]

Theorem (3.1.27):

1) $d(\omega+\eta)=d\omega+d\eta$.
2) If ω is a k-form and η is an l-form, then
$$d(\omega\wedge\eta)=d\omega\wedge\eta+(-1)^k\omega\wedge d\eta.$$
3) $d(d\omega)=0$. Briefly, $d^2=0$.
4) If ω is a k-form on $\mathbf{R}^m$ and $f:\mathbf{R}^n\to\mathbf{R}^m$ is differentiable, then $f^*(d\omega)=d(f^*\omega)$.

Proof:

(1) $$\begin{aligned}d(\omega+\eta)&=\frac{\partial(\omega+\eta)}{\partial x}dx+\frac{\partial(\omega+\eta)}{\partial x}dy+\frac{\partial(\omega+\eta)}{\partial x}dz\\&=\frac{\partial\omega}{\partial x}dx+\frac{\partial\omega}{\partial y}dy+\frac{\partial\omega}{\partial z}dz+\frac{\partial\eta}{\partial x}dx+\frac{\partial\eta}{\partial y}dy+\frac{\partial\eta}{\partial z}dz\\&=d\omega+d\eta.\end{aligned}$$

(2) Using (1), it is enough to prove (2) for $\omega=a\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}$ and $\eta=b\,dx^{j_1}\wedge\cdots\wedge dx^{j_l}$:

$$\begin{aligned}d(\omega\wedge\eta)&=d\left[\left(a\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}\right)\wedge\left(b\,dx^{j_1}\wedge\cdots\wedge dx^{j_l}\right)\right]\\&=d(ab)\left(dx^{i_1}\wedge\cdots\wedge dx^{i_k}\right)\wedge\left(dx^{j_1}\wedge\cdots\wedge dx^{j_l}\right)\\&=[(da)b+a(db)]\wedge\left(dx^{i_1}\wedge\cdots\wedge dx^{i_k}\right)\wedge\left(dx^{j_1}\wedge\cdots\wedge dx^{j_l}\right)\\&=da\wedge d\left(x^{i_1}\wedge\cdots\wedge dx^{i_k}\right)\wedge\left(b\,dx^{j_1}\wedge\cdots\wedge dx^{j_l}\right)\\&\qquad+(-1)^k\left(a\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}\right)\wedge\left(db\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}\right),\\&=d\omega\wedge\eta+(-1)^k a\,dx^{i_1}\wedge\cdots\wedge dx^{i_k}\wedge db\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}\\&=d\omega\wedge\eta+(-1)^k\omega\wedge d\eta.\end{aligned}$$

(3) Since
$$d\omega=\sum_{i_1<\cdots<i_k}\sum_{\alpha=1}^{n}D_\alpha\left(\omega_{i_1,\dots,i_k}\right)\cdot dx^\alpha\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k},$$
we have
$$d(d\omega)=\sum_{i_1<\cdots<i_k}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}D_{\alpha,\beta}\left(\omega_{i_1,\dots,i_k}\right)\cdot dx^\beta\wedge dx^\alpha\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}.$$

In this sum the terms
$$D_{\alpha,\beta}\left(\omega_{i_1,\dots,i_k}\right)\cdot dx^\beta\wedge dx^\alpha\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}$$
and
$$D_{\beta,\alpha}\left(\omega_{i_1,\dots,i_k}\right)\cdot dx^\alpha\wedge dx^\beta\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}$$

cancel in pairs.

(4) This is clear if ω is a 0-form. Suppose, inductively, that (4) is true when ω is a k-form. It suffices to prove (4) for a $(k+1)$-form of the type $\omega \wedge dx^i$. We have

$$\begin{aligned} f^*\left(d\left(\omega \wedge dx^i\right)\right) &= f^*\left(d\omega \wedge dx^i + (-1)^k \omega \wedge d\left(dx^i\right)\right) \\ &= f^*\left(d\omega \wedge dx^i\right) = f^*(d\omega) \wedge f^*\left(dx^i\right) \\ &= d\left(f^*\omega \wedge f^*\left(dx^i\right)\right) \qquad \text{by (2) and (3)} \\ &= d\left(f^*\left(\omega \wedge dx^i\right)\right). \ [2,13] \end{aligned}$$

A form ω is called closed if $d\omega = 0$ and exact if $\omega = d\eta$, for some η. Theorem (3.1.27) shows that every exact form is closed, and it is natural to ask whether, conversely, every closed form is exact. If ω is the 1-form $Pdx + Qdy$ on $\mathbf{R}^2$, then

$$\begin{aligned} d\omega &= (D_1Pdx + D_2Pdy) \wedge dx + (D_1Qdx + D_2Qdy) \wedge dy \\ &= (D_1Q - D_2P)dx \wedge dy. \end{aligned}$$

Thus, if $d\omega = 0$, then $D_1Q = D_2P$.

Example (3.1.28):

Let $g_1, g_2 : \mathbf{R}^2 \to \mathbf{R}$ be continuous. Define $f : \mathbf{R}^2 \to \mathbf{R}$ by

$$f(x, y) = \int_0^x g_1(t, 0)dt + \int_0^y g_2(x, t)dt.$$

a) Show that $D_2 f(x, y) = g_2(x, y)$.
b) How should f be defined so that $D_1 f(x, y) = g_1(x, y)$?
c) Find a function $f : \mathbf{R}^2 \to \mathbf{R}$ such that $D_1 f(x, y) = x$ and $D_2 f(x, y) = y$. find one such that $D_1 f(x, y) = y$ and $D_2 f(x, y) = x$.

Proof:

a) $D_2 f(x, y) = 0 + g_2(x, y) = g_2(x, y)$.
b) We should let

$$f(x, y) = \int_0^x g_1(t, y)dt + \int_0^y g_2(a, t)dt.$$

Where $t \in \mathbf{R}$ is a constant.

c) Let

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

Example (3.1.28) 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 $\mathbf{R}^2$, however, such a function may not exist. In the classical example (3.1.26) the form is usually denoted $d\theta$, since it equals $d\theta$ on the set $\{(x,y): x < 0, \text{or } x \geq 0 \text{ and } y \neq 0\}$, here θ is defined. Note, however, that θ cannot be defined continuously on all of $\mathbf{R}^2 - 0$. If $\omega = df$ for some function $f: \mathbf{R}^2 - 0 \rightarrow \mathbf{R}$, then $D_1 f = D_1\theta$ and $D_2 f = D_2\theta$, so $f = \theta +$ constant, showing that such an f cannot exist.

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

$$f(x) = \int_0^1 \frac{d}{dt} f(tx) dt$$

$$= \int_0^1 \sum_{i=1}^{n} D_i f(tx) \cdot x^i dt$$

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

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 $A \subset \mathbf{R}^n$ with the property that whenever $x \in A$, the line segment from 0 to x is contained in A; such an open set is called star-shaped with respect to 0 (Figure(3.5)). A somewhat involved calculation shows that 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: [13]

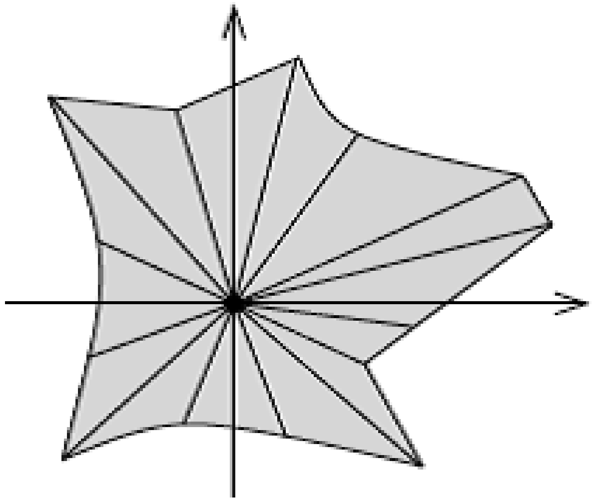


Figure (3.5): A star-shaped subset of R^2 [8]

Theorem (3.1.29): (Poincaré Lemma)

If $A \subset \mathbf{R}^n$ is an open set star-shaped with respect to 0, then every closed form on A is exact.

Proof:

We will define a function I from l-forms to $(l-1)$-forms, such that $I(0) = 0$ and $\omega = I(d\omega) + d(I\omega)$ for any form ω. It follows that $\omega = d(I\omega)$ if $d\omega = 0$. Let

$$\omega = \sum_{i_1 < \cdots < i_l} \omega_{i_1,\ldots,i_l} \, dx^{i_1} \wedge \cdots \wedge dx^{i_l}.$$

Since A is star-shaped we can define

$$I\omega(x) = \sum_{i_1 < \cdots < i_l} \sum_{\alpha=1}^{l} (-1)^{\alpha-1} \left(\int_0^1 t^{l-1} \, \omega_{i_1,\ldots,i_l}(tx) dt \right) x^{i_\alpha}$$
$$dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_\alpha}} \wedge \cdots \wedge dx^{i_l}.$$

(The symbol $\frown$ over dx^{i_α} indicates that it is omitted.) The proof that $\omega = I(d\omega) + d(I\omega)$ is an elaborate computation: We have,

$$d(I\omega) = l \cdot \sum_{i_1 < \cdots < i_l} \left(\int_0^1 t^{l-1}\, \omega_{i_1,\dots,i_l}(tx) dt \right) dx^{i_1} \wedge \cdots \wedge dx^{i_l}$$

$$+ \sum_{i_1 < \cdots < i_l} \sum_{\alpha=1}^{l} \sum_{j=1}^{n} (-1)^{\alpha-1} \left(\int_0^1 t^l D_j \left(\omega_{i_1,\dots,i_l} \right)(tx) dt \right) x^{i_\alpha}$$

$$dx^j \wedge dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_\alpha}} \wedge \cdots \wedge dx^{i_l}.$$

We also have

$$d\omega = \sum_{i_1 < \cdots < i_l} \sum_{j=1}^{n} D_j\left(\omega_{i_1,\dots,i_l}\right) \cdot dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_l}.$$

Applying I to the $(l+1)$-form $d\omega$, we obtain

$$I(d\omega) = \sum_{i_1 < \cdots < i_l} \sum_{j=1}^{n} \left(\int_0^1 t^l\, D_j\left(\omega_{i_1,\dots,i_l}\right)(tx) dt \right) x^j\, dx^{i_1} \wedge \cdots \wedge dx^{i_l}$$

$$- \sum_{i_1 < \cdots < i_l} \sum_{j=1}^{n} \sum_{\alpha=1}^{l} (-1)^{\alpha-1} \left(\int_0^1 t^l D_j \left(\omega_{i_1,\dots,i_l} \right)(tx) dt \right) x^{i_\alpha}$$

$$dx^j \wedge dx^{i_1} \wedge \cdots \wedge \widehat{dx^{i_\alpha}} \wedge \cdots \wedge dx^{i_l}.$$

