LINEAR ANALYSIS AND CALCULUS

Andrew M. Gleason

PART 3
MATHEMATICS 21

HARVARD UNIVERSITY

1972—73

Harvard University
Lecture Notes

LINEAR ANALYSIS AND CALCULUS

Andrew M. Gleason

Part III

Mathematics 21

Harvard University

1972-73

PART 3

CONTENTS

Chapter 8. Real functions of several real variables

CHAPTER 8

REAL FUNCTIONS OF SEVERAL REAL VARIABLES
OR FUNCTIONS FROM AN INNER PRODUCT SPACE TO $\mathbb{R}$

Your first experiences with calculus concerned functions from $\mathbb{R}$ (or part of $\mathbb{R}$) to $\mathbb{R}$, often called real functions of a real variable. Now we want to consider, for example, real functions of two real variables. Such functions abound in real life. The formula, area = length × width, for the area of a rectangle leads to a function of one real variable if, say, the width of the rectangle is held constant while the length varies, or if the length and width vary simultaneously in a dependent manner like $\ell = \sqrt{1 + w^2}$. But we may wish to study the situation when the length and width vary independently. Then we shall be concerned with the function

$$< x, y > \longmapsto xy$$

from $(0,\infty) \times (0,\infty)$ to $\mathbb{R}$. Of course this is a very simple function, so it is not likely that we shall learn very much about it using a general theory that we couldn't learn more directly. But one need not go beyond polynomial functions to find questions that would be hard to answer without using the concepts of calculus. For example, what is the minimum value of $x^3 + y^2 - x - 2y + xy$ considering only positive values of x and y?

In the study of functions of one variable, graphs provide valuable insights. For functions of two variables graphs are again valuable. With some difficulty we can make the graph of a function from $\mathbb{R}^2$ to $\mathbb{R}$. It will be a surface in three-dimensional space. If we are unwilling to make the three-dimensional model, we can visualize it or make a perspective drawing of it on paper. The geometric problem of finding a line tangent to a curve now becomes the problem of finding a plane tangent to a surface. Either of these problems is the analytic problem of finding a first degree approximation to a function. The geometric and analytic ways of looking at a problem often produce quite different insights, so we want to be flexible in choosing our point of view. Hence we shall often regard a function from $\mathbb{R}^2$ to $\mathbb{R}$ as a function from a plane or two-dimensional inner product space to $\mathbb{R}$.

We cannot draw graphs of functions of three or more variables (because we would need a space of dimension four, at least). Nevertheless, the geometric point of view remains useful, so we shall often think of a function from $\mathbb{R}^n$ to $\mathbb{R}$ as being defined on an n-dimensional inner product space. Of course, geometrical ideas in higher dimension are really only analogies. The ultimate test of our ideas must remain in the domain of analysis (ie., statements about the real numbers).

8.1 Foundations.

Although we shall not attempt to prove all the basic theorems concerning continuous functions, convergence, etc., it is important to get some intuitive feeling for the ideas that underlie such notions. In this whole chapter we shall deal with inner product spaces of finite dimension. Finite dimension is the crucial point here; with but few exceptions everything could be done without reference to an inner product. It is also fortunate that a serviceable intuition for the concepts can be gained by studying examples in dimensions two and three.

We begin by defining some useful properties of subsets of a space.

8.1.1 Definition. Let V be an inner product spzce, $x \in V$, and $\rho > 0$. the open ball of radius ρ about x is the set

$$\{ v \in V : \| v - x \| < \rho \}.$$

The closed ball of radius ρ about x is the set

$$\{ v \in V : \| v - x \| \leq \rho \}.$$

The term ball is very appropriate when V is three-dimensional, since then the sets are what we ordinarily think of as balls. In dimension two the word disk is commonly used in place of ball. In dimension one, the set is a segment or an interval, but one does not ordinarily describe an interval as having a radius.

The difference between an open ball and a closed ball is simply that the closed ball includes the "skin" while the open ball does not.

Note that the words open and closed have meanings that generalize their meanings in connections with intervals. The open ball of radius ρ about x in $\mathbb{R}$ is the open interval $(x - \rho, x + \rho)$ and the closed ball is the closed interval $[x - \rho, x + \rho]$. We shall extend the meaning of these words further in 8.1.3 and 8.1.4.

8.1.2 Definition. Let S be any set in an inner product space V and $s \in S$. Then s is called an interior point of S and S is said to be a neighborhood of s if and only if S contains some ball about s.

Examples. Suppose V is $\mathbb{R}$ with the usual inner product. Then $\|\lambda\| = |\lambda|$. Take S to be an interval $[a,b]$. We already know what an interior point of S is; it is any point of (a,b), that is, any point of S except an endpoint. Let us check that this agrees with the above definition. If $s \in (a,b)$, take ρ to be the smaller of $s-a$ and $b-s$. Then every v satisfying $\|v-s\| < \rho$, that is, every v between $s-\rho$ and $s+\rho$, lies in S, so s is an interior point. On the other hand, a is not an interior point of S, because no matter how small ρ may be, $v = a - \frac{1}{2}\rho$ satisfies $\|v-a\| < \rho$ and $v \notin S$; thus there is a point in the ball of radius ρ about a not in S. Similarly for b.

In the plane, let S be the closed unit disk, that is $\{ v : \|v\| \leq 1 \}$. The points of the unit circle are not interior points of S and all other points of S are. Say $s \in S$ and $\|s\| < 1$. Then take $\rho = 1 - \|s\|$. If $\|v-s\| < \rho$, we have

$$\|v\| = \|v - s + s\| \leq \|v - s\| + \|s\| < \rho + \|s\| = 1,$$

so $v \in S$. Hence S contains the ball of radius ρ about s, so s is an interior point of S. On the other hand, if $\|t\| = 1$, then $t \in S$, but t is not interior to S. For, if ρ is any positive number, then

$$v = (1 + \tfrac{1}{2}\rho)t$$

satisfies $\| v-t \| < \rho$ and $v \notin S$.

If S is the unit circle, or any other set we should ordinarily describe as a curve, then S has no interior points. Speaking somewhat loosely, a point of S is interior to S if and only if all of its near neighbors in V are also in S. Thus S contains a solid chunk of space surrounding each of it interior points. The definition explicitly says the chunk is to be round, but any solid part of space can be "pared down" to be round.

The definition doesn't specify whether the ball is question is open or closed. It doesn't matter. If S contains the closed ball about s of radius ρ, it contains also the open ball of the same radius. If S contains the open ball of radius ρ about s it contains also the closed ball of radius $\frac{1}{2}\rho$ about s.

8.1.3 <u>Definition</u>. Let V be an inner product space. A set S in V is called <u>open</u> if and only if each of its points is an interior point.

We should check immediately that this more general meaning of the word <u>open</u> is consistent with 8.1.1. Let S be an open ball in the sense of 8.1.1. Say

$$S = \{ v : \| v-x \| < \rho \}.$$

Suppose $y \in S$; we must show that y is an interior point of S. We know $\| y-x \| < \rho$, so $\theta = \rho - \| y - x \| > 0$. We claim that the open ball of radius θ about y is a subset of S. Indeed, if z is in this ball, that is, if $\| z - y \| < \theta$, then

$$\| z-x \| = \| z-y+y-z \| \le \| z-y\| + \| y-z \| < \theta + \| y-x \| = \rho ;$$

so $z \in S$.

There are lots of open sets besides open balls. Let w be a fixed non-zero vector in V. Let $H = \{ v : (v,w) > 0 \}$. H is the half-space of vectors on one side of the hyperplane $\{ v : (v,w) = 0 \}$. H is an open set. For suppose

$y \in H$. Then $(y,w) > 0$. So we may take $\rho = \frac{(y,w)}{\|w\|}$ and consider the ball of radius ρ about y. If z is in this ball, that is, if $\| z-y \| < \rho$, then by the Cauchy-Schwarz inequality

$$|(w,z-y)| \leq \| w \| \cdot \| z-y \| < \| w \| \rho = (w,y)$$

so

$$(w,z) = (w,y) + (w,z-y) > (w,y) - |(w,z-y)| > 0.$$

Thus $z \in H$. This proves that H contains an entire ball about y. Since y may be any point of H, H is open.

It is worth remarking that the whole of V is open and so is the null set. This is another example of the fact that our definitions are intended to apply even in apparently trivial situations.

8.1.4 Definition. Let S be a subset of an inner product space V. Then S is closed if and only if its complement (ie., the set of all points of V not in S) is open.

The set $\{ v : \| v-a \| \leq \beta \}$ is a closed set for any choice of $a \in V$ and $\beta \in \mathbb{R}$. For $\beta < 0$, it is empty. For $\beta = 0$, it contains only the single point a. For $\beta > 0$, we have already called the set a closed ball; it is left to you to check that it is indeed closed in the sense of 8.1.4.

Beware: Unlike a door, a set need not be either open or closed. A half-open interval on the line, say $[a,b)$, is neither. Moreover, a set can be both open and closed. In the present context, this is true only for the null-set and the whole space. The concepts open set and closed set have a much wider applicability than to subsets of inner product spaces. They lie at the foundation of the subject known as topology.

The ideas of open and closed are perhaps most readily understood in terms of the idea of boundary points. These are the points at which a set abuts its complement. Precisely, a point v is a boundary point of a set S if and only if each ball about v contains points of both S and its complement. Here v may or may not belong to S. A set is open if it contains none of its boundary points and closed if it contains all of them.

A good rule of thumb is that a set described by strong inequalities (ie., $<$ or $>$) is open, while one described by equalities and/or weak inequalities (ie., $\leq$ or $\geq$) is closed. Thus, in three-space

$$\{ < x, y, z > : x^3 + y^2 < z \}$$

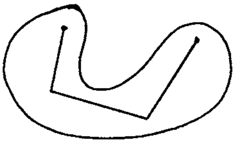
is open, while

$$\{ < x, y, z > : x \geq 0 \text{ and } x + y = z^2 \}$$

is closed.

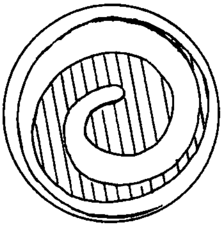
8.1.5 Definition. Suppose S is an open subset of an inner product space. The set S is connected if and only if every two points of S can be joined by a broken line in S. That is, given any two points a and b in S, there exists a sequence $x_1, x_2, \ldots, x_n$ of points such that each of the segments $ax_1, x_1x_2, \ldots, x_{n-1}x_n, x_nb$ lies in S. (By the segment uv we mean the set of points of the form $\lambda u + (1 - \lambda)v$ where $0 \leq \lambda \leq 1$.)

On the line a set is connected and open if and only if it is an open interval. In this case the broken line can always be taken straight. In higher dimensions, the situation can be more complicated, as indicated in the figure. If, for any two points a and b of S, the segment ab lies in S, then S is said to be convex.



Our formal definition of connected will be seen to agree with your intuitive idea of connected set for open sets. Our definition does not apply to a circle since it is not open. A more complicated definition of connected can be given that applies to arbitrary sets. It makes a circle connected and a line less a single point disconnected as you expect.

In most of our theoretical work with functions of several variables we shall confine ourselves to functions defined on open subsets of $\mathbb{R}^n$. This is analagous to considering only functions defined on open intervals in $\mathbb{R}$. There are, of course, cases in which it is desirable to discuss functions whose domains include one or more boundary points, just as it is often important to consider functions defined on closed intervals in $\mathbb{R}$. But boundary points can be very complicated in two or more dimensions. The figure suggests one of the unpleasant possibilities. (The long thin tail of the shaded open set spirals around infinitely often approaching the unit circle. Every point of the unit circle is a boundary point.)



When the boundary is smooth, it causes no more difficulty than endpoints on the line, but it is hard to give in advance a reasonable definition of a smooth boundary, and if we attempted to state our theorems so as to cover boundary points, we would spend an inordinate amount of energy on questions that are clearly of secondary importance. It is best therefore to ignore boundary questions in our general work and consider them separately whenever they arise explicitly.

8.1.6 Definition. Let E be an open subset of an inner product space V. and let f be a function from E to $\mathbb{R}$. We say that f is continuous at $v_0 \in E$ if and only if

$$(\forall \varepsilon > 0)(\exists \delta > 0) \quad \|v-v_0\| < \delta \implies |f(v) - f(v_0)| < \varepsilon.$$

We say that f is continuous if and only if it is continuous at each point of E.

This is just the old definition of continuous except that close in the domain is now measured in terms of the norm. It still means that you get approximately the right answer if you compute with approximately the right argument.

Because we assume that E is open, whenever δ is chosen small enough $\|v-v_o\| < \delta \Longrightarrow v \in E$. We shall always take δ this small; this guarantees that $f(v)$ will be defined.

We state without proof some standard theorems concerning the continuity of sums, products, etc. The proofs are almost identical with the proofs of the corresponding theorems for functions of a single variable.

8.1.7 Theorem. Let E be an open subset of an inner product space, and let f and g be two continuous functions from E to $\mathbb{R}$. Let $\lambda, \mu \in \mathbb{R}$. Then $\lambda f + \mu g$ and fg (defined pointwise as usual) are continuous. Provided g does not vanish at any point of E, f/g is continuous.

8.1.8 Theorem. Let E be an open subset of an inner product space, and let $f : E \longrightarrow \mathbb{R}$ be continuous. Let $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be continuous. Then $\varphi \circ f$ is continuous.

Example. It follows that functions defined by rational expressions in the coordinates are continuous provided we keep away from (ie., exclude from the domain) points at which division by zero is called for. Thus

$$\frac{x^3 + xy - y^4 - 12}{x + y}$$

defines a continuous function having domain

$$\{ < x, y > : x + y \neq 0 \}.$$

Then applying the second theorem, we see that

$$\sin \left(\frac{x^3 + xy - y^4 - 12}{x + y} \right)$$

is also continuous with the same domain. It follows that any function that is defined throughout an open set by a single formula involving only continuous functions is continuous. Thus

$$\sqrt{1 - x^2 - y^2}$$

defines a continuous function on any open subset of the interior of the unit disk (on the whole open unit disk if we like) but on no larger open set, since the formula doesn't make sense beyond the closed unit disk. The formula defines a function (which is indeed continuous) on the closed unit disk, but we are considering only functions with open domains.

8.1.9 Definition. A subset S of an inner product space is called bounded if and only if there is a number M such that $(\forall s \in S)\ \|s\| \leq M$.

In other words, S is bounded if and only if it is a subset of some ball. It doesn't matter whether this ball is centered at the origin, as in the definition, because the ball of radius M about v is itself a subset of the ball of radius $M + \|v\|$ about the origin.

Here is a theorem that is most conveniently stated for a function having a closed domain.

8.1.10 Theorem. Let X be a bounded closed subset of an inner product space and let $f : X \rightarrow \mathbb{R}$ be continuous. Then f achieves both a maximum and a minimum value.

To say that f achieves a maximum value means there is a point $x_0 \in X$ such that

$$(\forall x \in X)\quad f(x_0) \geq f(x).$$

It is essential that X be both bounded and closed.

In practice the function f will usually be defined on a set larger than just X. The maximum and minimum values guaranteed by the theorem are only maximum and minimum in competition with values taken by f on X.

Example. Let $f(x,y) = (x + 2y)(x^2 + 2y^2 - 9)$. Then f is defined on all of $\mathbb{R}^2$. Let X be the set where $x^2 + 2y^2 \leq 9$, a closed elliptical region. X is closed and bounded, so there must be points where f achieves a maximum and a minimum relative to X. Since f is zero on the entire boundary of X (ie., the ellipse $x^2 + 2y^2 = 9$), while f has both positive and negative values on X, both the maximum and the minimum values occur at interior points of X. The theorem tells us nothing about how to find these points. Since the function f happens to be differentiable, however, there is an effective method for finding them. We shall do so in section 8.2.

Consider the function g given by $g(x,y) = 1/xy$. It is defined on all of $\mathbb{R}^2$ except the two coordinate axes. Let X be the square with vertices at $< 1, 1 >$, $< 1, 2 >$, $< 2, 2 >$, and $< 2, 1 >$. X is a closed and bounded set, so g must achieve a maximum and a minimum relative to X. These are easily seen to occur at the vertices $< 1, 1 >$ and $< 2, 2 >$, respectively. Note that on the closed but unbounded set W where $x \geq 1$, $y \geq 1$, g does not achieve a minimum value although its values are all positive on W. On the bounded, but not closed set U where $0 < x \leq 1$, $0 < y \leq 1$, g does not achieve a maximum value.

After theorem 7.1.10 we remarked that the convergence of a sequence in a finite-dimensional inner product space V is not affected by how we choose the inner product in V. And this is equally true for convergence in terms of other norms on V, even norms that do not come from an inner product. A similar statement is true concerning the concepts we have introduced in this section. The question of which sets are open or closed or of which points are boundary points of a given set will not be affected by a change of norms. Similarly, if we change the meaning of $\| \ \|$ in definition 8.1.6 from one norm to another, we won't affect the continuity of the function f.

However, this does <u>not</u> mean that $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is continuous if it is continuous in each variable separately. Here is an example of a discontinuous function f such that

For each fixed b, $x \longmapsto f(x,b)$ is continuous $\mathbb{R} \longrightarrow \mathbb{R}$.

For each fixed a, $y \longmapsto f(a,y)$ is continuous $\mathbb{R} \longrightarrow \mathbb{R}$.

(This is what we mean by saying f is continuous in each variable separately.)

Let

$$f(x,y) = \frac{xy}{x^2 + y^2} \qquad \text{if } < x, y > \neq < 0, 0 >$$

$$f(0,0) = 0.$$

To check that f is continuous in x for a fixed value of y, say $y = b \neq 0$, note that

$$f(x,b) = \frac{bx}{x^2 + b^2}$$

which is continuous since the denominator is never zero. When $b = 0$, we have $f(x,0) = 0$ for all x; surely a continuous function. Similarly, f is continuous in y for each fixed value of x. However, f is not continuous. It is, of course, continuous on the domain $\mathbb{R}^2 - \{ < 0, 0 > \}$, since it is defined there by a single elementary formula. But it is not continuous at the origin. For points of the form $< \alpha, \alpha >$ with $\alpha \neq 0$, $f(\alpha,\alpha) = 1/2$. Since there are such points arbitrarily near the origin, while $f(0,0) = 0$, f is not continuous at the origin. In fact, f is constant except at the origin along each line through the origin. On the line $y = mx$, $f(x,y) = m/(1 + m^2)$, except for $< x, y > = < 0, 0 >$. Hence in any neighborhood of the origin, f takes on all values between $-1/2$ and $+1/2$.

Exercises.

1. Prove that the intersection of any two open sets is open. Your proof should need no modificiation if the intersection should be empty.

2. Prove that the intersection of any two closed set is closed. Is the null-set closed?

3. Prove that a plane in three-space is closed but has no interior points. The situation is the same for a linear subspace of dimension k in an inner product space of dimension n, if $k < n$.

4. Show that what you ordinarily think of as the interior of a triangular region in the plane is indeed the set of its interior points in the sense of 8.1.2.

5. How should the definition (8.1.6) of a continuous function be modified to encompass functions with non-open domains?

6. The example of the text of a function that is discontinuous but continuous in each variable separately can be made even more surprising. Let

$$f(x,y) = \frac{xy^{1/3}}{x^2 + y^{2/3}} \qquad \text{if} \quad < x, y > \neq < 0, 0 >$$

$$f(0,0) = 0.$$

Show that this function is continuous along <u>every</u> line in the plane, but is nevertheless discontinuous at the origin. (For the latter consider points of the form $< \alpha, \alpha^3 >$.)

7. Let $\| \ \|_1$ and $\| \ \|_2$ be any two norms on a finite dimensional vector space V. There is a theorem to the effect that there must exist a constant K such that

$$(\forall v \in V) \quad \| v \|_1 \leq K \| v \|_2.$$

(This isn't easy to prove, even if V is two-dimensional. Try it!) Use this result to prove that, if $f : V \longrightarrow \mathbb{R}$ is continuous using $\| \ \|_1$ in the definition, then it is also continuous using $\| \ \|_2$. Show also that a subset S of V that is open in the sense of $\| \ \|_1$ is also open in the sense of $\| \ \|_2$.

8.2 Partial Differentiation

8.2.1 Partial derivatives. The simplest way to apply the differential calculus to a function of two variable is to keep one of the variables constant and consider the function as depending on just one variable. Suppose, for example,

$$f(x,y) = 6x^3 + x^2y + 3y^2 + \sin xy. \tag{2}$$

We can fix the value of x temporarily, say $x = 2$, and consider instead

$$48 + 4y + 3y^2 + \sin 2y. \tag{3}$$

This function is differentiable and we can calculate its derivative in the usual manner and get

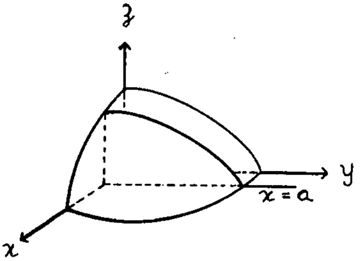
$$4 + 6y + 2 \cos 2y. \tag{4}$$

Geometrically, what we have done is to restrict the function to the line $x = 2$. If we think of the graph of the original function (2), that is the set of points $< x, y, z >$ in $\mathbb{R}^3$ such that

$$z = f(x,y)$$

which is a surface with one point above (or below) each point of the x-y plane, then we are looking at the plane $x = 2$ in $\mathbb{R}^3$. This plane cuts the surface in a curve. If we think of y and z as coordinates in this plane, the curve is the graph of the function (3), ie., the set of points $< y, z >$ such that

$$z = 48 + 4y + 3y^2 + \sin 2y.$$



Plane section of graph
(not the function of the text)

The derivative (4) tells us about the slopes of the lines tangent to the curve in the plane and we can use this information as usual. For example, since (4) is positive for $y > 0$, we know that f increases along the line $x = 2$ in the direction of increasing y. In fact, since (4) is zero for just one value of y, say y_0, is positive for $y > y_0$, and is negative for $y < y_o$, we know that, along the line $x = 2$, f has a minimum value at y_o; etc. (It is easy to see that $-\pi/4 < y_o < -2/3$.)

We could similarly analyze f along the line $x = 1$ by fixing $x = 1$; or along any other line of the form $x = a$. When we fix $x = a$, f becomes

$$6a^3 + a^2y + 3y^2 + \sin ay$$

and the derivative is given by

$$a^2 + 6y + a \cos ay.$$

Now it is clear that there is no reason to explicitly replace 'x' in (2) by 'a'. We can simply differentiate (2) directly treating x as a constant to get

$$(5) \qquad x^2 + 6y + x \cos xy$$

and we can regard this new expression as defining a new function on all of $\mathbb{R}^2$. This new function is called a partial derivative of f, in this case the partial derivative of f with respect to its second argument. We shall denote it

$$f_2' \quad \text{or} \quad D_2f.$$

Both 'f_2'' and 'D_2f' are symbols for a new function of two variables. We have

$$f_2'(x,y) = D_2f(x,y) = x^2 + 6y + x \cos xy$$

and we can substitute in this formula as we please, for example

$$f_2'(a,3) = a^2 + 18 + a \cos 3a$$

$$D_2f(0,0) = 0.$$

The partial derivative (5) is often called the partial derivative with respect to y and denoted

$$\frac{\partial f}{\partial y} \quad \text{or} \quad f_y'.$$

We shall reserve this notation for a slightly different, but closely related, situation. (See p. 8-52.)

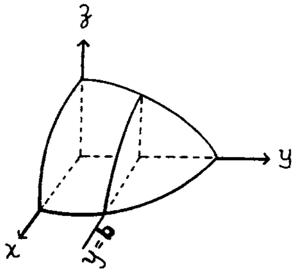
Similarly, we can treat y as a constant and differentiate with respect to x. Then we get

$$D_1 f(x,y) = f'_1(x,y) = 18x^2 + 2xy + y \cos xy,$$

called the partial derivative of f with respect to the first variable. This new partial derivative tells about the behavior of f along lines of the form $y = b$. More precisely, $f'_1(x,b)$ is the slope of a line tangent to the curve

$$z = f(x,b)$$

which we may think of as the intersection of the plane $y = b$ with the graph of f.



Plane section of graph
(not the function of the text)

Partial derivatives suffer from the fact that they look at f along lines only, and only along lines parallel to the axes at that. It is therefore a bit surprising that for the functions commonly encountered the information contained in the partial derivatives suffices to compute directly the behavior of f along any smooth curve. The criterion for this favorable situation is simply that the partial derivatives themselves are continuous functions. Any function defined by an elementary formula will fulfill this condition as long as the partial derivatives exist.

Functions of three or more variables (ie., functions from part of $\mathbb{R}^n$ to $\mathbb{R}$) are handled the same way. Partial derivatives are computed by keeping all but one of the variables constant. If f depends on three variables, we find f'_3 by keeping the first two variables constant and differentiating with respect to the third. This amounts to looking at f along lines parallel to the third coordinate axis. Suppose

$$f(x,y,z) = \sin (x^2 + y^3 + 2z).$$

Then

$$f'_1(x,y,z) = 2x \cos (x^2 + y^3 + 2z),$$

$$f'_2(x,y,z) = 3y^2 \cos (x^2 + y^3 + 2z),$$

$$f'_3(x,y,z) = 2 \cos (x^2 + y^3 + 2z).$$

Note that to compute partial derivatives of a function given explicitly by a formula you need no new rules for differentiation. Just think of all but one variable as constant while differentiating. After you have differentiated, the result is once again regarded as a function on all of the original domain (or something smaller if the derivative fails to exist at some points).

Example.

$$g(x,y) = \frac{x^{1/3}}{x + y} .$$

Here g is defined except on the line $x + y = 0$.

$$g_2'(x,y) = \frac{-x^{1/3}}{(x + y)^2}$$

for all $< x, y >$ in the original domain.

$$g_1'(x,y) = \frac{1}{3}\frac{x^{-2/3}}{x + y} - \frac{x^{1/3}}{(x + y)^2}$$

and now we must exclude all points of the second axis, $x = 0$, because of the factor $x^{-2/3}$. The first partial derivative of g does not exist at these points.

Exercises. Discuss the domains on which the following formulas define functions and calculate their several partial derivatives.

1. $f(x,y) = e^{x^2 - y^2}$

2. $g(x,y) = x \cos y - y^2$

3. $h(x,y) = \frac{xy}{x^2 + y^2}$

4. $f(x,y,z) = \sin (xy^2z^3)$

5. $g(x,y,z) = e^{x \cos y} \log (z + x^2)$

6. $h(x,y,z) = \sqrt[3]{xyz}$ (Discuss also where the partial derivatives are defined.)

