integrals

chapter 2

2.1

Non-Constant Forms

In Chapter 1 a constant force field is described by a 1-form

$$A\,dx + B\,dy + C\,dz,$$

where

A = work required for unit displacement in x-direction
B = work required for unit displacement in y-direction
C = work required for unit displacement in z-direction.

In a force field in which the force depends on the location (x, y, z) the quantities A, B, C depend on x, y, z, that is, $A = A(x, y, z)$, $B = B(x, y, z)$, $C = C(x, y, z)$. The expression

$$(1) \quad A(x, y, z)\,dx + B(x, y, z)\,dy + C(x, y, z)\,dz$$

can then be regarded as assigning to each point (x, y, z) of space the 1-form which describes the force field at that point.

Similarly, a non-constant flow is described by an expression of the form

$$(2) \quad A(x, y, z)\,dy\,dz + B(x, y, z)\,dz\,dx + C(x, y, z)\,dx\,dy$$

H.M. Edwards, *Advanced Calculus: A Differential Forms Approach*, Modern Birkhäuser Classics,
DOI 10.1007/978-0-8176-8412-9_2, © Harold M. Edwards 2014

assigning to each point (x, y, z) of space the 2-form which describes the flow at that point.

Henceforth an expression such as (1) will be called a *1-form on xyz-space*, and what was called a 1-form in Chapter 1 will be called a *constant* 1-form (a 1-form in which the functions A, B, C are constant). Similarly, a *2-form on xyz-space* will mean an expression of the form (2), a *1-form on the xy-plane* will mean an expression of the form

$$A(x, y)\,dx + B(x, y)\,dy,$$

a *3-form on xyz-space* will mean an expression of the form

$$A(x, y, z)\,dx\,dy\,dz,$$

and so forth. If there is danger of confusion the term 'variable' will be used in parentheses—e.g. (variable) 1-form, (variable) 2-form—to distinguish 1-forms from constant 1-forms, but henceforth '1-form' will *always* mean '(variable) 1-form'.*

**The mathematical terminology is unfortunately ambiguous, the term 'form' referring both to (variable) forms and to constant forms. [The terminology of physics makes a very useful distinction between vectors and vector fields which is exactly the distinction between constant forms and forms. A (variable) form is a certain kind of tensor field (namely a field of alternating covariant tensors) and physicists may prefer to think of forms as alternating covariant tensor fields.] Since constant forms are the exception rather than the rule, it is only reasonable to use the shortest term possible ('form') for the idea which occurs most frequently and to use the longer term ('constant form') for the exceptional cases.*

Exercises

1 *The central force field.* Newton's law of gravitational attraction states that the force exerted by a massive body (the sun) fixed at the origin $(0, 0, 0)$ on a particle in space is directed toward the sun and has magnitude proportional to the inverse square of the distance to the sun. Show that this means that the force field is described by the 1-form

$$\frac{kx}{r^3}\,dx + \frac{ky}{r^3}\,dy + \frac{kz}{r^3}\,dz$$

where k is a positive constant and $r = r(x, y, z) = \sqrt{x^2 + y^2 + z^2}$; that is, where

$$A(x, y, z) = kx/(x^2 + y^2 + z^2)^{3/2}, \text{ etc.}$$

[Show that for each fixed point (x, y, z) this 1-form represents a constant force which has the right direction (see Exercise 6, §1.1). Then use the fact that the magnitude of a force is measured by the amount of work required per unit displacement in the direction opposing the force.]

2 *Flow from a source (planar flow).* Find the 1-form describing a planar flow from a source at $(0, 0)$, assuming the flow is outward at all points (x, y) and has magnitude inversely proportional to the radius $r = r(x, y) = \sqrt{x^2 + y^2}$ (so that the flow is radially symmetric and the total flow across a circle of radius r about the origin is independent of r; that is, so there is no source between two circles of different radius). The magnitude of a flow is measured by the rate of flow across a line perpendicular to the direction of flow. Sketch the flow vectors.

3 *Flow from a source (spatial flow).* Find the 2-form describing flow in space from a source at (0, 0, 0), again assuming that the flow is *outward* at all points with a magnitude depending only on r, and assuming that there are no sources between spheres about the origin. [The surface area of the sphere of radius r is $4\pi r^2$.]

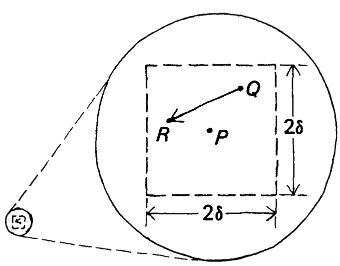
4 *Linear flows.* A linear flow is described by a (variable) 0-form assigning numbers to oriented points. Find the 0-form describing flow from a source at 0 on the line. How would a 0-form be described in general? What is a constant 0-form?

2.2 *Integration*

If the force field is not constant, then finding the amount of work required for a given displacement requires a process of *integration*. The essential idea is that the 1-form

(1) $A(x, y, z)\, dx + B(x, y, z)\, dy + C(x, y, z)\, dz$

which describes the force field gives the approximate amount of work required for small displacements. The amount of work required for a displacement which is not small can then be described as a limit of sums of values of (1).



It will be assumed that *the force field depends continuously on* (x, y, z); that is, it will be assumed that for any given point $P = (\bar{x}, \bar{y}, \bar{z})$, the values of the functions A, B, C at points near P differ only slightly from their values at P. This means that throughout a small neighborhood of P, say the cube $\{|x - \bar{x}| < \delta, |y - \bar{y}| < \delta, |z - \bar{z}| < \delta\}$, the force field (1) is practically equal to the constant force field

$$A(\bar{x}, \bar{y}, \bar{z})\, dx + B(\bar{x}, \bar{y}, \bar{z})\, dy + C(\bar{x}, \bar{y}, \bar{z})\, dz$$

so that the work required for a displacement QR inside this neighborhood is practically equal to the value of this constant 1-form on the oriented line segment QR.

This relationship between the 1-form (1) and work is expressed by writing

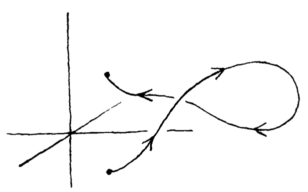
(2) *work required for small displacements* $\sim A(x, y, z)\, dx + B(x, y, z)\, dy + C(x, y, z)\, dz.$

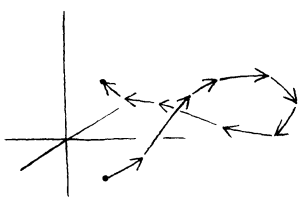
[Read $\sim$ as 'is approximately equal to'.] Thus, if the points $P = (\bar{x}, \bar{y}, \bar{z})$, $Q = (x_1, y_1, z_1)$, $R = (x_2, y_2, z_2)$ are close together, then the amount of work required to go from Q to R is approximately equal to

$$A(\bar{x}, \bar{y}, \bar{z})(x_2 - x_1) + B(\bar{x}, \bar{y}, \bar{z})(y_2 - y_1) + C(\bar{x}, \bar{y}, \bar{z})(z_2 - z_1),$$

and the closer together they are, the better the approximation.*

It is precisely in this context, of course, that dx, dy, dz are thought of as being 'infinitesimals'. The point of view taken here, however, is that dx, dy, dz are functions assigning numbers to directed line segments. What is 'infinitesimal', then, is the line segment on which they are evaluated. Instead of saying that (2) holds for small line segments, with the approximation improving for shorter line segments, it is often said simply that work required for 'infinitesimal' displacements = $A(x, y, z)\,dx + B(x, y, z)\,dy + C(x, y, z)\,dz$.





If S is any oriented curve, then an approximation to the amount of work required for the displacement S is found as follows: Approximate S by an oriented polygonal curve consisting of short straight-line displacements. The approximate amount of work required for each of these is found from (2), and the amount required for S is approximately equal to the sum of these values. The number found in this way is called an approximating sum; there are two approximations involved in the process: first, the approximation of the curve S by a polygonal curve, and second, the approximation of the amount of work required for each segment of the polygonal curve by (2). Since both approximations can be improved by taking the polygonal approximation to S to consist of more and shorter segments, it would be expected that the approximating sum could be made arbitrarily close to the true value by refining the approximation in this way. That is, the amount of work required for the displacement S should be equal to the *limiting value* of the approximating sums as the approximating curve is taken to consist of more and shorter segments fitting the curve S more and more closely. This limiting value (if it exists) is called the *integral* of the 1-form (1) over the curve S and is denoted

$$\int_S (A(x, y, z)\,dx + B(x, y, z)\,dy + C(x, y, z)\,dz)$$

or simply

$$\int_S (A\,dx + B\,dy + C\,dz),$$

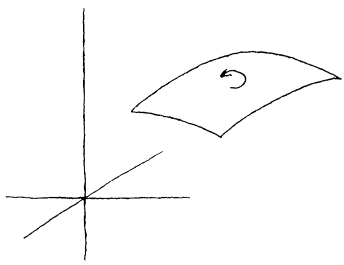
where it is understood that A, B, C are functions of x, y, z.