Adding, the triple sums cancel, and we obtain

$$d(I\omega) + I(d\omega) = \sum_{i_1 < \cdots < i_l} l \cdot \left(\int_0^1 t^{l-1}\, \omega_{i_1,\dots,i_l}(tx) dt \right) dx^{i_1} \wedge \cdots \wedge dx^{i_l}$$

$$+ \sum_{i_1 < \cdots < i_l} \sum_{j=1}^{n} \left(\int_0^1 t^l x^j D_j \left(\omega_{i_1,\dots,i_l} \right)(tx) dt \right) dx^{i_1} \wedge \cdots \wedge dx^{i_l}$$

$$= \sum_{i_1 < \cdots < i_l} \left(\int_0^1 \frac{d}{dt} \left[\left(t^l \omega_{i_1,\dots,i_l} \right)(tx) \right] dt \right) dx^{i_1} \wedge \cdots \wedge dx^{i_l}$$

$$= \sum_{i_1 < \cdots < i_l} \omega_{i_1,\dots,i_l}\, dx^{i_1} \wedge \cdots \wedge dx^{i_l}$$

$$= \omega.$$

Section (3.2): Geometry and The Fundamental Theorem

A singular n-cube in $A \subset \mathbf{R}^n$ is a continuous function $c: [0,1]^n \to A$. We let $\mathbf{R}^0$ 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 $\mathbf{R}^n$ is the standard n-cube $I^n: [0,1]^n \to \mathbf{R}^n$ defined by $I^n(x) = x$ 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 added, and multiplied by integers. For example

$$2(c_1 + 3c_4) + (-2)(c_1 + c_3 + c_2) = -2c_2 - 2c_3 + 6c_4. \text{ [13]}$$

The reason for introducing is that to every n-chain c we wish to associate a $(n-1)$-chain ∂c, which is called the boundary of c, and which is supposed to be the sum of the various singular $(n-1)$-cubes around the boundary of each singular n-cubes in c. In practice, it is convenient to modify this idea.

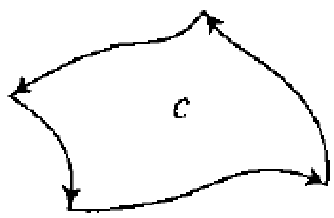
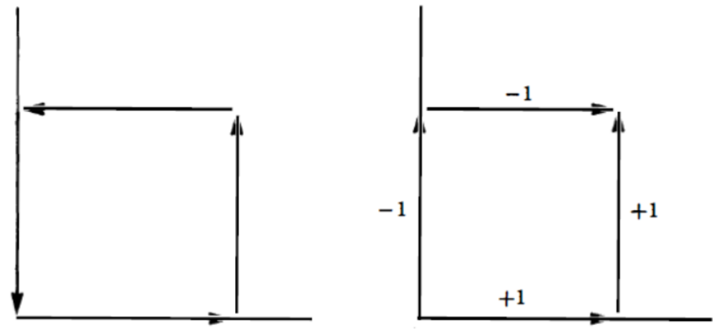


Figure (3.6)

The boundary of I^2, for example, will not be the sum of the four singular 1-cubes indicated in Figure (3.7(a)), but the sum, with the indicated coefficients, of a 1-cubes shown in Figure (3.7(b)).



(a) (b)

Figure (3.7)

For each i with $1 \leq i \leq n$ we first 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

$$I^n_{(i,0)}(x) = I^n\left(x^1, \ldots, x^{i-1}, 0, x^i, \ldots, x^{n-1}\right)$$

$$= \left(x^1, \ldots, x^{i-1}, 0, x^i, \ldots, x^{n-1}\right),$$

$$I^n_{(i,1)}(x) = I^n\left(x^1, \ldots, x^{i-1}, 1, x^i, \ldots, x^{n-1}\right)$$

$$= \left(x^1, \ldots, x^{i-1}, 1, x^i, \ldots, x^{n-1}\right).$$

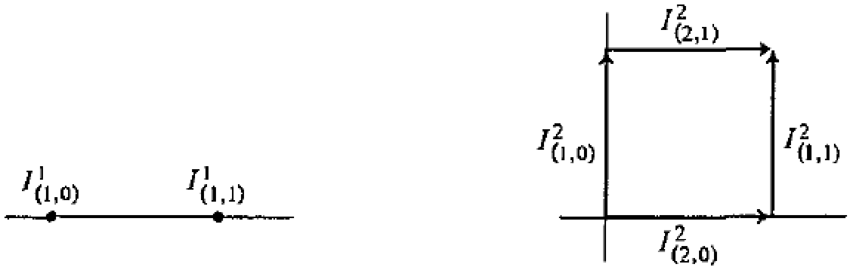


Figure (3.8)

We call $I^n_{(i,0)}$ the $(i,0)$-face of I^n and $I^n_{(i,1)}$ the $(i,1)$-face (Figure (3.8)). We then define

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

The (i,α)-face of a singular n-cube c is defined by

$$c_{(i,\alpha)} = c \circ \left(I^n_{(i,\alpha)}\right).$$

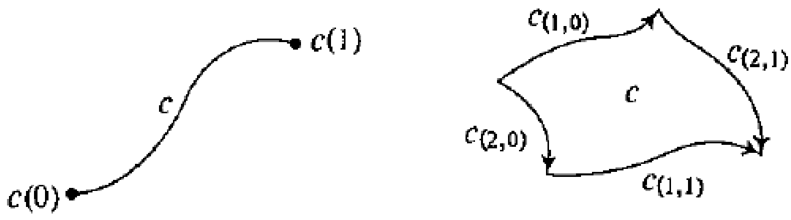


Figure (3.9)

Now we define

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

Finally, the boundary of an n-chain $\sum a_i c_i$ is define 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 research, we include here the one standard property of ∂.

Proposition (3.2.1):

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

Proof:

Let $i \leq j \leq n-1$, and consider $(I^n_{(i,\alpha)})_{(j,\beta)}$. For $x \in [0,1]^{n-2}$, we have from the definition

$$\begin{aligned}(I^n_{(i,\alpha)})_{(j,\beta)}(x) &= I^n_{(i,\alpha)}\left(I^{n-1}_{(j,\beta)}(x)\right) \\ &= I^n_{(i,\alpha)}\left(x^1, \dots, x^{j-1}, \beta, x^j, \dots, x^{n-2}\right) \\ &= I^n\left(x^1, \dots, x^{i-1}, \alpha, x^i, \dots, x^{j-1}, \beta, x^j, \dots, x^{n-2}\right).\end{aligned}$$

Similarly

$$\begin{aligned}(I^n_{(j+1,\beta)})_{(i,\alpha)}(x) &= I^n_{(j+1,\beta)}\left(I^{n-1}_{(i,\alpha)}(x)\right) \\ &= I^n_{(j+1,\beta)}\left(x^1, \dots, x^{i-1}, \alpha, x^i, \dots, x^{n-2}\right) \\ &= I^n\left(x^1, \dots, x^{i-1}, \alpha, x^i, \dots, x^{j-1}, \beta, x^j, \dots, x^{n-2}\right).\end{aligned}$$

Thus $(I^n_{(i,\alpha)})_{(j,\beta)} = (I^n_{(j+1,\beta)})_{(i,\alpha)}$ for $i \le j \le n-1$. It follows easily for any singular n-cube c that $(c_{(i,\alpha)})_{(j,\beta)} = (c_{(j+1,\beta)})_{(i,\alpha)}$ for $i \le j \le n-1$. Now

$$\begin{aligned}\partial(\partial c) &= \partial\left(\sum_{i=1}^{n}\sum_{\alpha=0,1}(-1)^{i+\alpha}c_{(i,\alpha)}\right)\\ &= \sum_{i=1}^{n}\sum_{\alpha=0,1}\sum_{j=1}^{n-1}\sum_{\beta=0,1}(-1)^{i+\alpha+j+\beta}(c_{(i,\alpha)})_{(j,\beta)}.\end{aligned}$$

In this sum $(c_{(i,\alpha)})_{(j,\beta)}$ and $(c_{(j+1,\beta)})_{(i,\alpha)}$ occur with opposite signs. Therefore all terms cancel out in pairs, and $\partial(\partial c) = 0$. Since the theorem is true for any singular n-cube, it is also true for singular n-chains. [12]

It is natural to ask whether proposition (3.2.1) 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 $c\colon [0,1] \to \mathbf{R}^2 - 0$ by $c(t) = (\sin 2\pi nt, \cos 2\pi nt)$, where n is a non-zero integer. Then $c(1) = c(0)$. [13]

Notice that for some n-chains c we have not only $\partial(\partial c) = 0$, but even $\partial c = 0$. For example, this is the case if $c = c_1 - c_2$, where c_1 and c_2 are two 1-cubes with $c_1(0) = c_2(0)$ and $c_1(1) = c_2(1)$. If c is just a singular 1-cube itself, then

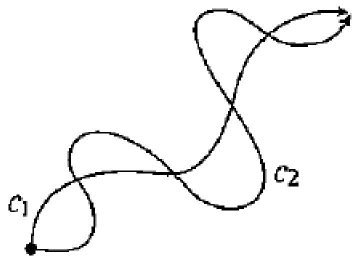


Figure (3.10)

$\partial c = 0$ precisely when $c(0) = c(1)$, i.e., when c is a "closed" curve.

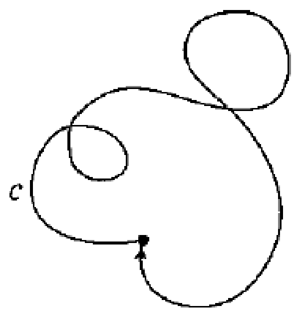


Figure (3.11)

In general, any n-chain c called closed if $\partial c = 0$. But there is no 2-chain c' in $\mathbf{R}^2 - 0$, with $\partial c' = c$. [12]

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.

