A Brief Review of Some Calculus
重温微积分

何憾*

September 12, 2020

Contents

*Here is my personal website

1 The Differential Calculus of Functions of Several Variables

1.1 Linear Transforms

Recall that a mapping A of a vector space X into a vector space Y is said to be a *linear operator* if

$$A\left(x_1 + x_2\right) = Ax_1 + Ax_2, \quad A(cx_1) = cA(x_1)$$

for all $x_1, x_2 \in X$ and all scalars c. Note that one often writes Ax instead of $A(x)$ if A is linear.

Linear operators of X into X are often called *linear transformations* on X. If A is a linear transformation on X which (i) is one-to-one and (ii) maps X onto X, we say that A is *invertible*. In this case we can define an operator

A^{-1} on X by requiring that $A^{-1}(Ax)=x$ for all $x\in X$. It is trivial to verify that we then also have $A\left(A^{-1}x\right)=x$, for all $x\in X$, and that A^{-1} is linear.

THEOREM. A linear operator A on a finite-dimensional vector space X is one-to-one if and only if the range of A is all of X.

Let $L(X,Y)$ be the set of all linear transformations of the vector space X into the vector space Y. Instead of $L(X,X)$, we shall simply write $L(X)$ If $A_1,A_2\in L(X,Y)$ and if c_1,c_2 are scalars, define $c_1A_1+c_2A_2$ by

$$\left(c_1A_1+c_2A_2\right)(x)=c_1A_1x+c_2A_2x \quad (x\in X)$$

It is then clear that $c_1A_1+c_2A_2\in L(X,Y)$.

If X,Y,Z are vector spaces, and if $A\in L(X,Y)$ and $B\in L(Y,Z)$, we define their product BA to be the composition of A and B :

$$(BA)x=B(Ax) \quad (x\in X)$$

Then $BA\in L(X,Z)$. Note that BA need not be the same as AB, even if $X=Y=Z$.

We equip the Euclid space $\mathbb{R}^n$ with the *inner product* $\langle\cdot,\cdot\rangle$ defined by

$$\langle x,y\rangle\equiv x\cdot y:=\sum_{k=1}^{n}x_ky_k\,,\quad\text{for all}\quad x,y\in\mathbb{R}^n\,.$$

This inner product induce a *norm* $\|\cdot\|$ on $\mathbb{R}^n$ defined by

$$\|x\|:=\sqrt{\langle x,x\rangle}=\left(\sum_{k=1}^{n}x_k^2\right)^{\frac{1}{2}}\,,\quad\text{for all}\quad x\in\mathbb{R}^n\,.$$

For $A\in L\left(\mathbb{R}^n,\mathbb{R}^m\right)$, define the (*operator*) *norm* $\|A\|$ of A to be the supremum of all numbers $\|Ax\|$, where x ranges over all vectors in $\mathbb{R}^n$ with $\|x\|\le 1$. In other words,

$$\|A\|:=\sup_{\|x\|\le 1}\|Ax\|\,.$$

Observe that the inequality

$$\|Ax\|\le\|A\|\|x\|\,,\quad\text{for all}\quad x\in\mathbb{R}^n\,.$$

Also, if λ is such that $\|Ax\|\le\lambda\|x\|$ for all $x\in\mathbb{R}^n$ then $\|A\|\le\lambda$. Moreover

(i) If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $\|A\| < \infty$, and A is a uniformly continuous mapping of $\mathbb{R}^n$ into $\mathbb{R}^m$.

(ii) If $A, B \in L(\mathbb{R}^n, \mathbb{R}^m)$ and c is a scalar, then

$$\|A + B\| \leq \|A\| + \|B\|, \quad \|cA\| = |c|\|A\|$$

With the distance between A and B defined as $\|A - B\|$, $L(\mathbb{R}^n, \mathbb{R}^m)$ is a complete metric space.