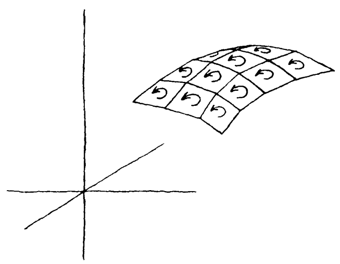
Similarly, let a flow in space be described by a (variable) 2-form $A\,dy\,dz + B\,dz\,dx + C\,dx\,dy$. Then, again assuming that A, B, C are continuous, the flow across a small polygon near $P = (\bar{x}, \bar{y}, \bar{z})$ is approximately equal to the value on this polygon of the constant 2-form

$$A(\bar{x}, \bar{y}, \bar{z})\, dy\, dz + B(\bar{x}, \bar{y}, \bar{z})\, dz\, dx + C(\bar{x}, \bar{y}, \bar{z})\, dx\, dy.$$

This is summarized by saying that

(3) *flow across small polygons* $\sim A(x, y, z)\, dy\, dz + B(x, y, z)\, dz\, dx + C(x, y, z)\, dx\, dy.$





Then for an arbitrary oriented surface S the flow across S is equal to the limit of approximating sums obtained by constructing a polygonal approximation to S in which all polygons are small, using (3) to find the approximate rate of flow across each polygon, and adding. The limit of the approximating sums (if it exists) is called the integral of the 2-form over S and is denoted

$$\int_S (A\, dy\, dz + B\, dz\, dx + C\, dx\, dy).$$

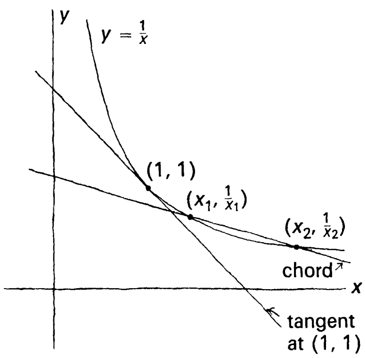
In general, an integral is formed from an *integrand*, which is a 1-form, 2-form, or 3-form, and a *domain of integration* which is, respectively, an oriented curve, a surface, or a solid. The integral is defined as the limit of approximating sums, and an approximating sum is formed by taking a finely divided polygonal approximation to the domain of integration, 'evaluating' the integrand on each small oriented polygon, and adding. The integrand is 'evaluated' on a small oriented polygon by choosing a point P in the vicinity of the polygon, by evaluating the functions A, B, etc. at P to obtain a constant form, and by evaluating the constant form on the polygon in the usual way (as in Chapter 1).

At this point two questions arise: How can this definition of 'integral' be made precise? How can integrals be evaluated in specific cases?

It is difficult to decide which of these questions should be considered first. On the one hand, it is hard to comprehend a complicated abstraction such as 'integral' without concrete numerical examples; but, on the other hand, it is hard to understand the numerical evaluation of an integral without having a precise definition of what the integral is. Yet, to consider both questions at the same time would confuse the distinction between the *definition* of integrals (as limits of sums) and the *method of evaluating* integrals (using the Fundamental Theorem of Calculus). This confusion is one of the greatest obstacles to understanding calculus and should be avoided at all costs. Therefore, all consideration of the evaluation of integrals is postponed to Chapter 3, and the remainder of this chapter is devoted solely to the question of the definition and elementary properties of integrals.

Exercises

1 Let $A\,dx + B\,dy + C\,dz = (1/r^3)[x\,dx + y\,dy + z\,dz]$ be the central force field with $k = 1$ (see Exercise 1, §2.1), and let S be the line segment from $(1, 0, 0)$ to $(2, 0, 0)$. Find the approximating sum to $\int_S (A\,dx + B\,dy + C\,dz)$ formed by dividing S into 10 segments of equal length and using the midpoint of each interval to evaluate $A\,dx + B\,dy + C\,dz$. [Express the answer as a number times a sum of reciprocals of integers; actual numerical evaluation of the sum is difficult.] Call this number $\sum(10)$. Similarly, let $\sum(n)$ be the approximating sum formed by dividing S into n equal segments (n = positive integer), and evaluating at midpoints. Express $\sum(n)$ as a number times a sum of reciprocals of integers. Suppose that a computing machine has been programmed to compute $\sum(n)$ for any n, the value being rounded to two decimal places. Find an upper bound for the magnitude of the difference $|\sum(10) - \sum(20)|$. [Each term of $\sum(10)$ corresponds to two terms of $\sum(20)$. To get an upper bound on the magnitude of the difference $|(1/x_1^2) - (1/x_2^2)|$ it suffices to observe that the slope of the chord of the graph of $1/x^2$ from $(x_1, (1/x_1^2))$ to $(x_2, (1/x_2^2))$ $(1 \le x_1 \le x_2)$ is greater (both are negative) than the slope of the tangent at $(1, 1)$ which gives



$$\frac{1}{x_1^2} - \frac{1}{x_2^2} \le 2(x_2 - x_1).]$$

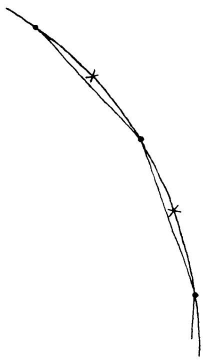
If $\sum(10)$ and $\sum(20)$ are both rounded to two decimal places, how great can the difference of the resulting numbers be? Find an integer N such that $|\sum(N) - \sum(mN)| < .005$ for all integers m. Show that moreover $|\sum(n) - \sum(mn)| < .005$ for all $n \ge N$ and all m. Show that $|\sum(N) - \sum(n)| < .01$ for all $n \ge N$. Conclude that the number produced by the computer for any $n \ge N$, no matter how large, will differ from the number it produces for N only by ± 1 in the last decimal place. This is what it means to say that the number $\sum(N)$ rounded to two decimal places represents the limiting value (the integral) with an accuracy of two decimal places.

2 *Computation with decimals and decimal approximations.* A number a (the approximate value) is said to represent a number t (the true value) with an accuracy of 3 decimal places if $|a - t| < .001$. Show that:

(a) If a represents t with an accuracy of 3 decimal places, if a_3 is a rounded to 3 decimal places, and if t_3 is t rounded to 3 decimal places, then a_3 differs from t_3 by at most ± 1 in the last (third) place.

(b) No matter how close a is to t, a_3 may still differ from t_3 by ± 1 in the last place.

3 Form approximating sums to $\int_S \frac{1}{x}\,dx$ in the same way as in Exercise 1; that is, divide the interval, S, into n equal segments and evaluate at midpoints. Suppose the computer has been programmed to find this number $\sum(n)$ rounded to three decimal places. Find an N such that $\sum(N)$ represents all $\sum(n)$ for $n \geq N$ (and hence represents the limiting value) with an accuracy of three decimal places.



4 Let S be the circle $\{(x, y): x^2 + y^2 = 1\}$ oriented counterclockwise, and let $A\,dx + B\,dy$ be the 1-form

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

of Exercise 2, §2.1, giving flow from a source at the origin. Estimate $\int_S (A\,dx + B\,dy)$ by using the inscribed regular n-gon to approximate S and evaluating $A\,dx + B\,dy$ for each segment at the midpoint of the corresponding arc of the circle (because this is easiest). Call the result $\sum(n)$. Express $\sum(n)$ explicitly in terms of the number $\sin\left(\frac{\pi}{n}\right)$. [The formula $\sin(x + y) - \sin(x - y) = 2 \cos x \sin y$ and the analogous formula for $\cos(x + y) - \cos(x - y)$ are used.] The formula

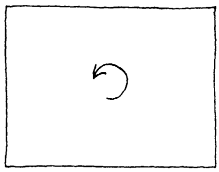
$$\lim_{x \to 0} [(\sin x)/x] = 1$$

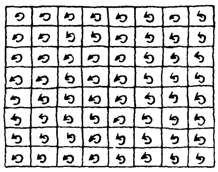
can be used to evaluate the limit as $n \to \infty$.

5 Let $x^3y^2z\,dx\,dy$ be a 2-form describing a flow in space and let S be the rectangle $\{(x, y, z): 0 \leq x \leq 1,\ 0 \leq y \leq 2,\ z = 1\}$ oriented counterclockwise. Let $\sum(n, m)$ be the approximating sum to $\int_S x^3y^2z\,dx\,dy$ obtained by dividing the x interval $0 \leq x \leq 1$ into n equal parts, dividing the y interval $0 \leq y \leq 2$ into m equal parts, orienting each of the mn rectangles counterclockwise, and evaluating the 2-form at the midpoint of each rectangle. Find an N such that $\sum(N, N)$ represents $\sum(n, m)$ with four-place accuracy for $n > N$, $m > N$.

2.3

Definition of Certain Simple Integrals. Convergence and the Cauchy Criterion





The greatest difficulty in giving a precise formulation of the informal definition of §2.2 lies in describing precisely what is meant by a 'finely divided polygonal approximation to the domain of integration'. This section is devoted to giving a precise formulation of the definition of §2.2 for integrals in which the domain of integration is a simple domain for which a 'finely divided polygonal approximation' can be described explicitly.