Definition (3.2.2):
Let ω is a k-form on $[0,1]^k$, Then ω can be written uniquely in the form

$$\omega = f dx^1 \wedge \cdots \wedge f dx^k.$$

We define the integral of ω over A by the equation

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

We could also write this as

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

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

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

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

A special definition must be made for $k = 0$. A 0-form ω is a function; if $c\colon \{0\} \to A$ 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 $Pdx + Qdy$ is a 1-form on $\mathbf{R}^2$ and $c: [0,1] \to \mathbf{R}^2$ is a singular 1-cube (a curve), then one can prove that

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

where $t_0, \dots, t_n$ 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 the 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 (4).

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). [13]

Theorem (3.2.3): (Stokes' Theorem)

If ω is a $(k-1)$-form on an open set $A \subset \mathbf{R}^n$ and c is a k-chain in A, then

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

Proof:

Most of the proof involves the special case where ω is a $(k-1)$-form on $\mathbf{R}^k$ and $c = I^k$. in this case, ω is a sum of $(k-1)$-forms of the type

$$f dx^1 \wedge \cdots \wedge \widehat{dx^\iota} \wedge \cdots \wedge dx^k,$$

and it suffices to prove the theorem for each of these. We now compute. First, a little notation translation shows that

$$\int_{[0,1]^{k-1}} {I^k_{(j,\alpha)}}^* \left(f dx^1 \wedge \cdots \wedge \widehat{dx^\iota} \wedge \cdots \wedge dx^k\right)$$

$$= \begin{cases} 0 & \text{if } j \neq i \\ \displaystyle\int_{[0,1]^k} f(x^1, \dots, \alpha, \dots, x^k) dx^1 \dots dx^k & \text{if } j = i. \end{cases}$$

Therefore

$$\int_{\partial I^k} f dx^1 \wedge \cdots \wedge \widehat{dx^\iota} \wedge \cdots \wedge dx^k$$

$$= \sum_{j=1}^{k} \sum_{\alpha=0,1} (-1)^{j+\alpha} \int_{[0,1]^{k-1}} {I^k_{(j,\alpha)}}^* \left(f dx^1 \wedge \cdots \wedge \widehat{dx^\iota} \wedge \cdots \wedge dx^k\right)$$

$$= (-1)^{i+1} \int_{[0,1]^k} f(x^1, \dots, 1, \dots, x^k) dx^1 \dots dx^k$$

$$+ (-1)^{i} \int_{[0,1]^k} f(x^1, \dots, 0, \dots, x^k) dx^1 \dots dx^k .$$

On the other hand,

$$\int_{I^k} d\left(f dx^1 \wedge \cdots \wedge \widehat{dx^\iota} \wedge \cdots \wedge dx^k\right) = \int_{[0,1]^k} D_i f \, dx^i \wedge dx^1 \wedge \cdots \wedge \widehat{dx^\iota} \wedge \cdots \wedge dx^k$$

$$= (-1)^{i-1} \int_{[0,1]^k} D_i f .$$

By Fubini's theorem and the fundamental theorem of calculus we have

$$\int_{I^k} d\big(f dx^1 \wedge \cdots \wedge \widehat{dx^\iota} \wedge \cdots \wedge dx^k\big)$$

$$= (-1)^{i-1} \int_0^1 \cdots \left(\int_0^1 D_i f(x^1, \ldots, x^k) dx^i \right) dx^1 \ldots \widehat{dx^\iota} \ldots dx^k$$

$$= (-1)^{i-1} \int_0^1 \cdots \int_0^1 [f(x^1, \ldots, 1, \ldots, x^k) - f(x^1, \ldots, 0, \ldots, x^k)]\, dx^1 \ldots \widehat{dx^\iota} \ldots dx^k$$

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

$$+(-1)^i \int_{[0,1]^k} f(x^1, \ldots, 0, \ldots, x^k) dx^1 \ldots dx^k.$$

Thus

$$\int_{I^k} d\omega = \int_{\partial I^k} \omega.$$

For an arbitrary singular k-cube, chasing through the definitions show that

$$\int_{\partial c} \omega = \int_{\partial I^k} c^*\omega.$$

Therefore

$$\int_c d\omega = \int_{I^k} c^*(d\omega) = \int_{I^k} d(c^*\omega) = \int_{\partial I^k} c^*\omega = \int_{\partial c} \omega\,.$$

Finally, if c is a k-chain $\sum a_i c_i$, we have

$$\int_c d\omega = \sum a_i \int_{c_i} d\omega = \sum a_i \int_{\partial c_i} \omega = \int_{\partial c} \omega\,.$$

Notice that Stokes' theorem not only uses the fundamental theorem of calculus, but actually becomes that theorem when $c = I^1$ and $\omega = f$.

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

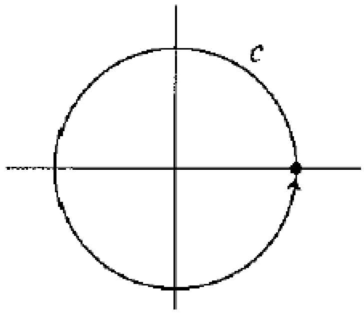


Figure (3.12)

As an application of Stokes' theorem, we show that the curve $c\colon [0,1] \to \mathbf{R}^2 - \{0\}$ define by although closed, is not ∂c^2 for any 2-chain c^2. If we did have $c = \partial c^2$, then we would have

$$\int_c d\theta = \int_{\partial c^2} d\theta = \int_{c^2} d(d\theta) = \int_{c^2} 0 = 0.$$

But a straightforward computation show that

$$\int_c d\theta = \int_c \frac{-y}{x^2+y^2}\, dx + \frac{x}{x^2+y^2}\, dy = 2\pi.$$

Although we used this calculation to show that c is not a boundary, we could just as well have used it to show that $\omega =$"$d\theta$" is not exact. For, if we had $\omega = df$ for some C^∞ function $f\colon \mathbf{R}^2 - \{0\} \to \mathbf{R}$, then we would have

$$2\pi = \int_c \omega = \int_c df = \int_{\partial c} f = \int_0 f = 0.$$

We were previously able to give a simpler argument to show that "$d\theta$" is not exact, but Stokes' theorem is the tool which will enable us to deal with forms on $\mathbf{R}^n - \{0\}$. For example, we will eventually obtain a 2-form ω on $\mathbf{R}^3 - \{0\}$,

$$\omega = \frac{xdy \wedge dz - ydx \wedge dz + zdx \wedge dy}{(x^2+y^2+z^2)^{3/2}}$$

Which is closed but not exact. For the moment we are keeping the origin of ω a secret, but a straightforward calculation shows that $d\omega = 0$. To prove that ω is not exact we will want to integrate it over a 2-chain which "fills up" the 2-sphere

$S^2 \subset \mathbf{R}^3 - \{0\}$. This is possible only when A is orientable; the reason will be clear from the next result, which is basic for our definition. [12]

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

a) It is trivial.
b) It is trivial because the terms appearing in it have been properly defined.
c) It has significant consequences. [13]

Chapter (4)

Stokes' Theorem on Manifolds

Section (4.1): Fields and Forms on Manifolds

In this section, we consider fields and forms on manifolds, the discussion of which requires illustration of manifolds-with-boundary.

Definition (4.1.1):

If U and V are open sets in $\mathbf{R}^n$, a differentiable function $h: U \to V$ with a differentiable inverse $h^{-1}: V \to U$ will be called a diffeomorphism.

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

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

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

In other words, $U \cap M$ is, "up to diffeomorphism," simply $\mathbf{R}^k \times \{0\}$ (see Figure (4.1). The two extreme cases of our definition should be noted: a point in $\mathbf{R}^n$ is a 0-dimensional manifold, and an open subset of $\mathbf{R}^n$ is an n-dimensional manifold.

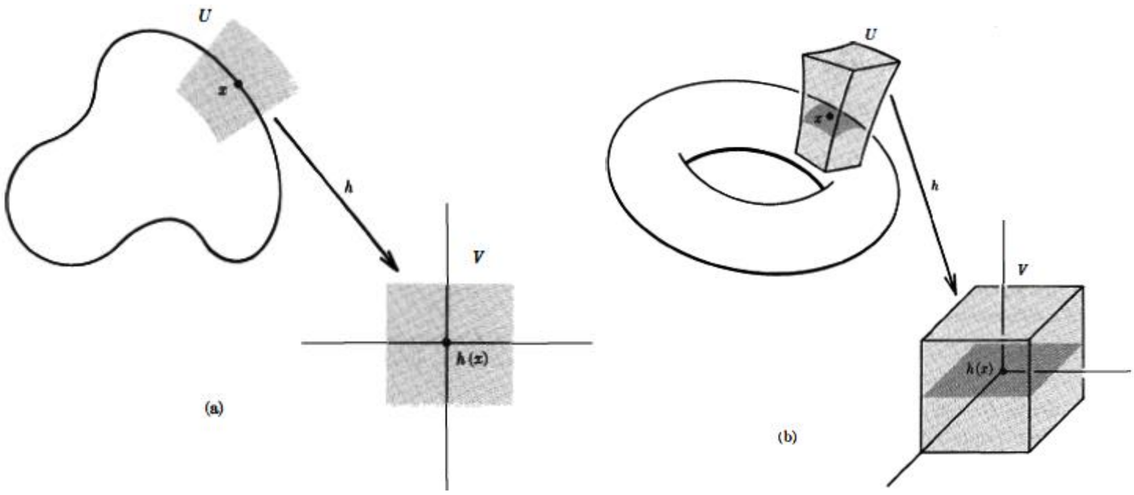


Figure (4.1): A one-dimensional manifold in $\mathbf{R}^2$ and a two-dimensional manifold in $\mathbf{R}^3$

One common example of an n-dimensional manifold is the n-sphere S^n, defined as $\{x \in \mathbf{R}^{n+1}: |x| = 1\}$. If we are unwilling to trouble our self with the details, we may instead use the following theorem, which provides many examples of manifolds (note that $S^n = g^{-1}(0)$, where $g: \mathbf{R}^{n+1} \to \mathbf{R}$ is defined by $g(x) = |x|^2 - 1$).

Theorem (4.1.2):

Let $A \subset \mathbf{R}^n$ be open and let $g: A \to \mathbf{R}^p$ 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 $\mathbf{R}^n$. [13]

Example (4.1.3):

Consider the case $k = 1$. If a is a coordinate patch on M, the condition that h' have rank 1 means merely that $h' \neq 0$. This condition rules out the possibility that M could have "cusps" and "corners." For example, let $h: \mathbf{R} \to \mathbf{R}^2$ be given by the equation $h(t) = (t^3, t^2)$, and let M be the image set of h. Then M has a cusp at the origin. (See Figure (4.2)) Here h is of class C^∞ and h^{-1} is continuous, but h' does not have rank 1 at $t = 0$.