(iii) If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $B \in L\left(\mathbb{R}^m, \mathbb{R}^k\right)$, then

$$\|BA\| \leq \|B\|\|A\|.$$

Since we now have metrics in the spaces $L(\mathbb{R}^n, \mathbb{R}^m)$, the concepts of open set, continuity, etc., make sense for these spaces. Our next theorem utilizes these concepts.

Theorem 1.1. Let $GL_n(\mathbb{R})$ be the set of all invertible linear transformations on $\mathbb{R}^n$. Then the following statements hold.

(i) If $A \in GL_n(\mathbb{R})$, $B \in L(\mathbb{R}^n)$, and

$$\|B - A\| < \frac{1}{\|A^{-1}\|},$$

then $B \in GL_n(\mathbb{R})$.

(ii) $GL_n(\mathbb{R})$ is an open subset of $L(\mathbb{R}^n)$, and the mapping $A \to A^{-1}$ is continuous on $GL_n(\mathbb{R})$. (This mapping is also obviously a $1-1$ mapping of $GL_n(\mathbb{R})$ onto $GL_n(\mathbb{R})$, which is its own inverse, so the inverse mapping is indeed a homeomorphism).

Proof. If $A = I$, it's easy to check that

$$B^{-1}\left(= \frac{I}{I - (I - B)}\right) = \sum_{m=0}^{\infty} (I - B)^m.$$

Similarly, in general case,

$$B^{-1}\left(=\frac{A^{-1}}{I-A^{-1}(A-B)}\right)=\sum_{m=0}^{\infty}A^{-(m+1)}(A-B)^m\,. \tag{1.1}$$

So (i) holds. Clearly (i) implies that $GL_n(\mathbb{R})$ is open. To show $A\mapsto A^{-1}$ is continuous on $GL_n(\mathbb{R})$, observe that for fixed $A\in GL_n(\mathbb{R})$, and $B\in GL_n(\mathbb{R})$,

$$B^{-1}-A^{-1}=B^{-1}(A-B)A^{-1}\,.$$

Hence

$$\left\|B^{-1}-A^{-1}\right\|\le\left\|B^{-1}\right\|\left\|A-B\right\|\left\|A^{-1}\right\|\,.$$

It follows from (1.1) that

$$\|B^{-1}\|\le\frac{\|A^{-1}\|}{1-\|A^{-1}(A-B)\|}\,.$$

This establishes the continuity assertion made in (ii), since

$$\left\|B^{-1}-A^{-1}\right\|\to 0\ \text{ as }\ \|B-A\|\to 0\,.$$ □

Suppose $\{x_1,\ldots,x_n\}$ and $\{y_1,\ldots,y_m\}$ are bases of finite dimensional vector spaces X and Y, respectively. Then every $A\in L(X,Y)$ determines a set of numbers a_{ij} such that

$$Ax_j=\sum_{i=1}^{m}a_{ij}y_i\quad(1\le j\le n)\,.$$

It is convenient to visualize these numbers in a rectangular array of m rows and n columns, called an m by n *matrix*:

$$[A]=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\cdots&\cdots&\cdots&\cdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}.$$

Observe that the coordinates a_{ij} of the vector Ax_j (with respect to the basis $\{y_1,\ldots,y_m\}$) appear in the j th column of $[A]$. The vectors Ax_j are therefore sometimes called the *column vectors* of $[A]$. With this terminology, the range of A is spanned by the column vectors of $[A]$.

If $x = \sum_1^n c_j x_j$, the linearity of A shows that

$$Ax = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} c_j \right) y_i \tag{1.2}$$

Thus the coordinates of Ax are $\sum_1^n a_{ij} c_j$. Suppose next that an m by n matrix is given, with real entries (a_{ij}). If A is then defined by (1.2), it is clear that $A \in L(X, Y)$ and that $[A]$ is the given matrix. Thus there is a natural one-to-one correspondence between $L(X, Y)$ and $M_{m\times n}(\mathbb{R})$, the set of all real m by n matrices. We emphasize, though, that $[A]$ depends not only on A but also on the choice of bases in X and Y. The same A may give rise to many different matrices if we change bases, and vice versa. We shall not pursue this observation any further, since we shall usually work with fixed bases.