Consider first the case of an integral $\int_R A\,dx\,dy$ in which $A\,dx\,dy$ is a (variable) 2-form on the xy-plane and in which the domain of integration R is a rectangle

$$R = \{a \le x \le b, c \le y \le d\}$$

in the xy-plane oriented counterclockwise. [If R is oriented clockwise then $\int_R A\,dx\,dy$ is defined to be $-\int_{-R} A\,dx\,dy$ where $-R$ denotes R with the opposite orientation. This is in accord with the usual definition* $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$ of integrals over intervals $\{a \le x \le b\}$ which are oriented from right to left.] In this case a 'finely divided polygonal approximation' to the domain of integration R can be obtained simply by drawing lines $x =$ const., $y =$ const. to divide R into subrectangles and by orienting each subrectangle counterclockwise in agreement with the orientation of R. An approximating sum corresponding to this 'finely divided approximation to R' is obtained by 'evaluating' $A(x, y)\,dx\,dy$ on each subrectangle and adding over all rectangles. To form such an approximating sum it is necessary to choose:

The notation $\int_a^b f(x)\,dx$ denotes, of course, the integral of the 1-form $f(x)\,dx$ over the interval $\{a \le x \le b\}$ oriented from a to b. Unfortunately there is no such convenient notation for indicating orientations of 2-dimensional integrals.

lines $x = x_i$, where $a = x_0 < x_1 < x_2 < \cdots < x_{m-1} < x_m = b$
lines $y = y_j$, where $c = y_0 < y_1 < y_2 < \cdots < y_{n-1} < y_n = d$
points P_{ij}, one in each of the mn rectangles R_{ij}

$$= \{(x, y)\colon x_{i-1} \le x \le x_i, y_{j-1} \le y \le y_j\}.$$

The approximating sum is then

$$\sum_{i,j} [A(P_{ij})][\text{oriented area of } R_{ij}] \tag{1}$$
$$= \sum_{i=1}^{m} \sum_{j=1}^{n} A(P_{ij})(x_i - x_{i-1})(y_j - y_{j-1}).$$

The choices x_i, y_j, P_{ij} will be denoted collectively by the letter α and the corresponding sum (1) by $\sum(\alpha)$.†

†*This sum is a number once A, R are given and all the choices α are made. It is helpful to imagine that a computer has been programmed to compute this number $\sum(\alpha)$ whenever it is provided with the choices α. See Appendix 1.*

To say that the approximating sums $\sum(\alpha)$ approach a limit as the approximation is refined means essentially that the choices α do not significantly affect the result $\sum(\alpha)$, provided only that the polygons R_{ij} on which the approximating sum $\sum(\alpha)$ is based are all small. Let $|\alpha|$ be the largest dimension of any of the rectangles specified by α, that is, $|\alpha| = \max(x_1 - x_0, x_2 - x_1, \ldots, x_m - x_{m-1}, y_1 - y_0, \ldots, y_n - y_{n-1})$. $|\alpha|$ is called the *mesh size* of α. For the limit to exist means that if the mesh size $|\alpha|$ is small the resulting sum $\sum(\alpha)$ is insensitive to the choices to a very high degree; that is, another approximating sum $\sum(\alpha')$ in which the mesh size is similarly small will differ very little from $\sum(\alpha)$. Specifically, convergence of the sums $\sum(\alpha)$ is defined by the Cauchy Convergence Criterion:

The integral $\int_R A\,dx\,dy$ is said to *converge* if it is true that given any margin for error ϵ there is a mesh size δ such that

> any two approximating sums $\sum(\alpha)$, $\sum(\alpha')$ in which the mesh sizes are both less than δ differ by less than the prescribed margin for error ϵ,

that is,

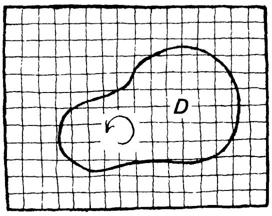
$$|\alpha| < \delta,\ |\alpha'| < \delta \text{ imply } \left|\sum(\alpha) - \sum(\alpha')\right| < \epsilon.$$

If this is the case then the *limiting value* can be defined to be the number* which is determined to within any margin of error in the obvious way. For example, if it is desired to find the limiting value with an accuracy of five decimal places (see Exercise 2, §2.2) set $\epsilon = .00001$, let δ be the corresponding mesh size, choose *any* α with $|\alpha| < \delta$, form $\sum(\alpha)$, and round to five decimal places. The number which results is determined, except for ± 1 in the last place, solely by the 2-form $A\,dx\,dy$ and the domain of integration R. If a different set of choices α is used, no matter how different they may be and, in particular, no matter how much finer a subdivision of R they may involve, the result will be the same except for at most ± 1 in the last place. The limiting value defined in this way is called the *integral* of $A\,dx\,dy$ over R and is denoted $\int_R A\,dx\,dy$.

It is an axiom of the real number system that a number is defined once such a process for computing it to any prescribed degree of accuracy has been given (see §9.1).

If this is not the case, that is, if the integral does not converge, then it is said that the integral *diverges* and that the limiting value $\int_R A\,dx\,dy$ *does not exist.*

This completes the definition of the integral of a 2-form $A(x, y)\,dx\,dy$ over an oriented rectangle R. The integral $\int_R A\,dx\,dy$ either is a *number*, defined above, or it does not exist. Only minor modifications are necessary to define the integral of a 3-form over an oriented rectangular parallelopiped or the integral of a 1-form over an oriented interval. The integral of a 2-form over a more general oriented domain D of the plane—for example, over the disk $D = \{(x, y)\colon x^2 + y^2 \leq 1\}$ oriented counterclockwise—can be defined by the simple trick of taking a rectangle R containing the domain D, setting the integrand equal to zero outside D, and proceeding as before. Assuming then that D is an oriented domain of the plane which can be enclosed in a rectangle R, the integral of a 2-form $A\,dx\,dy$ over D is defined by setting



$$A_D(x, y) = \begin{cases} A(x, y) & \text{if } (x, y) \text{ is in } D \\ 0 & \text{if } (x, y) \text{ is not in } D \end{cases}$$

and defining

$$\int_D A(x, y)\,dx\,dy = \int_R A_D(x, y)\,dx\,dy,$$

where the orientation of R is chosen to agree with that of D.* The integral of a 3-form over any oriented domain in space which can be enclosed in a rectangular parallelopiped, and the integral of a 1-form over any oriented domain of the line which can be enclosed in an interval, are defined by the same trick. Such domains are called *bounded* domains (that is, they are domains which stay within certain finite bounds). In summary, the integral of a k-form over a bounded, oriented domain of k-dimensional space has been defined for $k = 1, 2, 3$.

However, the definition begs a substantial part of the question, namely, *does the integral converge*? Generally speaking, one can say that in all reasonable cases the answer is yes. In particular the answer is yes in all cases which are considered in the following chapters, namely those cases in which the bounded domain D is *a region enclosed by a finite number of curves*† and the function A is *bounded*‡ and *continuous*§ on D.

An outline of the proof of the fact that these conditions on A and D guarantee the convergence of the integral is given below. This proof involves some rather difficult mathematical arguments, and one should not expect to understand it completely on the first or second reading.

**It is easy to show that convergence or divergence of the integral, as well as the limiting value in the case of convergence, is independent of the choice of the enclosing rectangle R (see Exercise 6).*

†Meaning 'reasonable' curves, that is, differentiable curves. The detailed statement of the theorem is given in Chapter 6.

‡This means that there is a number M such that $|A(P)| \leq M$ at all points P of D.

§This means that given a point $(\bar{x}, \bar{y})$ of D and a margin for error ϵ there is a distance δ such that the value of A varies by less than ϵ on the square $\{|x - \bar{x}| < \delta, |y - \bar{y}| < \delta\}$; that is, given $(\bar{x}, \bar{y})$ in D and $\epsilon > 0$ there is a $\delta > 0$ such that $|A(x, y) - A(\bar{x}, \bar{y})| < \epsilon$ whenever $|x - \bar{x}| < \delta$, $|y - \bar{y}| < \delta$. Intuitively it means that if (x, y) is near $(\bar{x}, \bar{y})$ then $A(x, y)$ is near $A(\bar{x}, \bar{y})$.

However, there is no better way to grasp the meaning of convergence of integrals than to study the proof of this theorem, particularly in concrete cases such as those of Exercises 1, 2, 3 at the end of this section, and Exercises 1, 3, 4, 5 of the preceding section.

Outline of proof

Let D be a bounded, oriented domain enclosed by a finite number of smooth curves. Let R be a rectangle containing D and let A be the function on R which is the given continuous function at points of D and which is 0 at points of R not in D (this function was denoted A_D above). The orientation of R, which is taken to agree (counterclockwise or clockwise) with that of D, can be assumed to be counterclockwise, since this affects only the sign of the result. It is to be shown that for such a function A on R the sums $\sum(\alpha)$ defined by (1) converge.

For the sums $\sum(\alpha)$ to converge means that $\sum(\alpha')$ can be made to differ arbitrarily little from $\sum(\alpha)$ by making $|\alpha|$ and $|\alpha'|$ small. In particular, if α' differs from α only in the choice of the points P'_{ij} and not in the choice of rectangles—that is, if $R'_{ij} = R_{ij}$—then

$$\left|\sum(\alpha) - \sum(\alpha')\right| = \left|\sum_{i,j} [A(P'_{ij}) - A(P_{ij})] \text{ area } (R_{ij})\right|$$

can be made small by making $|\alpha| = |\alpha'|$ small. Letting S denote the subdivision of R into subrectangles R_{ij} common to α and α', this implies that the *maximum* value can be made small, that is,

$$(2) \quad U(S) = \sum_{i,j} \left[\max_{P,P' \text{ in } R_{ij}} \{A(P') - A(P)\} \right] [\text{area } (R_{ij})]$$

can be made small by making the mesh size $|S|$ of S small. (Since the maximum is ≥ 0, all terms are positive and the absolute value signs may be dropped.) Specifically, if the integral converges, then given any margin for error ϵ there is a mesh size δ such that the number $U(S)$ defined by (2) is less than ϵ whenever the mesh size of the subdivision S is less than δ, i.e. $|S| < \delta$, implies $U(S) < \epsilon$. In brief, if the integral converges, then $U(S) \to 0$ as $|S| \to 0$.

The converse of this statement is also true; that is, *if* $U(S) \to 0$ as $|S| \to 0$ *then* the integral converges. One need only note that if α is based on a subdivision S and if α' is based on a subdivision S' which is a refinement* of S then it is still true that $|\sum(\alpha) - \sum(\alpha')| \leq U(S)$.