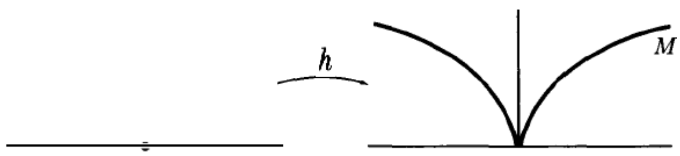


Figure (4.2)

Similarly, let $g: \mathbf{R} \to \mathbf{R}^2$ be given by $g(t) = (t^3, |t^3|)$, and let N be the image set of g. Then N bas a corner at the origin. (See Figure (4.3)) Here g is of class C^2 and g^{-1} is continuous, but g' does not have rank 1 at $t = 0$.

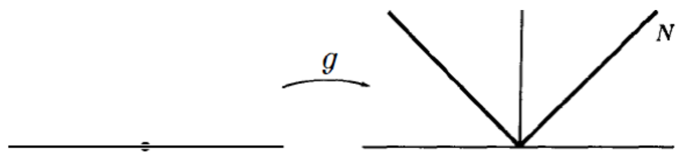


Figure (4.3)

Example (4.1.4):

Consider the case $k = 2$. The condition that $h'(a)$ have rank 2 means that the columns $\partial h/\partial x_1$ and $\partial h/\partial x_2$ of h' are independent at a. Note that $\partial h/\partial x_j$ is the velocity vector of the curve $f(t) = h(a + te_j)$ and is thus tangent to the surface M. Then $\partial h/\partial x_1$ and $\partial h/\partial x_2$ span a 2-dimensional “tangent plane” to M. See Figure (4.4).

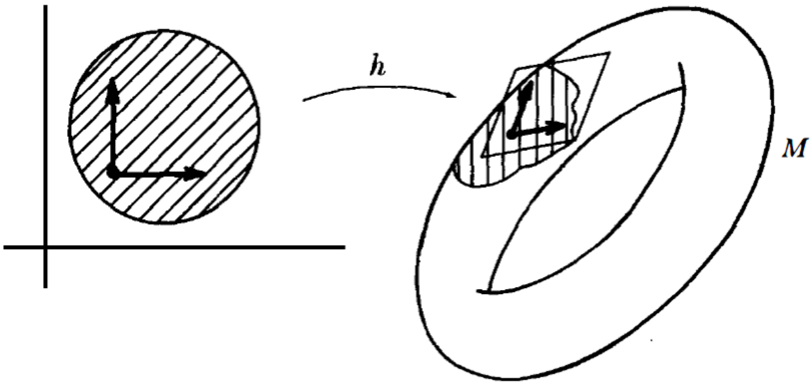


Figure (4.4)

As an example of what can happen when this condition fails, consider the function $h: \mathbf{R}^2 \to \mathbf{R}^3$ given by the equation

$$h(x, y) = (x(x^2 + y^2), y(x^2 + y^2), x^2 + y^2),$$

and let M be the image set of h. Then M fails to have a tangent plane at the origin. See Figure (4.5). The map h is of class C^∞ and h^{-1} is continuous, but h' does not have rank 2 at 0. [10]

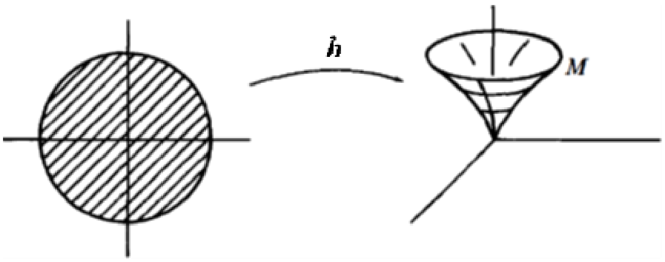


Figure (4.5)

Theorem (4.1.4):

A subset M of $\mathbf{R}^n$ 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 $W \subset \mathbf{R}^k$, and a $1-1$ differentiable function $f: W \to \mathbf{R}^n$ 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 (4.6)).]

Proof:

If M is a k-dimensional manifold in $\mathbf{R}^n$, choose $h: U \to V$ satisfying (M). Let $W = \{a \in \mathbf{R}^k : (a, 0) \in h(M)\}$ and define $f: W \to \mathbf{R}^n$ by $f(a) = h^{-1}(a, 0)$. Clearly $f(W) = M \cap U$ and f^{-1} is continuous. If $H: U \to \mathbf{R}^k$ is $H(z) = \left(h^1(z), \dots, h^k(z)\right)$, then $H\big(f(y)\big) = y$ for all $y \in W$; therefore $H'\big(f(y)\big) \cdot f'(y) = I$ and $f'(y)$ must have rank k.

Suppose, conversely, that $f: W \to \mathbf{R}^n$ satisfies condition (C). Let $x = f(y)$. Assume that the matrix $\left(D_j f^i(y)\right)$, $1 \leq i, j \leq k$ has a non-zero determinant. Define $g: W \times \mathbf{R}^{n-k} \to \mathbf{R}^n$ by $g(a, b) = f(a) + (0, b)$. Then $\det g'(a, b) = \det\left(D_j f^i(a)\right)$, so $\det g'(y, 0) \neq 0$. By Theorem (1.2.11) there is an open set V_1' containing $(y, 0)$ and an open set V_2' containing $g(y, 0) = x$ such that $g: V_1' \to V_2'$ has a differentiable inverse $h: V_2' \to V_1'$. Since f^{-1} is continuous, $\{f(a) : (a, 0) \in V_1'\} = U \cap f(W)$ for some open set U. Let $V_2 = V_2' \cap U$ and $V_1 = g^{-1}(V_2)$. Then $V_2 \cap M$ is exactly $\{f(a) : (a, 0) \in V_1\} = \{g(a, 0) : (a, 0) \in V_1\}$, so

$$
\begin{aligned}
h(V_2 \cap M) = g^{-1}(V_2 \cap M) &= g^{-1}(\{g(a, 0) : (a, 0) \in V_1\}) \\
&= V_1 \cap (\mathbf{R}^k \times \{0\}).
\end{aligned}
$$

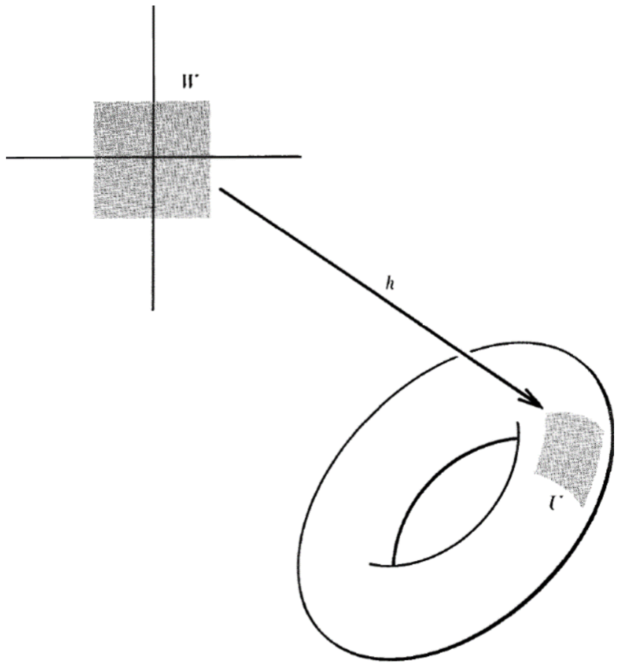


Figure (4.6)

One consequence of the proof of Theorem (4.1.4) should be noted. If $f_1: W_1 \to \mathbf{R}^n$ and $f_2: W_2 \to \mathbf{R}^n$ are two coordinate systems, then

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

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

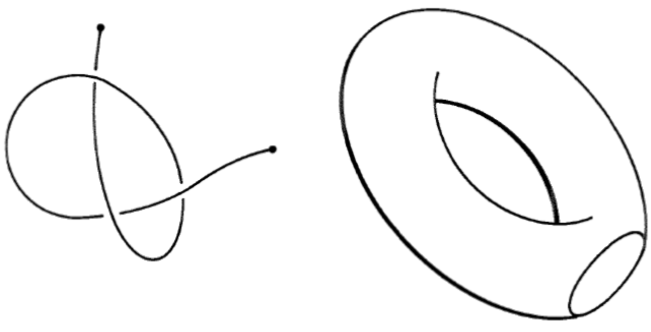
The half-space $\mathbf{H}^k \subset \mathbf{R}^k$ is defined as $\{x \in \mathbf{R}^k : x^k \geq 0\}$. A subset M of $\mathbf{R}^n$ is a k-dimensional manifold-with-boundary (Figure (4.7)) if for every point $x \in M$ either condition (M) or the following condition is satisfied:

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

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

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

It is important to note that conditions (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 $\mathbf{R}^k$, containing $h(x)$, into a subset of $\mathbf{H}^k$ which is not open in $\mathbf{R}^k$. Since $\det(h_2 \circ h_1^{-1}) \neq 0$. The set of all points $x \in M$ for which condition M' is satisfied is called the boundary of M and denoted ∂M.



(a) (b)

Figure (4.7): A one-dimensional and a two-dimensional manifold-with-boundary in $\mathbf{R}^3$.

Now we can describe tangent space and vectors fields on manifolds, and we will be concerned with orientation.

Let M be a k-dimensional manifold in $\mathbf{R}^n$ and let $f: W \to \mathbf{R}^n$ be a coordinate system around $x = f(a)$. Since $f'(a)$ has rank k, the linear transformation $f_*: \mathbf{R}^k{}_a \to \mathbf{R}^n{}_x$ is $1-1$, and $f_*(\mathbf{R}^k{}_a)$ is a k-dimensional subspace of $\mathbf{R}^n{}_x$. If $g: V \to \mathbf{R}^n$ is another coordinate system, with $x = g(b)$, then

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

Thus the k-dimensional subspace $f_*(\mathbf{R}^k{}_a)$ 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 (4.8)). In next section we will use the fact that there is a natural inner product T_x on M_x, induced by that on $\mathbf{R}^k{}_x$: if $v, w \in M_x$ define $T_x(v, w) = \langle v, w \rangle_x$.