Finally, suppose $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ are standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, and A is given by (1.2). The Schwarz inequality shows that

$$\|Ax\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} c_j \right)^2 \leq \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}^2 \cdot \sum_{j=1}^{n} c_j^2 \right) = \sum_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} a_{ij}^2 \|x\|^2 .$$

Thus

$$\|A\| \leq \left[\sum_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} a_{ij}^2 \right]^{1/2} .$$

We see that if the matrix elements a_{ij} are continuous functions of a parameter, then the same is true of A. In fact, the converse is true by the equivalence of the norms on finite dimensional vector space.

1.2 The Differential of a Function of Several Variables

1.2.1. Differentiability In order to arrive at a definition of the "derivative" of a function whose domain is $\mathbb{R}^n$ (or an open subset of $\mathbb{R}^n$), let us take another look at the familiar case $n = 1$, and let us see how to interpret the derivative in that case in a way which will naturally extend to $n > 1$.

If f is a real function with domain $(a, b) \subset \mathbb{R}^1$ and if $x \in (a, b)$, then $f'(x)$ is usually defined to be the real number

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

provided, of course, that this limit exists. Thus

$$f(x+h) - f(x) = f'(x)h + r(h) \tag{1.3}$$

where the "remainder" $r(h)$ is small, in the sense that

$$\lim_{h \to 0} \frac{r(h)}{h} = 0\,.$$

Note that (1.3) expresses the difference $f(x+h) - f(x)$ as the sum of the linear function that takes h to $f'(x)h$, plus a small remainder.

We can therefore regard the derivative of f at x, not as a real number, but as the *linear operator* on $\mathbb{R}^1$ that takes h to $f'(x)h$. Observe that every real number α gives rise to a linear operator on $\mathbb{R}^1$ the operator in question is simply multiplication by α. Conversely, every linear function that carries $\mathbb{R}^1$ to $\mathbb{R}^1$ is multiplication by some real number. It is this natural 1-1 correspondence between $\mathbb{R}^1$ and $L\left(\mathbb{R}^1\right)$ which motivates the preceding statements.

Let us next consider a function f that maps $(a, b) \subset \mathbb{R}^1$ into $\mathbb{R}^m$. In that case, $f'(x)$ was defined to be that vector $y \in \mathbb{R}^m$ (if there is one) for which

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} - y = 0\,.$$

We can again rewrite this in the form

$$f(x+h) - f(x) = hy + r(h)\,, \tag{1.4}$$

where $r(h)/h \to 0$ as $h \to 0$. The main term on the right side of (1.4) is again a linear function of h. Every $y \in \mathbb{R}^m$ induces a linear operator of $\mathbb{R}^1$ into

$\mathbb{R}^m$, by associating to each $h \in \mathbb{R}^1$ the vector $hy \in \mathbb{R}^m$. This identification of $\mathbb{R}^m$ with $L\left(\mathbb{R}^1, \mathbb{R}^m\right)$ allows us to regard $f'(x)$ as a member of $L\left(\mathbb{R}^1, \mathbb{R}^m\right)$.

Thus, if f is a differentiable mapping of $(a, b) \subset \mathbb{R}^1$ into $\mathbb{R}^m$, and if $x \in (a, b)$ then $f'(x)$ is the linear transformation of $\mathbb{R}^1$ into $\mathbb{R}^m$ that satisfies

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - f'(x)h}{h} = 0\,,$$

or, equivalently,

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - f'(x)h\|}{|h|} = 0\,.$$

We are now ready for the case $n > 1$.

Definition 1.1. A function $f : E \to \mathbb{R}^m$ defined on a set $E \subset \mathbb{R}^n$ is *differentiable* at the point $x \in E$, which is a limit point of E, if there exists a linear mapping $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{\substack{h \to 0 \\ x+h \in E}} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0\,. \tag{1.5}$$

The linear mapping A is called the *differential*, *tangent mapping*, or *derivative mapping* of the function $f : E \to \mathbb{R}^n$ at the point $x \in E$, and is usually denoted by the symbols $\mathrm{d}f(x), Df(x)$, or $f'(x)$.