**A subdivision S' is said to be a refinement of a subdivision S if it is obtained from S by adding more lines $x = $ const., $y = $ const.*

(Exercise 4.) Then if α, α' are any two sets of data there is a third set of data, α'', based on a refinement of both S and S' (merely take S'' to include all lines x = const., y = const. specified by either S or S'), and hence

$$\begin{aligned}|\textstyle\sum(\alpha) - \sum(\alpha')| &= |\textstyle\sum(\alpha) - \sum(\alpha'') + \sum(\alpha'') - \sum(\alpha')|\\ &\leq |\textstyle\sum(\alpha) - \sum(\alpha'')| + |\sum(\alpha'') - \sum(\alpha')|\\ &\leq U(S) + U(S').\end{aligned}$$

If $U(S) \to 0$ as $|S| \to 0$ this can be made small by making $|S|$ and $|S'|$ both small; that is, given ϵ, there is a mesh size such that $|S| < \delta$ and $|S'| < \delta$ implies $|\sum(\alpha) - \sum(\alpha')| < \epsilon$, and therefore the integral converges. Thus *the integral converges if and only if* $U(S) \to 0$ as $|S| \to 0$.

This important conclusion is perhaps more comprehensible when it is formulated as follows: The number $U(S)$ represents the 'uncertainty' of an approximating sum $\sum(\alpha)$ to $\int_D A\,dx\,dy$ based on the subdivision S. Any approximating sum based on any refinement of S differs from any approximating sum based on S by at most $U(S)$. Thus further refinement changes the result by at most the 'uncertainty' $U(S)$. The integral converges if and only if this uncertainty can be made small by making the mesh size small.

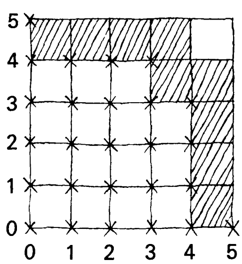
The problem, therefore, is to show that the assumptions on A and D are sufficient to guarantee that $U(S) \to 0$ as $|S| \to 0$. In order to do this, it is useful to decompose the sum (2) which defines $U(S)$ into two parts $U(S) = U_1(S) + U_2(S)$ as follows: A term of $U(S)$ corresponding to a rectangle R_{ij} of S is counted in the first sum $U_1(S)$ if R_{ij} is contained entirely in D, and counted in the second sum $U_2(S)$ if R_{ij} lies partly in D and partly outside D. (If R_{ij} lies entirely outside D then A is identically zero on R_{ij} and the corresponding term in $U(S)$ is zero.) It will be shown that the numbers $U_1(S)$ and $U_2(S)$ are both small when $|S|$ is small (but for quite different reasons).

The sum $U_2(S)$ is small because the total area of the rectangles R_{ij} which lie partly inside and partly outside D is small. More specifically, each term is at most $(A(P') - A(P))[\text{area }(R_{ij})] \leq 2M[\text{area }(R_{ij})]$ where M is a bound on the bounded function A, so the total $U_2(S)$ is at most $2M$ times the total area of such R_{ij}. It is intuitively clear that if the boundary of D consists of a finite number of reasonable curves then the total area of such rectangles R_{ij} can be made arbitrarily small by making the mesh size small. The rigorous proof of this statement must await a precise definition of 'reasonable' curves. (See Chapter 6 for the proof. For specific examples this statement can be proved directly—see Exercise 1.)

The sum $U_1(S)$ is small because $|A(P') - A(P)|$ is small (when P' is near P) by the continuity of A. Specifically, it can be shown that for every $\epsilon > 0$ there is a $\delta > 0$ such that $|A(P) - A(P')| < \epsilon$ whenever $|P - P'| < \delta$ (meaning that both the x-coordinates and the y-coordinates of P, P' differ by less than δ). In other words A is *uniformly* continuous on D (see §9.3.). Then, if $|S| < \delta$, it follows that $|U_1(S)|$ is at most ϵ times the total area of the rectangles R_{ij} inside D. Since ϵ is arbitrarily small and the total area of the rectangles in D is bounded (by the area of a rectangle containing D, for instance) it follows that $U_1(S)$ can be made small by making $|S|$ small. (Again, the rigorous proof is postponed to Chapter 6. In specific examples it can be proved directly that $U_1(S) \to 0$ as $|S| \to 0$; see, for example, Exercise 1, §2.2.)

This completes the outline of the proof that the integral $\int_D A\, dx\, dy$ converges.

Exercises



$U_5 = 9$
$C_5 = 15$
$A_5 = 3.12$

1 *Computation of π.* The number π is defined to be the area of the circle (disk) of radius 1, that is,

$$\pi = \int_D dx\, dy$$

where D is the disk $\{(x, y): x^2 + y^2 \le 1\}$. Given an integer n, draw a square grid of n^2 squares and $(n + 1)^2$ vertices, labeling the vertices with integer coordinates (p, q), $0 \le p \le n$, $0 \le q \le n$. Mark the vertices (p, q) for which $p^2 + q^2 \le n^2$ with an X. Shade all squares for which the inner vertex (p, q) has an X but for which the outer vertex $(p + 1, q + 1)$ does not. Let

U_n = # of shaded squares (U = uncertain)

C_n = # of squares whose vertices all have X's (C = certain)

$A_n = 4\left(\frac{1}{n^2} \cdot C_n + \frac{1}{n^2} \cdot \frac{1}{2} \cdot U_n\right)$ (A = approximation)

Find A_{10}. Find A_{20}. Show that any approximating sum to the integral $\int_D dx\,dy$ which is based on the subdivision by lines $x = \pm\frac{p}{n}$, $y = \pm\frac{q}{n}$ or on any refinement of this subdivision differs from A_n by at most $2n^{-2}U_n$. What accuracy (how many decimal places) does this estimate guarantee for the approximation $A_{10} \sim \pi$? What is the actual accuracy? What accuracy does it guarantee for A_{20}? What is the actual accuracy? Find a formula for U_n. [Count the crossings of the lines x = const. and y = const. separately.] How large would n have to be for this estimate to guarantee two-place accuracy? Note that the approximations are in fact more accurate than this estimate of the error would indicate. Explain this.

2 Many mathematicians, notably Karl Friedrich Gauss (1777–1855), have investigated the number N_r of points $(\pm p, \pm q)$ with integer coordinates contained in the circle of radius r (including points on the circle).

(a) Find N_5, $N_{\sqrt{7}}$.
(b) Show that N_r/r^2 is an approximating sum to $\pi = \int_D dx\,dy$. [Subdivide the plane by lines $x = \pm\frac{n}{r} + \frac{1}{2r}$, $y = \pm\frac{m}{r} + \frac{1}{2r}$ and evaluate at midpoints of squares.]
(c) Use the argument of Exercise 1 to prove that the number of squares which lie on the boundary of D is $\leq 8(r + \frac{1}{2})$.
(d) Show that any approximating sum based on any refinement of the subdivision of (b) differs from N_r/r^2 by less than $8(r + \frac{1}{2})/r^2$, and hence that the *limiting value* π differs from N_r/r^2 by less than $8(r + \frac{1}{2})/r^2$, i.e.

$$|\pi r^2 - N_r| < 8(r + \tfrac{1}{2}).$$

As was seen in the preceding problem, this estimate of the error is much too large; Gauss conjectured that the error is of the order of magnitude of $\sqrt{r}$, that is, that there is a number M such that $|\pi r^2 - N_r| < M\sqrt{r}$, but this conjecture has never been proved or disproved.*

**(Note added in 1980) G. H. Hardy 'Collected Papers, Vol. 2, p. 290) did disprove $|\pi r^2 - N_r| < M\sqrt{r}$. He conjectured that $< Mr^{(1/2)+\epsilon}$ was correct, and this has never been disproved. The statement about Gauss conjecturing $< M\sqrt{r}$ is not historically reliable.*

3 Show that the integral $\int_D dx\,dy$ defining π converges. [This is of course a special case of the theorem proved in the text, so it is a question of extracting the necessary parts of that proof. Take a fine subdivision by lines $x = \pm\frac{p}{n}$, $y = \pm\frac{q}{n}$, take $B_1, \ldots, B_N$ to be the squares which lie on the boundary, estimate their area, and estimate for *any* subdivision S the total area of the squares which touch one of the B_i in terms of the mesh size $|S|$. Show that this total area, and hence $U(S)$, can be made small by making $|S|$ small.]

4 Show that if $\sum(\alpha)$ is an approximating sum based on the subdivision S and if $\sum(\alpha')$ is any approximating sum based on a subdivision S' which is a refinement of S, that is, for which every rectangle R'_{ij} of S' is contained in or equal to some rectangle of S, then

$$\left|\sum(\alpha) - \sum(\alpha')\right| \leq U(S) = \sum_{\substack{\text{rectangles} \\ R_{ij} \text{ of } S}} \left[\max_{P, P' \text{ in } R_{ij}} \{A(P) - A(P')\}\right] [\text{area } R_{ij}]$$

[Lump together all terms of $\sum(\alpha')$ which correspond to the same rectangle R_{ij} of S.]

5 Show that if an integral converges and if L is the limiting value then any approximating sum $\sum(\alpha)$ based on the subdivision S is within $U(S)$ of the limiting value $|L - \sum(\alpha)| \leq U(S)$. [This observation was used in Exercise 2. It suffices to show that $|L - \sum(\alpha)| \leq U(S) + \epsilon$ for all $\epsilon > 0$. Note that by definition of L, $|L - \sum(\alpha')|$ can be made less than any $\epsilon > 0$ by making $|S'|$ small where S' is the subdivision of α'. Take S' to be a refinement of S.]

6 *Irrelevance of the enclosing rectangle.* Suppose that A_D is a bounded function which is zero outside the domain D, and suppose that the rectangles R, R' both contain D. Show that $\int_R A_D\, dx\, dy = \int_{R'} A_D\, dx\, dy$. [First show that one may as well assume that $R' \subset R$ and, in fact, that $D = R'$. Then since every approximating sum to $\int_{R'}$ is also an approximating sum to $\int_R$ it is easy to show:

$$\int_R \text{ converges implies } \int_{R'} \text{ converges;}$$

if both converge the limits are equal.

The only statement remaining is

$$\int_{R'} \text{ converges implies } \int_R \text{ converges.}$$

This is proved by showing that the rectangles of a subdivision S of R which lie on the boundary of R' make an insignificant contribution to $U(S)$.]