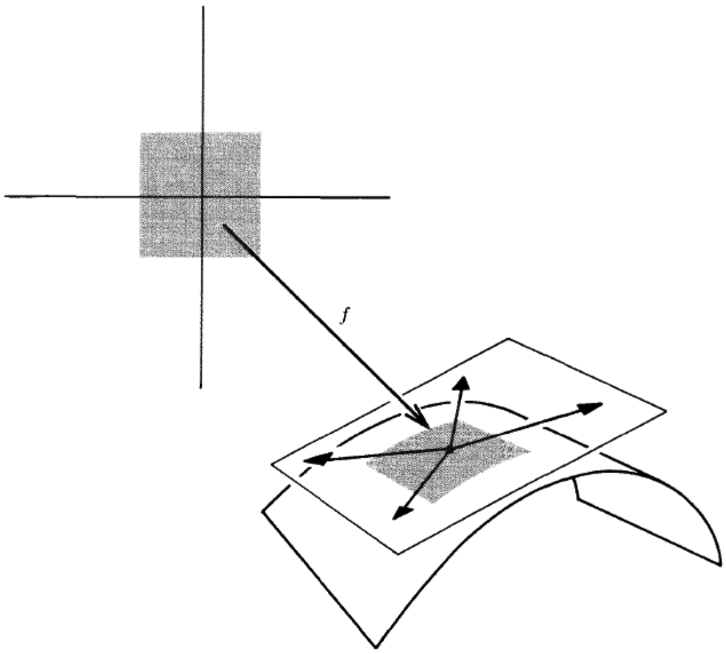


Figure (4.8)

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 $f: W \to \mathbf{R}^n$ is a coordinate system, there is a unique (differentiable) vector field G on W such that $f_*\big(G(a)\big) = F\big(f(a)\big)$ 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 $f_*\big(G(a)\big) = F\big(f(a)\big)$ 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 $g: V \to \mathbf{R}^n$ and $g^*\big(H(b)\big) = F\big(g(b)\big)$

for all $b \in V$, then the component functions of $H(b)$ must equal the component functions of $G\left(f^{-1}(g(b))\right)$, 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 $f: W \to \mathbf{R}^n$ 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 can be written as

$$\omega = \sum_{i_1 < \cdots < i_p} \omega_{i_1,\ldots,i_p}\, dx^{i_1} \wedge \cdots \wedge dx^{i_p}.$$

Here the functions $\omega_{i_1,\ldots,i_p}$ are defined only on M. The definition of $d\omega$ given previously would make no sense here, since $D_j\left(\omega_{i_1,\ldots,i_p}\right)$ has no meaning. Nevertheless, there is a reasonable way of defining $d\omega$.

Theorem (4.1.5):
There is a unique $(p+1)$-form $d\omega$ on M such that for every coordinate system $f: W \to \mathbf{R}^n$ we have

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

Proof:
If $f: W \to \mathbf{R}^n$ is a coordinate system with $x = f(a)$ and $v_1, \ldots, v_{p+1} \in M_x$, there are unique $w_1, \ldots, w_{p+1}$ in $\mathbf{R}^k{}_a$ such that $f^*(w_i) = v_i$. Define $d\omega(x)\big(v_1, \ldots, v_{p+1}\big) = d(f^*\omega)(a)\big(w_1, \ldots, w_{p+1}\big)$. One can check that this definition of $d\omega(x)$ does not depend on the coordinate system f, so that $d\omega$ is well-defined. Moreover, it is clear that $d\omega$ has to be defined this way, so $d\omega$ is unique.

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 (4.9)) provided that for every coordinate system $f: W \to \mathbf{R}^n$ and $a, b \in W$ the relation

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

holds if and only if

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

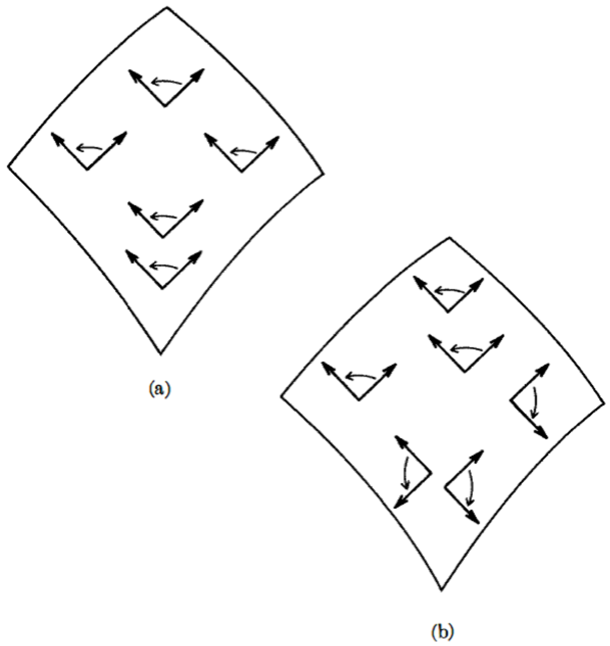


Figure (4.9): (a) consistent and (b) inconsistent choices of orientations

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

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

for one, and hence for every $a \in W$, then f is called orientation-preserving. If f is not orientation-preserving and $T: \mathbf{R}^k \to \mathbf{R}^k$ 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

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

implies that

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

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 (4.10)).

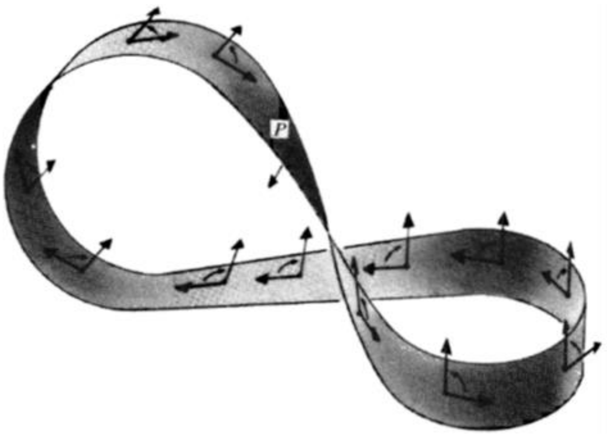


Figure (4.10): The Möbius strip, a non-orientable manifold. A basis begins at P, moves to the right and around, and comes back to P with the wrong orientation

Our definitions of vector fields, forms, and orientations can be made for manifolds-with-boundary also. If M is a k-dimensional 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 (4.11)). If $f: W \to \mathbf{R}^n$ 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 orientations $(\partial \mu)_x$, 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 $\mathbf{H}^k$ with the usual orientation, we find that the induced orientation on $\mathbf{R}^{k-1} = \{x \in \mathbf{H}^k : x^k = 0\}$ is $(-1)^k$ times the usual orientation.

If M is an oriented $(n-1)$-dimensional manifold in $\mathbf{R}^n$, a substitute for outward unit normal vectors can be defined, even though M is not necessarily the boundary of an n-dimensional manifold.

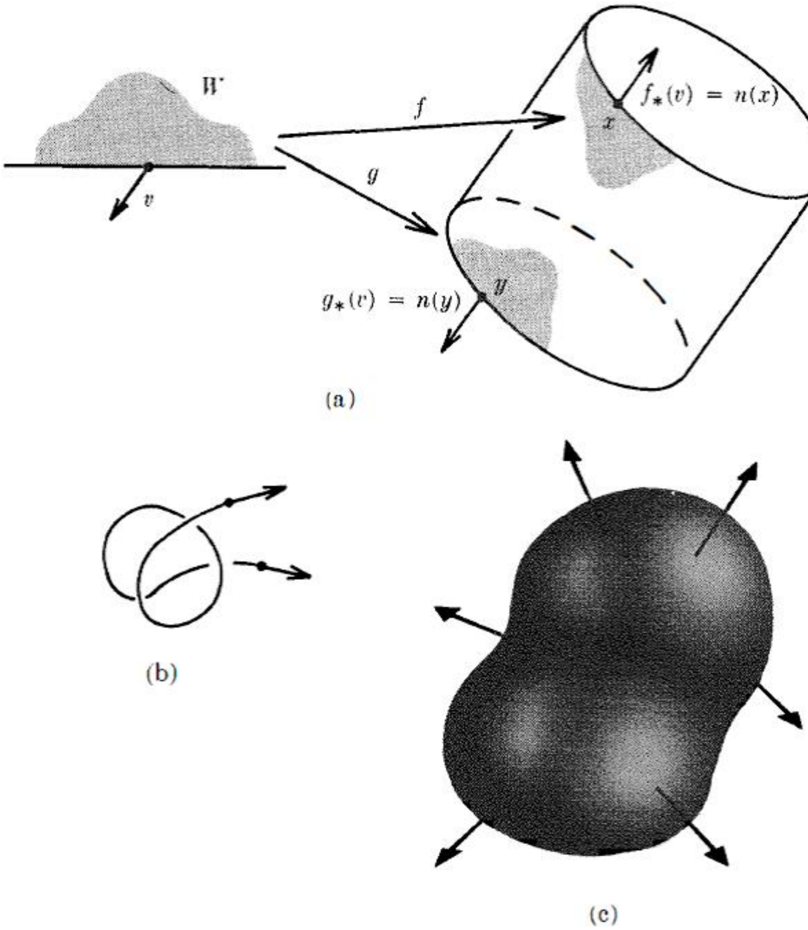


Figure (4.11): Some outward unit normal vectors of manifolds-with-boundary in $\mathbf{R}^3$.

If $[v_1, \ldots, v_{k-1}] = \mu_x$, we choose $n(x)$ in $\mathbf{R}^n{}_x$ so that $n(x)$ is a unit vector perpendicular to M_x and $[n(x), v_1, \ldots, v_{k-1}]$ is the usual orientation of $\mathbf{R}^n{}_x$. We still call $n(x)$ the outward unit normal to M (determined 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 we start to paint it on one side we end up painting it all over); in other words, choosing $n(x)$ 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. [13]

Let M be a k-dimensional manifold with nonempty boundary in $\mathbf{R}^n$ such that M has orientation μ_x. Suppose $x \in \partial M$. There exists a singular k-cube in M $c: I^k \to M$ such that:

c agrees with the orientation of M, $c(I^k)$ intersects ∂M precisely in the set $c(I_{1,1})$ while the rest of $c(I^k)$ lies in the interior of M, and $x = c(a)$ for some point a in the interior of $I_{1,1}$.