7. If $f(x,y) = \frac{x^2}{x + y}$, calculate f_1', f_2', f_{12}'' $(= (f_1')_2')$, and f_{21}''.

8. Let V be the linear space of all functions from $\mathbb{R}^2$ to $\mathbb{R}$. Show that the set W of those functions f for which $D_1 f$ is everywhere defined is a linear subspace of V and that $D_1 : W \rightarrow V$ is a linear operator.

8.2.6 Higher order partial derivatives. The partial derivatives just considered are called collectively first-order partial derivatives. Since the first-order partial derivatives of a function $f : E \longrightarrow \mathbb{R}$ are themselves functions from E to $\mathbb{R}$, we can differentiate them again to get what are called second-order partial derivatives. If f is a function of two variables, we shall have two first-order partial derivatives, f_1' and f_2', and these will have two partial derivatives each making four second-order partial derivatives of f, namely

$$f_{11}'' = (f_1')_1', \ f_{12}'', \ f_{21}'', \ f_{22}''.$$

Then there will be eight third-order partial derivatives, etc. A function of three real variables will have three first-order partials, nine second-order, 27 third-order, etc.

If f is a polynomial function, it is easy to see that $f_{12}'' = f_{21}''$, and a little experimentation with familiar functions suggests that this is usually the case. Indeed so. In 8.3.22 we shall prove a theorem to the effect that

$$f_{12}'' = f_{21}''$$

whenever these functions are continuous. Examples can be given for which this equation fails, but it is valid for all functions given by elementary formulas as long as the derivatives exist.

If f is a function of more than two variables, the theorem just mentioned is still applicable because f_{12}'' and f_{21}'' are both computed by keeping fixed the variables numbered $3, 4, \ldots, n$. If f_{12}'' should exist and be continuous when the variables are allowed to vary over the whole domain of f, then it will also be continuous when only the first two variables are allowed to vary.

The actual numbering to the variables doesn't matter, so we have, for example,

$$f_{34}'' = f_{43}'', \quad f_{25}'' = f_{52}'', \text{ etc.},$$

provided these derivatives are all continuous.

The theorem applies as well to partial derivatives of orders higher than the second. If the partial derivatives of f through order three are all continuous on the domain of f, then

$$f'''_{123} = f'''_{132} = f'''_{312} = f'''_{321} = f'''_{231} = f'''_{213}.$$

The first of these equations follows from applying the theorem for second-order partials to f'_1. Similarly for the third and fifth equalities. Since $f''_{13} = f''_{31}$ throughout the domain of f we must have $(f''_{13})'_2 = (f''_{31})'_2$, that is, $f'''_{132} = f'''_{312}$. Similarly, $f'''_{321} = f'''_{231}$.

Suppose E is an open subset of $\mathbb{R}^n$. A function $f : E \longrightarrow \mathbb{R}$ is said to be a function of class C^k if and only if all of its partial derivatives through order k exist at each point of E and are continuous on E. For such a function the order of successive differentiations is immaterial until k differentiations have been performed. Counting all ways of doing the differentiation, there are n^k partial derivatives of order k, but the equalities cut this down to $n(n+1)(n+2)\cdots(n+k-1)/k!$ distinct ones.

A function is of class C^∞ if and only if all its partial derivatives of every order exist (in which case they must be continuous).

The symbols 'C^k' and 'C^∞' are also used to denote the set of all functions of class C^k or C^∞. These sets are linear subspaces of the set of all functions from E to $\mathbb{R}$.

All of the so-called elementary functions (those compounded with addition, subtraction, multiplication, division, radicals, exponentials, logarithms, and trigonometric functions) are C^∞ except for "thin" sets in their domain. If no radicals are involved there are no exceptional points at all. For example, $\log(\sin x + e^{yz})$ is defined for those $< x, y, z >$ for which $\sin x + e^{yz} > 0$ and is C^∞ on the whole of this domain. Fractional power functions are not infinitely differentiable at zero. Consequently, compounds involving radicals need not be differentiable at points where a radicand is zero.

Exercises.

1. Calculate the four (three different) second-order partial derivatives of the functions having formulas

(a) $x^3e^x \sin y$ (b) $(x + y) \cos (x - y)$ (c) $x^2 + y^2$

2. Show that the functions with formulas $\log (x^2 + y^2)$, $\arctan \frac{y}{x}$, and $e^x \cos y$ are all solutions of the second-order partial differential equation

$$f''_{11} + f''_{22} = 0,$$

known as Laplace's equation.

3. Suppose g is a particular solution of the linear partial differential equation with constant coefficients

$$af''_{11} + bf''_{12} + cf''_{22} + rf'_1 + sf'_2 + tf = 0.$$

Show that g'_1 and g'_2 are also solutions (assuming $g \in C^3$).

4. Prove that $f''_{12} = f''_{21}$ if f is a polynomial function from $\mathbb{R}^2$ to $\mathbb{R}$.

5. Show that the following function f is C^1 and its mixed partial derivatives (f''_{12} and f''_{21}) exist everywhere, but $f''_{12}(0,0) \neq f''_{21}(0,0)$.

$$f(x,y) = \frac{x^3y}{x^2 + y^2} \quad \text{if} \quad < x, y > \neq < 0, 0 >$$

$$f(0,0) = 0$$

6. Can there be a C^2-function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ such that $f'_1(x,y) = \cos (1 + x^2y)$ and $f'_2(x,y) = \sin (1 + x^2y)$? Can there be a C^2- function $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ such that

$$g'_1(x,y) = y \sin x^2y^2 \quad \text{and} \quad g'_2(x,y) = x \sin x^2y^2 \text{ ?}$$

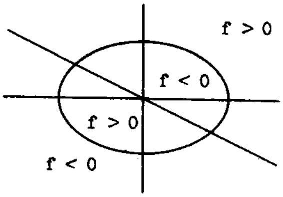
8.2.7 Maxima and minima. Consider the function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ given by

$$f(x,y) = (x^2 + 2y^2 - 9)(x + 2y).$$

It is clear that f is zero on the ellipse whose equation is

$$x^2 + 2y^2 = 9$$

and on the line $x + 2y = 0$. These curves divide the plane into four regions. The sign of f in each of these regions is shown on the sketch. Somwhere in the lower left half-elliptical region f must achieve a maximum, and somewhere in the upper right half it must achieve a minimum. (See the discussion on page 8-9.) Where are these points and what are these maximum and minimum values of f?



Suppose the maximum is achieved at $< a, b >$. Then, considered only along the line $x = a$, f must achieve a maximum for $y = b$. Therefore, the derivative of f along this line must vanish at $y = b$; that is

$$f_2'(a,b) = 0.$$

Similarly, f must have a maximum for $x = a$ along the line $y = b$, so

$$f_1'(a,b) = 0.$$

Thus we have two conditions that the unknown values a and b must satisfy and we may expect that these conditions will essentially determine a and b. Since

$$f_1'(x,y) = 3x^2 + 4xy + 2y^2 - 9$$

and

$$f_2'(x,y) = 2x^2 + 4xy + 12y^2 - 18,$$

the equations for a and b are

(8)
$$\begin{aligned} 3a^2 + 4ab + 2b^2 - 9 &= 0 \\ 2a^2 + 4ab + 12b^2 - 18 &= 0. \end{aligned}$$

If we double the first of these equations and subtract the second, we get

$$4a^2 + 4ab - 8b^2 = 0$$

that is,

$$4(a + 2b)(a - b) = 0.$$

Hence either $a = -2b$, in which case (substituting back in (8))

$$< a, b > = < \sqrt{6}, -\frac{1}{2}\sqrt{6} > \text{ or } < -\sqrt{6}, \frac{1}{2}\sqrt{6} >,$$

or $a = b$, in which case

$$< a, b > = < 1, 1 > \text{ or } < -1, -1 >.$$

The first two of these points are the points where the line $x + 2y = 0$ crosses the ellipse, so they are not the points we want. The others fall one in the lower left half-ellipse, the other in the upper right half. Since there is only one point in the lower left half at which both f_1' and f_2' vanish, this point must be the maximum point. Hence the largest value taken by f on the elliptical region is $f(-1,-1) = 18$.

The same argument shows that the minimum value of f on the upper right half-ellipse occurs also at a point at which both f_1' and f_2' vanish, so it must be at $< 1, 1 >$ and the minimum value of f on the elliptical region is $f(1,1) = -18$.

The argument we have just used is quite general and it leads to a solution of many maximum and minimum problems in several dimensions. We formalize the ideas for reference.

8.2.9 Definition. Suppose E is an open set in $\mathbb{R}^n$ and f is a function from E to $\mathbb{R}$. The point $p \in E$ is a local minimum point for f if and only if there is a neighborhood N of p such that

$$f(p) \leq f(q) \text{ for all } q \in N.$$

The point $r \in E$ is a local maximum point for f if and only if there is a neighborhood U of r such that

$$f(r) \geq f(q) \text{ for all } q \in U.$$

This means that p is a minimum point in comparison with its immediate neighbors only. It is quite possible that f takes a smaller value at some point s remote from p. An actual minimum point for f must also be a local minimum point, but not vice versa.

8.2.10 Theorem. If E is an open set in $\mathbb{R}^n$ and the function $f : E \to \mathbb{R}$ has a local minimum (or maximum) at $p \in E$ and if the partial derivatives $f_1', f_2', \ldots, f_n'$ exist at p, then these partial derivatives vanish at p.

The theorem gives us an often effective device for finding maximum and minimum points. Find the points, called critical points, at which all the first-order partial derivatives vanish. Usually there will be only a finite number of these and the required maximum or minimum point will be among them. There are of course other possibilities for the maximum or minimum. They might occur at some point where one of the partial derivatives fails to exist. Moreover, there is always the possibility that there is no maximum or minimum point. Recall Theorem 8.1.10 which guarantees the existence of a maximum when the domain considered is bounded and closed. The maximum might occur at a boundary point of the domain. At a boundary point Theorem 8.2.10 does not apply, since p need not be a maximum point in comparison with all of its immediate neighbors in every direction. However, we can be sure that either the maximum occurs in the interior (in which case Theorem 8.2.10 does apply) or at a boundary point. This is a direct generalization of the familiar case of a function of one variable. The maximum value of a differentiable function f over a bounded closed interval $[a,b]$ in $\mathbb{R}$ occurs either at an interior point c in which case $f'(c) = 0$ or at a boundary point, ie., either at a or at b.

Searching the boundary for a maximum or a minimum requires a different approach. We illustrate by an example.

Example. Find the maximum and minimum values of

$$f(x,y) = 23x^2 + 72xy + 2y^2$$

on the set where $x^2 + y^2 \leq 1$.

Since the set to be searched is bounded and closed, we are certain that maximum and minimum points exist. The partial derivatives are

$$f_1'(x,y) = 46x + 72y$$

$$f_2'(x,y) = 72x + 4y.$$

These are both zero only at $< 0, 0 >$. So if the maximum or the minimum value of f occurs at an interior point of the unit disk, it must be at the origin. However, $f(0,0) = 0$, $f(1,0) = 23$, and $f(1/2, - 1/2) = -47/4$, so it is clear that the origin is neither a maximum or a minimum point. These points must occur on the boundary.

Since the boundary is a smooth curve, we choose a parametrization for it, say $t \longmapsto < \cos t, \sin t >$. The required points must correspond to some value of t, so we consider the function g where

$$g(t) = f(\cos t, \sin t) = 23 \cos^2 t + 72 \sin t \cos t + 2 \sin^2 t \;.$$

$$= \frac{25}{2} + \frac{21}{2} \cos 2t + 36 \sin 2t.$$

We can find the maximum and minimum values of g by the familiar one-variable method.

$$g'(t) = - 21 \sin 2t + 72 \cos 2t.$$

This vanishes if $\tan 2t = 24/7$, that is

$$\frac{2 \tan t}{1 - \tan^2 t} = \frac{24}{7}$$

whence $\tan t = 3/4$ or $- 4/3$. Correspondingly,

$$< \cos t, \sin t > = \pm < \frac{4}{5}, \frac{3}{5} > \quad \text{or}$$

$$= \pm < \frac{3}{5}, - \frac{4}{5} >$$

Then f achieves it maximum value, 50, on the disk at the two boundary points $\pm < 4/5. 3/5 >$ and its minimum value, -25, at the points $\pm < 4/5, -3/5 >$.

There is another method, called the Lagrange multiplier method, that is useful in finding maxima or minima along curved boundaries. We shall study it in section 8.5. Compare our work with that of section 6.3.

In section 8.4 we shall take up a method for deciding (in most cases) whether a critical point is a local maximum point or a local minimum point. It is analogous to the second derivative test for functions of one variable.

Exercises.

1. Find the maximum and minimum values of

$$(x - 2y)(x^2 + y^2 - 15)$$

on the disk $x^2 + y^2 \leq 15$.

2. Find the maximum and minimum values of

$$x^2y^3(1 - x - y)$$

on the triangular domain where $x \geq 0$, $y \geq 0$, and $x + y \leq 1$.

3. Find the maximum and minimum values of

$$x^3y^2z(1 - x - y - z)$$

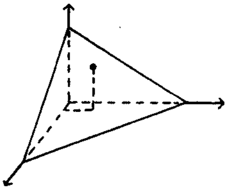
on the tetrahedral domain in three space where $x \geq 0$, $y \geq 0$, $z \geq 0$, and $x + y + z \leq 0$.

4. Find the points of the surface

$$z = \frac{1}{xy}$$

that are closest to the origin.

5. Consider all planes in three-space that pass through $< 1, 2, 3 >$ and meet the positive x-, y-, and z-axes. Which of these cuts off the tetrahedron of least volume ?



6. What are the maximum and minimum values of

$$x^2 + xy - y^2$$

on the unit square (the set where $0 \leq x \leq 1$ and $0 \leq y \leq 1$) ?

On the unit disk (the set where $x^2 + y^2 \leq 1$) ?

8.3 The total derivative or differential of a function from V to $\mathbb{R}$.

The basic idea of differentiation in higher dimensions is the same as it is in one dimension. Geometrically, it is drawing a tangent to the graph of the function. Analytically, it is approximating the function by a function of first degree. The tangent to a surface in three-space will be a plane, not a line, and in higher dimensions it will be a hyperplane or "flat" space of higher dimension. The dimension of the tangent to a surface of high dimension will always be the same as that of the surface. Of all the higher dimensional cases the only one we can visualize is that of a function from $\mathbb{R}^2$ to $\mathbb{R}$. We shall begin with a primarily geometric treatment of that case. Then we shall derive the same results more rigorously using analysis. The analytic argument has the clear advantage that it is applicable immediately to any dimension. The geometric approach, on the other hand, can provide us with images that make the analytic formulas highly plausible.

8.3.1 Inhomogeneous linear functions. If V is a vector space, the function $g : V \longrightarrow \mathbb{R}$ is called inhomogeneous linear or a function of the first degree if and only if it can be represented as the sum of a constant and a linear functional $\bar{g} : V \longrightarrow \mathbb{R}$. If $V = \mathbb{R}^2$, this means

$$g(x,y) = \alpha + \beta x + \gamma y$$

for some numbers α, β, and γ. Here the linear functional $\bar{g}$ is given by $\bar{g}(x,y) = \beta x + \gamma y$.

The graph of such a function is a hyperplane in $V \times \mathbb{R}$. If V has dimension n, then $V \times \mathbb{R}$ has dimension $n + 1$ and the graph is a coset of a linear subspace of dimension n. For example, if $V = \mathbb{R}^2$, then $V \times \mathbb{R} = \mathbb{R}^3$. The graph of g (above) is the set of all $< x, y, z >$ such that

$$z = \alpha + \beta x + \gamma y. \tag{2}$$

This is the $< 0, 0, \alpha >$-coset of the linear space spanned by $< 1, 0, \beta >$ and $< 0, 1, \gamma >$. Conversely, any plane not containing a line parallel to the z-axis has an equation of the form (2) and hence can be regarded as the graph of an inhomogeneous linear function from $\mathbb{R}^2$ to $\mathbb{R}$. This means that any plane in $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$ that is the graph of a function is the graph of an inhomogeneous linear function.

Exercise. Prove: If $g : V \longrightarrow \mathbb{R}$ is an inhomogeneous linear function, say $g = \alpha + \bar{g}$, where $\bar{g}$ is linear, then for any $v, b \in V$

$$g(v) = g(b) + \bar{g}(v - b).$$

8.3.3 The tangent plane. Suppose that f is a function $\mathbb{R}^2 \longrightarrow \mathbb{R}$. The graph of f, that is, the set of all $< x, y, z >$ in $\mathbb{R}^3$ such that

$$z = f(x,y)$$

is a surface S in $\mathbb{R}^3$. Let q be a point of S; say $q = < x_o, y_o, z_o >$. Finally let P be the plane tangent to S at q. (We assume it is intuitively clear what a tangent plane is. Also we assume that the tangent plane exists: Not every surface has tangent planes at every point. For example, what would you mean by a plane tangent to a cone at its vertex?) We want to find the equation of P. We assume it contains no line parallel to the z-axis, so it is the graph of some inhomogeneous linear function, that is, it is given by

$$z = \alpha + \beta x + \gamma y$$

for suitable constants α, β, and γ. How do we find these numbers?

The plane $y = y_o$ cuts S in a curve C passing through q. We make the plausible assumption that the line tangent to C at q lies in the plane P. (This cannot be made into a rigorous argument without a definition of the tangent plane, and we have none as yet.) We can regard x and z as coordinates in the plane $y = y_o$. Then C is just the graph

$$z = f(x, y_o)$$

The slope of the tangent to C at q is given by the derivative for $x = x_o$, that is,

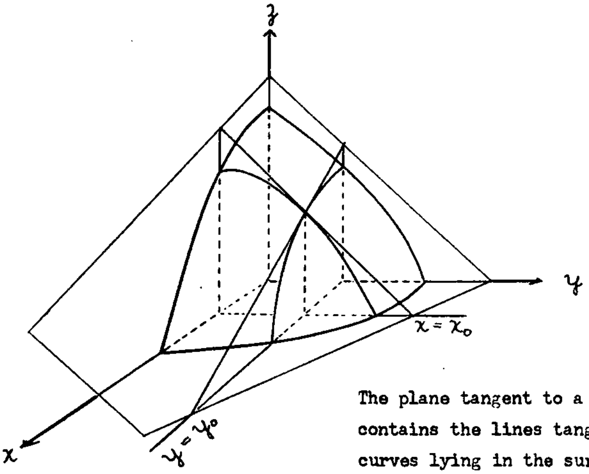
$$f_1'(x_o, y_o).$$

The equation of the tangent is therefore

(4) $$z = f_1'(x_o, y_o)(x - x_o) + z_o.$$

If (4) is interpreted in R^3, it describes a whole plane. To obtain just the tangent line we must adjoin the equation

(5) $$y = y_o.$$



The plane tangent to a surface contains the lines tangent to curves lying in the surface.

Similarly, the plane $x = x_o$ cuts the surface S in a curve Γ passing through q. Using z and y as coordinates in this plane, the equation of Γ is

$$z = f(x_o, y)$$

and the tangent to Γ at q has the equations

(6) $$z = f_2'(x_o, y_o)(y - y_o) + z_o,$$ $$x = x_o.$$

Now the plane described by

(7) $$z = f_1'(x_o, y_o)(x - x_o) + f_2'(x_o, y_o)(y - y_o) + z_o$$

contains both of the lines (4)-(5) and (6). Since there is just one plane passing through two intersecting lines in three-space, P must be the plane given by (7).

In the figure the surface S appears to be convex. Therefore it lies entirely on one side of its tangent plane. There are surfaces, however, which cross their tangent planes locally at every point. Consider, for example, a saddle. In some directions it curves towards you and in others, away. This means that the surface of the saddle lies partly on one side of its tangent plane and partly on the other.

We can look at this phenomenon analytically. A typical saddle-shaped surface is a hyperbolic paraboloid (see figure, page 6-96). An example is given by

$$z = x^2 - y^2. \tag{8}$$

The tangent plane to this surface (obtained from (7) using $f(x,y) = x^2 - y^2$) is

$$\begin{aligned} z &= 2x_0(x - x_0) - 2y_0(y - y_0) + z_0 \\ &= 2x_0x - 2y_0y - x_0^2 + y_0^2. \end{aligned}$$

(Because $< x_0, y_0, z_0 >$ is on the surface, $z_0 = x_0^2 - y_0^2$.) This plane lies below the surface (8) along the line $y = y_0$ since

$$x^2 - y_0^2 > 2x_0x - 2y_0^2 - x_0^2 + y_0^2$$

for all x except x_0. On the other hand the tangent plane lies above the surface (8) along the line $x = x_0$ since

$$x_0^2 - y^2 < 2x_0^2 - 2y_0y - x_0^2 + y^2$$

for all y except y_0.

Our derivation of (7) as the equation of the tangent plane amounts to this: If a tangent plane exists at $< x_0, y_0, z_0 >$, then (7) must be its equation; in particular, the first order partial derivatives f_1' and f_2' must exist at $< x_0, y_0 >$.

What about the converse? If $f_1'(x_0, y_0)$ and $f_2'(x_0, y_0)$ exist, will there be a tangent plane? Not necessarily. The function f of section 8.1, page 8-10, although it is not even continuous at $< 0, 0 >$, has both partial derivatives there. Since these partial derivatives are both 0, the only candidate for a tangent plane is $z = 0$ (ie., the x-y plane). We wouldn't want to call this plane tangent to the graph of f, because, for example, it seems to bear no relation at all to f over the line $x = y$. On this line f is constant 1/2 except at the origin where it is 0. Except for this one point the corresponding portion of the graph of f lies far away from our hypothetical tangent plane. This example shows us that we wouldn't want to define the tangent plane using partial derivatives and equation (7). The definition we shall adopt (in 8.3.17) says that the tangent plane at q to a surface S is a plane through q that lies *exceptionally close* to S near q. (Recall that the line tangent to a curve C through a point p is the line through p that lies closest to C in the immediate neighborhood of p.) When the tangent plane exists, it will of course be given by equation (7), but we must not expect the tangent plane to exist just because (7) makes sense.

8.3.9 The chain rule. Suppose now that $\varphi: \mathbb{R}^2 \to \mathbb{R}$ is a function whose graph is a surface S with a tangent plane at $q = < x_o, y_o, z_o >$ and that the parametric curve Γ

$$t \longmapsto < f(t), g(t), h(t) > \tag{10}$$

lies in S and passes through q at the time $t = t_o$. Analytically, this means that

$$h(t) = \varphi(f(t), g(t)) \tag{11}$$

for all t, and $f(t_o) = x_o$, $g(t_o) = y_o$, $h(t_o) = z_o$.

We think of (10) as a motion and regard its velocity vector at time t_o as emanating from q. Since this vector is tangent to Γ, it lies in the plane tangent to S at q. The parametric description of the line tangent to Γ is

$$t \longmapsto < x_o, y_o, z_o > + (t - t_o) < f'(t_o), g'(t_o), h'(t_o) >$$

and the tangent plane to S at q has the equation

$$z - z_o = \varphi_1'(x_o, y_o)(x - x_o) + \varphi_2'(x_o, y_o)(y - y_o).$$

The fact that the line tangent to Γ lies in the plane tangent to S becomes

$$(t - t_o)h'(t_o) = \varphi_1'(x_o, y_o)(x - x_o)(t - t_o)f'(t_o) + \varphi_2'(x_o, y_o)(t - t_o)g'(t_o)$$

for all t. Cancelling out the $(t - t_o)$, this is equivalent to

$$(12) \qquad h'(t_o) = \varphi_1'(x_o, y_o)f'(t_o) + \varphi_2'(x_o, y_o)g'(t_o).$$

This is a formula for the derivative of h given by (11). In fact, if f and g are any differentiable functions with $f(t_o) = x_o$, $g(t_o) = y_o$, we can define a function h by (11) and then the parametric curve Γ given by (10) will lie in S the graph of φ. Then (12) follows from our belief that the tangent to a curve in S must lie in the plane tangent to S at the same point.

Since the choice of t_o in (12) is arbitrary (except that we require that S have a tangent plane at $< f(t_o), g(t_o), h(t_o) >$), we can replace t_o by t; ie., if h is given by (11) then

$$(13) \qquad h'(t) = \varphi_1'(f(t), g(t))f'(t) + \varphi_2'(f(t), g(t))g'(t)$$

for all t. This is a two-dimensional form of the chain rule for finding the derivative of a composite function.

There is a nice interpretation of this chain rule. Suppose that a plane has been heated irregularly ao that its temperature varies from point to point. Put coordinates on the plane as usual and say that $\varphi(x,y)$ is the temperature at the point $< x, y >$. Now imagine an observer moving in the plane along the parametrized curve

$$t \longmapsto < f(t), g(t) >.$$

Then h(t) is the temperature observed at the time t. Then (13) says that the rate h'(t) of temperature variation at any time is the sum of two contributions, one $\varphi_1'(f(t), g(t))f'(t)$ due to the component of the velocity in the x-direction and another $\varphi_2'(f(t), g(t))\cdot g'(t)$ due to the component in the y-direction.

Let us describe what we have done in more general terms. Given a function $F : \mathbb{R} \longrightarrow \mathbb{R}^2$ and a function $\varphi : \mathbb{R}^2 \longrightarrow \mathbb{R}$, we form the composite function

$$\varphi \circ F : \mathbb{R}^2 \longrightarrow \mathbb{R}.$$

Granting that φ and F are differentiable, we would like a formula for the derivative of $\varphi \circ F$.

We can express F with component functions f and g, that is

$$F(t) = < f(t), g(t) >.$$

Then $(\varphi \circ F)(t) = \varphi(f(t), g(t)) = h(t)$, so $\varphi \circ F = h$. Thus (13) is the desired formula for the derivative of $\varphi \circ F$.

Note that (13) is a direct generalization of the chain rule for the derivative of a composite function $k = \psi \circ f$ where ψ and f are both functions from $\mathbb{R}$ to $\mathbb{R}$. According to the usual chain rule

$$k'(t) = \psi'(f(t))f'(t).$$

Equation (13) is much like this one, the principal difference being that (13) has two terms corresponding to the two arguments of φ. It is easy to guess the formula for the derivative of a composite function of the form

$$\mathbb{R} \longrightarrow \mathbb{R}^n \longrightarrow \mathbb{R}.$$

If the first step is given by

$$t \longmapsto < f_1(t), f_2(t), \ldots, f_n(t) >$$

and the second is φ, then the composite is

$$h(t) = \varphi(f_1(t), f_2(t), \ldots, f_n(t))$$

and the derivative is given by

$$\begin{aligned} h'(t) = {} & (D_1\varphi)(f_1(t), f_2(t), \ldots, f_n(t))f_1'(t) + \\ & (D_2\varphi)(f_1(t), f_2(t), \ldots, f_n(t))f_2'(t) + \\ & \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad + \\ & (D_n\varphi)(f_1(t), f_2(t), \ldots, f_n(t))f_n'(t) \end{aligned} \tag{14}$$

Note that here f_1' means the ordinary derivative of the function f_1. Since this function has only one argument, it doesn't have partial derivatives. We have used the D_i-notation for the i-th partial derivative of φ to reduce the possibility of confusion in the interpretation of subscripts.