7 The definition of 'integral' given in the text was first given by Bernhard Riemann (1826–1866) in whose honor this notion of integration is called 'Riemann integration' and the approximating sums are called 'Riemann sums'. Riemann gave the following necessary and sufficient condition for the convergence of the integral $\int_R A\, dx\, dy$ where R is a rectangle and A is an *arbitrary* function on R:

Riemann's criterion. For any positive number σ and any subdivision S of R, let $s(S, \sigma)$ be the total area of those rectangles of S for which the variation $U_{ij} > \sigma$, i.e. in which there are points P, P' with $A(P) - A(P') > \sigma$. Then the integral $\int_R A\,dx\,dy$ converges if and only if

(1) A is bounded on R and
(2) for every $\sigma > 0$ the total area $s(S, \sigma)$ of those rectangles on which the variation is $> \sigma$ can be made small by making the mesh size $|S|$ small; that is, given $\sigma > 0$ and $\epsilon > 0$ there is a $\delta > 0$ such that $s(S, \sigma) < \epsilon$ whenever $|S| < \delta$. In short, for every $\sigma > 0$, $s(S, \sigma) \to 0$ as $|S| \to 0$.

Prove Riemann's criterion. [From the argument of the text it suffices to show that $U(S) \to 0$ as $|S| \to 0$ if and only if (1) and (2) hold. To prove 'only if', show that if A is unbounded then for any fixed S the maximum $U(S)$ is in fact infinite; whereas if (2) is false for some σ then $U(S) > \sigma \cdot s(S, \sigma)$ does not go to zero as $|S| \to 0$. To prove 'if', split $U(S)$ into two sums, one consisting of terms in which $U_{ij} \leq \sigma$ and the other in which $U_{ij} > \sigma$, make the first sum small by making σ small, and make the second sum small by using (1) and (2).]

8 In the text the condition 'the integral diverges' is defined to mean the negation of the statement 'the integral converges'. State precisely what this condition says about the approximating sums $\sum(\alpha)$; in other words, reformulate the *denial* of the Cauchy Criterion as a positive statement. Show that $\int_R A\,dx\,dy$ diverges when R is a rectangle and when $A(x, y)$ is the function

$$A(x, y) = \begin{cases} 1 & \text{if } x, y \text{ are both rational numbers} \\ 0 & \text{if one or both coordinates are irrational.} \end{cases}$$

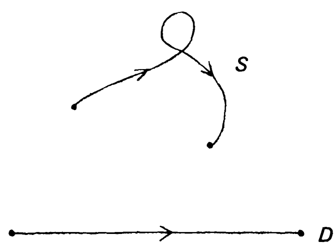
If L is a number with the property "Given $\epsilon > 0$ and given $\delta > 0$ there exists an approximating sum $\sum(\alpha)$ to $\int_R A\,dx\,dy$ with $|L - \sum(\alpha)| < \epsilon$ and $|\alpha| < \delta$" what values can L have? [Use the fact that every interval contains both rational and irrational numbers to show that L can be any number between 0 and the oriented area of R.]

9 Suppose that a, b, c, d are four numbers such that the lines $ax + by = \text{const.}$ and $cx + dy = \text{const.}$ are not parallel. Then a bounded domain can also be subdivided using lines $ax + by = \text{const.}$ and $cx + dy = \text{const.}$ Describe how to form an approximating sum to $\int_D A\,dx\,dy$ based on such a subdivision. Define the *mesh size* of such a subdivision to be largest dimension (in x- or y-direction) of any parallelogram of the subdivision. Sketch a proof of the fact that if D is a rectangle and if A is continuous then these approximating sums converge. The proof that their limit is $\int_D A\,dx\,dy$ is examined in Exercise 10.

10 Given an arbitrary subdivision of the plane into polygons, describe how to form an approximating sum to $\int_D A\,dx\,dy$ based on the subdivision. Define 'mesh size'. Sketch a proof of the fact that if D is a rectangle and if A is continuous then these approximating sums converge. Show that their limit is $\int_D A\,dx\,dy$. Why is it necessary to restrict to polygonal subdivisions? The approximating sums of the text were based on rectangular subdivisions because they are easiest to describe (being defined simply by the numbers x_i, y_j) and because it is quite difficult to define precisely what is meant by 'an arbitrary subdivision of the plane into polygons'.

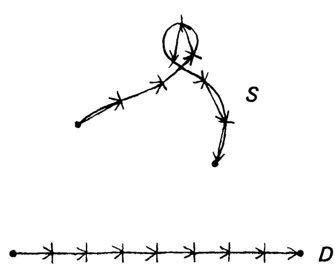
2.4 *Integrals and Pullbacks*

In §2.2 the integral of a k-form over an oriented k-dimensional domain was described as a limit of approximating sums, an approximating sum being formed by constructing a finely divided oriented polygonal domain which approximates the domain of integration, 'evaluating' the k-form on each polygon of this polygonal domain, and adding; the 'evaluation' involves first evaluating the k-form at some point near the polygon to obtain a constant k-form, and then proceeding as in Chapter 1.



In §2.3 this description was made the basis of a rigorous definition for cases in which the domain of integration is a rectangle (interval, rectangular parallelopiped) by describing explicitly what is meant by a 'finely divided polygonal approximation' to a rectangle. This definition was extended to arbitrary bounded domains in k-space by the simple trick of enclosing such a domain in a k-dimensional rectangle, setting the integrand equal to zero outside the domain, and proceeding as before.

The definition of §2.3 does not apply, however, to integrals over curves in the plane, over curves in space, or over surfaces in space. Such integrals will be defined in this section by assuming that the domain of integration can be described *parametrically* and by defining the integral as an integral over the parameter space.



Let the curve S in the xy-plane represent a displacement of a particle. Then S can be described by a pair of functions $(x(t), y(t))$ giving the coordinates of the particle as functions of time, these functions being defined for the time interval $D = \{a \le t \le b\}$ during which the displacement S occurs. A curve represented in such a way is called a *curve defined by a parameter*, as opposed to a *curve defined by an equation*. (For example, the curve $\{(\cos t, \sin t)\colon 0 \leq t \leq 2\pi\}$ is a curve defined by a parameter, whereas the set $\{(x, y)\colon x^2 + y^2 = 1\}$ is a curve defined by an equation.) Then a polygonal approximation to S can be constructed by giving a subdivision $a = t_0 < t_1 < t_2 \ldots < t_{n-1} < t_n = b$ of the parameterizing time interval D and drawing the polygonal curve from $(x(t_0), y(t_0))$ to $(x(t_1), y(t_1))$ to ... to $(x(t_n), y(t_n))$. The corresponding sum approximating $\int_S (A\,dx + B\,dy)$ is $\sum_{i=1}^{n} (A\,\Delta x_i + B\,\Delta y_i)$ where $\Delta x_i = x(t_i) - x(t_{i-1})$, $\Delta y_i = y(t_i) - y(t_{i-1})$ and where the functions A, B are evaluated at some point in the neighborhood of the line segment from $(x(t_{i-1}), y(t_{i-1}))$ to $(x(t_i), y(t_i))$—at the point $(x(\hat{t}_i), y(\hat{t}_i))$ where $\hat{t}_i$ is a point in the ith interval $t_{i-1} \leq \hat{t}_i \leq t_i$, for instance. Thus $\int_S (A\,dx + B\,dy)$ is approximately

$$\begin{aligned} &\sum_{i=1}^{n} [A(x(\hat{t}_i), y(\hat{t}_i))\,\Delta x_i + B(x(\hat{t}_i), y(\hat{t}_i))\,\Delta y_i] \\ &\qquad = \sum_{i=1}^{n} \left[A\,\frac{\Delta x_i}{\Delta t_i} + B\,\frac{\Delta y_i}{\Delta t_i}\right]\Delta t_i \end{aligned} \tag{1}$$

where $\Delta t_i = t_i - t_{i-1}$ = value of dt on the directed interval $t_{i-1}t_i$. But the sum on the right is approximately the integral of the 1-form

$$\left[A(x(t), y(t))\,\frac{dx}{dt}(t) + B(x(t), y(t))\,\frac{dy}{dt}(t)\right] dt$$

over the oriented interval from a to b. This integral has been defined and converges provided that the functions $A(x, y)$, $B(x, y)$ are continuous, which is assumed already, and provided that the functions $\frac{dx}{dt}(t)$, $\frac{dy}{dt}(t)$ exist and are continuous, which will be assumed henceforth. Therefore $\int_S A\,dx + B\,dy$ can be *defined* to be $\int_D \left[A\,\frac{dx}{dt} + B\,\frac{dy}{dt}\right] dt$ and equation (1) shows that the resulting number corresponds to the intuitive description of $\int_S (A\,dx + B\,dy)$ given in §2.2.

A slightly different way of arriving at the same result is to interpret $A\,dx + B\,dy$ as representing 'work' and to ask how much work is done during a short time interval. This is approximately

$$\left[A(x(\hat{t}_i), y(\hat{t}_i))\frac{\Delta x_i}{\Delta t_i} + B(x(\hat{t}_i), y(\hat{t}_i))\frac{\Delta y_i}{\Delta t_i}\right]\Delta t_i$$

and the shorter the time interval the better the approximation; that is,

$$\begin{matrix}\text{work done during} \\ \text{short time intervals}\end{matrix} \sim \left[A\frac{dx}{dt} + B\frac{dy}{dt}\right]dt.$$

The work done during a time interval which is not short is then found by integration

$$\begin{aligned}\int_S (A\,dx + B\,dy) &= \text{work done during the time interval } D \\ &= \int_D \left[A\frac{dx}{dt} + B\frac{dy}{dt}\right]dt.\end{aligned}$$

Seen in this light, what is involved is a pullback operation in which a 1-form on the t-line is obtained from the 1 form $A\,dx + B\,dy$ on the xy-plane and a map of the t-line to the xy-plane. The 1-form $\left(A\dfrac{dx}{dt} + B\dfrac{dy}{dt}\right)dt$ is is called the pullback of the (variable) 1-form under the (non-affine) map $x = x(t)$, $y = y(t)$.