(Recall that $I_{1,1}$ is the set of $(u_1, u_2, \dots, u_k) \in I^k$ such that $u_1 = 1$.) The existence of such a cube c is easily seen by starting with a coordinate patch $g: U \to M$ which covers x (where U is open in the half-space, $\mathbf{H}^k$) and constructing an appropriate map $h: I^k \to \mathbf{H}^k$ such that c can be taken to be $g \circ h$. See Figure (4.12).

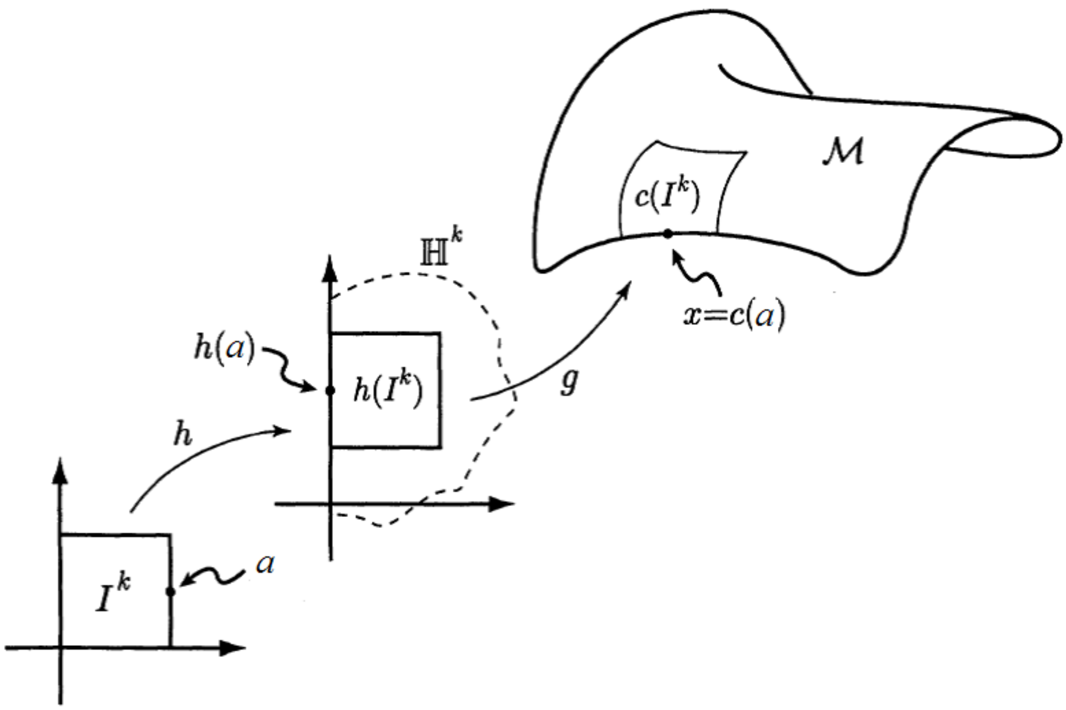


Figure (4.12)

We know the induced orientation on ∂I^k at the point a is $e_2 \wedge e_3 \wedge \cdots \wedge e_k$, so we define the induced orientation on M at x to be

$$\alpha[f_*((e_2)_a), \dots, f_*((e_k)_a)] = \partial \mu_x$$

where α is a positive scalar chosen in such a way as to ensure $\partial \mu_a$ will be a unit vector.

Example (4.1.6):

Let $M = \{(x_1, x_2) \in \mathbf{R}^2 : x_1^2 + x_2^2 \leq 1\}$, a unit disk, and we endow this 2-manifold with the orientation $dx_1 \wedge dx_2$. Then $\partial M = \{(x_1, x_2) \in \mathbf{R}^2 : x_1^2 + x_2^2 = 1\}$. We want to find the induced orientation at a point $x_0 = (x_{01}, x_{02}) \in \partial M$.

There exists θ_0 such that $(x_{01}, x_{02}) = (\cos(\theta_0), \sin(\theta_0))$. There exist $\delta > 0$ and $r_0 \in (0,1)$ such that the map $g(r, \theta) = (r \cos\theta, r \sin\theta)$ is a diffeomorphism on the rectangle $[r_0, 1] \times [\theta_0 - \delta, \theta_0 + \delta]$ which carries the rectangle into M,

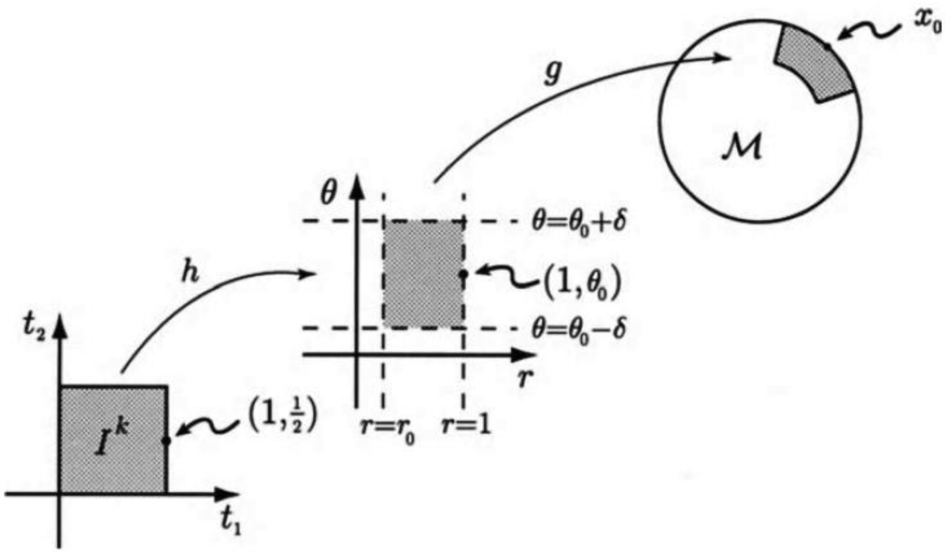


Figure (4.13)

Carries one edge of the rectangle into ∂M, and carries $(1, \theta_0)$ to x_0. The map $h(t_1, t_2) = \big(r_0 + t_1(1 - r_0), \theta_0 + \delta(2t_2 - 1)\big)$ is a diffeomorphism which carries the unit square $[0,1]^2$ onto the rectangle $[r_0, 1] \times \theta_0 - \delta, \theta_0 + \delta]$ and takes $(1, 1/2)$ to the point $(1, \theta_0)$.

The composition of these two maps defines a cube $c: [0,1]^2 \to M$ so that one edge of the unit square goes into ∂M and $(1, 1/2)$ maps to x_0. See Figure (4.13).

It is straightforward to calculate that

$$[c'(t_1, t_2)](e_1 \wedge e_2) = 2\delta\big(r_0 + t_1(1 - r_0)\big)(1 - r_0)\big((e_1 \wedge e_2)\big),$$

and since $2\delta\big(r_0 + t_1(1 - r_0)\big) > 0$, we see that c is orientation-preserving. Now the induced orientation of ∂I^2 at $(1, 1/2)$ is e_2, so we can find the induced orientation of ∂M at x_0 by calculating $[c'(1, 1/2)](e_2) = -2\delta \sin\theta_0 \ e_1 + 2\delta \cos\theta_0 \ e_2$. Hence $\partial \mu_{x_0} = -x_{02} e_1 + x_{01} e_2$. [9]

Section (4.2): Stokes' Theorem

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 $f: W \to \mathbf{R}^n$ 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.

Theorem (4.2.1):

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

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

Proof:

We have

$$\int_{c_1} \omega = \int_{[0,1]^k} c_1{}^*(\omega) = \int_{[0,1]^k} (c_2^{-1} \circ c_1)^* c_2{}^*(\omega).$$

(Here $c_2^{-1} \circ c_1$ is defined only on a subset of $[0,1]^k$ and the second equality depends on the fact that $\omega = 0$ outside of $c_1([0,1]^k) \cap c_2([0,1]^k)$.) It therefore suffices to show that

$$\int_{[0,1]^k} (c_2^{-1} \circ c_1)^* c_2{}^*(\omega) = \int_{[0,1]^k} c_2{}^*(\omega) = \int_{c_2} \omega.$$

If $c_2{}^*(\omega) = f dx^1 \wedge \cdots \wedge dx^k$ and $c_2^{-1} \circ c_1$ is denoted by g, then by Theorem (3.1.25) we have

$$\begin{aligned}(c_2^{-1} \circ c_1)^* c_2{}^*(\omega) &= g^*(f dx^1 \wedge \cdots \wedge dx^k)\\ &= (f \circ g) \cdot \det g' \cdot dx^1 \wedge \cdots \wedge dx^k\\ &= (f \circ g) \cdot |\det g'| \cdot dx^1 \wedge \cdots \wedge dx^k,\end{aligned}$$

since $\det g' = \det(c_2^{-1} \circ c_1)' > 0$. The result now follows from Theorem (2.2.10).

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 (4.2.1) shows $\int_M \omega$ does not depend on the choice of c. Suppose now that ω is an arbitrary k-form on M. There is an open cover $\mathcal{O}$ of M such that for each $U \in \mathcal{O}$ 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 (2.2.5) (this is certainly true if M is compact). An argument similar to that in Theorem (2.2.5) shows that $\int_M \omega$ does not depend on the cover $\mathcal{O}$ 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 $(k-1)$-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. [13]

Theorem (4.2.2): (Stokes' Theorem)

If M is an oriented k-dimensional manifold-with-boundary and ∂M is given the induced orientation, and ω is an $(k-1)$-form on M with compact support, then

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

Proof:

Suppose first that there is an orientation-preserving singular k-cube in $M - \partial M$ such that $\omega = 0$ outside of $c([0,1]^k)$. By Theorem (3.2.3) and the definition of $d\omega$ we have

$$\int_c d\omega = \int_{[0,1]^k} c^*(d\omega) = \int_{[0,1]^k} d(c^*\omega) = \int_{\partial I^k} c^*\omega = \int_{\partial c} \omega.$$

Then

$$\int_M d\omega = \int_c d\omega = \int_{\partial c} \omega = 0,$$

since $\omega = 0$ on ∂c. On the other hand, $\int_{\partial M} \omega = 0$ since $\omega = 0$ on ∂M.