A glance at (1.5) shows that f is *continuous* at any point at which f is differentiable. Moreover, there are some remarks of this definition.

Remark 1.1. There is an obvious uniqueness problem which has to be settled before we go any further. Suppose E and f are as in the definition $x \in E$, and (1.5) holds with $A = A_1$ and with $A = A_2$. Then $A_1 = A_2$. Indeed, if $B = A_1 - A_2$, the inequality

$$\|Bh\| \leq \|f(x+h) - f(x) - A_1 h\| + \|f(x+h) - f(x) - A_2 h\|\ .$$

shows that $\|Bh\|/\|h\| \to 0$ as $h \to 0$. Hence $\|B\| = 0$, which implies $B = 0$.

Remark 1.2. The relation (1.5) can be rewritten in the form

$$f(x+h) - f(x) = f'(x)h + r(h)\,, \tag{1.6}$$

where the remainder $r(h) = o(h)$ as $h \to 0$, $x + h \in E$; in other words

$$\lim_{\substack{h\to 0\\ x+h\in E}} \frac{\|r(h)\|}{\|h\|} = 0\,.$$

In the future we shall mostly be dealing with the case when E is a open set in $\mathbb{R}^n$. So if $x \in E$, then for any sufficiently small displacement h from x the point $x + h$ will also belong to E. We may interpret (1.6) by saying that for fixed x and small h, the left side of (1.6) is approximately equal to $f'(x)h$, that is, to the value of a linear transformation applied to h.

Remark 1.3. If f is differentiable at every $x \in E$, (so every point in E is a limit point of E, for example E is open), we say that f is *differentiable* in E. Then For every $x \in E$, $f'(x)$ is linear transformation of $\mathbb{R}^n$ into $\mathbb{R}^m$. But f' is also a function: f' maps E into $L(\mathbb{R}^n, \mathbb{R}^m)$.

Remark 1.4. We remark that the differential is defined on the displacements h from the point $x \in E$. To emphasize this, we attach a copy of the vector space $\mathbb{R}^n$ to the point $x \in \mathbb{R}^n$ and denote it $T_x\mathbb{R}^n$. The space $T_x\mathbb{R}^n$ can be interpreted as a set of vectors attached at the point $x \in \mathbb{R}^n$. The vector space $T_x\mathbb{R}^n$ is called the *tangent space* to $\mathbb{R}^n$ at $x \in \mathbb{R}^n$.

The origin of this terminology will be explained below. The value of the differential on a vector $h \in T_x\mathbb{R}^n$ is the vector $f'(x)h \in T_{f(x)}\mathbb{R}^m$ attached to the point $f(x)$ and approximating the increment $f(x + h) - f(x)$ of the function caused by the increment h of the argument x. Thus $\mathrm{d}f(x)$ or $f'(x)$ is a linear transformation $f'(x) : T_x\mathbb{R}^n \to T_{f(x)}\mathbb{R}^m$.

Example 1.1. We have defined derivatives of functions carrying $\mathbb{R}^n$ to $\mathbb{R}^m$ to be linear operators of $\mathbb{R}^n$ into $\mathbb{R}^m$. What is the derivative of such a linear operator? The answer is very simple.

If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ and if $x \in \mathbb{R}^n$, then

$$A'(x) = A\,. \tag{1.7}$$

Note that x appears on the left side of (1.7) but not on the right. Both sides of (1.7) are members of $L\left(\mathbb{R}^n, \mathbb{R}^m\right)$, whereas $Ax \in \mathbb{R}^m$. The proof of (1.7) is a triviality, since

$$A(x+h) - Ax = Ah$$

by the linearity of A. With $f(x) = Ax$, the numerator in (1.5) is thus 0 for every $h \in \mathbb{R}^n$. In (1.6), $r(h) = 0$.