The justification for defining

$$\int_S (A\,dx + B\,dy + C\,dz) = \int_D \left[A\frac{dx}{dt} + B\frac{dy}{dt} + C\frac{dz}{dt}\right]dt,$$

when S is a parametric curve $\{(x(t), y(t), z(t): t \text{ in } D\}$ in which the functions x, y, z have continuous derivatives is exactly the same. The 1-form

$$\left[A\frac{dx}{dt} + B\frac{dy}{dt} + C\frac{dz}{dt}\right]dt$$

on the line is called the pullback of the (variable) 1-form $A\,dx + B\,dy + C\,dz$ under the (non-affine) map $x = x(t)$, $y = y(t)$, $z = z(t)$.

Similarly, let S be a surface which is given in the form

$$S = \{(x(u, v), y(u, v), z(u, v): (u, v) \text{ in } D\}$$

where $x(u, v)$, $y(u, v)$, $z(u, v)$ are given functions of two variables u, v defined on a domain D of the uv-plane.

Such a surface is called a surface defined by parameters, as opposed to a surface defined by an equation. (For example, the surface

$$S = \left\{(\cos\theta\cos\varphi, \sin\theta\cos\varphi, \sin\varphi): 0 \le \theta \le 2\pi, -\frac{\pi}{2} \le \varphi \le \frac{\pi}{2}\right\}$$

is a surface defined by parameters, whereas the set $\{(x, y, z): x^2 + y^2 + z^2 = 1\}$ is a surface defined by an equation.) The integral of a 2-form over such a surface

$$\int_S [A\,dy\,dz + B\,dz\,dx + C\,dx\,dy]$$

S will be *defined* to be the integral of a pulled-back 2-form over the parameter space D. This 2-form on D is to assign to a small polygon in D the approximate value of $A\,dy\,dz + B\,dz\,dx + C\,dx\,dy$ on its image.

To find such a 2-form on D, consider a particular point $(\bar{u}, \bar{v})$ in D. Defining the partial derivatives, as usual, by

$$\frac{\partial x}{\partial u}(\bar{u}, \bar{v}) = \lim_{u \to \bar{u}} \frac{x(u, \bar{v}) - x(\bar{u}, \bar{v})}{u - \bar{u}}$$

$$\frac{\partial x}{\partial v}(\bar{u}, \bar{v}) = \lim_{v \to \bar{v}} \frac{x(\bar{u}, v) - x(\bar{u}, \bar{v})}{v - \bar{v}}$$

it is to be expected that the approximation

$$x(u, v) \sim x(\bar{u}, \bar{v}) + \frac{\partial x}{\partial u}(\bar{u}, \bar{v})(u - \bar{u}) + \frac{\partial x}{\partial v}(\bar{u}, \bar{v})(v - \bar{v})$$

holds for (u, v) near $(\bar{u}, \bar{v})$; that is, the given function $x(u, v)$ is well approximated by the affine function on the right. The given functions $y(u, v)$ and $z(u, v)$ can be approximated in the same way, leading to the conclusion that the map $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ defining the surface is well approximated near $(\bar{u}, \bar{v})$ by the affine map

$$\begin{aligned} x &= x(\bar{u}, \bar{v}) + \frac{\partial x}{\partial u}(\bar{u}, \bar{v})(u - \bar{u}) + \frac{\partial x}{\partial v}(\bar{u}, \bar{v})(v - \bar{v}) \\ (2) \quad y &= y(\bar{u}, \bar{v}) + \frac{\partial y}{\partial u}(\bar{u}, \bar{v})(u - \bar{u}) + \frac{\partial y}{\partial v}(\bar{u}, \bar{v})(v - \bar{v}) \\ z &= z(\bar{u}, \bar{v}) + \frac{\partial z}{\partial u}(\bar{u}, \bar{v})(u - \bar{u}) + \frac{\partial z}{\partial v}(\bar{u}, \bar{v})(v - \bar{v}). \end{aligned}$$

The image of a small polygon in D under the actual map is nearly its image under the affine map (2), and the value of $A\,dy\,dz + B\,dz\,dx + C\,dx\,dy$ on its image is therefore nearly the value of the pullback of

$$(3) \quad A(\bar{x},\bar{y},\bar{z})\,dy\,dz + B(\bar{x},\bar{y},\bar{z})\,dz\,dx + C(\bar{x},\bar{y},\bar{z})\,dx\,dy$$

on the polygon itself (where $\bar{x} = x(\bar{u},\bar{v})$, etc.). It is reasonable to define the pullback of $A\,dy\,dz + B\,dz\,dx + C\,dx\,dy$ under the map $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ to be the 2-form in uv whose value at any point $(\bar{u}, \bar{v})$ is the pullback of the constant form (3) under the affine map (2).

This (variable) 2-form in uv is easily computed. One need only set

$$\begin{aligned} dx &= \frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial v}\,dv \\ dy &= \frac{\partial y}{\partial u}\,du + \frac{\partial y}{\partial v}\,dv \\ dz &= \frac{\partial z}{\partial u}\,du + \frac{\partial z}{\partial v}\,dv \\ x &= x(u, v) \\ y &= y(u, v) \\ z &= z(u, v), \end{aligned}$$

substitute these expressions in $A(x, y, z)\,dy\,dz + B(x, y, z)\,dz\,dx + C(x, y, z)\,dx\,dy$, and simplify using the usual rules $du\,du = 0$, $du\,dv = -dv\,du$. For example, the pullback of $x\,dy\,dz - y^2\,dx\,dy$ under the map

$$\begin{aligned} x &= e^u + v \\ y &= u + 2v \\ z &= \cos u \end{aligned}$$

is

$$\begin{aligned} &(e^u + v)(du + 2\,dv)(-\sin u\,du) - (u + 2v)^2(e^u\,du + dv)(du + 2\,dv) \\ &\qquad = [(e^u + v)2\sin u + (u + 2v)^2(1 - 2e^u)]\,du\,dv. \end{aligned}$$

Note that the pullbacks of 1-forms above were found by the same method: The pullback of $A(x, y)\,dx + B(x, y)\,dy$ was found merely by performing the substitutions $x = x(t)$, $y = y(t)$, $dx = \frac{dx}{dt}\,dt$, $dy = \frac{dy}{dt}\,dt$ to 'express $A\,dx + B\,dy$ in terms of t'.

Thus the pullback of a (variable) k-form under a (not necessarily affine) map can be defined by these computational rules

$$(4)\quad dx = \frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial v}\,dv,\ du\,du = 0,\ du\,dv = -dv\,du, \text{etc.},$$

and it is reasonable to interpret the resulting k-form in uv as the k-form whose value on a small k-dimensional figure in uv-space is equal to the value of the given k-form on the image of this figure under the given map.

Using the definition of pullback by the computational rules (4), one can define the integral of a k-form over a k-dimensional domain which is defined by parameters to be the integral of the pullback over the parameterizing domain D. [Assuming that the partial derivatives of the given map exist and are continuous, the pullback is a continuous k-form on a k-dimensional space; hence the integral is defined and converges whenever the parameterizing domain D is reasonable in the sense of §2.3.]

To prove that this definition of integrals over curves and surfaces has all the desired properties is a rather long task (see Chapter 6). What is important for the moment is the computation of pullbacks and an intuitive understanding of the relation of this operation to the notion of integration as described in §2.2.

Exercises

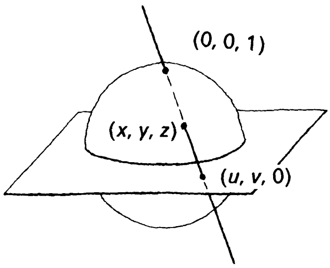
1 Find the pullbacks of

(a) $x\,dy\,dz$ under $x = \cos uv$, $y = \sin uv$, $z = uv^2$
(b) $xy\,dz\,dx$ under $x = u \cos v$, $y = u + v$, $z = u \sin v$
(c) $z^3\,dx\,dy$ under $x = e^u + v$, $y = e^u - v$, $z = 2$

2 If $x\,dy + y\,dx$ gives the work required for small displacements in the plane, and if $(x, y) = (\sqrt{t}, t)$ for $t > 0$ gives position as a function of time, find rate of work as a function of time (=work done/time elapsed as elapsed time $\to 0$). Graph the motion by drawing the curve along which the particle moves and labeling points with the time at which they are passed. Write the amount of work done between times $t = 1$ and $t = 4$ as an integral.

3 The mapping

$$x = 2u/(u^2 + v^2 + 1)$$
$$y = 2v/(u^2 + v^2 + 1)$$
$$z = (u^2 + v^2 - 1)/(u^2 + v^2 + 1)$$



arises from the *stereographic projection* of the sphere $x^2 + y^2 + z^2 = 1$ onto the plane $z = 0$. Specifically, the point (x, y, z) given by these formulas is the unique point of the sphere which lies on the line through $(u, v, 0)$ and $(0, 0, 1)$. Check this fact and show that every point of the sphere except $(0, 0, 1)$ corresponds to exactly one point of the plane. Find the pullback under this map of the 2-form $x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$. [The computation is long but the answer is simple.]

4 Find the pullback of $x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$ under the map

$$x = \cos\theta\cos\varphi,\ y = \sin\theta\cos\varphi,\ z = \sin\varphi$$

giving spherical coordinates on $x^2 + y^2 + z^2 = 1$.