8.3.15 Approximation by inhomogeneous linear functions. Now we shall look at the same ideas in analytic terms. Instead of finding a plane tangent to the graph of f, we find an inhomogeneous linear function that approximates f very well.

Suppose to start with that f is a function from $\mathbb{R}^2$ to $\mathbb{R}$, say,

$$f(x,y) = 2xy + 2x - y + 5.$$

This is a continuous function so, near the origin, it can be approximated by its value $f(0,0) = 5$. The error in this approximation will generally be of the same order of magnitude as the deviation of the argument from $< 0, 0 >$. For example, if $x = y = .001$, then

$$|f(x,y) - 5| = |5.001002 - 5| = .001002$$

and this is about the same size as $\| < x, y > \| = (.001)\sqrt{2}$.

However, if we include the first degree terms in our approximation, that is if we approximate f by g where $g(x,y) = 5 + 2x - y$, then the error

$$|f(x,y) - g(x,y)|$$

is a good deal smaller than $\| < x, y > \|$. For example, as long as $\| < x, y > \| < .001$,

$$\begin{aligned} |f(x,y) - g(x,y)| &= |2xy| = 2|x|\cdot|y| \\ &\leq 2(.001)\| < x, y > \| \end{aligned}$$

(because $|x| < .001$ and $|y| < \| < x, y > \|$).

Suppose instead we wanted to approximate f near $< 1, 1 >$. Since $f(1,1) = 8$, we can approximate f roughly near $< 1, 1 >$ by the constant function 8. We can do much better if we allow some first degree terms. Since

$$f(x,y) = 8 + 4(x - 1) + (y - 1) + 2(x - 1)(y - 1)$$

we take

$$h(x,y) = 8 + 4(x - 1) + (y - 1).$$

If we use h in place of f for $< x, y >$ near $< 1, 1 >$, the error will be small relative to $\| < x, y > - < 1, 1 > \|$. In fact,

$$\begin{aligned}|f(x,y) - h(x,y)| &= |2(x - 1)(y - 1)| \\ &\leq 2\delta \| < x, y > - < 1, 1 > \|,\end{aligned}$$

provided $\| < x, y > - < 1, 1 > \| < \delta$.

Start again with a new function, say

$$f(x,y) = xy^2,$$

and suppose we want to approximate f near $< 1, -1 >$. Temporarily put $x = 1 + u$, $y = - 1 + v$. Then

$$\begin{aligned}f(x,y) &= (1 + u)(- 1 + v)^2 \\ &= 1 + u - 2v - 2uv + v^2 + uv^2.\end{aligned}$$

We take as approximator

$$g(x,y) = 1 + u - 2v = 1 + (x - 1) - 2(y + 1).$$

The error in our approximation will be

$$\begin{aligned}|f(x,y) - g(x,y)| &= | - 2uv + v^2 + uv^2| \\ &\leq 2|u|\cdot|v| + |v|^2 + |u|\cdot|v|^2 \\ &\leq 4\delta \| < u, v > \|\end{aligned}$$

provided $\| < u, v > \| < \delta < 1$. We can rewrite this

$$|f(x,y) - g(x,y)| \leq 4\delta \| < x, y > - < 1, -1 > \|$$

as long as $\| < x, y > - < 1, -1 > \| < \delta < 1$. Again the error is small relative to the deviation of the argument from the chosen point $< 1, -1 >$, at least for small deviations.

It should be clear how we can find similar approximations for any polynomial function near any prescribed point. For more complicated functions the existence of good first degree approximations is less obvious; in fact there are functions which cannot be so approximated near some points. We take the existence of such an approximation as the definition of differentiability.

Let f be a real-valued function defined on some neighborhood of a point $< a, b >$ in $\mathbb{R}^2$. We say that f is <u>differentiable at</u> $< a, b >$ if and only if there exists an inhomogeneous linear function $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ such that

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x,y) \quad \| < x, y > - < a, b > \| < \delta \implies \tag{16}$$
$$|f(x,y) - g(x,y)| \leq \varepsilon \| < x, y > - < a, b > \|.$$

Compare this with 5), page 7-5.

There can be at most one such function g. (The proof is left to you. There are suggestions in exercise 6.)

The connection between this definition and our previous work lies in the following fact.

8.3.17 The plane tangent to the graph of f at the point $< a, b, f(a,b) >$ is the graph of g.

There is no way we can prove this statement, because we still have no definition of the tangent plane. In fact we shall adopt this statement as the definition of the tangent plane.

This begins to sound like some sort of logical hocus-pocus. In a sense we are free to define <u>tangent plane</u> as we please, but this is not really so. The point is that the analytic definition of tangent plane does indeed capture the geometric idea of tangent plane. We may not be able to prove it, but you should convince yourself that it is true. We have introduced the idea of tangent planes into this analytical discussion just to build a bridge to the more intuitive, but for many more vivid, realm of geometry. Technical proofs about differentiation will all be carried out in analytical terms using (16). As far as proofs are concerned, we could as well omit all references to geometry.

Our next step is to show by analytical argument that the function g of (16) is given by

$$g(x,y) = f(a,b) + f_1'(a,b)(x - a) + f_2'(a,b)(y - b). \tag{18}$$

This is the analytical version of the geometrical argument leading to (7).

To begin with we know that

$$g(x,y) = \alpha + \beta x + \gamma y$$

for some α, β, and γ. If (16) is true,. then $g(a,b)$ must be $f(a,b)$. (No matter what ϵ and δ are, the choice $x = a$, $y = b$ fulfills the condition $\| < x, y > - < a, b > \| < \delta$, so $|f(a,b) - g(a,b)| \leq \epsilon \cdot 0$.)
Hence

$$f(a,b) = \alpha + \beta a + \gamma b,$$

so we can write

$$g(x,y) = f(a,b) + \beta(x - a) + \gamma(y - b).$$

It remains to show that $\beta = f_1'(a,b)$ and $\gamma = f_2'(a,b)$.

To this end consider pairs $< x, y >$ of the form $< a + h, b >$. Given $\epsilon > 0$ and the corresponding δ from (16), if $0 < |h| < \delta$, we shall have $\| < a + h, b > - < a, b > \| = |h| < \delta$ and therefore

$$|f(a+h,b) - g(a+h,b)| \leq \epsilon |h|.$$

Substituting the value of $g(a+h,b)$ from (18) and dividing through by $|h|$,

$$\left| \frac{f(a+h,b) - f(a,b)}{h} - \beta \right| \leq \epsilon .$$

But this says precisely that β is the derivative at a of the function

$$x \longmapsto f(x,b),$$

and this is exactly $f_1'(a,b)$. Thus $\beta = f_1'(a,b)$. Similarly, $\gamma = f_2'(a,b)$.

We come now to the chain rule. If

$$h(t) = \varphi(f(t), g(t))$$

then

$$h'(t) = \varphi_1'(f(t), g(t))f'(t) + \varphi_2'(f(t), g(t))g'(t),$$

provided that f, g, and φ are differentiable (f and g in the usual sense, φ in the sense of (16)). This can be proved analytically, but the argument is quite lengthy and we defer it to p. 8-48.

There is an important special case of the chain rule. Suppose $t \longmapsto \langle f(t), g(t) \rangle$ is a uniform rectilinear motion; that is, $f(t) = a + tu$, $g(t) = b + tv$, where a, b, u, and v are constants. Then

$$h(t) = \varphi(a + tu, b + tv)$$

and

$$h'(0) = \varphi_1'(a,b)u + \varphi_2'(a,b)v$$

is called the *derivative of* φ *at* $\langle a, b \rangle$ *along the vector* $\langle u, v \rangle$. The partial derivative $\varphi_1'(a,b)$ is just the special case $u = 1$, $v = 0$. When $\langle u, v \rangle$ is a unit vector, we can think of the parameter t as representing distance instead of time. Hence, in this case, the derivative of φ along $\langle u, v \rangle$ can be interpreted as the rate of change of φ with respect to distance in a certain direction. Such a derivative is often called a *directional derivative*.

The derivative of φ at a fixed point $\langle a, b \rangle$ along a vector $\langle u, v \rangle$ depends linearly on the choice of v. This focusses our attention on the linear operator

$$\langle u, v \rangle \longmapsto \varphi_1'(a,b)u + \varphi_2'(a,b)v$$

from $\mathbb{R}^2$ to $\mathbb{R}$. This linear operator is called the *differential of* φ *at* $\langle a, b \rangle$. We denote it $d\varphi(a,b)$. Since we have coordinates, we can conveniently represent it by the row vector

$$\| \varphi_1'(a,b) \quad \varphi_2'(a,b) \|.$$

Since there is a row vector at every point $\langle a, b \rangle$, there is a function

$$d\varphi : \langle a, b \rangle \longmapsto \| \varphi_1'(a,b) \quad \varphi_2'(a,b) \|.$$

This function $d\varphi$, called the *differential of* φ, is defined on the domain of φ (assuming, of course, that φ is differentiable at each point) and its values are row vectors. Such a function is called a *covector field* or a *differential form*. (See also §7.4) The components of $d\varphi$ are the first order partial derivatives of φ in order.

Think of $< u, v >$ as the column vector $\left\| \begin{matrix} u \\ v \end{matrix} \right\|$. Then the derivative of φ along $\left\| \begin{matrix} u \\ v \end{matrix} \right\|$ at $< a, b >$ is

$$d\varphi(a,b) \cdot \left\| \begin{matrix} u \\ v \end{matrix} \right\|,$$

an ordinary matrix product since $d\varphi(a,b)$ is a row vector of length 2.

With this notation we can express the chain rule very neatly. Suppose E is an open set in $\mathbb{R}^2$ and $\varphi : E \to \mathbb{R}$ is a differentiable function. Suppose I is an interval in $\mathbb{R}$ and $F : I \longrightarrow E$ is differentiable. Then $\varphi \circ F : I \to \mathbb{R}$ is differentiable and its derivative is

$$(\varphi \circ F)' = d\varphi \cdot F'.$$

(Recall that F' is a column vector.) We must know where to evaluate these vectors. F' is to be evaluated at t and $d\varphi$ at $F(t)$.

8.3.19 Approximate calculation. Look back at our definition of differentiability. The essential point is that a function is differentiable at $< a, b >$ if it can be well approximated by a suitable function of degree one near $< a, b >$. When it exists this first degree approximation is given by (18) which we can write

$$g(x,y) = f(a,b) + df(a,b) \cdot \left\| \begin{matrix} x-a \\ y-b \end{matrix} \right\|.$$

Note that $df(a,b)$ is the first degree part of the approximating function g.

To focus on the approximation aspect we might write

$$f(x,y) \sim f(a,b) + df(a,b) \cdot \left\| \begin{matrix} x-a \\ y-b \end{matrix} \right\| \tag{20}$$

provided $\left\| \begin{matrix} x-a \\ y-b \end{matrix} \right\|$ is small. So far the view we have taken of this approximation has been that we know $f(x,y)$ and $f(a,b)$ and the approximation condition (16) determines the row vector $df(a,b)$. But we often look at the situation the other way about. We know $f(a,b)$ and $df(a,b)$ and we use (20) to estimate $f(x,y)$.

Suppose, for example, that $f(x,y) = x\sqrt{x + y^2}$. Then $f(9,4) = 45$. Suppose we want to estimate $f(9.1, 3.9)$. We have

$$df(x,y) = \left\| \sqrt{x + y^2} + \frac{x}{2\sqrt{x + y^2}} \quad \frac{xy}{\sqrt{x + y^2}} \right\|$$

$$df(9,4) = \| 5.9 \quad 7.2 \|$$

Hence

$$f(9.1,\ 3.9) \sim 45 + \| 5.9 \quad 7.2 \| \cdot \left\| \begin{matrix} 0.1 \\ -0.1 \end{matrix} \right\| = 44.87$$

The true value is 44.8677 to four decimals.

If we had used the familiar one-dimensional approximation to estimate $f(9.1,\ 4)$ we would have considered

$$g(x) = x\sqrt{x + 16}.$$

Then $g'(9) = 5.9$, so $g(9.1) \sim 45 + (5.9)(0.1)$. To obtain $f(9,\ 3.9)$ we would have considered

$$h(y) = 9\sqrt{9 + y^2},$$

$h'(4) = 7.2$, $h(3.9) \quad 45 + (7.2)(-0.1)$. Note that the two-dimensional approximation procedure simply accumulates the two changes due to small variations in the two arguments. This is quite a general fact. Small changes in the value of a function due to small changes in the arguments are additive in the first approximation. This simply reflects the fact that, in the first approximation, the change in value is linear in the argument changes.

The two-dimensional approximation suffers from the same disadvantage that the one-dimensional one does. There is no estimate of the size of the error in the approximation. It is possible to get such estimates using second derivatives and we shall do this in §8.4.

Exercises.

1. Find the equations of the planes tangent to the following surfaces in $\mathbb{R}^3$ at the points indicated.

   (a) $xyz = 6$ at $< 1, 2, 3 >$
   (b) $z = x/y$ at $< 6, 2, 3 >$
   (c) $z = e^{x-y^2}$ at $< 1, 1, 1 >$
   (d) $x^2 + 2y^2 - 3z^2 = 3$ at $< 2, 1, 1 >$
   (e) $z = \sin xy$ at $< 0, 2, 0 >$
   (f) $z = x^3 -2xy + y^4$ at $< 1, 1, 0 >$

2. Planes are drawn tangent to the surface given by $z = xy + 2y + x^3$ at the points $< 0, 0, 0 >$ and $< 1, 1, 4 >$. At what angle do they intersect?

3. The sphere $x^2 + y^2 + z^2 = 6$ and the surface $z = 2xy$ meet in a curve. What line is tangent to this curve at the point $< 1, 1, 2 >$?

4. Justify the rule that the relative error in a product is approximately the sum of the relative errors in the factors. (The relative error means the error divided by the true value.)

5. There is a differentiable function f defined near $< 1, 1 >$ such that $f(1,1) = 1$ and for all $< x, y >$, $z = f(x,y)$ is a solution of

$$z^5 - xyz^2 + xz - y = 0.$$

   Find the partial derivatives of f at $< 1, 1 >$ and use them to estimate $f(1.1, 0.9)$.

6. Show that if h is an inhomogeneous linear function that satisfies

$$(\forall \varepsilon > 0)(\exists \delta > 0) \quad \|< x, y >\| < \delta \implies |h(x,y)| \leq \varepsilon \|< x, y >\|$$

   then h is everywhere zero. Use this to prove that there can be at most one inhomogeneous linear function g that satisfies (16). Note that there is no loss of generality to take $< a, b > = < 0, 0 >$.

7. Prove the formula $d(fg) = fdg + gdf$, where f and g are differentiable functions from $\mathbb{R}^2$ to $\mathbb{R}$. Note that fdg is pointwise the product of a scalar and a row vector.

We shall now state the definitions and prove the principal theorems concerning the differentiation of functions from an n-dimensional inner product space to $\mathbb{R}$.

8.3.21 Definition Let E be an open set in a finite dimensional inner product space V. Let $a \in V$ and let f be a function from E to $\mathbf{R}$. Then f is said to be differentiable at a if and only if there is a linear functional $h : V \longrightarrow \mathbb{R}$ such that

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall v \in V) \quad \|v - a\| < \delta \implies$$

$$| f(v) - f(a) - h[v-a] | \leq \quad \|v - a\|$$

Here and subsequently we have used [ ] to indicate where a linear functional is acting on a vector. Before continuing the definition it is important to have in mind the fact that the linear functional h, if it exists at all, is unique. The proof of this fact in the general case is essentially the same as in the two-dimensional case, so we omit it. See exercise 6, page 8-39. The linear functional h is called the differential of f at a. We shall usually write it $df(a)$. If $V = \mathbb{R}^n$ or if V has dimension n and a linear coordinate system has been introduced, we shall take $df(a)$ to be a row vector of length n. In this case $h[v-a]$ means the matrix product of the row vector h and the column vector $v - a$.

The function f is said to be differentiable if and only if it is differentiable at each point of E. In this case df the differential of f is a function from E to V^* or to the set of row vectors.

This is an extension of definition (16) for $\mathbb{R}^2$ written in vector notation. The inhomogeneous linear function there is

$$g(v) = f(a) + h[v - a].$$

When f is differentiable and V is $\mathbb{R}^n$, it follows from the same arguments as on page 8-35 that

$$df(a) = \| f_1'(a) \quad f_2'(a) \quad \cdots \quad f_n'(a) \|$$

Hence the function df is the row vector of partial derivatives of f:

$$df = \| f_1' \quad f_2' \quad \cdots \quad f_n' \|.$$

We have seen that, even in dimension two, the existence of its partial derivatives is not sufficient to guarantee that f is differentiable. However, the following theorem gives us a way of checking that f is differentiable by inspecting its partial derivatives. If they are continuous, f is differentiable. This criterion shows immediately that any function given by a single formula in the coordinates involving only differentiable functions is itself differentiable. (Radicals can cause trouble at a point where a radicand is zero.) Hence the question of differentiability can be ignored in the vast majority of cases.

8.3.22 *Theorem*. *Let* E *be an open set in* $\mathbb{R}^n$. *Let* $a \in E$ *and let* f *be a function from* E *to* $\mathbb{R}$. *Suppose that the partial derivatives* $f_1', f_2', \ldots, f_n'$ *are defined in a neighborhood of* a *and are continuous at* a. *Then* f *is differentiable at* a *and*

$$df(a) = \| f_1(a) \quad f_2(a) \quad \cdots \quad f_n(a) \|.$$

Proof. We shall give the proof only for $n = 2$. The proof for larger values of n involves no additional ideas. We shall also assume that $a = < 0, 0 >$. This is no real loss of generality, but it makes the formulas look a lot less complicated.

The idea of the proof is to estimate the difference between f and the alleged good linear approximation of f using the mean value theorem. Since the mean value theorem, in the form that we know it, applies only to functions of one variable, we break the difference into two parts in each of which only one argument actually varies. (If the proof were for $\mathbb{R}^n$ there would be n parts.)

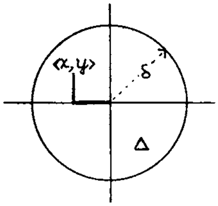
We must estimate the difference

$$D = f(x,y) - f(0,0) - xf_1'(0,0) - yf_2'(0,0).$$

We have to show that $|D|$ is appropriately small whenever $\|< x, y >\|$ is sufficiently small.

First we represent D as the sum

$$D = f(x,y) - f(x,0) - yf_2'(0,0)$$
$$+ f(x,0) - f(0,0) - xf_1'(0,0)$$



corresponding to steps along the broken line shown. We shall show that these parts are separately small.

Let a positive ϵ be given. Choose $\delta > 0$ so small that

(23) The partial derivatives f_1' and f_2' exist at all points of the disk Δ of radius δ about $a = < 0, 0 >$, and

(24) For any $< u, v > \in \Delta$,

$$|f_1(u,v) - f_1(0,0)| < \epsilon/2 \quad \text{and}$$
$$|f_2(u,v) - f_2(0,0)| < \epsilon/2.$$

We can do this because the derivatives exist in a neighborhood of a and are continuous at a by hypothesis.

Now suppose $\|< x, y >\| < \delta$, that is, $< x, y > \in \Delta$. We shall prove that

(25) $$|f(x,y) - f(x,0) - yf_2'(0,0)| \leq \epsilon|y|/2$$

and

(26) $$|f(x,0) - f(0,0) - xf_1'(0,0)| \leq \epsilon|x|/2.$$

Consider the function g defined on $[0,y]$ (if y is negative, it will be on $[y,0]$) by

$$g(t) = f(x,t).$$

For values of t in this interval, $< x, t > \in \Delta$, so by (23) g is differentiable at all points of this interval and

$$g'(t) = f_2'(x,t).$$

(This is the definition of the partial derivative f_2'.) Hence we can apply the mean value theorem to g. There is a number v between 0 and y such that

$$g(y) - g(0) = g'(v)(y - 0).$$

That is,

$$f(x,y) - f(x,0) = yf_2'(x,v).$$

Hence

$$\begin{aligned} |f(x,y) - f(x,0) - yf_2'(0,0)| &= |y| \cdot |f_2'(x,v) - f_2'(0,0)| \\ &\leq \varepsilon |y|/2. \end{aligned}$$

The last step is by condition (24) using the fact that $< x, v > \in \Delta$. Thus (25) is established.

The mean value theorem also applies to the function $t \longmapsto f(t,0)$ and we get

$$f(x,0) - f(0,0) = xf_1'(u,0)$$

for some u between 0 and x. Therefore,

$$\begin{aligned} |f(x,0) - f(0,0) - xf_1'(0,0)| &= |x| \cdot |f_1'(u,0) - f_1'(0,0)| \\ &\leq \varepsilon |x|/2. \end{aligned}$$

Again we used condition (24). This is (26).

Finally we have

$$\begin{aligned} |D| &\leq |f(x,y) - f(x,0) - yf_2'(0,0)| + \\ &\qquad |f(x,0) - f(0,0) - xf_1'(0,0)| \\ &\leq \frac{\varepsilon}{2}\left(|x| + |y|\right) \leq \varepsilon \| < x, y > \|. \end{aligned}$$

The last inequality follows because $|x| \leq \sqrt{x^2 + y^2} = \| < x, y > \|$ and similarly, $|y| \leq \| < x, y > \|$.

Thus we have proved

$$(\forall < x, y > \in \mathbb{R}^2) \quad \| < x, y > - < 0, 0 > \| < \delta \implies$$
$$|f(x,y) - f(0,0) - xf_1'(0,0) - yf_2'(0,0)| \leq \varepsilon \| < x, y > - < 0, 0 > \|.$$

Since we showed how to get the appropriate δ given any positive ε, we have proved that f is differentiable at $< 0, 0 >$ and that its differential is as claimed. □

If E is an open set in V and $f : E \rightarrow \mathbb{R}$ is a differentiable function, its differential is a function from E to V^*; in coordinates, it is a function from E to the space of row vectors. Naturally we prefer that this function be continuous. It will be continuous if and only if, when expressed in coordinates, its components are continuous. Since these components are just the partial derivatives, we have the following important fact.

8.3.27 *Theorem*. *Let f be a real-valued function defined on an open subset E of $\mathbb{R}^n$. The differential of f exists and is a continuous function from E to $\mathbb{R}^{n*}$ (the set of n-long row vectors) if and only if the partial derivatives $f'_1, f'_2, \ldots, f'_n$ are defined and continuous on E.*

Proof. It follows from the previous theorem that if the partial derivatives are defined and continuous on E, then f is differentiable at each point of E and its differential, being given by the partial derivatives, is continuous. (Note how the fact that E is open enters here. Theorem 8.3.22 would not be applicable at a boundary point of E. Fortunately, no point of E is a boundary point.)

Conversely, we know (although we haven't given a proof in the general case) that at any point where f is differentiable all its partial derivatives exist and df consists of these partial derivatives assembled into a row vector. Hence, if f differentiable at each point of E, it partial derivatives are defined on all of E. Moreover, if df is continuous, its components must be continuous; ie., the partial derivatives must be continuous. □

On page 8-18 we defined a function to be C^k if all of its partial derivatives through order k exist and are continuous. We mentioned the fact that, as long as the derivatives involved exist and are continuous, the order of differentiation is immaterial, and we shall now prove a theorem to this effect. As we noted before, the theorem really concerns only two variables, so we state it only for that case.

8.3.28 *Theorem. Let* f *be a* C^1*-function from an open set* E *in* $\mathbb{R}^2$ *to* $\mathbb{R}$. *Suppose* f''_{12} *is defined and continuous at each point of* E. *Then* f''_{21} *is defined at each point of* E *and* $f''_{21} = f''_{12}$.

Proof. To prove this we should pick an arbitrary point $< a, b >$ of E and show that $f''_{21}(a,b)$ exists and equals $f''_{12}(a,b)$. There is, however, no real loss of generality and the proof is easier to read, if we assume $a = b = 0$; so we make this assumption. Then the problem is to prove

$$\lim_{h \to 0} \frac{f'_2(h,0) - f'_2(0,0)}{h} = f''_{12}(0,0) \tag{29}$$

since the limit, if it exists, is $f''_{21}(0,0)$. We can write this

$$\lim_{h \to 0} \frac{1}{h}\left(\lim_{k \to 0} \frac{f(h,k) - f(h,0)}{k} - \lim_{k \to 0} \frac{f(0,k) - f(0,0)}{k}\right)$$

$$= \lim_{h \to 0} \lim_{k \to 0} \frac{1}{hk}\big(f(h,k) - f(h,0) - f(0,k) + f(0,0)\big) .$$

Here the inner limit is known to exist because f is C^1, but the outer limit is not yet known to exist. However, we are given that

$$f''_{12}(0,0) = \lim_{k \to 0} \frac{f'_1(0,k) - f'_2(0,0)}{k}$$

$$= \lim_{k \to 0} \lim_{h \to 0} \frac{1}{hk}\big(f(h,k) - f(0,k) - f(h,0) + f(0,0)\big)$$

with both limits existing. Comparing, we see that our two expressions differ only in the order in which the limits are taken. So we must prove that in this case it doesn't matter in what order the limits are taken. It is easy to give examples in which an iterated limit exists in one order but not in the other, or in which an iterated limit exists in either order but the two limits are different. So there is something non-trivial to prove here. The problem of reversing an iterated limit is typical of analysis. Many important theorems simply assert that under suitable circumstances an iterated limit may be reversed. For example, theorem 4.6.10 says a power series may be differentiated term-by-term; this means that the limit associated with the infinite sum and the limit associated with differentiation may, in the case of convergent power series, be taken in either order.