1.2.2. Partial Derivatives We again consider a function f that maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$. Let $\{e_1, \ldots, e_n\}$ and $\{u_1, \ldots, u_m\}$ be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. We define the *partial derivative* or *partial differential* of f at x by

$$\partial_j f(x) = \lim_{\substack{t \to 0 \\ t \in \mathbb{R}}} \frac{f\left(x + te_j\right) - f(x)}{t}$$

provided the limit exists in $\mathbb{R}^m$. Equivalently, $\partial_j f(x)$ is the linear mapping in $L(\mathbb{R}, \mathbb{R}^m)$ so that

$$f(x + te_j) - f(x) = \partial_j f(x) t + o(t) \text{ as } t \to 0\,, t \in \mathbb{R}\,.$$

Writing $f\left(x_1, \ldots, x_n\right)$ in place of $f(x)$, we see that $\partial_j f$ is the derivative mapping of f with respect to x_j, keeping the other variables fixed. The notation

$$\frac{\partial f}{\partial x_j}$$

is therefore often used in place of $\partial_j f$. The components of f are the real functions $f_1, \ldots, f_m$ defined by

$$f(x) = \sum_{i=1}^{m} f_i(x) u_i \quad (x \in E) \tag{1.8}$$

Similarly partial derivative of each component $f_i : E \to \mathbb{R}$ is defined by

$$\left(\partial_j f_i\right)(x) = \lim_{\substack{t \to 0 \\ t \in \mathbb{R}}} \frac{f_i\left(x + te_j\right) - f_i(x)}{t}\,.$$

Remark 1.5. In many cases where the existence of a derivative is sufficient when dealing with functions of one variable. However, continuity or at least boundedness of the partial derivatives is needed for functions of several variables. For example, the functions f and g described in Exercise 1.1 are not continuous, although their partial derivatives exist at every point of $\mathbb{R}^2$. Even for continuous functions, the existence of all partial derivatives does not imply differentiability, see Exercise 1.2 and Theorem 1.3.

However, if f is known to be differentiable at a point x, then its partial derivatives exist at x, and they determine the linear transformation $f'(x)$ completely:

Theorem 1.2. Suppose f maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, and f is differentiable at a point $x \in E$. Then the partial derivatives $(\partial_j f_i)(x)$ exist, and

$$f'(x)e_j = \sum_{i=1}^{m} (\partial_j f_i)(x)\, u_i \quad (1 \le j \le n)\,. \tag{1.9}$$

Here, $\{e_1, \ldots, e_n\}$ and $\{u_1, \ldots, u_m\}$ are the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.

Proof. Fix j. Since f is differentiable at x,

$$f(x + te_j) - f(x) = f'(x)(te_j) + r(te_j)\,,$$

where $\|r(te_j)\|/t \to 0$ as $t \to 0$. The linearity of $f'(x)$ shows therefore that

$$\lim_{t\to 0} \frac{f(x + te_j) - f(x)}{t} = f'(x)e_j\,.$$

If we now represent f in terms of its components, then

$$\lim_{t\to 0} \sum_{i=1}^{m} \frac{f_i(x + te_j) - f_i(x)}{t} u_i = f'(x)e_j\,.$$

It follows that each quotient in this sum has a limit, as $t \to 0$, so that each $(\partial_j f_i)(x)$ exists, and then the desired result follows. □

Let $[f'(x)]$ be the matrix that represents $f'(x)$ with respect to our standard bases. Then $f'(x)e_j$ is the j th column vector of $[f'(x)]$, and (1.9) shows therefore that the number $(\partial_j f_i)(x)$ occupies the spot in the i th row and j th column of $[f'(x)]$. Thus

$$\left[f'(x)\right] = [\partial_j f_i(x)] = \begin{bmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \cdots & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{bmatrix}.$$

The matrix $[f'(x)]$ is sometimes called the *Jacobi matrix* of f at x. When it is a square matrix, we say $\det[f'(x)]$ is the *Jacobian* of f at x.