Suppose next that there is an orientation-preserving singular k-cube in M such that $c_{(k,0)}$ is the only face in ∂M, and $\omega = 0$ outside of $c([0,1]^k)$. Then

$$\int_M d\omega = \int_c d\omega = \int_{\partial c} \omega = \int_{\partial M} \omega \,.$$

In general, there is an open cover $\mathcal{O}$ of M and a partition of unity Φ for M subordinate to $\mathcal{O}$ such that for each $\varphi \in \Phi$ the form $\varphi \cdot \omega$ is of one of the two sorts already considered. We have

$$0 = d(1) = d\left(\sum_{\varphi\in\Phi} \varphi\right) = \sum_{\varphi\in\Phi} d\varphi,$$

so

$$\sum_{\varphi\in\Phi} d\varphi \wedge \omega = 0.$$

Since ω has compact support, this is really a finite sum, and we conclude that

$$\sum_{\varphi\in\Phi}\int_M d\varphi \wedge \omega = 0.$$

Therefore

$$\begin{aligned}\int_M d\omega &= \sum_{\varphi\in\Phi}\int_M \varphi \cdot d\omega = \sum_{\varphi\in\Phi}\int_M d\varphi \wedge \omega + \varphi \cdot d\omega \\ &= \sum_{\varphi\in\Phi}\int_M d(\varphi \cdot \omega) = \sum_{\varphi\in\Phi}\int_{\partial M} \varphi \cdot \omega \\ &= \int_{\partial M} \omega. \ [12]\end{aligned}$$

Let M be a k-dimensional manifold (or manifold-with-boundary) in $\mathbf{R}^n$, 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 $\mathbf{R}^3$. 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

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

Thus

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

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.

Theorem (4.2.3):

Let M be an oriented two-dimensional manifold (or manifold-with-boundary) in $\mathbf{R}^3$ 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) $$n^3 dA = dx \wedge dy.$$

Proof:

Equation (1) is equivalent to the equation

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

This is seen by expanding the determinant by minors along the bottom row. To prove the other equations, let $z \in \mathbf{R}^3{}_x$. Since $v \times w = \alpha n(x)$ for some $\alpha \in \mathbf{R}$, we have

$$\langle z, n(x)\rangle \cdot \langle v \times w, n(x)\rangle = \langle z, n(x)\rangle\alpha = \langle z, \alpha n(x)\rangle = \langle z, v \times w\rangle.$$

Choosing $z = e_1, e_2$, and e_3 we obtain (2), (3), and (4).

A word of caution: if $\omega \in \Lambda^2(\mathbf{R}^3{}_a)$ 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 $c: [0,1] \to \mathbf{R}^n$ 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 $c: [0,1]^2 \to \mathbf{R}^n$, 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 (4.12). Many definitions of surface area have been proposed, disagreeing with each other, but all agreeing with our definition for differentiable surfaces.

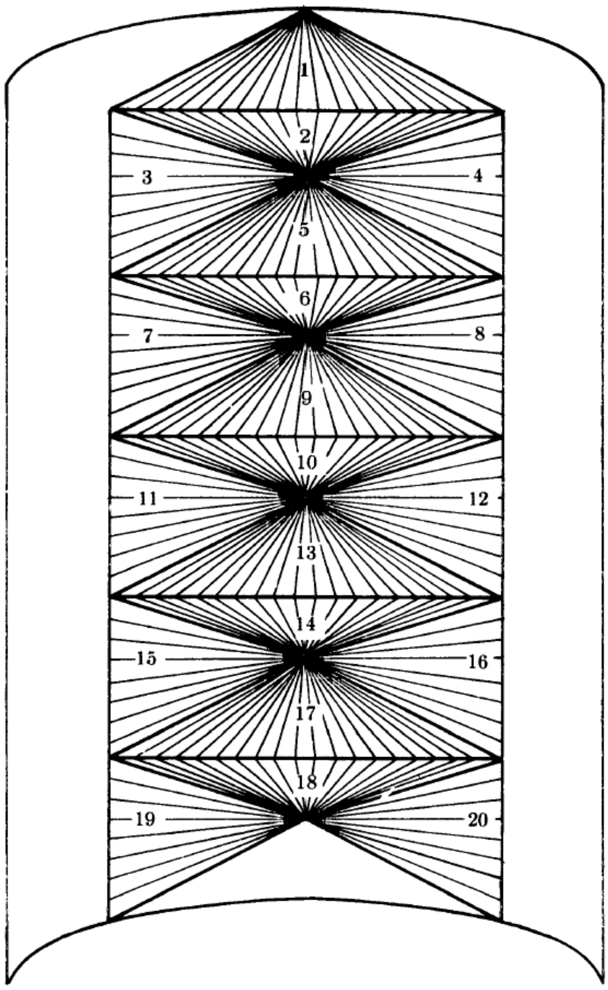


Figure (4.12): A surface containing 20 triangles inscribed in a portion of a cylinder. If the number of triangles is increased sufficiently, by making the bases of triangles 3, 4, 7, 8, etc., sufficiently small, the total area of the inscribed surface can be made as large as desired.

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

Theorem (4.2.4): (Green's Theorem)

Let $M \subset \mathbf{R}^2$ be a compact two-dimensional manifold-with-boundary. Suppose that $\alpha, \beta : M \to \mathbf{R}$ are differentiable. Then

$$\int_{\partial M} \alpha dx + \beta dy = \int_{M} (D_1\beta - D_2\alpha)\, dx \wedge dy = \iint_{M} \left(\frac{\partial \beta}{\partial x} - \frac{\partial \alpha}{\partial y}\right) dx\, dy.$$

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

Proof:

This is a very special case of Theorem (4.2.2), since $d(\alpha\, dx + \beta\, dy) = (D_1\beta - D_2\alpha)dx \wedge dy$.

Theorem (4.2.5): (Divergence Theorem)

Let $M \subset \mathbf{R}^3$ 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

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

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

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

Proof:

Define ω on M by $\omega = F^1\, dy \wedge dz + F^2\, dz \wedge dx + F^3 dx \wedge dy$. Then $d\omega = \operatorname{div} F\, dV$. According to Theorem (4.2.3), on ∂M we have

$$n^1 dA = dy \wedge dz,$$
$$n^2 dA = dz \wedge dx,$$
$$n^3 dA = dx \wedge dy.$$

Therefore on ∂M we have

$$\begin{aligned}\langle F, n\rangle\, dA &= F^1 n^1\, dA + F^2 n^2\, dA + F^3 n^3\, dA \\ &= F^1\, dy \wedge dz + F^2\, dz \wedge dx + F^3\, dx \wedge dy \\ &= \omega.\end{aligned}$$

Thus, by Theorem (4.2.2) we have

$$\int_M \operatorname{div} F\, dV = \int_M d\omega = \int_{\partial M} \omega = \int_{\partial M} \langle F, n\rangle\, dA.$$

Theorem (4.2.6): (Stokes' Theorem)

Let $M \subset \mathbf{R}^3$ 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

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

Proof:

Define ω on M by $\omega = F^1 \, dx + F^2 \, dy + F^3 \, dz$. Since $\nabla \times F$ has components $D_2F^3 - D_3F^2, D_3F^1 - D_1F^3,\ D_1F^2 - D_2F^1$, it follows, as in the proof of Theorem (4.2.5), that on M we have

$$\langle (\nabla \times F), n \rangle \, dA = (D_2F^3 - D_3F^2) dy \wedge dz + (D_3F^1 - D_1F^3) dz \wedge dx$$
$$+ (D_1F^2 - D_2F^1) dx \wedge dy$$
$$= d\omega.$$

On the other hand, since $ds(T) = 1$, on ∂M we have

$$T^1 \, ds = dx,$$
$$T^2 \, ds = dy,$$
$$T^3 \, ds = dz.$$

(These equations may be checked by applying both sides to T_x, for $x \in \partial M$, since T_x is a basis for $(\partial M)_x$.)

Therefore on ∂M we have

$$\begin{aligned} \langle F, T \rangle ds &= F^1T^1 ds + F^2T^2 ds + F^3T^3 ds \\ &= F^1 \, dx + F^2 \, dy + F^3 \, dz \\ &= \omega. \end{aligned}$$

Thus, by Theorem (4.2.2), we have

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

Theorems (4.2.5) and (4.2.4) are the basis for the names $\operatorname{div} F$ and $\operatorname{curl} F$. 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 $\operatorname{div} F = 0$ expresses the fact that the fluid is incompressible. If M is a disc, $\int_{\partial M} \langle F, T \rangle \, ds$ 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.

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 $\mathbf{R}^n$ which are, up to diffeomorphism, locally a portion of $\mathbf{R}^k$ which is bounded by pieces of $(k-1)$-planes. [13]

Section (4.3): Maxwell's Equations

Maxwell's equations link together three vector fields and a real-valued function. Let

$$E = E(x, y, z, t) = \big(E_1(x, y, z, t), E_2(x, y, z, t), E_3(x, y, z, t)\big)$$

and

$$B = B(x, y, z, t) = \big(B_1(x, y, z, t), B_2(x, y, z, t), B_3(x, y, z, t)\big)$$

be two vector fields with spacial coordinates (x, y, z) and time coordinate t. Here E represents the electric field while B represents the magnetic field. The third vector field is

$$j(x, y, z, t) = j_1(x, y, z, t),\ j_2(x, y, z, t), j_3(x, y, z, t),$$

which represents the current (the direction and the magnitude of the flow of electric charge). Finally, let

$$\rho(x, y, z, t)$$

be a function representing the charge density. Let c be a constant. (Here c is the speed of light in a vacuum.) Then these three vector fields and this function satisfy Maxwell's equations if

$$\operatorname{div}(E) = \rho$$

$$\operatorname{curl}(E) = -\frac{\partial B}{\partial t}$$

$$\operatorname{div}(B) = 0$$

$$c^2 \operatorname{curl}(B) = j + \frac{\partial E}{\partial t}.$$

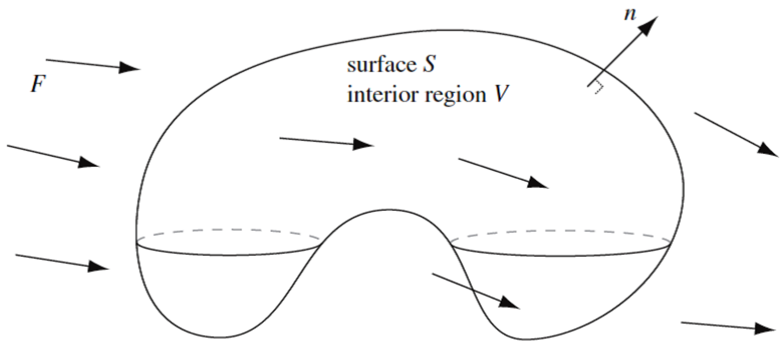


Figure (4.13)

We can reinterpret these equations in terms of integrals via various Stokes-type theorems. For example, if V is a compact region in space with smooth boundary surface S, as in Figure (4.13), then for any vector field F we know from the Divergence Theorem that

$$\int\int_S F \cdot n \, dA = \int\int\int_V \operatorname{div}(F) \, dx \, dy \, dz,$$

where n is the unit outward normal of the surface S.