The proof is accomplished by an ingenious application of the mean value theorem. Suppose $k \neq 0$ and put

$$g(t) = f(t,k) - f(t,0).$$

Granting that k is small enough so that everything lies in E, this is the difference of two functions each of which is differentiable for t near 0 since it is given that f_1' exists throughout E. Hence if $|h|$ is small

(30) $$g(h) - g(0) = hg'(uh)$$

for some number u between 0 and 1. Now

$$g'(t) = f_1'(t,k) - f_1'(t,0)$$

for all small t, so

(31) $$g'(uh) = f_1'(uh,k) - f_1'(uh,0) = k f_{12}''(uh,vk)$$

where v is between 0 and 1, by a second application of the mean value theorem using the fact that f_{12}'' exists. Hence if $|h|$ and $|k|$ are small enough but not zero

(32) $$f(h,k) - f(h,0) - f(0,k) + f(0,0) = hk f_{12}''(uh,vk)$$

since the left hand member is $g(h) - g(0)$. The continuity of f_{12}'' shows that this is nearly $hkf_{12}''(0,0)$ so the desired result follows easily.

The detailed argument is as follows. We must prove (29), that is

(33) $$(\forall\ \varepsilon > 0)(\exists\ \delta > 0)(\forall h)\ 0 < |h| < \delta \implies \left|\frac{f_2'(h,0) - f_2'(0,0)}{h} - f_{12}''(0,0)\right| < \varepsilon.$$

Let $\varepsilon > 0$ be given. Here is the recipe for choosing δ. Since E is open, we can find a disk centered at $<0, 0>$ that lies in E. We know that f''_{12} is a continuous function, so we can choose a smaller disk Δ centered at $< 0, 0 >$ so that

$$(34) \qquad |f_{12}(x,y) - f_{12}(0,0)| < \varepsilon/2$$

whenever $< x, y > \in \Delta$. Let δ be half the radius of Δ. Then if $|x| < \delta$ and $|y| < \delta$, $< x, y > \in \Delta$.

Now we must prove an inequality, the last part of (33), involving an arbitrary real number h satisfying $0 < |h| < \delta$. Let such an h be fixed.

For any real k satisfying $0 < |k| < \delta$, the function g is defined for $|t| \le h$ and is differentiable. Hence (30) is valid for some $u \in (0,1)$. Since the segment from $< uh, 0 >$ to $< uh, k >$ lies entirely in Δ and therefore in E, and since f''_{12} exists at all points of E, we can apply the mean value theorem to f'_1, regarded as a function of its second argument alone, and we obtain (31) where $< uh, vk >$ is a point of Δ. Now by (34)

$$|f''_{12}(uh,vk) - f''_{12}(0,0)| < \varepsilon/2.$$

We know this even though we don't know what u and v are; all we need is that they are both between 0 and 1.

Using (32) divided through by hk, we have

$$\left|\frac{1}{h}\left(\frac{f(h,k) - f(h,0)}{k} - \frac{f(0,k) - f(0,0)}{k}\right) - f''_{12}(0,0)\right| \le \varepsilon/2$$

This inequality is true for all values of k with $0 < |k| < \delta$. As $k \longrightarrow 0$, the left member has a limit because f'_2 exists. We have therefore

$$\left| \frac{1}{h}\left(f'_2(h,0) - f'_2(0,0)\right) - f''_{12}(0,0)\right| \le \varepsilon/2 < \varepsilon.$$

This is the inequality we set out to prove.

This proves that (33) and hence (29) is true. □

8.3.35 Theorem. Suppose E is an open set in a finite dimensional inner product space V and $f : E \longrightarrow \mathbb{R}$ is a differentiable function. Suppose I is an open interval in $\mathbb{R}$ and $g : I \longrightarrow E$ is differentiable. Then $f \circ g : I \longrightarrow \mathbb{R}$ is differentiable and its derivative is given by

$$(f \circ g)'(t) = df(g(t))[g'(t)].$$

(Remember, $g'(t)$ is a vector in V, $df(g(t))$ is in V^*, and $[\ ]$ indicates the action of a member of V^* on an element of V.)

Proof. We need only prove this for a fixed (but arbitrarily chosen) value of t, say t_o. For brevity's sake we introduce $v_o = g(t_o)$ and $h = df(g(t_o)) = df(v_o)$. Since $g'(t_o)$ is a column vector and h is a row vector, they have norms. From the Cauchy-Schwarz inequality it follows that

$$(36) \qquad |h[v]| \leq \|h\| \cdot \|v\|$$

for any $v \in V$. (To derive this formally, use 5.4.18.)

We want to prove that $(f \circ g)'(t_o) = h[g'(t_o)]$. This is the same as showing that

$$t \longmapsto f(g(t_o)) + (t - t_o)h[g'(t_o)]$$

is the best linear approximation of $f \circ g$ near t_o. Hence we must estimate

$$D = f(g(t)) - f(g(t_o)) - (t - t_o)h[g'(t_o)].$$

We break D into two parts which give the errors due to approximating f and g, respectively.

$$D = f(g(t)) - f(g(t_o)) - h[g(t) - g(t_o)]$$
$$+ h[g(t) - g(t_o) - (t - t_o)g'(t_o)].$$

For t near t_o the first of these parts is much less than $\|g(t) - g(t_o)\|$ which is itself of the same order of magnitude as $|t - t_o|$. We shall show directly that the second part is much less than $|t - t_o|$. Altogether we find $|D| \leq \varepsilon|t - t_o|$ if t is near enough to t_o. Now for the details.

Let a positive ε be given. Because f is differentiable at v_o, there is a positive δ_1 such that

$$(37)\qquad (\forall v \in V)\ \| v - v_o \| < \delta_1 \Longrightarrow$$
$$|f(v) - f(v_o) - h[v - v_o]| \leq \frac{\varepsilon}{2(\varepsilon + \|g'(t_o)\|)} \| v - v_o \|$$

Because g is differentiable at t_o it is also continuous there, so there is a positive δ such that $|t - t_o| < \delta \Longrightarrow$

$$(38)\qquad \| g(t) - g(t_o) \| < \delta_1 \quad \text{and}$$

$$(39)\qquad \| g(t) - g(t_o) - (t - t_o)g'(t_o) \| < \frac{\varepsilon}{2(1 + \|h\|)} |t - t_o|$$

Now let t be any number satisfying $|t - t_o| < \delta$. Then

$$\begin{aligned} \| g(t) - g(t_o) \| &= \| g(t) - g(t_o) - (t - t_o)g'(t_o) + (t - t_o)g'(t_o) \| \\ (40)\qquad &\leq \| g(t) - g(t_o) - (t - t_o)g'(t_o) \| + |t - t_o|\cdot\|g'(t_o)\| \\ &\leq (\varepsilon + \| g'(t_o) \|) |t - t_o| \end{aligned}$$

Because (38) is true, we can replace v by $g(t)$ in (37). Remember that $v_o = g(t_o)$. With the aid of (40) the resulting inequality simplifies to

$$(41)\qquad |f(g(t)) - f(g(t_o)) - h[g(t) - g(t_o)]| \leq \frac{\varepsilon}{2}|t - t_o|$$

By (36) and then (39) we have

$$\begin{aligned} | h[g(t) - g(t_o) - (t-t_o)g'(t_o)] | &\leq \|h\|\cdot\|g(t) - g(t_o) - (t-t_o)g'(t_o)\| \\ &\leq \frac{\varepsilon}{2} |t - t_o|. \end{aligned}$$

Putting this together with (41) we have

$$|D| \leq \varepsilon |t - t_o|.$$

Thus we have proved

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall t)\ |t - t_o| < \delta \Longrightarrow$$
$$|f(g(t)) - f(g(t_o)) - (t - t_o)h[g'(t_o)]| \leq \varepsilon |t - t_o|.$$

This is precisely the statement that $(f\circ g)'(t_o) = h[g'(t_o)]$. Since t_o was chosen arbitrarily, this completes the proof. □

When V is $\mathbb{R}^n$ (or when coordinates have been introduced into V) then $g : I \longrightarrow E$ has components $g_1, g_2, \cdots, g_n$ and g' is the column vector with components $g_1', g_2', \ldots, g_n'$. The differential of f is the row vector

$$\| D_1 f \quad D_2 f \quad \cdots \quad D_n f \|$$

and

$$(f \circ g)'(t) = (D_1 f) g_1'(t) + (D_2 f) g_2'(t) + \cdots + (D_n f) g_n'(t).$$

We have left out the arguments of the partial derivative functions $D_i f$. They are all to be evaluated at $g(t)$. This is the formula we guessed on page 8-31. It can be written in full without arguments as follows

$$(f \circ g)' = ((D_1 f) \circ g) g_1' + ((D_2 f) \circ g) g_2' + \cdots + ((D_n f) \circ g) g_n',$$

but is most commonly abbreviated

$$(f \circ g)' = (D_1 f) g_1' + (D_2 f) g_2' + \cdots + (D_n f) g_n'.$$

Formulas in partial differentiation are often quite long if written out in full so they are often abbreviated. To understand them you must think about what has been omitted.

Exercises

1. What is the natural domain of the function given by
$$f(x,y) = \sqrt[3]{x^2 - y^2} \ ?$$
At what points is it differentiable? Same questions for g :
$$g(x,y) = \sqrt[3]{x^4 - y^4}.$$

2. Suppose that $f : V \to \mathbb{R}$ is differentiable at v_o with differential h there. Let w be a fixed vector in V and consider the function $g(t) = f(v_o + tw)$ By calculation from the definitions without use of the chain rule, show that $g'(t) = h[w]$.

3. Show that the function f given by
$$f(x,y) = \frac{x^3}{x^2 + y^2}$$
has a derivative along every vector at $< 0, 0 >$ but is not differentiable at that point.

4. With the notations of Theorem 8.3.35 and assuming that f and g are C^2, derive a formula for $(f\circ g)''$.

5. Let v be a fixed vector field on $\mathbb{R}^2$, ie., a function from $\mathbb{R}^2$ to the set of two high column vectors. If f is a differentiable function $\mathbb{R}^2 \longrightarrow \mathbb{R}$, we can differentiate f at any point $< a, b >$ along $v(a,b)$. The result is $df(a,b)[v(a,b)]$, a number. Hence $df[v]$ denotes a function from $\mathbb{R}^2$ to $\mathbb{R}$. Assuming v is continuous check that

$$T : f \longmapsto df[v]$$

is a linear function from C^1 to C^0 satisfying

$$(*) \qquad T(f\cdot g) = f\cdot(Tg) + g\cdot(Tf).$$

(The products, represented by $\cdot$, are to be taken pointwise.) A linear operator satisfying this identity is called a <u>derivation</u>. Note that partial differentiation is a special case corresponding to a constant vector field. (Which?) It is an interesting theorem that the only derivations from C^∞ to C^∞ are of the kind just constructed where v is a C^∞ vector field.

6. Consider the functions f and g given by

$$f(x,y) = e^{-x}\sin y$$

$$g(x,y) = \frac{1}{1 + y\,e^{-x}}$$

Consider the iterated limits

$$\lim_{x\to\infty}\lim_{y\to\infty} f(x,y) \qquad\qquad \lim_{y\to\infty}\lim_{x\to\infty} f(x,y)$$

$$\lim_{x\to\infty}\lim_{y\to\infty} g(x,y) \qquad\qquad \lim_{y\to\infty}\lim_{x\to\infty} g(x,y).$$

Explain what happens.

7. Prove that if $f : V \longrightarrow \mathbb{R}$ is differentiable at the point $v \in V$, it is also continuous at v.

8.3.42 Leibniz' notation. When we take Cartesian coordinates in a geometric plane P, we are fixing two functions x and y from P to $\mathbb{R}$. The coordinates of a point $q \in P$ are then $x(q)$ and $y(q)$. It is important to note the difference between this interpretation of 'x' and 'y' and the one you may be more familiar with. An equation like $x^2 + 2y^2 = 1$ for curve is often interpreted as referring to

$$\{ < x, y > : x^2 + 2y^2 = 1 \}, \tag{43}$$

a set of points in $\mathbb{R}^2$. In the new usage we should interpret it as

$$\{ q : x(q)^2 + 2y(q)^2 = 1 \}, \tag{44}$$

a set of points in P. In (43) 'x' and 'y' are dummy or pattern variables, serving only to tell how to test whether a given ordered pair of numbers is in the set or not. In the latter 'q' is a dummy while 'x' and 'y' refer to specific functions defined geometrically in term of axes in P. When dummies are used in a mathematical expression, you can always replace them by different letters as long as no confusion of symbols is thereby introduced. For example,

$$\{ < u, v > : u^2 + 2v^2 = 1 \}$$

refers to the same set as (43). On the other hand, if we write

$$\{ q : u(q)^2 + 2v(q)^2 = 1 \}$$

the presumption is that 'u' and 'v' represent functions from P to $\mathbb{R}$ probably different from x and y, so this set is probably different from the one given by (44). To go even further to illustrate the distinction, note that

$$\{ < y, x > : y^2 + 2x^2 = 1 \}$$

is the same as (43), but

$$\{ q : y(q)^2 + 2x(q)^2 = 1 \}$$

is different from (44).

We have generally used 'x', 'y', 'z', and 't' as dummies when we define functions. Thus we might define $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ by

$$F(x,y) = x^2 + 2y^2.$$

Here 'x' and 'y' are dummies as we can see by noting that the formula

$$F(u,v) = u^2 + 2v^2$$

has exactly the same meaning. However, if 'x' and 'y' are the names of functions from P to $\mathbb{R}$, then

$$x^2 + 2y^2$$

is a function from P to $\mathbb{R}$; specifically the function

$$q \longmapsto x(q)^2 + 2y(q)^2.$$

(Here 'q' is the dummy.)

The distinction we are making here is often glossed over, but you must make it in order to understand the extremely useful Leibniz notation for partial derivatives.

Suppose $f : P \to \mathbb{R}$ is any function and x and y are the usual coordinate functions. Since a point q of P is completely determined when $x(q)$ and $y(q)$ are known, there must exist a function $F : \mathbb{R}^2 \to \mathbb{R}$ such that

$$f(q) = F(x(q),y(q))$$

for all $q \in P$. This is usually abbreviated $f = F(x,y)$. Assuming F is differentiable, then '$\frac{\partial f}{\partial x}$' denotes a new function from P to $\mathbb{R}$ given by

$$\frac{\partial f}{\partial x}(q) = (D_1F)(x(q),y(q))$$

which we may abbreviate

$$\frac{\partial f}{\partial x} = (D_1F)(x,y).$$

(This is not the only way the symbol $\frac{\partial f}{\partial x}$ is used, but it is the commonest.) Similarly,

$$\frac{\partial f}{\partial y} = (D_2F)(x,y).$$

This is Leibniz' notation for the derivative. You are familiar with it, of course, for functions of one variable. (If $y = f(x)$, then $\frac{dy}{dx} = f'(x)$.) It is very commonly used in applications of mathematics. Often applications are concerned with functions defined on space, like the temperature or the pressure at a given point. Let such a function be f and, in terms of some coordinate functions x, y, and z, say $f = F(x,y,z)$. Although the ultimate objective may be to find F (ie., a method for computing f), it is f that has a direct real-world interpretation. Similarly, the derivatives of F are only rules for computation, but $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$, being functions defined on space, often have an important physical interpretation.

8.3.46 Change of coordinate systems. Suppose f is a function defined on a plane P and we have two different coordinate systems on P. Then f has partial derivatives in both systems and it is important to know how they are related.

As a first example, suppose we have a linear coordinate system with coordinate functions x and y and a second linear coordinate system with coordinate functions u and v. We assume that both systems have the same origin. Then we can express x and y linearly in terms of u and v. To be definite, say

$$x = 2u + v$$
$$y = 3u + v.$$

(These are equations connecting functions on P; for example, $x(q) = 2u(q) + v(q)$.) Suppose we know $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. How can we find $\frac{\partial f}{\partial u}$ and $\frac{\partial f}{\partial v}$?

There is a function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ such that $f = F(x,y)$ and a function $G : \mathbb{R}^2 \longrightarrow \mathbb{R}$ such that $f = G(u,v)$. Therefore

$$G(u,v) = F(2u + v,\ 3u + v). \tag{44}$$

Although this is really a relation between functions defined on P, we can think of it as a formula for G in terms of F. Thinking of v as fixed we can differentiate using the chain rule to find G_1' :

$$\begin{aligned} G_1'(u,v) &= F_1'(2u + v,\ 3u + v)\cdot 2 + F_2'(2u + v,\ 3u + v)\cdot 3 \\ &= 2F_1'(x,y) + 3F_2'(x,y) \end{aligned}$$

In the Leibniz notation this becomes

$$\frac{\partial f}{\partial u} = 2\,\frac{\partial f}{\partial x} + 3\,\frac{\partial f}{\partial y}\,.$$

If we differentiate (44) with respect to v keeping u fixed we get

$$G_2'(u,v) = F_1'(2u + v,\ 3u + v) + F_2'(2u + v,\ 3u + v)$$

which becomes

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

Now let us do a more complicated case. Assume x and y form a Cartesian coordinate system (ie., the axes are perpendicular and their scales are the same). We introduce polar coordinates ρ and θ into P taking the positive x-axis as the initial ray as usual. Then ρ and θ are new functions from P to $\mathbb{R}$. (Actually θ isn't defined at the origin and there is some ambiguity elsewhere, but it doesn't matter for the present considerations.) And x and y are related to ρ and θ by

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

Then

$$f = F(\rho\cos\theta,\ \rho\sin\theta)$$

and

$$(47)\qquad \begin{aligned}\frac{\partial f}{\partial \rho} &= F_1'(\rho\cos\theta,\ \rho\sin\theta)\cos\theta + F_2'(\rho\cos\theta,\ \rho\sin\theta)\sin\theta\\ &= \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta\,.\end{aligned}$$

Similarly,

$$(48)\qquad \frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x}(-\rho\cos\theta) + \frac{\partial f}{\partial y}(\rho\sin\theta)$$

We can easily generalize these examples. Suppose u and v are any two functions on P and that x and y can be expressed in terms of u and v, say

$$x = \varphi(u,v)$$
$$y = \psi(u,v)$$

where φ and ψ are differentiable functions. Then

$$f = F(\varphi(u,v), \psi(u,v))$$

$$\frac{\partial f}{\partial u} = F_1'(\varphi(u,v), \psi(u,v))\,\varphi_1'(u,v) + F_2'(\varphi(u,v), \psi(u,v))\,\psi_1'(u,v)$$

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

Similarly,

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

Note how the Leibniz notation obviates the necessity of any notation at all for the intermediate functions F, φ, and ψ.

This new form of the chain rule extends immediately to higher dimensions. If, for example, f can be expressed differentiably in terms of x, y, and z and x, y, and z can in turn be expressed differentiably in terms of u, v, and w, then f can be expressed in terms of u, v, and w and

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial u}$$

and there are similar formulas for $\frac{\partial f}{\partial v}$ and $\frac{\partial f}{\partial w}$.

Since $\frac{\partial f}{\partial x}$ is a function on space, we can start over again and differentiate it. We obtain the second order partial derivatives

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) \quad \text{and} \quad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) .$$

These are usually abbreviated

$$\frac{\partial^2 f}{\partial x^2} \quad \text{and} \quad \frac{\partial^2 f}{\partial y \partial x} .$$

According to Theorem 8.3.28 we will usually have

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$$

Calculating second order partial derivatives in one system of coordinates in terms of the partial derivatives in another is a problem of frequent occurrence that requires careful attention. We illustrate by showing that, for any function u on P, the Laplacian

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$$

expressed in polar coordinates is

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho}$$

Formulas (47) and (48) are valid for any smooth function f on P. In particular they are valid for $f = u$, for $f = \partial u/\partial x$, and for $f = \partial u/\partial y$. Hence

$$(49) \qquad \frac{\partial u}{\partial \rho} = \frac{\partial u}{\partial x}\cos\theta + \frac{\partial u}{\partial y}\sin\theta$$

$$\frac{\partial}{\partial \rho}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial^2 u}{\partial x^2}\cos\theta + \frac{\partial^2 u}{\partial y \partial x}\sin\theta$$

$$\frac{\partial}{\partial \rho}\left(\frac{\partial u}{\partial y}\right) = \frac{\partial^2 u}{\partial x \partial y}\cos\theta + \frac{\partial^2 u}{\partial y^2}\sin\theta$$

Now we can differentiate (49). Remember that $\frac{\partial}{\partial \rho}\cos\theta = \frac{\partial}{\partial \rho}\sin\theta = 0$, since θ is treated as a constant when calculating $\frac{\partial}{\partial \rho}$.

$$(50) \qquad \begin{aligned} \frac{\partial^2 u}{\partial \rho^2} &= \frac{\partial}{\partial \rho}\left(\frac{\partial u}{\partial x}\right)\cos\theta + \frac{\partial}{\partial \rho}\left(\frac{\partial u}{\partial y}\right)\sin\theta \\ &= \frac{\partial^2 u}{\partial x^2}\cos^2\theta + 2\frac{\partial^2 u}{\partial x \partial y}\sin\theta\cos\theta + \frac{\partial^2 u}{\partial y^2}\sin^2\theta . \end{aligned}$$

Formula (48) can be written

$$\frac{\partial u}{\partial \theta} = -y\frac{\partial u}{\partial x} + x\frac{\partial u}{\partial y}$$

Hence

$$\begin{aligned} \frac{\partial^2 u}{\partial \theta^2} &= -\frac{\partial y}{\partial \theta}\left(\frac{\partial u}{\partial x}\right) - y\frac{\partial}{\partial \theta}\left(\frac{\partial u}{\partial x}\right) \\ &\qquad + \frac{\partial x}{\partial \theta}\left(\frac{\partial u}{\partial y}\right) + x\frac{\partial}{\partial \theta}\left(\frac{\partial u}{\partial y}\right) \\ &= -x\frac{\partial u}{\partial x} - y\left(-y\frac{\partial^2 u}{\partial x^2} + x\frac{\partial^2 u}{\partial y \partial x}\right) \\ &\qquad - y\frac{\partial u}{\partial y} + x\left(-y\frac{\partial^2 u}{\partial x \partial y} + x\frac{\partial^2 u}{\partial y^2}\right) \\ &= y^2\frac{\partial^2 u}{\partial x^2} - 2xy\frac{\partial^2 u}{\partial x \partial y} + x^2\frac{\partial^2 u}{\partial y^2} - \left(x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y}\right) \end{aligned}$$

Divide through by ρ^2 and add (50). Remember $x = \rho \cos \Theta$, $y = \rho \sin \Theta$.

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \theta^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - \frac{1}{\rho^2}\left(x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y}\right).$$

Transposing the last term and using (49) we get

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

A function u is called harmonic if it satisfies Laplace's equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0,$$

where x and y are Cartesian coordinate functions on the plane. Harmonic functions on three space are those that satisfy the three dimensional Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0,$$

where x, y, and z are three-dimensional Cartesian coordinate functions. Many important physical functions are harmonic in three-space, for example, gravitational or electrical potentials in empty space. A great deal has been discovered about solutions of Laplace's equation.

Let us determine all harmonic functions on the plane that depend only on the distance from a fixed point. We naturally take that point as the pole of a polar coordinate system. If u is the function, the Laplace equation is

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho} = 0.$$

We are asking that u be a function of ρ alone, that is

$$u = H(\rho)$$

Then the partial derivatives of u with respect to Θ are zero, and those with respect to ρ are given by the ordinary derivatives of H. Our equation becomes

$$H''(\rho) + \frac{1}{\rho}H'(\rho) = 0.$$

This is a second order linear equation. It may be regarded temporarily as a first order linear equation for H' and solved by the methods of §3.4. We find

$$H'(\rho) = \frac{b}{\rho}$$

and hence

$$u = a + b \log \rho$$

where a and b are constants. Note that only the constant solutions are smooth at the pole.

Exercises

1. Suppose $f = u^3 + 3uv - v^2$ where u and v can be expressed differentiably in terms of some functions x and y. Find $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial x \partial y}$, and $\frac{\partial^2 f}{\partial y^2}$ (in terms of u, v, $\frac{\partial u}{\partial x}$, etc.).

2. If $g = H(\cos y, \sin x)$ find $\frac{\partial g}{\partial x}$ and $\frac{\partial g}{\partial y}$.

3. Given that
$$\frac{\partial f}{\partial x} = f + \frac{\partial f}{\partial y}$$
show that
$$\frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} = f + 2\frac{\partial f}{\partial y}.$$

4. Suppose that $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a differentiable function and that x and y are the usual coordinate functions on $\mathbb{R}^2$. Show that
$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy.$$

5. Suppose that x, y and u, v are two different Cartesian coordinate systems in the plane with the same scale. If f is a C^2-function on the plane (this means f can be expressed in terms of x and y with a C^2-function from $\mathbb{R}^2$ to $\mathbb{R}$), show that
$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial v^2}.$$

The two coordinate systems need not have the same origin. Note that if this were not true, the Laplacian of a function would not often be of any physical significance. Actually, except for constant multiples, the Laplacian is the only second order differential operator with this invariance property. That fact explains its ubiquitous character.

6. Find all solutions of Laplace's equation in the plane that can be written in the form $H(\rho)\cos n\theta$ or $K(\rho)\sin n\theta$ where n is an integer. (After finding a second order ordinary differential equation for H, try $H(\rho) = \rho^t$.)

These solutions are extremely important because Laplace's equation is linear. The set of all solutions is therefore a linear subspace of the set of C^2-functions. This subspace is infinite dimensional, so it doesn't have a basis in the sense of 2.4.1, but the solutions found in this exercise span the set of solutions in the sense of §5.5: every solution of Laplace's equation can be represented as a limit (in a suitable sense) of finite linear combinations of these simple solutions. The theory of Fourier series figures most importantly in representing the solutions. Use Fourier series to solve the following problem.

Find a continuous functions f from the closed unit disk in the plane to $\mathbb{R}$ that satisfies Laplace's equation in the open unit disk and has preassigned (continuous) values on the unit circle. Don't worry about convergence questions. Just assume that the solution is a convergent infinite sum of the solutions found above.

8.4 Taylor's series.

In chapter 4 we saw that elementary functions of one variable can usually be expanded in a power series. The facts are similar for functions of several variables.

We shall show that the second degree terms in the Taylor's series for a function at a critical point determine, in most cases, whether the critical point is a maximum, a minimum, or a saddle point.

8.4.1 Polynomial functions. Let P be a plane. Choose a Cartesian coordinate system for it and let x and y be the coordinate functions. Any function from P to $\mathbb{R}$ of the form

$$\alpha + \beta x + \gamma y + \delta x^2 + \cdots + \xi x^n + \eta x^{n-1} y + \cdots + \sigma y^n$$

is called a <u>polynomial function on</u> P. The degree of this polynomial function is the highest total degree of non-zero terms occurring (after everything has been properly collected and simplified, of course). For example, $x^5 y^7$ has degree 12. The degree of $x(xy - y^3) + xy^3$ is three.