5 Write the integral of Exercise 4, §2.2, as an integral over $\{0 \le \theta \le 2\pi\}$.

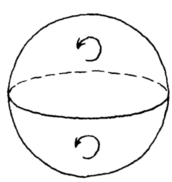
6 Find the pullback of $dx\,dy$ under the map

$$x = r\cos\theta,\ y = r\sin\theta$$

giving polar coordinates on the xy-plane. What is the approximate area of a ring-shaped region $\{r_1^2 \le x^2 + y^2 \le r_2^2\}$? What are the orientations when $r < 0$?

2.5

Independence of Parameter



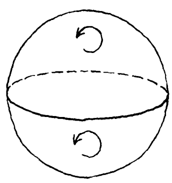
When the domain of integration is described parametrically the integral can be defined, as in the previous section, to be the integral of the pullback of the integrand over the parameterizing domain. However, it is frequently necessary—as will be seen in Chapter 3—to deal with integrals over oriented domains which are not described in this way, to deal, for example, with integrals of 2-forms over the sphere $S = \{x^2 + y^2 + z^2 = 1\}$. If the sphere is oriented by the convention 'counterclockwise as seen from the outside' (a mathematical description of this convention is given below) then the intuitive description of the integral given in §2.2 is applicable but the exact definition of §2.4 is not. The solution of this problem is simply to parameterize the domain of integration, but this raises some very difficult questions:

(i) When can a domain be parameterized?
(ii) What, precisely, does it mean to say that a parametric domain is a parameterization of a given domain?

(iii) If a domain is parameterized in two different ways, is the integral the same?

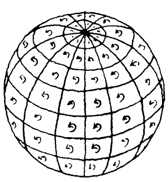
Rigorous answers to these questions will not be given until Chapter 6. In general, the first two questions are not important in practice, and the answer to the third question, which is very important in practice, is 'yes' under very broad assumptions.

For example, consider some parameterizations of the sphere $\{x^2 + y^2 + z^2 = 1\}$. A very common one is given by spherical coordinates

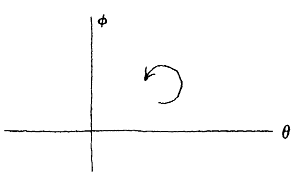


$$\begin{aligned} x &= r \cos\theta \cos\varphi \\ y &= r \sin\theta \cos\varphi \\ z &= r \sin\varphi \end{aligned}$$

on the rectangle $r = 1,\ 0 \le \theta \le 2\pi,\ -\frac{\pi}{2} \le \varphi \le \frac{\pi}{2}$. Denoting this rectangle by D and the sphere by S, a mapping from D to S has been given which can be seen geometrically to cover all of S; in fact, the lines θ = const. are the meridians of longitude, φ = const. are the parallels of latitude, and these coordinates are the usual ones for locating points on the earth's surface.



To integrate a 2-form over this parametric surface it is necessary to orient the parameterizing domain $D = \left\{r = 1, 0 \le \theta \le 2\pi, -\frac{\pi}{2} \le \varphi \le \frac{\pi}{2}\right\}$ and this is to be done in such a way that it gives the orientation 'counterclockwise as seen from the outside' to the sphere. A few pictures will suffice to show that the correct orientation of D is 'counterclockwise' when θ, φ are drawn as shown, but the same result can be reached without the aid of pictures as follows: An orientation of the $\theta\varphi$-plane can be described by specifying either $d\theta\, d\varphi$ or $d\varphi\, d\theta$ and saying that a triangle is positively oriented if the value of the specified 2-form on it is positive; thus $d\theta\, d\varphi \leftrightarrow$ counterclockwise and $d\varphi\, d\theta \leftrightarrow$ clockwise. Similarly, an orientation of the sphere can be described by specifying a non-zero 2-form and by saying that a small triangle on the sphere is positively oriented if the value of the specified 2-form is positive. In this particular example the 2-form $x\, dy\, dz + y\, dz\, dx + z\, dx\, dy$ describes the given orientation of the sphere; it is $dx\, dy$ at $(0, 0, 1)$, $dy\, dz$ at $(1, 0, 0)$, $-dx\, dy$ at $(0, 0, -1)$, $dz\, dx$ at $(0, 1, 0)$, etc.—all of which can be seen to describe the orientation 'counterclockwise as seen from the



outside' near the points in question. The pullback of $x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$ to $\theta\varphi$ therefore will show how the parameter space should be oriented; this pullback is of the form $f(\theta, \varphi)\,d\theta\,d\varphi$ where $f(\theta, \varphi)$ is a function whose values in the rectangle

$$D = \left\{0 \leq \theta \leq 2\pi, -\frac{\pi}{2} \leq \varphi \leq \frac{\pi}{2}\right\}$$

are all positive (see Exercise 4, §2.4). Thus the orientation of the parameterized sphere by $d\theta\,d\varphi$ agrees with the orientation of the sphere by $x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$ which is the orientation described verbally by the phrase 'counterclockwise as seen from outside'.

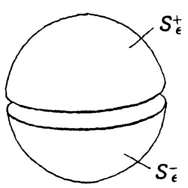
A second method of parameterizing the sphere is to take x, y as coordinates, that is, to project onto the xy-plane. Each point inside the disk

$$D = \{(x, y): x^2 + y^2 \leq 1\}$$

corresponds to two points on the sphere, one on the upper hemisphere and one on the lower hemisphere (except for points on the equator) leading to the parameterization of S by *two* parametric surfaces

$$S^+ = \{(x, y, \sqrt{1 - x^2 - y^2}): x^2 + y^2 \leq 1\}$$
$$S^- = \{(x, y, -\sqrt{1 - x^2 - y^2}): x^2 + y^2 \leq 1\}.$$

Since the orientation of S is described by $dx\,dy$ at $(0, 0, 1)$ and by $dy\,dx$ at $(0, 0, -1)$ these parametric surfaces should be oriented by orienting the domain $D = \{(x, y): x^2 + y^2 \leq 1\}$ using $dx\,dy$ for S^+ and $dy\,dx$ for S^-. This method of parameterizing S has the disadvantage that the parameterizing mappings do *not* have continuous partial derivatives at the equator so that the pullbacks of a 2-form on S are not defined at points of the boundary of D. The parameterization can nonetheless be used to find the integral $\int_S (A\,dy\,dz + B\,dz\,dx + C\,dx\,dy)$ of a 2-form over S by integrating over



$$S_\epsilon^+ = \{(x, y, \sqrt{1 - x^2 - y^2}): x^2 + y^2 \leq 1 - \epsilon\}$$

oriented $dx\,dy$ and

$$S_\epsilon^- = \{(x, y, -\sqrt{1 - x^2 - y^2}): x^2 + y^2 \leq 1 - \epsilon\}$$

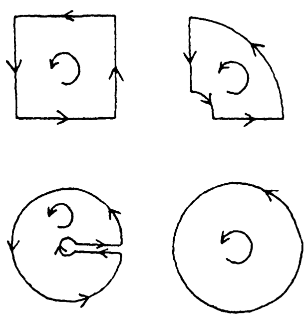
oriented $dy\,dx$, adding the results, and letting $\epsilon \to 0$.

A third parameterization of the sphere is given by the stereographic projection

$$\left\{\left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, \frac{u^2+v^2-1}{u^2+v^2+1}\right) : \text{all } u, v\right\}$$

(see Exercise 3, §2.4). The reader will easily find the orientation of the uv-plane which corresponds to the given orientation of S (Exercise 5). This parameterization has the disadvantage that the point $(0, 0, 1)$ is omitted. It can nonetheless be used to find the integral of a 2-form over S by integrating the pullback over $u^2 + v^2 \leq A$ and letting $A \to \infty$. Alternatively, the stereographic projection can be used to parameterize the lower hemisphere $\{u^2 + v^2 \leq 1\}$ and its mirror image used to parameterize the upper hemisphere.

These various parametric representations lead to various methods of computing 'the number $\int_S (A\,dy\,dz + B\,dz\,dx + C\,dx\,dy)$'. The fact that all of them result in the same value *requires proof* because no definition of 'the number $\int_S (A\,dy\,dz + B\,dz\,dx + C\,dx\,dy)$' has been given other than these methods of computing it. This fact, that integrals over domains can be computed using any parameterization of the domain, is called the principle of independence of parameter.



Another example of the principle of independence of parameter is contained in the rule for conversion to polar coordinates in a double integral. For example, if D is the disk $D = \{(x, y): x^2 + y^2 \leq 1\}$ oriented counterclockwise, then D is parameterized in polar coordinates

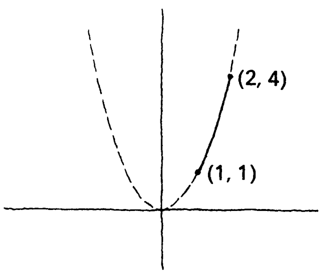
$$x = r\cos\theta$$
$$y = r\sin\theta$$

by the rectangle $0 \leq r \leq 1$, $0 \leq \theta \leq 2\pi$ oriented counterclockwise. Therefore the integral of $A(x, y)\,dx\,dy$ over the disk is equal to the integral of the pullback

$$A(r\cos\theta, r\sin\theta)(\cos\theta\,dr - r\sin\theta\,d\theta)(\sin\theta\,dr + r\cos\theta\,d\theta) = A(r\cos\theta, r\sin\theta)r\,dr\,d\theta$$

over the rectangle. The orientation $dx\,dy$ corresponds to the orientation $dr\,d\theta$ because $r > 0$.

Similarly, the integral of a 1-form over an oriented curve is independent of the choice of parameter. For example, if S is the curve $x^2 = y$ between the points



(1, 1) and (2, 4) oriented from the first toward the second, then the integral of $A\,dx + B\,dy$ over S can be computed using either the parameterization $S = \{(x, x^2): 1 \leq x \leq 2\}$ or the parameterization $S = \{(\sqrt{y}, y): 1 \leq y \leq 4\}$ oriented by dx and dy respectively.