In words, this theorem says that the divergence of a vector field measures how much of the field is flowing out of a region.

Then the first of Maxwell's equations can be restated as

$$\int\int_S E \cdot n \, dA = \int\int\int \operatorname{div}(E) \, dx \, dy \, dz$$

$$= \int\int\int \rho(x, y, z, t) \, dx \, dy \, dz$$

$=$ total charge inside the region V.

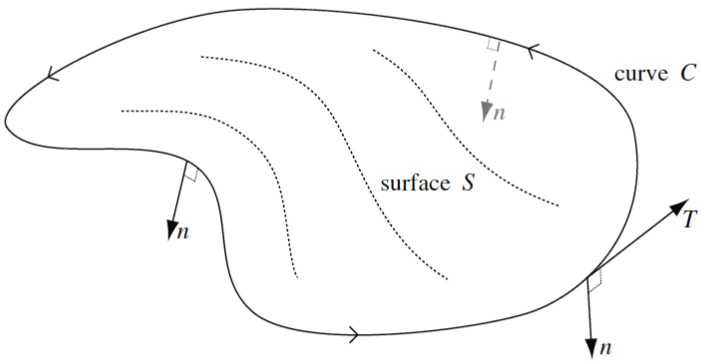


Figure (4.14)

Likewise, the third of Maxwell's equations is:

$$\int\int_S B \cdot n \, dA = \int\int\int \mathrm{div}(B)\, dx\, dy\, dz$$

$$= \int\int\int 0 \, dx\, dy\, dz$$

$$= 0$$

= There is no magnetic charge inside the region V.

This is frequently stated as "There are no magnetic monopoles," meaning there is no real physical notion of magnetic density.

The second and fourth of Maxwell's equations have similar integral interpretations. Let C be a smooth curve in space that is the boundary of a smooth surface S, as in Figure (4.14). Let T be a unit tangent vector of C. Choose a normal vector field n for S so that the cross product $T \times n$ points into the surface S.

Then the classical Stokes' theorem states that for any vector field F, we have

$$\int_C F \cdot T \, ds = \int\int_S \mathrm{curl}(F) \cdot n \, dA\,.$$

This justifies the intuition that the curl of a vector field measures how much the vector field F wants to twirl.

Then the second of Maxwell's equations is equivalent to

$$\int_C E \cdot T \, ds = -\int\int_S \frac{\partial B}{\partial t} \cdot n \, dA$$

Thus the magnetic field B is changing in time if and only if the electric field E is curling.

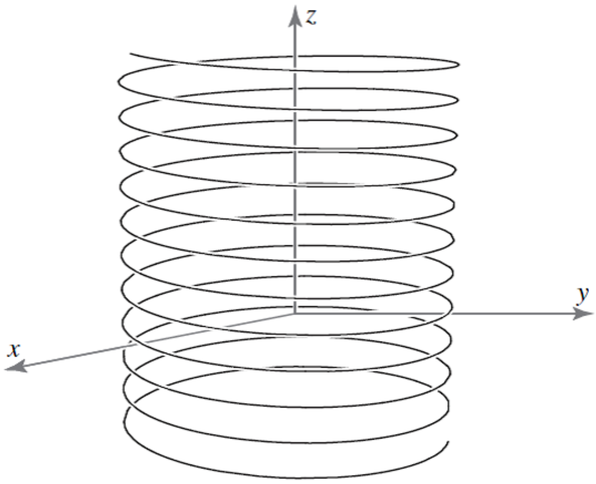


Figure (4.15)

This is the mathematics underlying how to create current in a wire by moving a magnet. Consider a coil of wire, centered along the z-axis (i.e., along the vector $k = (0,0,1)$).

The wire is coiled (almost) in the xy-plane. Move a magnet through the middle of this coil. This means that the magnetic field B is changing in time in the direction k. Thanks to Maxwell, this means that the curl of the electric field E will be non-zero and will point in the direction k. But this means that the actual vector field E will be "twirling" in the xy-plane, making the electrons in the coil move, creating a current.

This is in essence how a hydroelectric dam works. Water from a river is used to move a magnet through a coil of wire, creating a current and eventually lighting some light bulb in a city far away.

The fourth Maxwell equation gives

$$c^2 \int_C B \cdot T \, ds = \int\int_S \left(j + \frac{\partial B}{\partial t}\right) \cdot n \, dA.$$

Here current and a changing electric field are linked to the curl of the magnetic field.

In the following, we will review and list some of the standard notations that people use. The symbol ∇ is pronounced "nabla" (sometimes ∇ is called "del"). Let

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$$

$$= i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z}$$

where $i = (1,0,0)$, $j = (0,1,0)$, and $k = (0,0,1)$. Then for any function $f(x,y,z)$, we set the gradient to be

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$$

$$= i\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y} + k\frac{\partial f}{\partial z}.$$

For a vector field

$$F = F(x,y,z) = \big(F_1(x,y,z), F_2(x,y,z), F_3(x,y,z)\big)$$

$$= (F_1, F_2,\ F_3)$$

$$= \ F_1 \cdot i + F_2 \cdot j + F_3 \cdot k,$$

define the divergence to be:

$$\nabla \cdot F = \mathrm{div}(F)$$

$$= \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}.$$

The curl of a vector field in this notation is

$$\nabla \times F = \mathrm{curl}(F)$$

$$= \det\begin{pmatrix} i & j & k \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ F_1 & F_2 & F_3 \end{pmatrix}$$

$$= \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \left(\frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial z}\right), \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right).$$

Using the nabla notation, Maxwell's equations have the form

$$\nabla \cdot E = \rho$$
$$\nabla \times E = -\frac{\partial B}{\partial t}$$
$$\nabla \cdot B = 0$$
$$c^2\, \nabla \times B = j + \frac{\partial E}{\partial t}$$

Though these look like four equations, when written out they actually form eight equations. [4]

We can make all of this look much simpler by making the following definitions. First, we define a 2-form called the Faraday, which simultaneously describes both the electric and magnetic fields:

$$\begin{aligned} \mathrm{F} = {} & E_x\, dx \wedge dt + E_y\, dy \wedge dt + E_z\, dz \wedge dt \\ & + B_x\, dy \wedge dz + B_y\, dz \wedge dx + B_z\, dx \wedge dy. \end{aligned}$$

A direct calculation shows the first group of Maxwell's equations is equivalent to the condition

$$dF = 0.$$

The second group of equations is expressed in an analogous way, given a few additional algebraic developments. On forms of degree 2 we introduce the operator $*$ by the relations

$$\begin{aligned} {}^*(dx \wedge dt) &= dy \wedge dz & \qquad {}^*(dy \wedge dz) &= -dx \wedge dt \\ {}^*(dy \wedge dt) &= dz \wedge dx & \qquad {}^*(dz \wedge dx) &= -dy \wedge dt \\ {}^*(dz \wedge dt) &= dx \wedge dy & \qquad {}^*(dx \wedge dy) &= -dz \wedge dt \end{aligned}$$

Next, we define the "dual" 2-form, called the Maxwell:

$$\begin{aligned} {}^*\mathrm{F} = {} & E_x dy \wedge dz + E_y dz \wedge dx + E_z\, dx \wedge dy \\ & + B_x\, dt \wedge dx + B_y\, dt \wedge dy + B_z\, dt \wedge dz. \end{aligned}$$

The second group of Maxwell's equations can then be written

$$d({}^{*}\mathrm{F}) = 0$$

in a vacuum, and

$$d({}^{*}\mathrm{F}) = c^2\ {}^{*}\mathrm{J}$$

in the presence of electric charges. We denote by J the current, defined by

$$J = J_x\, dx + J_y\, dy + J_z\, dz + \rho\, dt.$$

We observe that since $d^2 = 0$, the 3-form ${}^{*}J$ is closed. This property may be written

$$\mathrm{div}\, J + \frac{\partial \rho}{\partial t} = 0,$$

which is a statement of the conservation of electric charge.

Maxwell's four vector equations now reduce to

$$d\mathrm{F} = 0,$$

$$d^{*}\mathrm{F} = c^{2\,*}\mathrm{J}$$

The differential form version of Maxwell's Equation has a huge advantage over the vector formulation: It is coordinate-free! A 2-form such as F is an operator that "eats" pairs of vectors and "spits out" numbers. The way it acts is completely geometric; that is, it can be defined without any reference to the coordinate system (t, x, y, z). This is especially poignant when one realizes that Maxwell's Equations are laws of nature that should not depend on a manmade construction such as coordinates. [6, 1]

References

1. Bachman, D. *A Geometric Approach to Differential Forms.* New York: Birkhäuser, 2012.
2. Boothby, W. M. *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Amsterdam: Academic Press, 2002.
3. Do Carmo, M. P. *Differential Forms and Applications*. Berlin: Springer-Verlag, 1994.
4. Garrity, T. A. *Electricity and Magnetism for Mathematicians: A Guided Path from Maxwell's Equations to Yang-Mills*. New York: Cambridge University Press, 2015.
5. Godinho, L. and Natário, J. *An Introduction to Riemannian Geometry: With Applications to Mechanics and Relativity*. Switzerland: Springer International Publishing, 2014.
6. Lafontaine, J. *An Introduction to Differential Manifolds*. Switzerland: Springer International Publishing, 2015.
7. Lang, S. *Introduction to Differentiable Manifolds*. New York: Springer-Verlag, 2002.
8. Lee, J. M. *Introduction to Smooth Manifolds*. New York: Springer, 2002.
9. Mikusinski, P. and Taylor M. D. *An Introduction to Multivariable Analysis from Vector to Manifold*. Boston: Birkhäuser, 2002.
10. Munkres, J. R. *Analysis on Manifolds*. Redwood City: Addison-Wesley Publishing Company, 1991.
11. Rudin, W. *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1976.
12. Spivak, M. *A Comprehensive Introduction to Differential Geometry*, Volume 1. Houston: Publish or Perish, 1999.
13. Spivak, M. *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. New York: Addison-Wesley, 1965.