There are two useful ways to arrange the terms of a polynomial. In one we write first the term of degree zero (ie., the constant term), then the terms of degree one, then those of degree two, etc. Then our function is represented as the sum of a constant, a linear form, a quadratic form, a cubic form, etc. (The word "form" is often used to describe a polynomial function that is homogeneous, that is all the terms are of the same degree.) For example,

$$\underset{\text{constant}}{2} + \underset{\text{linear}}{3x + 2y} + \underset{\text{quadratic}}{x^2 - y^2} + \underset{\text{cubic}}{3x^2 y - y^3}.$$

This representation is often convenient when we want to consider points near the origin. At these point x and y are both small, so the quadratic terms are generally smaller than the linear terms, the cubic terms are generally smaller than the quadratic terms, etc.

Sometimes it is better to write (or imagine) the terms in a two-dimensional array

$$\begin{array}{llll} 2 & +3x & +x^2 & +0 \\ 2y & +0 & +3x^2y & \\ -y^2 & +0 & & \\ -y^3 & & & \end{array}$$

In theoretical work with a "general" polynomial g, we usually write it

$$g = \sum_{p,q} \alpha_{p,q} x^p y^q .$$

Sometimes it is better to put $\beta_{p,q} = p!\ q!\ \alpha_{p,q}$. Then

$$g = \sum_{p,q} \beta_{p,q} \frac{x^p}{p!} \frac{y^q}{q!} .$$

The advantage of this way of writing it becomes apparent when we differentiate.

$$\frac{\partial g}{\partial x} = \sum_{p,q} \beta_{p,q} \frac{x^{p-1}}{(p-1)!} \frac{y^q}{q!}$$

where now the sum involves only values of $p \geq 1$. The general derivative is

$$\frac{\partial^{r+s} g}{\partial x^r \partial y^s} = \sum \beta_{p,q} \frac{x^{p-r}}{(p-r)!} \frac{y^{q-s}}{(q-s)!}$$

where the sum is now restricted to indices $p \geq r$, $q \geq s$.

The value of this derivative at the origin is easy to get. Since x and y vanish at the origin, only the constant term (corresponding to $p = r$, $q = s$) is not zero, so

$$\frac{\partial^{r+s} g}{\partial x^r \partial y^s}(0) = \beta_{r,s} .$$

Since we are using the Leibniz notation, the argument of a partial derivative is a point. Hence we have denoted the origin here by 0. In sums involving partial derivatives of various orders, it is understood that the zero-th derivative is the function itself.

From this expression it is clear that there is a polynomial of degree n (at most) whose partial derivatives at the origin of all orders up to n have prescribed values. Furthermore, this polynomial is unique.

Everything we have done here has an immediate and evident generalization to higher dimensions.

8.4.2 Taylor polynomials. Let f be a real valued function defined on a neighborhood of the origin and differentiable there. Then there is a polynomial g of degree at most one that approximates f well at 0. That means $|f - g|$ is small relative to $\sqrt{x^2 + y^2}$. More precisely, for a point v near 0, $|f(v) - g(v)|$ is much less than $\sqrt{x(v)^2 + y(v)^2} = \|v\|$. In full detail

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall v) \quad \|v\| < \delta \implies$$

$$|f(v) - g(v)| \leq \epsilon\|v\| .$$

This is, of course, just the definition of differentiability.

We know what the polynomial g is. It is

$$f(0) + \frac{\partial f}{\partial x}(0)\, x + \frac{\partial f}{\partial y}(0)\, y.$$

(This is just formula 8.3(18) converted to Leibniz notation with $a = b = 0$.) We can describe it as follows: g is the polynomial of degree at most one that has the same value as f and the same first partial derivatives as f at the origin.

To get an even better approximation of f we should try a polynomial of higher degree. There is a unique polynomial g of degree two at most that has the same value and partial derivatives as f at the origin through partial derivatives of order two; that is

$$\frac{\partial^{p+q} g}{\partial x^p \partial y^q}(0) = \frac{\partial^{p+q} f}{\partial x^p \partial y^q}(0)$$

for $p + q \leq 2$. We require, of course, that f has second order partial derivatives. In fact we shall assume that f is C^2. With this hypothesis we shall prove that $|f - g|$ is small relative to $x^2 + y^2$; that is

$$(\forall \epsilon > 0)(\exists\ \delta > 0)(\forall v) \quad \|v\| < \delta \implies$$
$$|f(v) - g(v)| \leq \ \|v\|^2 .$$

If f is a function of class C^k near the origin, the polynomial g of degree at most k such that (3) is true for $p + q \leq k$ is called the k-th Taylor polynomial for f at the origin.

We can also define the Taylor polynomials for f at other points. The k-th Taylor polynomial for f at v_o is the unique polynomial g of degree at most k such that

$$\frac{\partial^{p+q}g}{\partial x^p \partial y^q}(v_o) = \frac{\partial^{p+q}f}{\partial x^p \partial y^q}(v_o)$$

for $p + q \leq k$. It can be written explicitly as a sum. If the coordinates of v_o are a and b (ie., $x(v_o) = a$, $y(v_o) = b$), then

$$g = \sum \frac{\partial^{p+q}f}{\partial x^p \partial y^q}(v_o) \frac{(x-a)^p}{p!} \frac{(y-b)^q}{q!} ,$$

the sum being taken over all non-negative p, q with $p + q \leq k$.

This is a generalization of the one dimensional case and extends immediately to more dimensions. For example, in dimension three the k-th Taylor polynomial for a function f at v_o is

$$\sum \frac{\partial^{\,p+q+r}f}{\partial x^p \partial y^q \partial z^r}(v_o) \frac{(x-a)^p}{p!} \frac{(y-b)^q}{q!} \frac{(z-c)^r}{r!} ,$$

the sum being taken over all non-negative p, q, r with $p + q + r \leq k$. Here a, b, and c are the coordinates of v_o.

Exercises.

1. **Find the second Taylor polynomial for the following functions at the points indicated.**

(a) $1 + x^2 + xy + y^2$ at $< 0, 0 >$.　　(c) $\exp(x + xy)$ at $< 0, 0 >$

(b) $\tan(x + y)$ at $< 0, 1 >$　　(d) $\arcsin(x - y)$ at $< 0, 0 >$

8.4.3 Taylor's formula with remainder. Just as in the case of one variable the crucial step in showing that a function is actually approximated by its Taylor polynomials is to show that the error can be written in terms of higher order derivatives.

As is frequently the case, the results we want are easier to state and discuss in terms of the Leibniz notation but easier to prove in terms of the $F, F_1', \dots$ notation.

Theorem. Let G *be an open set in* $\mathbb{R}^2$ *and let* $F : G \longrightarrow \mathbb{R}$ *be a function of class* C^2. *Suppose* $< a, b >$ *and* $< c, d >$ *are two points of* G *such that the segment* S *joining them lies wholly in* G. *Then there is a point* $< r, s >$ *on* S *such that*

$$\begin{aligned} F(c,d) \;=\; & F(a,b) + (c-a)F_1'(a,b) + (d-b)F_2'(a,b) \\ & + \tfrac{1}{2!}(c-a)^2F_{11}''(r,s) + (c-a)(d-b)F_{12}''(r,s) + \tfrac{1}{2!}(d-b)^2F_{22}''(r,s). \end{aligned} \tag{4}$$

Proof. For brevity let $h = c - a$, $k = d - b$. Define a function φ of one variable by

$$\varphi(t) \;=\; F(a + th,\, b + tk).$$

Then φ is defined at least for $0 \leq t \leq 1$ and

$$\varphi'(t) = hF_1'(a+th,\, b+tk) + kF_2'(a+th,\, b+tk)$$

$$\varphi''(t) = h^2F_{11}''(a+th,\, b+tk) + 2hkF_{12}''(a+th,\, b+tk) + k^2F_{22}''(a+th,\, b+tk).$$

According to the extended theorem of mean value

$$\varphi(1) = \varphi(0) + \varphi'(0) + \tfrac{1}{2!}\varphi''(\xi)$$

where ξ is some number between 0 and 1. Using our formulas for φ and its derivatives and putting $r = a + \xi h$, $s = b + \xi k$, the result is (4). Note that $< r, s >$ is on the segment joining $< a, b >$ to $< c, d >$. □

For reference purposes we state without proof the generalization of this theorem to arbitrary dimension.

8.4.5 Theorem. Let G be an open set in $\mathbb{R}^n$ and let $F : G \to \mathbb{R}$ be a function of class C^2. Suppose $a = \langle a_1, a_2, \ldots, a_n \rangle$ and $c = \langle c_1, c_2, \ldots, c_n \rangle$ are two points of G such that the segment S joining them lies wholly in G. Then there is a point r on S such that

$$F(c) = F(a) + \sum_{i=1}^{n} (c_i - a_i) F_i'(a)$$

$$+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (c_i - a_i)(c_j - a_j) F_{ij}''(r). \ \square$$

The theorem extends as well to higher degree Taylor polynomials. If F is of class C^p, the function φ of the proof is C^p and the extended theorem of mean value tells us that

$$(6) \quad \varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2!}\varphi''(0) + \ldots + \frac{1}{(p-1)!}\varphi^{(p-1)}(0) + \frac{1}{p!}\varphi^{(p)}(\xi).$$

where ξ is some number between 0 and 1.

We can calculate the derivatives of φ as we did before, but the notation soon becomes awkward. So we introduce the differentiation operators D_1 and D_2. Then we have

$$\varphi'(t) = hD_1F + kD_2F = (hD_1 + kD_2)F$$

$$\varphi''(t) = h^2D_1^2F + 2hkD_1D_2F + k^2D_2^2F = (hD_1 + kD_2)^2F$$

and in general

$$\varphi^{(i)}(t) = (hD_1 + kD_2)^iF$$

where all the derivatives are to be evaluated at $\langle a + th, b + tk \rangle$. Because the operators D_1 and D_2 commute, that is, $D_1D_2 = D_2D_1$ (as long as the function F has continuous derivatives), we can treat expressions like $(hD_1 + kD_2)^i$ as if they were ordinary polynomials. Putting these values into (6) we obtain

$$
\begin{aligned}
F(c,d) &= F(a,b) + hF_1'(a,b) + kF_2'(a,b) \\
&\quad + \frac{h^2}{2!}\, F_{11}''(a,b) + hkF_{12}''(a,b) + \frac{k^2}{2!}\, F_{22}''(a.b) \\
&\quad + \cdots \qquad\qquad (7) \\
&\quad + \frac{1}{(p-1)!}\Big((hD_1 + kD_2)^{p-1}F\Big)(a,b) \\
&\quad + \frac{1}{p!}\Big((hD_1 + kD_2)^{p}F\Big)(a+\xi h,\ b+\xi k).
\end{aligned}
$$

This is valid for any function F of class C^p provided the line segment joining $< a, b >$ to $< c, d > = < a + h, b + k >$ lies in the domain of F.

Formula (7) is called *Taylor's formula with remainder*. There is, of course, a similar formula for functions of more variables.

The number ξ occurring in (7) depends on a, b, c, and d. Only rarely is there any reasonable way to compute ξ. Hence we usually only estimate the last term of (7)

Suppose this last term is expanded as a sum.

$$\sum_{i=0}^{p} \frac{h^i}{i!}\, \frac{k^{p-i}}{(p-i)!}\left(D_1^i D_2^{p-i} F\right)(a + \xi h,\ b + \xi k)$$

Since we are assuming that the p-th order derivatives of F are continuous, all the derivatives appearing here will have values rather close to their values at $< a, b >$ provided h and k are small, no matter what ξ is.

We shall use this argument to justify the claim made at the top of page 8-64 concerning the approximation of a function by its second Taylor polynomial.

The second Taylor polynomial for F at $< a, b >$ evaluated at $< c, d >$ is

$$
\begin{aligned}
&F(a,b) + (c-a)F_1'(a,b) + (d-b)F_2'(a,b) \\
&\quad + \frac{1}{2!}(c-a)^2F_{11}''(a,b) + (c-a)(d-b)F_{12}''(a,b) + \frac{1}{2!}(d-b)^2F_{22}''(a,b).
\end{aligned}
$$

This differs from the right side of (4) only in that the second order derivatives are evaluated in different places. The difference we want to estimate is therefore

$$E = \frac{1}{2} h^2 \big(F''_{11}(r,s) - F''_{11}(a,b)\big) + hk\big(F''_{12}(r,s) - F''_{12}(a,b)\big) + \frac{1}{2} k^2 \big(F''_{22}(r,s) - F''_{22}(a,b)\big). \tag{8}$$

Given $\epsilon > 0$, we can choose a positive δ so small that the disk Δ of radius δ about $< a, b >$ lies wholly in G and

$$|F''_{ij}(u,v) - F''_{ij}(a,b)| < \epsilon \tag{9}$$

for $i, j = 1, 2$ and any choice of $< u, v >$ in Δ.

Now if $< c, d > \in \Delta$, the line segment from $< a, b >$ to $< c, d >$ lies in G and $< r, s >$ lies in G. Hence (9) is applicable to each of the terms in (8) and we get

$$|E| \leq \frac{1}{2} h^2 \epsilon + |hk| \epsilon + \frac{1}{2} k^2 \epsilon$$

$$\leq \epsilon (h^2 + k^2) = \epsilon \| < c, d > - < a, b > \|^2 .$$

(The penultimate step because $|hk| \leq (h^2 + k^2)/2$.)

To get this inequality to look like our claim on page 8-64 we switch back to Leibniz notation. Suppose $f = F(x,y)$ and v_o is a point with coordinates a and b. Let g be the second Taylor polynomial for f at v_o. Finally take $c = x(v)$, $d = y(v)$. Then E becomes $f(v) - g(v)$ and

$$|f(v) - g(v)| \leq \epsilon \|v - v_o\|^2$$

provided $\|v - v_o\| < \delta$.

8.4.10 Taylor's series. If F is a function of class C^∞, then at any point F will have Taylor polynomials of every degree and we might reasonably hope that these polynomials will converge to F. We are led, therefore, to consider the double power series

$$\sum_{i,j=0}^{\infty} \frac{(x-a)^i}{i!} \frac{(y-b)^j}{j!} \left(D_1{}^i D_2{}^j F\right)(a,b).$$

This series is known as the <u>Taylor's series for</u> F <u>at</u> < a, b >. As we have mentioned in chapter 4, even in the case of functions of one variable, the

Taylor's series of a function need not converge; and even if it does converge, it need not converge to $F(x,y)$. However, for the functions commonly encountered, it will converge to $F(x,y)$ at least for small values of $|x-a|$ and $|y-b|$. The convergence will be absolute and various formal manipulations of series, such as term-by-term differentiation, will be valid, just as in the case of one variable.

Since a convergent double (or triple or higher) power series will also make sense for complex values of the variables, Taylor's series lead naturally to the theory of functions of several complex variables, one of the most active areas of mathematical research today.

When we want to find the Taylor's series for a function of several variables it is often easier to get it by formal manipulations than by calculating derivatives. An example will make the ideas clear.

Find the Taylor's series for $\log(\cos x + \sin y)$ through terms of degree three at $< 0, 0 >$. (This is the same as the third Taylor polynomial at $< 0, 0 >$.)

We know that

$$\cos x = 1 - \frac{x^2}{2} + \text{terms of degree} \geq 4$$

$$\sin y = y - \frac{y^3}{6} + \text{terms of degree} \geq 5$$

$$\log (1 + z) = z - \frac{z^2}{2} + \frac{z^3}{3} + \text{terms of degree} \geq 4.$$

(These are just one variable Taylor expansions.) So put $z = y - \frac{x^2}{2} - \frac{y^3}{6} + \text{terms}$ of degree ≥ 4 in the last of these series. Since

$$z^2 = y^2 - x^2y + \text{terms of degree} \geq 4$$

and

$$z^3 = y^3 + \text{terms of degree} \geq 4$$

we find

$$\begin{aligned}\log (\cos x + \sin y) &= \log \left(1 + y - \frac{x^2}{2} - \frac{y^3}{6} + \cdots\right) \\ &= y - \frac{x^2}{2} - \frac{y^3}{6} - \frac{1}{2}(y^2 - x^2y) + \frac{1}{3} y^3 \\ &\qquad + \text{terms of degree} \geq 4 \\ &= y - \frac{1}{2} x^2 - \frac{1}{2} y^2 + \frac{1}{2} x^2y + \frac{1}{6} y^3 \\ &\qquad + \text{terms of degree} \geq 4.\end{aligned}$$

Because we replaced z by a series that begins with first degree terms (ie., the constant term is zero), the successive powers of z begin with terms of higher and higher degree in x and y. Terms of degree at least four in z produce only terms of degree at least four in x and y. Hence they may be neglected if our goal is only to find the terms of degree three or less.

Taylor polynomials are often useful in practical computation. When they are found by formal manipulations, as above, one gets no easy way to estimate the error, however.

Exercises.

1. Using the method of formal power series manipulation, find the third degree Taylor polynomial for the following functions at $< 0, 0 >$ or $< 0, 0, 0 >$.

   (a) $\cosh(x - y + xy)$

   (b) $\arcsin(x + y - xz)$

   (c) $\dfrac{\sin(x + y)}{\cos(x - y)}$

   (d) $\log(x + \cos y)$

   (e) $\exp(xy - \sin z)$

   (f) $(1 + x)^y$

   (g) $\displaystyle\int_0^1 \exp(\sqrt{1 + x + ty})\,dt$

   (h) $\displaystyle\int_0^1 \frac{\log(1 + xt)}{1 - yt}\,dt$

2. In 1(h), for what values of x and y would you expect the Taylor series to converge?

3. The equation $x^5 - 5\alpha x^2 + 5\beta x - 1$ has the root $x = 1$ for $\alpha = \beta = 0$. There is a C^∞ function f of two variables such that $x = f(\alpha, \beta)$ is a root of the above equation for any small α and β, and such that $f(0,0) = 1$. Find the Taylor polynomial of degree two for f at $< 0, 0 >$. (Try the method of undetermined coefficients. Compare 4.6.31. You need not prove that the series converges.)

4. Suppose F is a polynomial function of degree at most three. Show that in formula (7), p. 8-67, with $p = 2$ you can always take $\xi = 1/3$.

8.4.11 Analysis of critical points. In §8.2 (p. 8-20 ff) we looked at the problem of finding the maximum and minimum of a function of several variables. We showed that a local maximum or minimum of a function F can occur only at

(a) a critical point, that is, a point where all the first order partial derivatives of F vanish, or

(b) a boundary point of the domain considered, or

(c) a point at which F is non-differentiable.

Usually we deal with everywhere differentiable functions, so case (c) does not often arise.

Even in one dimension a critical point need not be either a local maximum or a local minimum. For example, x^3 has a critical point at 0, but it is neither a local maximum point nor a local minimum point for x^3. One test for the existence of a local extreme value in one dimension is to examine the second derivative. Suppose a is a critical point of F (ie., $F'(a) = 0$). If $F''(a) > 0$, then a is a strict local minimum point, that is, $F(a) < F(x)$ for all x near a but different from a. If $F''(a) < 0$, then a is a strict local maximum point. If $F''(a) = 0$, the test fails; we cannot decide on the basis of this information alone whether F has a local extreme value. We shall develop a similar test for functions of several variables using the second order partial derivatives at a critical point.

First, let us look at the one dimensional case from the point of view of Taylor's series. Expand F at a critical point a.

$$F(a + h) = F(a) + \frac{1}{2} h^2 F''(a) + \cdots .$$

(The linear term is omitted because $F'(a) = 0$.) If $F''(a) > 0$, the term $\frac{1}{2} h^2 F''(a)$ will be positive for all values of h except 0. For small values of h this term, although small, will still be larger than the higher degree terms omitted, since the latter all have the factor h^3. Hence

$$F(a + h) > F(a)$$

for all small but non-zero h (ie., $|h|$ small) and a is a strict local minimum point for F. Similarly, if $F''(a) < 0$, a will be a strict local maximum point. If $F''(a) = 0$, then the second degree terms do not control the local behavior of F and the test fails.

Now suppose F is a function of two variables with a critical point at $< a, b >$. The Taylor's series for F at $< a, b >$ begins

$$F(a + h, b + k) = F(a,b) + \frac{1}{2}\left(h^2F''_{11} + 2hkF''_{12} + k^2F''_{22}\right) + \cdots$$

where the derivatives are all to be evaluated at $< a, b >$. (The linear terms are omitted because $F'_1(a,b) = F'_2(a,b) = 0$.) Because all subsequent terms in the series involve h and k to at least degree three, we expect the variation of F near $< a, b >$ to be essentially controlled by the second degree terms. Omitting the factor 1/2, these terms are a quadratic form in h and k called the <u>Hessian form of</u> F <u>at</u> $< a, b >$. The behavior of F near $< a, b >$ is in most cases determined by the Hessian form. We shall show that

(a) If the Hessian form of F at $< a, b >$ is positive definite, F has a strict local minimum at $< a, b >$.

(b) If the Hessian form is negative definite, F has a strict local maximum at $< a, b >$.

(c) If the Hessian form takes both positive and negative values, then $< a, b >$ is neither a local maximum or a local minimum, but some kind of saddle point.

There remains the possibility that the Hessian form is semi-definite, but not definite. In this case the test fails.

These conclusions are equally valid in higher dimensions. The quadratic terms in the Taylor series for F at a critical point, again leaving out the factor 1/2, constitute the Hessian form of F. Its matrix is

$$M = \begin{Vmatrix} F''_{11} & F''_{12} & \cdots & F''_{1n} \\ F''_{21} & F''_{22} & \cdots & F''_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ F''_{n1} & F''_{n2} & \cdots & F''_{nn} \end{Vmatrix}$$

all derivatives being evaluated at the critical point. This matrix is known as the <u>Hessian matrix</u>. Assuming F is C^2, it is symmetric.

Although our discussion involved reference to terms in the Taylor's series for F of degrees higher than two and hence to derivatives of F of orders higher than two, we can prove the statements above on the hypothesis that F is merely C^2.

8.4.12 Theorem. Let E be an open set in $\mathbb{R}^n$ and let $F : E \quad \mathbb{R}$ be a C^2-function. Suppose a is a critical point for F and that the Hessian form for F at a is H. Then

(a) If H is positive definite, a is a strict local minimum point for F.

(b) If H is negative definite, a is a strict local maximum point for F.

(c) If H takes both positive and negative values, a is neither a maximum nor a minimum point for F but some kind of saddle point.

Proof. Let $a = < a_1, a_2, \ldots, a_n >$. Suppose H is positive at the point $k = < k_1, k_2, \ldots, k_n >$.

Consider the function φ of one real variable defined by

$$\varphi(t) = F(a_1 + tk_1, a_2 + tk_2, \ldots, a_n + tk_n).$$

(This is F along the parametrized line $t \mapsto a + tk$.) We know that

$$\varphi'(0) = \sum k_i F_i(a) = 0$$

because a is a critical point of F, and

$$\varphi''(0) = \sum k_i k_j F_{ij}(a) = H(k) > 0.$$

Hence φ has a strict local minimum point at 0. This means $\varphi(t) > \varphi(0)$ for all sufficiently small but non-zero t. But this is the same as

$$F(a + tk) > F(a).$$

Thus, F takes values larger than $F(a)$ at points arbitrarily near to a. Hence, a is certainly not a local maximum point for F.

Similarly, if H takes a negative value somewhere, there is a line through the origin along which F has a strict local maximum at the origin. Hence the origin is not a local minimum point for F.

This proves (c).

We next prove (a). Assume H is positive definite. Then the Hessian matrix M is positive definite. According to theorem 6.3.12, there is a positive number $\mathcal{E}$ such that:

> If M' is any $n \times n$ symmetric matrix, each of whose entries is within $\mathcal{E}$ of the corresponding entry of M, then M' is also positive definite.

Because the second order partial derivatives of F are all continuous by hypothesis, we can choose δ so small that if b is any point with $\|b - a\| < \delta$ then $b \in E$ and

$$|F''_{ij}(b) - F''_{ij}(a)| < \mathcal{E}$$

for all i and j.

Now consider any point $c = < c_1, c_2, \ldots, c_n >$ with $0 < \|c - a\| < \delta$. By Theorem 8.4.5

$$F(c) = F(a) + \frac{1}{2} \sum_{i,j} (c_i - a_i)(c_j - a_j)F_{ij}(b)$$

where b is some point on the segment joining a to c. (Remember the first degree terms are zero because a is a critical point.) This point b will satisfy $\|b - a\| < \delta$ and hence the matrix

$$M' = \begin{Vmatrix} F''_{11}(b) & F''_{12}(b) & \cdots & F''_{1n}(b) \\ F''_{21}(b) & F''_{22}(b) & \cdots & F''_{2n}(b) \\ \cdots & \cdots & \cdots & \cdots \\ F''_{n1}(b) & F''_{n2}(b) & \cdots & F''_{nn}(b) \end{Vmatrix}$$

is positive definite. The sum in (11) is the quadratic form corresponding to M' evaluated at c - a. Since $c - a \neq 0$, the sum is positive. Hence

$$F(c) > F(a).$$

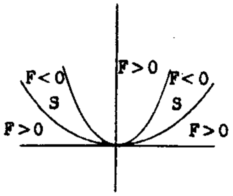
This proves that a is a strict local minimum for F. This finishes the proof of (a). The proof of (b) is similar. □

Remark. It is tempting to use the reasoning of the first part of the proof to prove (a) as follows. If H is positive definite, then F has a strict local minimum at a along every straight line through a. Hence a is a local minimum point for F.

That this last conclusion is a non-sequitur can be seen from the following example. Let

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

Then F has a critical point at the origin and a strict local minimum point at the origin along every line through the origin. Nevertheless, the origin is not a local minimum point for F. In fact, F is negative everywhere in the region S between the parabolas $y = x^2$ and $y = 3x^2$. Hence F has a strict local maximum point at the origin along the parabola $y = 2x^2$. The trouble is that no line through the origin can penetrate S immediately.



The Hessian form of a function at a critical point can be classified according to the scheme of §6.3. Properties of the Hessian form are often ascribed directly to the critical point. Thus a critical point is said to be degenerate if the Hessian form is degenerate. The criterion for this is the vanishing of the Hessian determinant (ie., the determinant of the Hessian matrix). The index of a critical point is the index of the Hessian form at that point.