Plausible as these statements are in specific cases, it is rather difficult to give a precise definition of the statement that two parameterized surfaces are parameterizations of the same surface, and more difficult still to prove rigorously that integrals are independent of parameter. Until these subjects are dealt with carefully (in Chapter 6) the notion of the integral of a form over an oriented surface is *not defined* until a particular parameterization of the oriented surface is given. However, this is strictly a technical difficulty; integrals *are* independent of parameter, and the informal description of integrals in §2.2 is nearer to their true meaning than is the precise definition of integrals over parameterized surfaces given in the preceding section.

Exercises

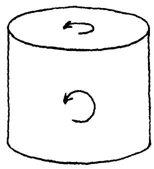
1 Parameterize the surface $x + y + z = 1$. Orient the parameterization so as to agree with the orientation given on the original surface by $dy\,dz$; by $dz\,dx$; by $dx\,dy$. Sketch the given surface showing its orientation.

2 Find the pullbacks of $dy\,dz$, $dz\,dx$, and $dx\,dy$ under each of the parameterizations of the sphere considered—spherical coordinates, projection on the xy-plane, and stereographic projection.

3 Let $\int_a^b f(x)\,dx$ be a given integral. Let $x = x(u)$ be a 'parameterization' of the interval $\{a \leq x \leq b\}$ by a new parameter u on an interval $\{\alpha \leq u \leq \beta\}$. That is, let $x(u)$ be a differentiable function establishing a one-to-one correspondence between points of the interval $\{\alpha \leq u \leq \beta\}$ and points of the interval $\{a \leq x \leq b\}$. State the principle of independence of parameter in this case. Pay particular attention to the orientation. Apply this to the integral $\int_a^b x^n\,dx$ $(a > 0, b > 0)$ when $x = e^u$.

4 In order to find the correct orientation of spherical coordinates $\theta\varphi$ it is not necessary to compute the pullback of $x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$ at all points, since the pullback at any one point is sufficient to determine the sign. The point $(x, y, z) = (1, 0, 0)$, $(\theta, \varphi) = (0, 0)$ is particularly simple. Find the pullback at this point.

5 Find the correct orientation of the parameterization by stereographic coordinates. [Find the pullback of $x\,dy\,dz + y\,dz\,dx + z\,dx\,dy$ at the point $(0, 0, -1)$.]



6 Parameterize the three pieces of the boundary of the cylinder $\{(x, y, z): x^2 + y^2 \leq 1, -1 \leq z \leq 1\}$. Orient each of the three pieces by the rule 'counterclockwise as seen from the outside'.

7 Parameterize the surface obtained by rotating the circle $(x - 2)^2 + z^2 = 1, y = 0$ about the z-axis. Orient the parameter space to agree with the orientation of the given surface by the rule 'counterclockwise as seen from the outside'. [First parameterize the given circle and rotate about the z-axis using cylindrical coordinates (r, θ, z). Convert to (x, y, z) coordinates using $x = r\cos\theta$, $y = r\sin\theta$, $z = z$.] This surface is called a *torus*.

2.6

Summary. Basic Properties of Integrals

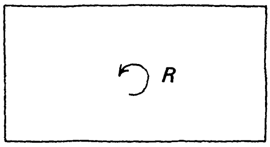
Chapter 1 was devoted to constant k-forms and to their evaluation on simple k-dimensional figures such as oriented line segments, triangles, parallelograms, and cubes. This chapter has been devoted to (variable) k-forms and to their evaluation on oriented k-dimensional domains. The 'value' of a k-form on an oriented k-dimensional domain has been defined as a limit of sums, that is, as an *integral.*

The precise definition of the notion of 'integral' is difficult for two reasons—first, because it involves the notion of 'limit', and second, because it involves the notion of 'oriented k-dimensional domain'. The notion of 'limit' is defined by the Cauchy Convergence Criterion, which was discussed in detail in §2.3. The problem of defining 'oriented k-dimensional domain' is more difficult and has been avoided entirely by restricting consideration to *specific* domains such as rectangles, disks, spheres, etc., and to domains which are parameterized by such domains.*

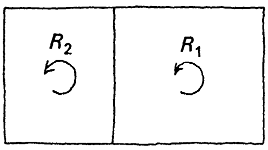
**The precise definition of 'k-dimensional domain', which is given in Chapters 5 and 6, depends essentially on the Implicit Function Theorem.*

The following properties of integrals are all immediate consequences of the definition of integrals as limits of sums. They are stated specifically for integrals of the form $\int_R A(x, y)\,dx\,dy$ where R is an oriented rectangle in the xy-plane and $A(x, y)$ is a continuous function defined at all points of R, but they all have analogs which are true for integrals of k-forms on k-dimensional domains in the general cases described in §2.2:

(i) If the orientation of the domain of integration is reversed the integral changes sign: $\int_{-R} A\,dx\,dy = -\int_R A\,dx\,dy$.



(ii) If the domain of integration is divided into two (or more) smaller rectangles oriented in accordance with the orientation of the original, then the integral over the whole is the sum of the integrals over the parts:

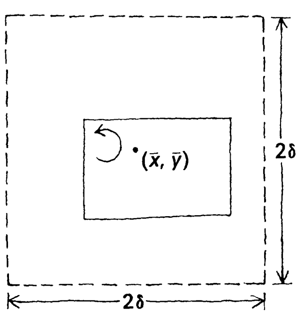


$$\int_{R_1+R_2} A\,dx\,dy = \int_{R_1} A\,dx\,dy + \int_{R_2} A\,dx\,dy.$$

(iii) If the integrand is multiplied by a constant the integral is multiplied by the same constant: $\int_R cA(x, y)\,dx\,dy = c\int_R A(x, y)dx\,dy$.

(iv) If the integrand is a sum of two (or more) terms then the integral of the sum is the sum of the integrals: $\int_R (A_1 + A_2)\,dx\,dy = \int_R A_1\,dx\,dy + \int_R A_2\,dx\,dy$.

If the rectangle R is subdivided into a very large number of very small pieces, then, by (ii), $\int_R A\,dx\,dy$ is the sum of the integrals over the individual pieces. The integral over a very small rectangle is roughly the value of A on the rectangle times the oriented area of the rectangle. This is only roughly true because, of course, A does not have *a* value on the rectangle but many values. However, the assumption that A is continuous is the assumption that A is nearly constant on sufficiently small rectangles, so that the integral over such a rectangle is nearly 'the' value of A times the oriented area. Specifically, the definition of continuity of a function easily implies the following:



(v) Given a point $(\bar{x}, \bar{y})$ and given $\epsilon > 0$ there is a $\delta > 0$ such that if R is any rectangle containing $(\bar{x}, \bar{y})$ and contained in the square $\{|x - \bar{x}| < \delta, |y - \bar{y}| < \delta\}$, then $\int_R A\,dx\,dy$ differs from $A(\bar{x}, \bar{y})$ times the oriented area of R by less than ϵ times the area of R

$$\left|\int_R A\,dx\,dy - A(\bar{x}, \bar{y})\int_R dx\,dy\right| < \epsilon \cdot \left|\int_R dx\,dy\right|.$$

A useful way to remember this statement is by means of the formula

$$\lim_{R \to P} \frac{\int_R A\,dx\,dy}{\int_R dx\,dy} = A(P)$$

where the rectangle R is thought of as shrinking down to the point P. This statement about integrals over small rectangles, together with the subdivision property (ii), is the substance of the intuitive idea of 'integral' as it is described in §2.2.

Another type of formula which is frequently useful is the formula for a double integral as an iterated integral

$$(1)\qquad \int_R A(x, y)\,dx\,dy = \int_c^d \left[\int_a^b A(x, y)\,dx\right] dy$$

where $A(x, y)$ is a continuous function on the rectangle $R = \{a \le x \le b, c \le y \le d\}$, where R is oriented counterclockwise, and where $\int_a^b A(x, y)\,dx$ is considered as a function of y. For the proof of this formula see Exercise 2.

The integral of a form over a domain which is described parametrically is defined in terms of the pullback of the form under the parameterizing map. The pullback operation, which is a simple generalization of the pullback operation for constant forms under affine maps, is studied further in Chapter 5.

Exercises

1 Prove the properties (i)–(v) of $\int_R A\,dx\,dy$.

2 Prove that if $A(x, y)$ is continuous on $R = \{a \le x \le b, c \le y \le d\}$ then the integral on the right side of (1) converges and the formula (1) holds. [Use the fact, stated at the end of §2.3, that for every $\epsilon > 0$ there is a $\delta > 0$ such that $|A(x, y) - A(\bar{x}, \bar{y})| < \epsilon$ whenever (x, y), $(\bar{x}, \bar{y})$ are points of R such that $|x - \bar{x}| < \delta$, $|y - \bar{y}| < \delta$. (This is Theorem 2 of §9.4.) Then if $\sum(\alpha)$ is an approximating sum to the integral on the right side of (1) based on a subdivision of $\{c \le y \le d\}$ finer than δ, it follows that $\sum(\alpha)$ differs by less than $\epsilon(b - a)(d - c)$ from an approximating sum to $\int_R A\,dx\,dy$, which in turn differs by less than $\epsilon(b - a)(d - c)$ from $\int_R A\,dx\,dy$. Thus $\sum(\alpha) \to \int_R A\,dx\,dy$ as was to be shown.]

Springer

http://www.springer.com/978-0-8176-8411-2

Advanced Calculus
A Differential Forms Approach
represented by his heir Betty Rollin, H.M.E.
2014, XIX, 508 p. 102 illus., Softcover
ISBN: 978-0-8176-8411-2
A product of Birkhäuser Basel