Suppose F is a C^2-function on $\mathbb{R}^n$ and a is a critical point of F. Let H be the Hessian form for F at a. Let P_1 be the a-coset of some linear subspace P of $\mathbb{R}^n$ (ie., P_1 is a "flat" space through a). Consider F restricted to the set P_1. Then a is still a critical point for F and it is easy to check that the Hessian form for the restricted F at a is just H restricted to P. (For P of dimension one, this was established in the first part of the proof of 8.4.12. The argument given there extends to the general case.)

Now suppose P has been so chosen that H restricted to P is positive definite. Then F restricted to P_1 has a strict local minimum at a relative to P_1. By the same argument, if N is a linear subspace on which H is negative definite, and N_1 is the a-coset of N, then F has a strict local maximum at a relative to N_1.

To make practical use of Theorem 8.4.12 we need some way to decide whether the Hessian form is positive definite, negative definite, or indefinite. This is provided by Theorem 6.3.8. We illustrate with an example.

Example. Find the critical points of the following function on $\mathbb{R}^3$ and discuss their nature.

$$f = x^2 + xy + y^2 + xz + z^2 - \frac{1}{18} z^3.$$

The first order partial derivatives of f are

$$2x + y + z$$

$$x + 2y$$

$$x + 2z - \frac{1}{6} z^2.$$

The critical points are found by setting all three of these expressions equal to zero and solving for x, y, and z. The first two equations lead to $x = -2z/3$, and we find that there are two critical points $< 0, 0, 0 >$ and $<-\frac{16}{3}, \frac{8}{3}, 8 >$.

The Hessian matrix (at a general point) is

$$\left\Vert \begin{matrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 - \frac{1}{3} z \end{matrix} \right\Vert .$$

The sequence of determinants (as in 6.3.8) is $1, \; 2, \; 3, \; 4 - z$.

At both critical points the Hessian determinant is not zero, so both critical points are non-degenerate.

At the first critical point, $< 0, 0, 0 >$, the sequence is $1, 2, 3, 4$. There are no changes of sign, so the Hessian is positive definite and there is a strict local minimum point.

At the second critical point, $< -\frac{16}{3}, \frac{8}{3}, 8 >$, the sequence is $1, 2, 3, -4$. There is one change of sign so the index is one. The critical point is a saddle point. The Hessian form is positive definite on the linear subspace spanned by $< 1, 0, 0 >$ and $< 0, 1, 0 >$, that is the x-y plane. Correspondingly, f has a strict local minimum at the critical point relative to the plane whose equation is $z = 8$. Since the Hessian form is negative definite along the z-axis ($= \text{sp}\ \{< 0, 0, 1 >\}$), f has a strict local maximum at the critical point along the line $x = -16/3, \; y = 8/3$.

Exercises

1. Find the critical points of the following functions on $\mathbb{R}^2$ and discuss their nature.

   (a) $\frac{1}{x} + xy + \frac{8}{y}$ (b) $x^2 + \frac{4}{xy^2} + y^2$

   (c) $\sin x + \sin y$ (d) $xy + \tan x + \tan y$

   Do the same for the following functions on $\mathbb{R}^3$.

   (e) $x^2 + y^2 + 3z^2 - 2z^3$ (f) $x^3 + xy - xz + y^2 - z^2$

   (g) $xyz + \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$

2. Show that $x^4 + x^2y + y^2$ has a degenerate critical point at $< 0, 0 >$ that is, nevertheless, a strict global minimum point.

3. Calculate the Hessian form of the example on page 8-75 and show that it is degenerate. Note that the anomalous behavior appears along curves tangent to the subspace of degeneracy of the Hessian.

4. The Hessian form of a C^2-function can be defined at any point as the second degree terms in the second Taylor polynomial. Discuss, in terms of the Hessian form, whether the graph of a function F lies above or below its tangent plane near the point of tangency.

5. If f is a quadratic form, show that at every point the Hessian form of f is $2f$.

6. Suppose f is C^2 on all of $\mathbb{R}^n$ and at every point the Hessian of f is positive definite. Prove that: If v and w are two distinct points in $\mathbb{R}^n$ and $0 < t < 1$, then

$$f(tv + (1-t)w) < tf(v) + (1-t)f(w).$$

   A function satisfying this inequality is called strictly convex. Hint: Reduce to the one-dimensional case.

8.5 A geometric view of functions, the implicit function theorem.

By considering the ideas of level lines and level surfaces, we can acquire valuable insights into the nature of functions of several variables. These ideas are particularly useful in understanding the complex of results known collectively as the implicit function theorem. Roughly, this tells us when we can "solve" the equation $F(x,y) = 0$ for y in terms of x.

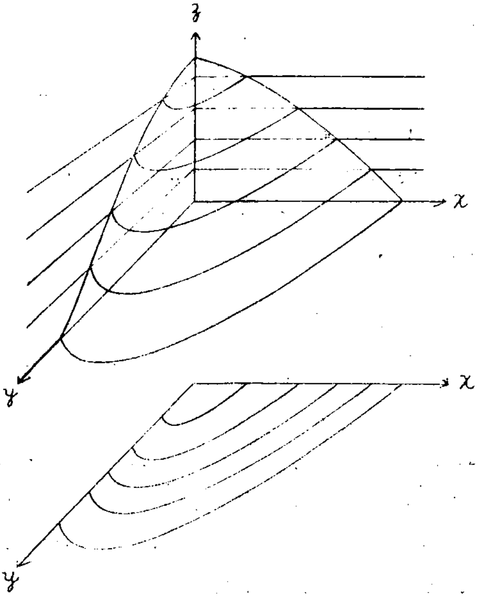
8.5.1 Level lines. Suppose f is a real-valued function defined on a plane (or an open subset of a plane). For each real number α, consider the set

$$\{ p : f(p) = \alpha \}.$$

If f is a reasonable function, these sets will be smooth curves, with perhaps an occasional singularity. Curves corresponding to nearly equal values of α will be more or less parallel. These curves are called level curves for f, sometimes contour lines or contour curves. Functions of two variables are often depicted by drawing a few representative level curves.

Contour lines are often used on maps to show the elevation of the terrain. The one hundred feet above sea level contour line shows where the shore would be if the sea rose one hundred feet. Maps usually show contour lines for equal intervals of elevation, say for one hundred feet, two hundred feet, three hundred feet, etc. The spacing of the contour lines on the map then tells whether the hills are steep or gentle. If the contour lines are close together it means that we go up a lot in a short linear distance. The hills are steep. When the contour lines are far apart, the hills are gentle. Similar considerations apply to the level lines of a function.

The α-level curve for a function can be obtained from its graph as follows: Find the curve where the graph intersects the horizontal plane $z = \alpha$ and "drop" this curve onto the x-y plane.



Above: Surface cut by equally spaced horizontal planes.

Below: Corresponding level lines in the x-y-plane.

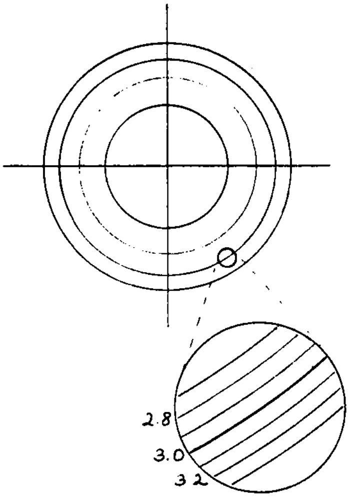
If f is a first degree function on a plane, its level curves are all straight lines. Geometrically this is because the graph of f is a non-horizontal plane and it will meet any horizontal plane in a line. Analytically this can be seen as follows: In coordinates $f = a + bx + cy$ where b and c are not both zero. A level curve for f has the equation

$$a + bx + cy = \alpha,$$

and this represents a straight line. Moreover, the lines corresponding to equally spaced levels are equally spaced parallel lines. The level difference divided by the distance between level lines is the tangent of the angle between the graph of f and the horizontal plane.

When we turn to more complicated functions, the level curves are usually some sort of smooth curves. Nearby level curves are usually in some sense parallel. This concept is a bit vague, but it is clearly illustrated in the following example.

Consider the function $x^2 + y^2$ on the plane. Its level curves for positive values are concentric circles. The origin is a degenerate level curve. If we look at this family of curves at a point p other than the origin with a high-powered microscope, the curves will appear to be parallel straight lines. The field of view would be small but much magnified. We would see circular arcs of such large apparent radius that they would look straight. This is almost equivalent to saying that $x^2 + y^2$ is differentiable at p. A function is differentiable at p if, when looked at in a sufficiently small neighborhood of p, it becomes indistinguishable from a first degree function. It is important, however, that p not be a critical point. No matter what magnification we use, if we look at the origin (a critical point for $x^2 + y^2$) we will see the degenerate level curve surrounded by concentric circles.



Level curves for $x^2 + y^2$.

Above: Levels 1, 2, 3, and 4.

Below: Ten times magnified view of small inset circle, showing additional level curves.

The same basic ideas apply to functions of three variables. If f is a real-valued function defined on space, the level sets

$$\{ p : f(p) = \alpha \}$$

will usually be smooth surfaces and are therefore called <u>level surfaces</u>. The level surfaces for a first degree function will be parallel planes with equally spaced values of α corresponding to equally spaced planes. For more complicated, but still differentiable, functions the level surfaces will be curved, but if viewed with a microscope they will appear like parallel planes, the resemblance increasing as the field of view diminishes and the magnification increases, all provided we are not looking at a critical point.

Level surfaces for a function on three-space are particularly valuable because they can be directly visualized whereas the graph of such a function cannot (because it takes four dimensions).

The idea of level surfaces remains sensible for functions of more than three variables even though we can no longer visualize them. The level surfaces for a function of four variables will be curved three-dimensional surfaces in four-space. In thinking about such things we are forced more than ever to rely on the analytic definitions. Because the analytic definitions serve so well to describe our intuitive conceptions of curves and surfaces in three-space, we can feel reasonably confident that our perceptions of three-space will provide valid insights into the nature of higher-dimensional space.

We shall give a geometric argument that shows that level lines are smooth curves. Suppose S is a smooth surface in three-space. By this we mean that S has a well-defined tangent plane T_q at every point q and T_q moves continuously with q. Let H be a plane that cuts S in a curve C. We would like to show that C is a smooth curve. Suppose q is a point of C such that $T_q \neq H$. Then $H \cap T_q$ is a line. Since S hugs closely to T_q near q, $H \cap T_q$ will be tangent to $H \cap S = C$ at q. This shows that C has a tangent at every point q except those for which $T_q = H$. Moreover, since T_q moves continuously with q, $H \cap T_q$ moves continuously with q. Thus, C is a smooth curve (has a continuously turning tangent line) as long as we avoid points at which $H = T_q$.

An analytical version of this argument will be part of the proof of theorem 8.5.12

Suppose now that S is the graph in $\mathbb{R}^3$ of some C^1-function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$. If q is a point of S, say $q = < a, b, F(a,b) >$, the equation of the plane T_q tangent to S at q is

$$z = F(a,b) + F_1'(a,b)(x - a) + F_2'(a,b)(y - b).$$

Since the coefficients here are continuous functions of a and b (because F is C^1), T_q moves continuously with q. Let H be a horizontal plane cutting S. If $q \in H \cap S$, the condition that $T_q \neq H$ is that at least one of the coefficients $F_1'(a,b)$ and $F_2'(a,b)$ is not zero; that is

$$dF(a,b) \neq 0.$$

Now the curve $C = H \cap S$ is, except for being "dropped" onto the x-y-plane, a level curve for F. We conclude that the level curves for F are smooth curves except possibly at points where dF vanishes.

Now consider an arbitrary point q of the surface S and its tangent plane T_q. Let L be any line in T_q passing through q. We can choose the plane H so that $H \cap T_q = L$. Then L will be the line tangent to $H \cap S$ at q. Hence we conclude that every line through q in T_q is tangent to some smooth curve in S.

8.5.2 The gradient of a function. Let f be a real-valued C^1-function defined on an open set E of an inner product space V. Then df is a differential form on E, that is, a function from E to V^*. If coordinates are at hand, df becomes a functions whose values are row vectors. There is no convenient way to visualize a row vector geometrically. However, because of the inner product in V, it is possible to replace a row vector by a column vector, and a column vector has a simple geometrical representation.

According to Theorem 5.4.18 (p. 5-66) for each linear functional g on V there is a vector w in V such that

$$(\forall v \in V) \qquad g[v] = (w,v).$$

Now $df(p)$ is a linear functional on V, so there is a member of V that represents it. This vector is called the gradient of f at p. We write it $\nabla f(p)$. (The symbol '∇' is pronounced "del".) The gradient of f is then a function ∇f from E to V. It is a vector field in the sense of 7.4.1.

The defining relation for the gradient is

$$(\forall v \in V) \qquad df(p)[v] = (\nabla f(p),\, v)$$

and the chain rule for computing the derivative of $f \circ g$ where g is a parametric curve in V becomes

$$(f \circ g)'(t) = (\nabla f(g(t)),\, g'(t)).$$

(See 8.3.35.)

How do we find this vector field computationally? If we are using orthonormal coordinates, just transpose the row vector df to obtain the column vector ∇f. Thus, ∇f is the column vector whose components are the partial derivatives of f, if the coordinates are orthonormal. To see this, note that, if coordinates are ortho-normal, the inner product of two vectors v and w is the same as the matrix product $v^T w$. For example

$$\left(\begin{Vmatrix} v_1 \\ v_2 \\ v_3 \end{Vmatrix}, \begin{Vmatrix} w_1 \\ w_2 \\ w_3 \end{Vmatrix} \right) = v_1w_1 + v_2w_2 + v_3w_3 = \begin{Vmatrix} v_1 & v_2 & v_3 \end{Vmatrix} \cdot \begin{Vmatrix} w_1 \\ w_2 \\ w_3 \end{Vmatrix}$$

The stress on orthonormal coordinates is quite necessary because when other systems of coordinates are used, the gradient has quite a different appearance. (See exercises 15 and 16.) It is important to realize that the gradient vector field of a function is independent of coordinate systems, it is only its representation that changes with the coordinates. Thinking in terms of orthonormal coordinates, we see right away that the gradient of a C^1-function is a continuous vector field.

Why bother with both the differential and the gradient of a function? We could certainly get along with just one of them. Since the gradient is a vector field, we can visualize it as an arrow diagram as on page 7-54. This can be very helpful. More important, however, are the many physical interpretations of the gradient. We shall touch briefly on these below.

The gradient is defined only when an inner product is available. Of course, we can always impose an inner product on a vector space, but if an inner product is imposed that is not germane to the situation under study, the gradient of a function is not likely to be useful.

Let us find the relation between the gradient vector field of a function and its graph and level curves. Suppose f is a real C^1-function defined on a plane, and let p be a point of the plane. The directional derivatives of f at p (p. 8-36) are the derivatives of f along unit vectors u. They are given by

$$df(p)[u] = (\nabla f(p), u).$$

Since $\|u\| = 1$, the Cauchy-Schwarz inequality tells us that

$$-\|\nabla f(p)\| \leq (\nabla f(p), u) \leq \|\nabla f(p)\| .$$

with the upper equality holding if and only if $\nabla f(p)$ and u have the same direction. Hence, of all directions at p, the one in which f increases fastest is the direction of $\nabla f(p)$, and $\|\nabla f(p)\|$ is the directional derivative in this direction.

Think of the graph of f as a three-dimensional landscape over the x-y-plane. Let q be the point of the graph over the point p of the x-y-plane. Then q is a point on a hillside. Our last result says that the gradient vector at p points in the direction of the steepest ascent of the hill at q, while the length (norm) of the gradient tells how steep the hill is. If the gradient is zero at p, there is no (instantaneous) rise or fall of the hill in any direction. Such a point could be a hill-top (maximum point for f) a pit-bottom (minimum point for f), or a pass between two hills (saddle point for f). Note that the gradient of f is zero exactly when the differential of f is zero, so the gradient vector vanishes exactly at the critical points of f.

Along a line through p orthogonal to $\nabla f(p)$ the directional derivative is zero. This means that f is not varying (instantaneously) along this line In terms of hills it means, if the steepest line on a hill is North-South, the slope is zero in the East-West direction. In the very simplest case, f is a first degree function and its graph is a non-horizontal plane. The level curves for f are parallel lines. The gradient of f is a constant vector field (ie., $\nabla f(p)$ is independent of p) perpendicular to these lines. The gradient vectors are always orthogonal to the level curves. Suppose $t \mapsto g(t)$ is an arc-length parametrization of the level curve through p with $g(0) = p$. Then $f \circ g$ is a constant function, so

$$(f \circ g)'(0) = (\nabla f(p), g'(0)) = 0.$$

But $g'(0)$ is a unit vector tangent to the level curve at p. This shows that the gradient vector at p is orthogonal to the level curve through p.

We can now make more precise our previous claim that nearby level curves are nearly parallel. Suppose that $\nabla f(p) \neq 0$. Since ∇f is a continuous vector field, in a small neighborhood of p, ∇f is almost constant. This implies that all values of ∇f near p are almost parallel to one another. (Note that this would not follow if $\nabla f(p) = 0$.) Hence the level curves, being at each point orthogonal to the gradient at that point, are almost parallel near p.

For a function f defined on three-space the results are essentially the same. The gradient vector at p points in the direction in which f increases fastest. Assuming $\nabla f(p) \neq 0$, the level set for f through p is a smooth surface near p and its tangent plane at p is orthogonal to $\nabla f(p)$. All the planes tangent to level surfaces of f at points near p will be nearly parallel to one another and this will give the level surfaces themselves the appearance of being parallel. If $\nabla f(p) = 0$, p is a critical point for f, and the level set for f through p need not be a smooth surface at all (in fact, usually it will not be, as we shall see). As before any smooth curve lying in the level

passing through p and lying in the level set of p will have its tangent vector orthogonal to $\nabla f(p)$, but this tells us nothing since $\nabla f(p) = 0$.

8.5.3 Level curves and surfaces near a non-degenerate critical point. Suppose f is a real-valued C^1-function defined on a plane. We have seen that, near a point p where df (or ∇f) does not vanish, the level curves are smooth and roughly parallel. Let g be the first Taylor polynomial for f at p and think of the level sets for g. Since g is a first degree function, these level sets are a family of parallel straight lines. The level curves for f itself can be obtained by bending those for g slightly. The level curve for g through p is the line tangent to the level curve for f through p. By slightly bending this line we can make it (locally) the level curve of f. If we think of the plane as made of rubber, it is easy to see how we can deform a small piece of the plane near p so as to make the level lines for g coincide with the corresponding level lines for f. There is, in fact, a theorem that makes this statement quite precise.

The situation in three or more dimensions is similar. Let f be a real-valued C^1-function, let p be a non-critical point for f, and let g be the first Taylor polynomial for f at p. It is possible to deform space slightly near p, with the distortion getting less and less as we approach p, so that the level surfaces for g, which are planes, become the corresponding level surfaces for f.

Now let f be a C^2-function, let p be a non-degenerate critical point for f, and let g be the second Taylor polynomial for f at p. Again it is possible to deform space near p slightly so that the level sets for g become those for f.

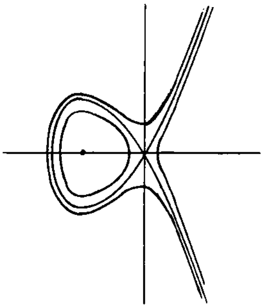
Suppose for a moment that space is two-dimensional (ie., a plane), and to avoid excessive notation, suppose the critical point is the origin and that f vanishes there. Then the second Taylor polynomial for f at the origin is just a quadratic form, g.

If g is positive definite, its level sets are a one point set, the origin, surrounded by concentric ellipses. Therefore, near the origin the level sets for f consist of a one point set, again the origin, surrounded by curves that are slightly deformed concentric ellipses. The one point level set through the origin is consistent with the fact that f has a strict local minimum at the origin. If we look at a neighborhood of the origin with a microscope, as the magnification increases, the more nearly will the level sets for f resemble the level sets for g.

If g is negative definite, the picture is essentially the same. The level sets for g are the origin and concentric ellipses (The ellipses now correspond to negative values of g.), while those for f are the origin and slightly deformed concentric ellipses.

Now assume that g has index one. Then the picture is quite different. The level set for g through the origin consists of a pair of crossed lines. Each other level set consists of two disconnected parts and is a hyperbola. We can get the level sets for f by deforming the picture slightly, keeping the origin fixed. The level set for f through the origin will consist of two crossed curves, each tangent to one of the lines of the level set for g. The other level sets for f will resemble the hyperbolas for g. They will be disconnected near the origin, but they may be reconnected at some remote point.

An example will show how to use these facts. The function $y^2 - x^2(x + 3)$ has two critical points in the plane, both non-degenerate. At $< -2, 0 >$ it has a strict local minimum point with value -4. This is shown in the figure as a dot. At $< 0, 0 >$ there is a critical point of index one. The level set through the origin therefore consists locally of two crossed curves. Since the second Taylor polynomial is $y^2 - 3x^2$, the curves are tangent to the two lines given by $y^2 - 3x^2 = 0$; that is, $y = \pm\sqrt{3}\ x$. It turns out that the crossed curves eventually join together to make a single curve that crosses itself. The figure also shows the level curves for values + 1 and - 1. The former is connected, while the latter consists of a closed curve surrounding the minimum point and an infinite arc in the right half plane.



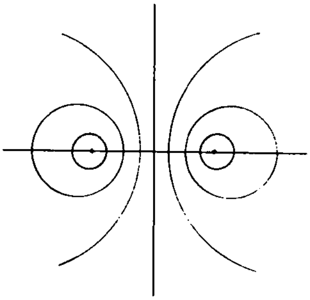
From this information it is easy to describe all the level sets. For values less than -4, there will be one arc in the right half-plane. At level -4 the minimum point at $< -2, 0 >$ appears as well. For values between -4 and 0, there will be a closed loop surrounding the minimum point and an arc in the right half-plane. These two parts coalesce for value 0 to make the curve that crosses itself at the origin. The level sets for all positive values will be roughly like the one shown for $+1$.

The values taken by f at its critical points are called <u>critical values</u>. Generally speaking, the level sets corresponding to nearby values have the same rough shape, but there is an abrupt change of shape when a critical value is passed. In the example, the level sets were connected for values less than -4, but a new piece appeared as we passed the critical value -4. The level sets remained in two pieces until we got to the critical value 0, at which point the two pieces coalesced.

Gross changes in shape can also occur when the level curves "go to infinity." An example of this phenomenon is given by the function

$$\frac{2x}{1 + x^2 + y^2}$$

There are critical points at $< -1, 0 >$ and $< +1, 0 >$, both non-degenerate. The former is a minimum and the latter a maximum. The level curves are all circles except the line $x = 0$. At this level (value 0) the shape of the level curves changes abruptly although there is no critical point.



In general if you can find the level sets for the critical values of f, the rest can easily be sketched. All other level sets are curves with no singularities. Between two consecutive critical values the level curves must make a smooth transition from one critical level set to the other. This in itself is usually enough to determine their appearance to a satisfactory level of accuracy.

What we have done applies only to non-degenerate critical points. Near a degenerate critical point the level sets may have a very complicated structure. In fact, no complete analysis of the structure of functions near a degenerate critical point is known.

Level surfaces for functions defined on three-space are in a way even more important than level curves in the plane, because level surfaces provide the only way we can visualize functions on three space. Fortunately, the ideas are essentially the same as in the plane. We stick to functions of class C^2 at least with only non-degenerate critical points.

There are basically only two kinds of critical points. Those of index zero and three, corresponding to minimum and maximum points of f, are one kind, while those of index one and two, corresponding to saddle points, are the other. At a critical point of index zero or three there is a one point level set surrounded by slightly deformed ellipsoids.

As we saw in chapter six (p. 6-92 ff) the critical level set for a quadratic form of index one or two is a quadric cone. The region inside the two nappes of the cone is packed with hyperboloids of two sheets, one sheet in each nappe. The region outside the nappes is wrapped with hyperboloids of one sheet. If the index is one, the hyperboloids inside the nappes correspond to the values less than the critical value. If the index is two, these hyperboloids correspond to values greater than the critical value. According to the general result the level sets for f will be slightly deformed versions of those for its second Taylor polynomial. In particular the critical level set for f will be a deformed quadric cone. Although otherwise a smooth surface, it pinches down to a point as it passes through the critical point.

The actual determination of the level sets for a given f may be quite a difficult task, but as in the case of two variables it is easier if we keep in mind the general facts about their structure. The first thing is to find the critical points of f and then the level sets corresponding to the critical values. All remaining level sets are smooth surfaces with no singularities. We illustrate with an easy example.

Let $f = y^2 + z^2 - x^2(x + 3)$. There are critical points at $< -2, 0, 0 >$ and $< 0, 0, 0 >$. They are both non-degenerate and have indices 0 and 1, respectively, with critical values -4 and 0. If we fix x, that is, confine our attention to a plane of the form $x = \lambda$, we see that the level sets become circles with center on the x-axis. Therefore the level sets are all surfaces of revolution with axis the x-axis. In fact they are the surfaces obtained by revolving the level curves of the example of page 8-89. The 0-level set is worthy of particular note. It is a surface with a bubble pinched off at the origin. The conical point at the origin is characteristic of the critical point of index one.

8.5.4 Fall lines. Let f be a function defined on the plane and let us consider once again the graph of f as a landscape spread out over the x-y-plane. If a drop of water is spilled on a hillside it will run down the hill taking the steepest path down. Let us neglect any tendency of the water to "coast" because of acquired velocity. Then the path of the water will always be exactly along the line of steepest descent. The path of the water projected down on the x-y-plane will always have the direction of the negative of the gradient of f. This condition becomes a differential equation satisfied by the projected paths that in reasonable cases completely determines the paths. The paths are called <u>fall lines</u> for f.

Suppose x and y are cartesian coordinates in a two-dimensional inner product space V. Let $f = x^2 - y^2$ We shall find the fall lines for f.

We seek parametrized curves $g : \mathbb{R} \longrightarrow V$ such that, for any t, $g'(t)$ the tangent vector to the curve has the direction of $-\nabla f(g(t))$. While we are at it, let us make

$$g'(t) = -\nabla f(g(t)). \tag{5}$$

This means we are prescribing the rate at which the fall line is to be described as well.

Write g in components, say

$$g(t) = \begin{Vmatrix} u(t) \\ v(t) \end{Vmatrix}.$$

Now

$$f = \begin{Vmatrix} 2x \\ -2y \end{Vmatrix}$$

so (5) becomes

$$\begin{Vmatrix} u'(t) \\ v'(t) \end{Vmatrix} = \begin{Vmatrix} -2u(t) \\ 2v(t) \end{Vmatrix}$$

Therefore

$$\begin{Vmatrix} u(t) \\ v(t) \end{Vmatrix} = \begin{Vmatrix} ae^{-2t} \\ be^{2t} \end{Vmatrix}$$

for suitable constants a and b which are determined by where the motion starts. We can eliminate t here and conclude that the motion takes place along the curve with equation

$$xy = ab.$$

If $ab \neq 0$, this is a hyperbola. It falls in two parts and the motion takes place along just one of them. The fall lines of f are therefore, for the most part, half hyperbolas. A motion that starts at the origin will be stationary, and a motion that starts elsewhere on a coordinate axis will take place on a half line from the origin.

The curves we have found have the property that at every point they are orthogonal to the level curves of f. They are therefore also known as the orthogonal trajectories of the level curves.

8.5.6 Potential fields. It frequently happens in physics that a body has potential energy by virtue of its position alone. The function that tells the potential energy of a given body in a given place is called the potential function. There is always a tendency for bodies to move so as to reduce their potential energy, so a body in a potential field will experience a force that tends to move it as quickly as possible to a point of lower potential energy. The magnitude of this force is proportional to the rate at which the potential energy falls with distance. Hence if units are chosen correctly, the force is exactly the negative of the gradient of the potential energy function. Many, but not all, of the force fields that arise in physical problems are the negative gradient of some potential function. Force fields that do arise in this manner are called conservative, because the principle of the conservation of energy (potential plus kinetic) applies to bodies moving under the influence of such force fields. (It does not apply to other force fields unless reckoning is made also of the energy necessary to maintain the field.)

An extremely important example of a conservative force field is the gravitational field surrounding a heavy body. (See exercise 10.)

Exercises. Assume throughout that x and y are orthonormal coordinates on a plane and x, y, and z are orthonormal coordinates on three-space.

1. Find a parametric representation for the line normal to the surface in space given by $x^3 + y^3 + \sin xyz = 0$ at the point $< 1, -1, 0 >$.

2. Find an equation for the plane tangent to $x^4 - xy^2 + 2y^3 = 2z^2$ at the point $< 1, 1, 1 >$.

3. At what points of the plane does the gradient vector of $x^2y + 2y^2$ point toward the origin? (We think of the gradient vector of f as running from v to $v + \nabla f(v)$.)

4. At what points of the plane are the level curves for $x^3 + y^3$ perpendicular to those for xy ?

5. At what points of the plane are the level curves for $x^3 - y^3$ tangent to those for $x^2 + xy$?

6. Show that, although $x^3 + y^3$ has a degenerate critical point, all of its level sets are in fact smooth curves.

7. What is the maximum value of the directional derivative of the function
$$\frac{x}{1 + x^2 + y^2}$$
considering all directions at all points of the plane?

8. If the gradient of a function on three-space always points towards the origin (see ex. 3.), show that the function is constant on spheres with center at the origin.

9. Find the fall lines for the function $x^2 + 2y^2$.

10. The potential energy of a body of mass m in the gravitational field of a fixed body of mass M is $-KMm/\rho$ where ρ is the distance between them. Show that the gravitational force on the first body is the negative of the gradient of this potential function.

11. Is the set defined in three-space by

$$x^3 + y^3 + \sin xyz = 1$$

everywhere a smooth surface?

12. Sketch the level curves in the plane for

(a) $\sin x + \sin y$

(b) $\frac{1}{x^2 + y^2} + 2x$

(c) $\frac{1}{x^2 + y^2} + x^2$

13. Suppose f is a real valued function defined on all of the plane and that all of its level sets are straight lines. Show that there are numbers a and b and a function g of one variable such that

$$f = g(ax + by).$$

What happens if f is defined on less than the whole plane, but all its level sets are straight line segments?

14. The plane curve defined by $2x^3 + 4x^2 + 4xy + 4y^2 = 1$ crosses itself. Where is the crossing and at what angle does it cross?

15. Suppose $b_1, b_2, \ldots, b_n$ is a basis of an inner product space V and $x_1, x_2, \ldots, x_n$ are the coordinate functions on V associated with this basis. Let M be the matrix of the inner product referred to this basis. Show that in this coordinate system the gradient of a C^1-function f is given by

$$M^{-1} \left\| \frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \qquad \frac{\partial f}{\partial x_n} \right\|^T$$

16. Suppose v and w are continuous vector fields defined on a subset E of two dimensional inner product space V. Suppose that $v(p)$ and $w(p)$ are linearly independent at every point p. Show that any continuous vector field u on E can be written $u = gv + hw$ (pointwise) where g and h are continuous functions from E to $\mathbf{R}$. (Continued next page.)

When computations are done in a two-dimensional inner product space using polar coordinates, it is customary to refer vector fields defined on $E = V - \{\text{origin}\}$ to the basis vector fields $\underset{\sim}{r}$ and $\underset{\sim}{\Theta}$ (these are usually written in bold-face type) defined as follows: For $p \in E$, $\underset{\sim}{r}(p)$ is the unit vector with the direction of p; ie., $\underset{\sim}{r}(p) = p/\|p\|$. (Remember p itself is a vector.) $\underset{\sim}{\Theta}(p)$ is the unit vector obtained from $\underset{\sim}{r}(p)$ by rotating it 90° in the positive direction.

Show that

$$\underset{\sim}{r} = \begin{Vmatrix} \cos\Theta \\ \sin\Theta \end{Vmatrix} \qquad \underset{\sim}{\Theta} = \begin{Vmatrix} -\sin\Theta \\ \cos\Theta \end{Vmatrix}$$

where the column vectors are with respect to the usual Cartesian coordinates.

Show that the gradient of a C^1-function f is given by

$$\nabla f = \frac{\partial f}{\partial \rho}\underset{\sim}{r} + \frac{1}{\rho}\frac{\partial f}{\partial \Theta}\underset{\sim}{\Theta} .$$

8.5.7 The implicit function theorem. Given a function F of two real variables the question often arises, can we solve

$$F(x,y) = 0$$

for y ? The implicit function theorem gives us valuable information about this problem. It is important to get in mind, however, what exactly we mean by "solving for y."

Consider first some particular cases.

(8) $$x^3 + xy + 5x + 4y = 0$$

(9) $$y^7 + y - x = 0$$

(10) $$\sin y - x = 0$$

It is easy to solve (8) for y.

$$y = -\frac{x^3 + 5x}{x + 4}$$

provided $x \neq -4$; there is no y satisfying (8) if $x = -4$. Here we have solved for y in the best possible sense. We have a formula for y in terms of x using only the familiar operations. We certainly cannot expect to do this well in a general context.

Equation (9) is more difficult. Fix an x temporarily. As y increases from $-\infty$ to $+\infty$, $y^7 + y$ also increases from $-\infty$ to $+\infty$. Hence, by the intermediate value theorem, there is a unique value of y such that $y^7 + y = x$; moreover, this value is unique. Hence for each real number x there is a unique y such that (9) is true. This determines y as a function of x. In technical terms, $\{ <x, y> : y^7 + y - x = 0 \}$ is a function. Not a familiar function to be sure, but a function. We offer no way to compute it other than to solve (9) afresh for each new value of x using your favorite algorithm for finding the roots of polynomial equations. This is the kind of information that the implicit function theorem gives: there exists a function that solves the given equation. It will also tell us that the solution function is differentiable. For example, the solution of (9) is C^∞.

Another complication arises with equation (10). For some values of x there are no values of y that satisfy (10); for others there are infinitely many. To express y as a function of x means that we must assign a unique value of y to each x. So we arbitrarily discard values of y outside $[-\pi/2, \pi/2]$. For each value of x in $[-1, 1]$ there is a unique y in $[-\pi/2, \pi/2]$ such that (10) holds. This defines a real-valued function with domain $[-1, 1]$. This has become a familiar function, usually written arcsin. When we write

$$y = \arcsin x \tag{11}$$

we have solved (10) for y, but only in a limited sense, because there are many pairs $<x, y>$, for example, $<0, \pi>$, that satisfy (10) but not (11).

We must be prepared for this possibility in any general theorem on solving for y

The best way to understand the theorem is through level sets. Given F, the set

$$S = \{ < x, y > : F(x,y) = 0 \}$$

is just the 0-level set for F in $\mathbb{R}^2$. For a general F there is no guarantee that S is not empty. Hence the theorem will assume that we have a point $< a, b >$ of S in hand. Then we ask, is the part of S near $< a, b >$ the graph of a function? We have given geometric arguments to show that near $< a, b >$, S is a curve provided $dF(a,b) \neq 0$. Now we shall give an analytic proof that it is the graph of a function provided $F_2'(a,b) \neq 0$.

If a piece of S is the graph of a function g, then

$$y = g(x)$$

can be regarded as the result of solving $F(x,y) = 0$ for y in terms of x. The theorem will not say how large the domain of g will be, only that it will be some interval around a. There will be no claim that g can be expressed in terms of familiar functions. Although we cannot claim that our solution is unique globally, it is the only solution that is continuous and satisfies $g(a) = b$.

8.5.12 *Theorem*. *Let* E *be an open set in* $\mathbb{R}^2$ *and let* $F : E \to \mathbb{R}$ *be a* C^1*-function*. *Suppose* $< a, b > \in E$, $F(a,b) = 0$, *and* $F_2'(a,b) \neq 0$. *Then*:

Existence: *There exists an open interval* I *in* $\mathbb{R}$ *such that* $a \in I$ *and* *a* C^1*-function* $g : I \to \mathbb{R}$ *such that* $g(a) = b$ *and*

$$(\forall x \in I) \quad F(x,g(x)) = 0.$$

Uniqueness: *If* J *is a subinterval of* I *containing* a *and* $h : J \to \mathbb{R}$ *is a continuous function such that* $h(a) = b$ *and*

$$(\forall x \in J) \quad F(x,h(x)) = 0,$$

then h *and* g *agree on* J.

Proof. Since $F_2'(a,b) \neq 0$, we shall assume that $F_2'(a,b)$ is actually positive. (If it is negative, consider the function $-F$ instead.) Set $F_2'(a,b) = 2\alpha$, where $\alpha > 0$. Since F_2' is continuous, there is an open disk Δ about $P = < a, b >$ such that $F_2'(x,y) > \alpha$ for all $< x, y > \in \Delta$. Say the radius of Δ is 2δ.

Consider the figure; some of the notation is defined there.

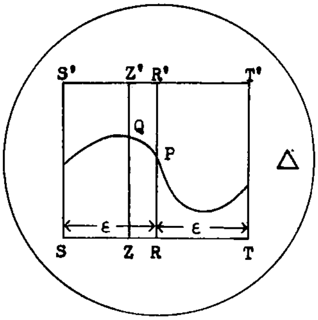
$P = < a, b >$

$R = < a, b - \delta >$ $\qquad$ $R' = < a, b + \delta >$

$S = < a - \varepsilon, b - \delta >$ $\qquad$ $S' = < a - \varepsilon, b + \delta >$

$T = < a + \varepsilon, b - \delta >$ $\qquad$ $T' = < a + \varepsilon, b + \delta >$

$Z = < x, b - \delta >$ $\qquad$ $Z' = < x, b + \delta >$



Because F_2' is strictly positive on Δ, F is strictly increasing along the segment RR'. Since $F(P) = 0$, $F(R) < 0$ and $F(R') > 0$.

If ε is a sufficiently small positive number then

S, S', T, and T' are all in Δ,

F is negative at each point of the segment ST, and

F is positive at each point of the segment $S'T'$.

(The last two condtions by the continuity of F.)

Choose any point x of $I = (a - \varepsilon, a + \varepsilon)$. Then x is the abscissa of a vertical segment ZZ'. We know that $F(Z) < 0$ and $F(Z') > 0$, so by continuity $F(Q) = 0$ for some Q on the segment ZZ'. Since F_2' is strictly positive on ZZ', F increases strictly along ZZ', so the point Q is unique. We define $g(x)$ as the ordinate of Q. Then $F(x, g(x)) = 0$.

This construction is valid for any $x \in I$, so the required function $g : I \to \mathbb{R}$ has been found. (It should be clear that the construction of g can be given in purely analytic terms. The notation is just messier and the ideas a bit harder to follow.) We must show that g is C^1, but first we shall prove the uniqueness.

Suppose h is a continuous function defined on a subinterval J of I satisfying $h(a) = b$ and $F(x, h(x)) = 0$ for all $x \in J$. We give an indirect proof that h agrees with g.

Suppose for some $x \in J$, $h(x) \neq g(x)$. Then $h(x)$ does not lie in $[b - \delta, b + \delta]$ because there is only one number y (namely, $y = g(x)$) such that $F(x,y) = 0$ and $y \in [b - \delta , b + \delta]$. Say that $h(x) > b + \delta$. By the continuity of h, there must be a point x' between a and x (so $x' \in J$) such that $h(x') = b + \delta$; that is, the graph of h crosses the segment $S'T'$ at $< x', b + \delta >$. But F is positive at this point, so $F(x', h(x')) \neq 0$, contrary to the assumption about h. This proves that $h(x) > b + \delta$ is impossible. Similarly, $h(x) < b - \delta$ is impossible, for then the graph of h would contain a point of ST and at this point F is not zero. Altogether this shows that $h(x) \neq g(x)$ is impossible. Thus h agrees with g on J.

Now we shall prove that g is differentiable at a. In fact we shall prove that

(13) $$g'(a) = -\frac{F_1'(a,b)}{F_2'(a,b)} .$$

Let F^* be the first Taylor polynomial for F at $< a, b >$.

$$F^*(x,y) = (x - a)F_1'(a,b) + (y - b)F_2'(a,b).$$

(Recall that $F(a,b) = 0$.)

Because $F_2'(a,b) \neq 0$, we can solve $F^*(x,y) = 0$ for y. Let the result be $y = g^*(x)$. Then

(14) $$g^*(x) = -\frac{F_1'(a,b)}{F_2'(a,b)}(x - a) + b$$

and

(15) $$F^*(x, g^*(x)) = 0$$

for all x.

(To see the relation between this and the geometric argument of p.8-82, temporarily introduce a third coordinate. Then $z = F^*(x,y)$ is the equation of the tangent plane T_q to the surface S given by $z = F(x,y)$ at $q = < a, b, 0 >$. The intersection of T_q with the x-y-plane H is the line L given by $F^*(x,y) = 0$ which is the same as $y = g^*(x)$. We shall prove that g^* approximates g near a in the sense of 8.3(16), p. 8-34, rewritten for one dimension. This proves that L is tangent to the graph of g as we claimed on geometric grounds.)

We want to show that $|g(x) - g^*(x)|$ goes to zero faster than $|x - a|$ as $x \longrightarrow a$. Given $\eta > 0$, we must find $\xi > 0$ and prove the inequality

(16) $$|g(x) - g^*(x)| \leq \eta|x - a|$$

for all x with $|x - a| < \xi$.

First, choose $\zeta > 0$ so that

(17) $$\|< x, y > - < a, b >\| < \zeta \quad \Longrightarrow \quad |F(x,y) - F^*(x,y)| \leq \frac{\alpha\eta}{M}\|< x, y > - < a, b >\| ,$$

where

$$M = \sqrt{1 + \frac{F_1'(a,b)^2}{F_2'(a,b)^2}} .$$

We can do this because F is differentiable at $< a, b >$.

Now let ξ be the smallest of ε, ζ/M, and $2\delta/M$.

Let x be any number such that $|x - a| < \xi$. From (14) it follows that

(18) $$\|< x, g^*(x) > - < a, b >\| = M|x - a|.$$

(Pythagorean theorem.) Now (18) and $M|x - a| < M\xi \leq \zeta$ imply that we can take $y = g^*(x)$ in (17) and get

$$|F(x, g^*(x)) - F^*(x, g^*(x))| \leq \frac{\alpha\eta}{M}\|< x, g^*(x) > - < a, b >\|.$$

Using (15) and (18), this becomes

(19) $$|F(x, g^*(x)| \leq \alpha\eta|x - a|.$$

From (18) and $M|x - a| < M\xi \leq 2\delta$, we deduce that $< x, g^*(x) > \in \Delta$. Since $|x - a| < \epsilon$, $g(x)$ is defined and $< x, g(x) > \in \Delta$. Hence the segment connecting $< x, g(x) >$ to $< x, g^*(x) >$ lies in Δ.

Recall that $F(x, g(x)) = 0$ and apply the mean value theorem.

$$\begin{aligned} |F(x, g^*(x))| &= |F(x, g(x)) - F(x, g^*(x))| \\ &= |F_2'(x, \Theta)(g(x) - g^*(x))| \\ &= F_2'(x, \Theta)|g(x) - g^*(x)| \\ &\geq \alpha\, |g(x) - g^*(x)|, \end{aligned} \tag{20}$$

where Θ is between $g(x)$ and $g^*(x)$. The last two steps follow from the fact that $F_2' > \alpha$ on all of Δ.

Comparing (19) and (20), we have

$$|g(x) - g^*(x)| \leq \eta|x - a|.$$

Since x was arbitrary except for the requirement $|x - a| < \xi$, this proves (16). And this shows that g is differentiable at a with derivative given by (13).

The argument just given applies with minor changes at any point $x \in I$, so g is differentiable at any point of I with

$$g'(x) = -\frac{F_1'(x, g(x))}{F_2'(x, g(x))}. \tag{21}$$

As a differentiable function g is continuous. Hence (21) shows that g' is a combination of continuous functions with the denominator never zero, so g' is continuous on all of I. We have proved that g is C^1.□

8.5.21 <u>Corollary</u>. <u>If</u> F <u>in the theorem is</u> C^k $(1 \leq k \leq \infty)$, <u>then</u> g <u>is also</u> C^k.

Proof. Assume F is C^k. We shall show by induction on p that g is C^p for $1 \leq p \leq k$. Suppose g is C^p where $p < k$. Then (21) exhibits g' as a combination of functions of class C^p and C^{k-1}. Since $p \leq k - 1$, this shows that g' is C^p. But then g is C^{p+1}.

If k is finite, this gives us induction as far as k, and we conclude g is C^k, If k is infinite, we conclude that g is C^p for every integer p. But this is what it means to be C^∞. □

The theorem we have just proved extends to any number of variables. Since the proof is about the same as for two variables, we shall only state the result

8.5.23 *Theorem*. *Let* E *be an open set in* $\mathbb{R}^{n+1}$ *and let* $F : E \to \mathbb{R}$ *be a* C^k*-function*, $1 \leq k \leq \infty$. *Suppose* $< a_1, a_2, \ldots, a_n, b > \in E$, $F(a_1,a_2,\ldots,a_n,b) = 0$, *and* $F'_{n+1}(a_1,a_2,\ldots,a_n,b) \neq 0$. *Then*:

Existence: *There exists an open ball* U *about* $< a_1, a_2, \ldots, a_n >$ *in* $\mathbb{R}^n$ *and a* C^k*-function* $g : U \to \mathbb{R}$ *such that* $g(a_1,a_2,\ldots,a_n) = b$ *and*

$$(\forall < x_1, x_2, \ldots, x_n > \in U)$$

$$F(x_1,x_2,\ldots,x_n,g(x_1,x_2,\ldots,x_n)) = 0$$

Uniqueness: *If* V *is a connected open subset of* U *containing* $< a_1, a_2, \ldots, a_n >$ *and* $h : V \to \mathbb{R}$ *is a continuous function such that* $h(a_1,a_2,\ldots,a_n) = b$ *and*

$$(\forall < x_1, x_2, \ldots, x_n > \in V)$$

$$F(x_1,x_2,\ldots,x_n,h(x_1,x_2,\ldots,x_n)) = 0,$$

then h *and* g *agree on* V. □

Once we know that the function g exists and is differentiable, we can easily calculate its derivative using the chain rule. If we differentiate

$$F(x_1,x_2,\ldots,x_n,g(x_1,x_2,\ldots,x_n)) = 0$$

with respect to x_j, keeping the other x's fixed, we get

$$F'_j + F'_{n+1} \cdot g'_j = 0.$$

Hence

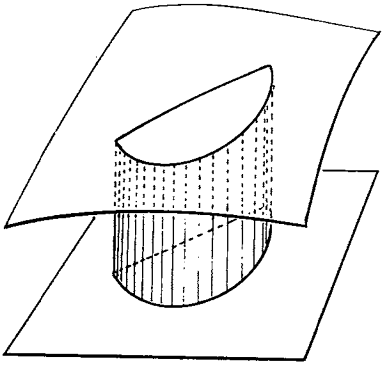
$$g'_j = - F'_j / F'_{n+1}.$$

(21) is just a special case of this formula.

8.6 Multiple integration.

The idea of integrating a function of one variable over a line interval generalizes in a very natural way to give us integrals of a function of two variables over a region in a plane and integrals of a function of three variables over a region in space.

8.6.1 Volume under a surface. Although the definition of a multiple integral is purely analytic it is most easily motivated by considerations of geometric volume just as the ordinary one-dimensional integral is motivated by area.



Suppose that f is a continuous real-valued function defined on the plane. Let S be a bounded closed region in the plane with boundary consisting of a finite number of smooth curves and corners. Such a region we shall call an <u>ordinary region</u>. For example, S might be a semi-circle, as in the figure. We assume the f is positive at every point of S.

The graph of f will be a surface in three-space. We want to know the volume of the solid V whose base is S, whose top surface is the part of the graph of f lying over S, and whose side walls are vertical over the boundary of S.

To find this volume we use the same reasoning as we used to find the area under a curve. First cut the region S up into a large number of ordinary regions which overlap only along their boundaries. Call them $T_1, T_2, \ldots, T_n$

Over each T_i stands a solid W_i with base T_i, top surface part of the graph and sidewalls vertical over the boundary of T_i.

Since the T's overlap only along boundaries which have area zero,

$$\text{Area } S = \text{Area } T_1 + \text{Area } T_2 + \dots + \text{Area } T_n.$$

Similarly, the solids W_i overlap only on surfaces which have volume zero, so

$$\text{Vol } V = \text{Vol } W_1 + \text{Vol } W_2 + \dots + \text{Vol } W_n.$$

Let us get inequalities for the volumes of the W's. Since f is a continuous function on the bounded closed set T_i, it has a maximum and a minimum value on T_i. Let these be M_i and m_i, respectively. Then

$$m_i \text{ Area } T_i \leq \text{Vol } W_i \leq M_i \text{ Area } T_i$$

because W_i contains a solid of fixed height m_i standing over T_i and W_i fits inside a solid of fixed height M_i standing over T_i.

Adding up these inequalities, we get

$$(2) \qquad \sum m_i \text{ Area } T_i \leq \text{Vol } V \leq \sum M_i \text{ Area } T_i$$

The left and right hand sums here are called the lower and upper Riemann sums, respectively, corresponding to the subdivision of S. They are analogous to the Riemann sums for a one-dimensional definite integral.

If S is carved up into sufficiently small pieces T_i, then the left and right members of (2) will differ by very little, in fact we can make the difference as small as we please by making the T's small enough. Hence (2) gives us a means of calculating Vol V as accurately as we please. Since the same limiting process comes up in many contexts, there is a notation for the unique number that fits between all the lower sums and all the upper sums. It is

$$\iint_S f \, dA$$

It is called the <u>double integral of</u> f <u>over the region</u> S.

Now we shall formalize some of these ideas.

8.6.3 Definition. A region in the plane will be called ordinary if and only if it is bounded, closed, and its boundary consists of a finite number of smooth curves and corners. This is not a standard term.

8.6.4 Area. The hardest part of the theory behind double integrals is the notion of area. We shall assume that we can assign to each ordinary region a non-negative number called its area in such a way that

(a) If an ordinary region S is subdivided into two ordinary regions T and U which share only boundary points, then

$$\text{Area } S = \text{Area } T + \text{Area } U.$$

(b) If two ordinary regions are congruent they have the same area.

(c) A square of unit edge has area one.

It is possible to prove that this can be done, and furthermore that it can be done in only one way.

Once we have established or assumed the existence of an area function we can define the double integral in a purely analytic way.

8.6.5 The double integral. Let S be an ordinary region in the plane and let $f : S \longrightarrow \mathbb{R}$ be continuous.

For each subdivision of S into ordinary regions $T_1, T_2, \ldots, T_n$ which overlap only along their boundaries form the Riemann upper sum

$$U(T_1, T_2, \ldots, T_n) = \sum M_i \text{ Area } T_i$$

where M_i is the largest value of f on T_i, and the Riemann lower sum

$$L(T_1, T_2, \ldots, T_n) = \sum m_i \text{ Area } T_i$$

where m_i is the least value of f on T_i.

It can be proved that every lower sum is less than or equal to every upper sum, even when these sums come from different subdivisions. Moreover, by choosing the T's small enough we can make

$$U(T_1, T_2, \ldots, T_n) - L(T_1, T_2, \ldots, T_n)$$

as small as we please. It follows from the nested interval principle (p. 4-21) that there is a unique real number I such that

$$L(T_1, T_2, \ldots, T_n) \leq I \leq U(T_1, T_2, \ldots, T_n)$$

for every choice of the subdivision $T_1, T_2, \ldots, T_n$. This number I is called the double integral of f over S and denoted

$$\iint\limits_S f \, dA.$$

There are other notations in common use. When x and y are Cartesian coordinates in the plane the double integral is commonly written

$$\iint\limits_S f \, dxdy$$

and when a formula for f in terms of x and y is at hand it is usually written out in the integral; eg.,

$$\iint\limits_S \frac{1}{x + y} \sin xy \, dx \, dy.$$

This notation looks ahead to the fact that double integrals are usually evaluated by performing two successive ordinary integrations.

The proof that upper sums and lower sums are eventually close together is instructive. For a fixed subdivision $T_1, T_2, \ldots, T_n$

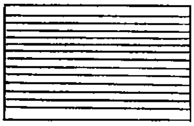
$$\begin{aligned} U - L &= \sum (M_i - m_i) \text{ Area } T_i. \\ &\leq D \sum \text{Area } T_i = D \text{ Area } S, \end{aligned}$$

where D is the largest of the numbers $M_i - m_i$.

Given $\varepsilon > 0$, it is possible to choose $\delta > 0$ so that

$$|f(p) - f(q)| < \frac{\varepsilon}{\text{Area } S}$$

whenever the points p and q are within δ of one another. Hence, if we choose the T's so that any two points in the same T_i are within δ of one another we shall certainly have $M_i - m_i < \varepsilon/\text{Area } S$, for all i. Then $D < \varepsilon/\text{Area } S$, and $U - L < \varepsilon$. This shows that, in order to make $U - L$ small we must make the T's small, not in the sense of area, but small in the sense of their linear dimensions. If we only made them small in area, they could all be long and thin. Then all of the numbers $M_i - m_i$ might be large.



Sometimes a more general type of Riemann sum is important. After choosing a subdivision of S, pick one point p_i in each part T_i and form the Riemann sum

$$R = \sum f(p_i) \text{ Area } T_i.$$

Since it is easy to see that R must come between the upper and lower sums for this subdivision, that is,

$$L \leq R \leq U,$$

R must be a good approximation to $\iint_S f \, dA$ whenever all the T's are small.

Exercise. Let S be the unit square in $\mathbb{R}^2$; ie., $S = \{ < x, y > : 0 \leq x, y \leq 1 \}$ Let $f(x,y) = xy$. Calculate the uppser and lower Riemann sums for f over S using the subdivision of S into n^2 small squares given by the lines $x = i/n$, $y = j/n$, $i, j = 1, 2, \ldots, n - 1$. Evaluate

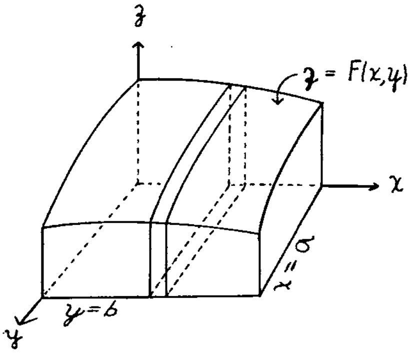
$$\iint_S xy \, dx \, dy.$$

8.6.6 Conversion to an iterated integral. The process of subdividing and computing Riemann sums is even more impractical for double integrals than it is for the one dimensional case. Double integrals are usually evaluated by converting them to two successive ordinary integrals.

Take a simple case to begin. Suppose S is the rectangular region between the y-axis and the line $x = a$ and between the x-axis and the line $y = b$, where x and y are Cartesian coordinates. Say the integrand is positive on S. Then the double integral

$$\iint_S F(x,y)\, dA$$

represents the volume of the solid shown.



We can also get this volume by the familiar technique of "slicing." We review it briefly. Cut the solid into thin slices by planes parallel to the y-z-plane, say $x = x_i$, $i = 1, 2, \ldots, n$. The volume of the slice between $x = x_{i-1}$ and $x = x_i$ is approximately

$$(x_i - x_{i-1}) \times B(x_i)$$

where $B(x_i)$ is the area of the cross-section made by the plane $x = x_i$. Hence the volume required is about

$$\sum B(x_i)(x_i - x_{i-1}). \tag{7}$$

This is a Riemann sum for the integral

$$\int_0^a B(x)\, dx. \tag{8}$$

As the slices are made thinner, the sums (7) converge both to the volume and to the integral, so the volume is given by the integral (8).

Now the cross-sectional area function itself can be obtained as an integral. The area in the plane $x = \lambda$ is given by

$$B(\lambda) = \int_0^b F(\lambda, y)\, dy.$$

Putting this together with our previous results we have

$$\iint_S F(x,y)\, dA = \int_0^a \left[\int_0^b F(x,y)\, dy \right] dx$$

where it is understood that the inner integration is to be carried out with x fixed.

Example. The double integral of the last exercise. S is the unit square.

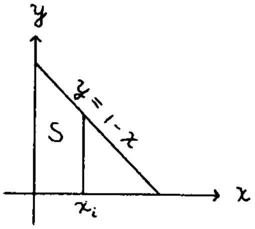
$$\begin{aligned} \iint_S xy\, dA &= \int_0^1 \left[\int_0^1 xy\, dy \right] dx \\ &= \int_0^1 \left[\frac{1}{2} xy^2 \Big|_{y=0}^{1} \right] dx \\ &= \int_0^1 \frac{1}{2} x\, dx = \frac{1}{4}. \end{aligned}$$

Although the argument we have just given for the equality of a double integral and an iterated integral depends on geometrical consideration of volume, it can be made purely analytic. Furthermore, it is valid for all continuous integrands, they need not be positive. The argument could just as well have been made by slicing the other way, that is, by planes parallel to the x-z-plane. Then our conclusion would be

$$\iint_S F(x,y)\, dA = \int_0^b \left[\int_0^a F(x,y)\, dx \right] dy$$

where it is now understood that the inner integration is carried out with y held constant.

If the region S is not a rectangle with sides parallel to the axes, the same ideas will work, but the area of a cross-section determined by a plane $x = x_1$ will be an integral whose limits may depend on x_1.



Suppose S is the triangular region shown. Let us calculate

$$\iint_S xy \ dA.$$

When we slice in the plane $x = x_1$, the section will lie above the segment shown in the figure. It will actually be a triangle with vertices at $< x_1, 0, 0 >$, $< x_1, 1-x_1, 0 >$, and $< x_1, 1-x_1, x_1(1-x_1) >$, since the intersection of the plane $x = x_1$ with the curved surface $z = xy$ happens to be straight. We can find the area methodically as an integral, however. It is

$$\int_0^{1-x_1} x_1 y \ dy = \frac{1}{2} x_1(1 - x_1)^2.$$

The solid in question extends from $x = 0$ to $x = 1$, so the overall volume is

$$\int_0^1 \frac{1}{2} x(1 - x)^2 \ dx = \frac{1}{24}$$

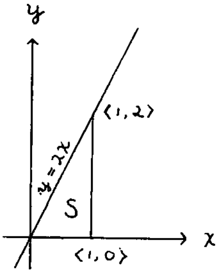
Usually one converts a double integral directly into an iterated integral without explicit consideration of the cross-sections. The only problem is to determine the proper limits for the two definite integrals. When S is the triangular region just considered

$$\iint_S F(x,y) \ dA = \int_0^1 \left[\int_0^{1-x} F(x,y) \ dy\right] dx.$$

The large brackets are more often than not omitted. Note that only the region S enters into the determination of the limits. Also note that the limits of the inner integral may involve the variable of the outer integration, but the limits of the outer integral are numbers (which might appear as letters); they do not involve the variables of integration.

In the above problem, if we decided to slice by planes parallel to the x-z-plane, then x would be the variable of the inner integration. The limits for the inner integration would be 0 and 1 - y, and the limits of the outer integration would be 0 and 1.

Suppose S is the triangular region shown here. We have a choice of two ways to convert a double integral over S into an iterated integral. If we keep x fixed at first and integrate with respect to y, then y varies from 0 to 2x. In the second integration x varies from 0 to 1. Hence

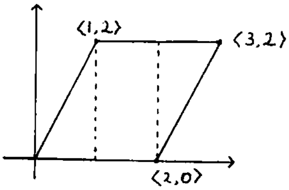


$$\iint_S F(x,y)\,dA = \int_0^1 \int_0^{2x} F(x,y)\,dy\,dx.$$

If we start the other way, then y is fixed for the inner integration and x varies from $\frac{1}{2}y$ to 1. In the second integration y varies from 0 to 2. So

$$\iint_S F(x,y)\,dA = \int_0^2 \int_{\frac{1}{2}y}^1 F(x,y)\,dx\,dy.$$

It is often necessary to cut the region into pieces in order to represent a double integral conveniently as an iterated integral. If S is the parallogram shown here, then



$$\iint_S F(x,y)\,dA = \int_0^2 \int_{\frac{1}{2}y}^{2+\frac{1}{2}y} F(x,y)\,dx\,dy$$

$$= \left(\int_0^1 \int_0^{2x} + \int_1^2 \int_0^2 + \int_2^3 \int_{2x-4}^2\right) F(x,y)\,dy\,dx.$$

This somewhat curious notation is often used to avoid repeating the integrand.

Since we have two ways (and, as we shall see presently, many more) to convert a double integral into an iterated integral, it may happen, and often does, that one way leads to easier computations than the other. Therefore, if you want to compute the value of a double integral, it will often pay to look at both ways to convert to an iterated integral. Often a problem starts as an iterated integral and can be simplified by converting it to a double integral and then back to an iterated integral the other way. The process of reversing the order of integration is valid whenever the integrand is continuous and the region for the corresponding double integral is ordinary. Careful attention must be paid to the limits of integration when reversing the order of integration. Always make a diagram of the two-dimensional region.

Example:

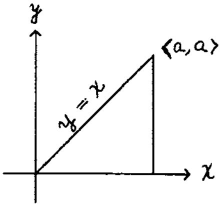
$$\int_0^a \int_y^a \sqrt{a^2 - x^2}\, dx\, dy.$$

We can perform the inner integration to get

$$\frac{a^2}{2}\int_0^a \left(\text{arc cos}\, \frac{y}{a} - \frac{y}{a^2} \sqrt{a^2 - y^2}\ \right) dy.$$

There is a good deal of work required to finish this. (Integrate the first term by parts.) However, the original integral can be converted to

$$\int_0^a \int_0^x \sqrt{a^2 - x^2}\, dy\, dx.$$

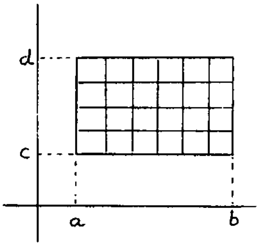


The first integration is now easy. We get

$$\int_0^a x\sqrt{a^2 - x^2}\, dx = -\frac{1}{3}(a^2 - x^2)^{3/2}\Big|_0^a = \frac{a^3}{3}.$$

We shall now sketch briefly the analytic argument behind the conversion of a double integral into an iterated integral. We restrict ourselves to the particularly easy case of a rectangular region with sides parallel to the axes. The definition of a double integral takes no position on how the region is to be cut up into smaller regions. Any way will do as long as all the little regions are small in their longest dimension. One obvious way to subdivide S is by a grid of lines parallel to the axes.

Let $x = x_1$, $x = x_2$, ..., $x = x_{n-1}$ be the lines of division parallel to the



y-axis, and let $y = y_1, y = y_2, \ldots, y = y_{m-1}$ be the lines parallel to the x-axis. Put $x_0 = a, x_n = b, y_o = c, y_m = d.$

An intermediate Riemann sum for the double integral may be found by taking the area of each little rectangle, multiplying by the value of F at the upper right corner of the rectangle, and adding. This gives

$$R = \sum_{i,j} F(x_i, y_j)(x_i - x_{i-1})(y_j - y_{j-1}).$$

If we sum this doubly-indexed set of numbers first by i then by j we get

$$R = \sum_j \left(\sum_i F(x_i, y_j)(x_i - x_{i-1})\right)(y_j - y_{j-1}).$$

Here the inner sum is a Riemann sum for calculating

$$\int_a^b F(x, y_j)\, dx.$$

Hence

$$R \sim \sum_j \left(\int_a^b F(x, y_j)\, dx\right)(y_j - y_{j-1}).$$

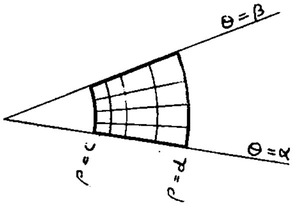
The latter is a Riemann sum for calculating

$$\int_c^d \left(\int_a^b F(x,y)\, dx\right) dy$$

Thus Riemann sums for the double integral are close to Riemann sums for the iterated integral. Hence

$$\iint_S F(x,y)\, dA = \int_c^d \int_a^b F(x,y)\, dx\, dy.$$

8.6.9 Polar coordinates. Another systematic way to subdivide a region is by a fine grid of polar coordinate lines. This will be a particularly useful way when the region S is conveniently described in polar coordinates. For the



moment suppose that S is bounded by the rays at angles α and β . and the circular arcs at distances c and d from the origin. Let the integrand be $f : S \longrightarrow \mathbb{R}$. We know that f can be expressed in terms of the polar coordinate functions ρ and θ; say $f = G(\rho, \theta)$.

We subdivide the region S with rays at $\theta = \theta_1$, $\theta = \theta_2$, $\ldots$, $\theta = \theta_{n-1}$, and with circular arcs at $\rho = \rho_1$, $\rho = \rho_2$, $\cdots$, $\rho = \rho_{m-1}$. Put $\theta_0 = \alpha$, $\theta_n = \beta$, $\rho_o = c$, $\rho_n = d$. Form the Riemann sum using the corner value for f on each piece. Then

$$R = \sum_{i,j} G(\rho_i, \theta_j)\, A_{ij} \,, \tag{10}$$

where A_{ij} is the area of the quasi-rectangular region bounded by the lines $\theta = \theta_{j-1}$ and $\theta = \theta_j$ and the circular arcs $\rho = \rho_{i-1}$ and $\rho = \rho_i$. The area of this piece is

$$\begin{aligned} A_{ij} &= \tfrac{1}{2}\,(\theta_j - \theta_{j-1})(\rho_i^{\,2} - \rho_{i-1}^{\,2}) \\ &= \tfrac{1}{2}\,(\rho_i + \rho_{i-1})(\theta_j - \theta_{j-1})(\rho_i - \rho_{i-1}). \end{aligned}$$

(Recall that the area of a pie-shaped piece of a circle is one-half the central angle (in radians) times the square of the radius.) If we put this into (10) and take the sum on j first, we get

$$R = \sum_i \tfrac{1}{2}(\rho_i + \rho_{i-1})\Bigl(\sum_j G(\rho_i, \theta_j)(\theta_j - \theta_{j-1})\Bigr)(\rho_i - \rho_{i-1})$$

Here the inner sum is a Riemann sum for

$$\int_\alpha^\beta G(\rho_i, \theta)\, d\theta \,,$$

hence

$$R \sim \sum_i \tfrac{1}{2}(\rho_i + \rho_{i-1})\left(\int_\alpha^\beta G(\rho_i, \theta)\, d\theta\right)(\rho_i - \rho_{i-1}).$$

If the subdivision is fine enough, $\frac{1}{2}(\rho_i + \rho_{i-1})$ will be very nearly ρ_i, so

$$R \sim \sum_i \left(\rho_i \int_\alpha^\beta G(\rho_i, \theta)\, d\theta\right)(\rho_i - \rho_{i-1}).$$

The latter is a Riemann sum for

$$\int_c^d \left(\int_\alpha^\beta \rho\, G(\rho, \theta)\, d\theta\right) d\rho$$

If the errors at each step are carefully accounted for, this becomes a proof that

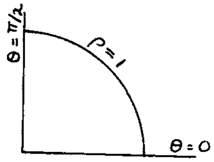
$$\iint_S f\, dA = \int_c^d \int_\alpha^\beta \rho\, G(\rho, \theta)\, d\theta\, d\rho.$$

We can do the integrations in the other order if we want and the result is

$$\int_\alpha^\beta \int_c^d G(\rho, \theta)\rho\, d\rho\, d\theta\,.$$

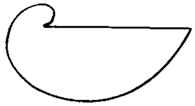
The mnemonic is "In polar coordinates, $dA = \rho\, d\rho\, d\theta$." In the non-rigorous, but sometimes helpful, formulation of calculus with infinitesimals, one says that the little quasi-rectangular regions, when made infinitely small, become true rectangles with dimensions $d\rho$ and $\rho d\theta$ and area $\rho\, d\rho\, d\theta$. Compare this with "In Cartesian coordinates, $dA = dxdy$."

Example. Find the integral of $\rho^2 \sin\theta$ over the first quadrant of the unit disk. Change it to an iterated integral in polar coordinates. The limits are 0 and 1 for ρ, and 0 and $\pi/2$ for θ. So we have



$$\int_0^1 \int_0^{\pi/2} \rho^2 \sin\theta\, \rho\, d\theta\, d\rho = \int_0^1 \rho^3\, d\rho = \frac{1}{4}.$$

It is not necessary that the boundaries of the region of integration be polar coordinate lines, we can convert to an iterated integral as long as we choose the limits of integration correctly. As before, the variable for the second (outer) integration may appear in the limits of the first integration, but neither variable of integration should appear in the limits of the outer integral.

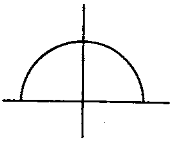


Example. Find the area swept out by the radius in generating one turn of Archimedes' spiral, $\rho = a\theta$. It is easy to see from the definition of the double integral that the area of S is just the integral of the constant function 1 over S, Since the curve is given in polar coordinates, we convert to an iterated integral in polar coordinates. Here ρ varies from 0 to $a\theta$, and then θ varies from 0 to 2π. So the required area is

$$\int_0^{2\pi}\int_0^{a\theta} \rho \; d\rho \, d\theta = \int_0^{2\pi} \frac{1}{2} a^2\theta^2 \, d\theta = \frac{4}{3}\pi^3 a^2.$$

The integrand of a double integral may be specified in terms of Cartesian coordinate functions, but nevertheless it may pay to convert it to an iterated integral using the polar coordinate grid. In making this conversion it is important to realize that the integrand must be expressed in terms of polar coordinate functions. Remember that intrinsically the integrand is supposed to be a function from S to R directly. The fact that it is presented in terms of x and y does not alter the situation.

Example.. Find the double integral of $\sqrt{x^2 + y^2}$ over the upper half of the unit disk.



We could set this up as an iterated integral using the Cartesian grid.

$$\int_{-1}^{+1}\int_0^{(1-x^2)^{1/2}} \sqrt{x^2 + y^2} \; dy \, dx.$$

The other integral looks no better. But if we notice that the integrand will be simply ρ when expressed in polar coordinates, we convert to polar coordinates and get

$$\int_0^{\pi}\int_0^1 \rho^2 \, d\rho \, d\theta = \frac{\pi}{3}$$

Switching from Cartesian to polar coordinates enables us to evaluate the one dimensional definite integral

$$\int_0^{\infty} e^{-x^2} dx$$

which is very important in probability theory. Call this integral I. (It is known that the indefinite integral of e^{-x^2} cannot be expressed as a combination of elementary functions, so none of the ordinary methods can possibly evaluate I.) Then

$$\begin{aligned} I^2 &= \int_0^{\infty} e^{-x^2} dx \int_0^{\infty} e^{-y^2} dy \\ &= \int_0^{\infty}\int_0^{\infty} e^{-x^2} dx \, e^{-y^2} dy \\ &= \int_0^{\infty}\int_0^{\infty} e^{-x^2-y^2} dx \, dy \end{aligned}$$

Now this last integral is the iterated integral corresponding to the double integral of $e^{-x^2-y^2}$ over the entire first quadrant. Since this region is unbounded, it is not an ordinary region and our theory does not apply directly. This is an <u>improper</u> double integral, and has to be considered as a limit of double integrals over large ordinary regions. The essential ideas are the same as in one dimension. (See p. 4-55.) One cannot convert improper double integrals to iterated integrals freely, but as is frequently the case, everything works out nicely when the integrand is everywhere positive as in this case. Hence we have

$$I^2 = \iint_S e^{-x^2-y^2} dA$$

where S is the infinite first quadrant.

Now we convert this double integral to an iterated integral in polar coordinates. We get

$$I^2 = \int_0^{\pi/2}\int_0^{\infty} e^{-\rho^2}\rho \quad d\rho\, d\theta.$$

The ρ which has appeared in the integrand is just what we need to be able to complete the evaluation.

$$\int_0^{\infty} e^{-\rho^2}\rho\, d\rho \quad = - \frac{1}{2} e^{-\rho^2} \Big|_0^{\infty} = \frac{1}{2}$$

Hence $I^2 = \pi/4$, $I = \frac{1}{2}\sqrt{\pi}$. A truly remarkable result.

8.6.11 The directed double integral. The Riemann integral on the line is defined in a manner strictly analogous to the definition of the double integral. It is usually replaced very soon by the directed integral on the line. This is a special case of the line integral discussed in chapter seven. The essential feature is that integration is conceived as having a direction along the line and the sign of an integral changes if the direction of integration is reversed. Thus

$$\int_a^b F(x)\ dx = - \int_b^a F(x)\ dx.$$

There is an analogous directed double integral in which one assigns an orientation to the region of integration. When the orientation is reversed, the sign of the integral changes. Except for the sign, the directed double integral agrees with the double integral we have been studying.

There is also a theory of double integrals where the region of integration may be on a curved surface. These are called surface integrals and they are most commonly taken as directed, just as the most common form of line integral on curves is directed; that is, they reverse sign when the orientation of the region of integration is reversed. We shall not study them here.

8.6.12 Surface areas. A useful application of double integrals is to the computation of surface areas.

Let Cartesian coordinates be chosen as usual in space, and imagine the x-y-plane to be horizontal. By vertical projection we mean the linear map

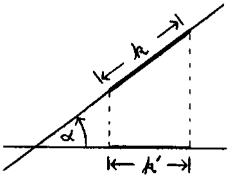
$$< r, s, t > \mapsto < r, s, 0 >.$$

This drops (or lifts) points vertically into the x-y-plane. We want to determine the effect of this transformation on surface area.

Consider first a non-vertical plane P. Non-vertical means it is the graph of some function of degree at most one. Let S be an ordinary region in P, and let S' be its vertical projection into the x-y-plane. If P is horizontal, then S' is congruent to S and therefore has the same area as S. Suppose that P is not horizontal. Then P is given by an equation

$$z = a + bx + cy$$

where b and c are not both zero. P meets the x-y-plane in a line L. If S is a rectangular region in P with two sides parallel to L, then S' is also rectangular with two sides parallel to L. Suppose the sides of S parallel to L have length h and the others length k. Then S' has dimensions h and k' where $k' = k \cos \alpha$ and α is the angle between P and the horizontal. (The figure shows a cross-section in a plane perpendicular to L.) Hence we have



$$\text{Area } S' = (\cos \alpha) \text{ Area } S.$$

From this it follows that the same relation relation holds for any ordinary region S in P, because any such region can be almost filled with tiny squares having sides parallel and perpendicular to L. The projections of these squares will almost fill S', etc. Putting the factor $\cos \alpha$ on the other side of the equation we have

$$\text{Area } S = \frac{1}{\cos \alpha} \text{ Area } S'$$

for any ordinary region S and its image S'.

The triples $< 0, 0, 1 >$ and $< -b, -c, 1 >$, regarded as column vectors, are orthogonal to the x-y-plane and the plane P, respectively. Since the angle between two planes is the angle between their normals,

$$\cos \alpha = \frac{1}{\sqrt{1 + b^2 + c^2}} .$$

So the relation between areas becomes

$$\text{(13)} \qquad \text{Area } S = \sqrt{1 + b^2 + c^2} \text{ Area } S' .$$

Now consider a C^1-function f defined on all or part of the x-y-plane. and let S' be an ordinary region in the domain of f. The graph G of f is a smooth surface in space. The set S of points of G lying over (or under) points of S' form a two-dimensional region on G which we may appropriately call an ordinary region on G, since it will be bounded by a finite number of smooth curves and corners. We want to find the area of S.

Divide S' into small ordinary regions $T_1', T_2', \ldots, T_n'$, each so small that df is practically constant on each of them. (This means that both partial derivatives of f are practically constant on each T_i'.) Above each T_i' is an ordinary region T_i on G.

At a point p of G there is a tangent plane. Its equation is

$$\text{(14)} \qquad z = a + \frac{\partial f}{\partial x}(p')\, x + \frac{\partial f}{\partial y}(p')\, y$$

where p' is the projection of p and a is a constant whose value is irrelevant at the moment. If we let p vary within one of the regions T_i, p' will vary in T_i'. Our choice of the T's then shows that all of the tangent planes (14) are virtually parallel. Thus T_i is almost on a plane. It seems plausible therefore that

$$\text{(15)} \qquad \text{Area } T_i \sim \sqrt{1 + \left(\frac{\partial f}{\partial x}(p')\right)^2 + \left(\frac{\partial f}{\partial y}(p')\right)^2} \text{ Area } T_i',$$

no matter how p' is chosen in T_i'. (This is in accordance with (13).) If we add up these inequalities for all indices i, we get

$$\text{(16)} \qquad \text{Area } S \sim \sum_i \sqrt{1 + \left(\frac{\partial f}{\partial x}(p_i')\right)^2 + \left(\frac{\partial f}{\partial y}(p_i')\right)^2} \text{ Area } T_i'.$$

Here p_i' is just any point chosen in T_i'. This sum is a Riemann sum for

$$\text{(17)} \qquad \iint_{S'} \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\, dA.$$

As the regions T_i' are made smaller, the tangent planes (14) at points within a single T_i become more nearly parallel, hence the errors in the approximations (15) become relatively less and the total error in (16) becomes arbitrarily small. We conclude that the area of S is given by the integral (17).

A few words are in order about the argument for (15). It cannot be made rigorous since we do not have a definition of surface area. How should we define it? Perhaps we should just define it in terms of the double integral (17). But we cannot make up definitions for a concept like surface area arbitrarily. If mathematics is to have any relevance to the real world, we must be sure that the technical definitions of concepts which, like surface area, purport to model reality do have a plausible relation to our perceptions. Hence, if a definition of surface area is offered and it turns out that we cannot justify the foregoing arguments with it, there would be good reason to suspect that the definition is inappropriate. Technical definitions for smooth surfaces have been worked out and it can be shown that there is really only one way to assign area to each ordinary region on a smooth surface so that various plausible requirements are satisfied. For rough surfaces, for example, surfaces that are the graphs of merely continuous functions, the situation is not yet completely understood.

As an example we shall calculate the area of a sphere, say the sphere with equation

$$x^2 + y^2 + z^2 = a^2.$$

First we restrict ourselves to the upper hemisphere. Since that is not an ordinary region on the graph of a C^1-function, we restrict further to the part lying over the disk of radius b ($< a$) centered at the origin. This is an ordinary region on the graph of the C^1-function

$$f = \sqrt{a^2 - x^2 - y^2}.$$

Then

$$\frac{\partial f}{\partial x} = -\frac{x}{f} \qquad \frac{\partial f}{\partial y} = -\frac{y}{f}$$

and we want

$$\iint \sqrt{1 + \left(\frac{x}{f}\right)^2 + \left(\frac{y}{f}\right)^2}\, dA$$

taken over the small disk. The integrand simplifies to a/f. Since

$$f = \sqrt{a^2 - \rho^2}$$

we convert to an iterated integral via the polar coordinate grid. We get

$$\int_0^{2\pi}\int_0^b \frac{a}{\sqrt{a^2 - \rho^2}}\,\rho\, d\rho\, d\theta = 2\pi a \left(-\sqrt{a^2 - \rho^2}\right)\Big|_0^b$$

$$= 2\pi a\,(a - \sqrt{a^2 - b^2})$$

Now we let $b \to a$ and we find that the area of the hemisphere is $2\pi a^2$. The area of the whole sphere is $4\pi a^2$, a familiar result.

We could have taken the integral from the beginning over the disk in the plane of radius a, but that would have been an improper integral since the integrand isn't defined at boundary points and is unbounded as we approach the boundary. We did what one always does in dealing with improper integrals, namely, integrate over a slightly smaller region and take the limit as the region gets larger. In simple cases the intermediate steps are usually elided.