Since the matrix $[f'(x)]$ and the linear operator $f'(x)$ are essentially the same thing, so we will not distinguish between the two, both are denoted by $f'(x)$. Then, if $h = \sum_1^n h_j e_j$ is any vector in $\mathbb{R}^n$, by (1.9) we have

$$f'(x)h = \begin{bmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \cdots & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}.$$

DEFINITION 1.2. A differentiable mapping f of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$ is said to be *continuously differentiable* in E if f' is a continuous mapping of E into $L(\mathbb{R}^n, \mathbb{R}^m)$.

More explicitly, it is required that to every $x \in E$ and to every $\varepsilon > 0$ corresponds a $\delta > 0$ such that

$$\left\|f'(y) - f'(x)\right\| < \varepsilon$$

if $y \in E$ and $\|x - y\| < \delta$. If this is so, we also say that f is a C^1-mapping, or that $f \in C^1(E, \mathbb{R}^m)$.

THEOREM 1.3. Suppose f maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$. Then $f \in C^1(E, \mathbb{R}^m)$ if and only if the partial derivatives $\partial_j f_i$ exist and are continuous on E for all $1 \leq i \leq m$, $1 \leq j \leq n$.

Proof. Assume first that $f \in C^1(E, \mathbb{R}^m)$. By (1.9), for each $1 \leq i \leq m$, $1 \leq j \leq n$ and $x \in E$,

$$\partial_j f_i(x) = \langle f'(x)e_j, u_i \rangle.$$

Hence

$$\partial_j f_i(y) - \partial_j f_i(x) = \langle \left[f'(y) - f'(x)\right] e_j, u_i \rangle \,,$$

and since $\|u_i\| = \|e_j\| = 1$, it follows that

$$\begin{aligned} \|(\partial_j f_i)(y) - (\partial_j f_i)(x)\| &\leq \left\| \left[f'(y) - f'(x)\right] e_j \right\| \\ &\leq \left\| f'(y) - f'(x) \right\| \,. \end{aligned}$$

Thus $\partial_j f_i$ is continuous.

For the converse , fix $x \in E$ and $\varepsilon > 0$. Since E is open, there is an open ball $B(x, r) \subset E$, with center at x and radius r, and the continuity of the functions $\partial_j f_i$ shows that r can be chosen so that for all $\|y - x\| < r$ and for all i, j,

$$\|\partial_j f_i(y) - \partial_j f_i(x)\| < \frac{\varepsilon}{mn} \,. \tag{1.10}$$

Suppose $h = \sum_1^n h_j e_j$, $\|h\| < r$, put $v_0 = 0$, and $v_k = h_1 e_1 + \cdots + h_k e_k$ for $1 \leq k \leq n$. Then

$$f_i(x + h) - f_i(x) = \sum_{k=1}^{n} \left[f_i\left(x + v_k\right) - f_i\left(x + v_{k-1}\right)\right] \,. \tag{1.11}$$

Since $\|v_k\| \leq \|h\| < r$ for $1 \leq k \leq n$ and since $B(x, r)$ is convex, the segments with end points $x + v_{k-1}$ and $x + v_k$ lie in $B(x, r)$. Since $v_k = v_{k-1} + h_k e_k$, the mean value theorem shows that the k th summand (1.11) is equal to

$$h_k \, \partial_k f_i \left(x + v_{k-1} + \theta_k h_k e_k\right)$$

for some $\theta_k \in (0, 1)$, and this differs from $h_k \, \partial_k f_i(x)$ by less than $|h_k| \varepsilon/(mn)$ using (1.10). By (1.11) it follows that, for all h such that $\|h\| < r$,

$$\left| f_i(x + h) - f_i(x) - \sum_{j=1}^{n} h_j \left(\partial_j f_i\right)(x) \right| \leq \frac{1}{nm} \sum_{j=1}^{n} |h_j| \, \varepsilon \leq \frac{1}{m} \|h\| \varepsilon \,,$$

and hence

$$\left\| f(x + h) - f(x) - \sum_{i=1}^{m} \sum_{j=1}^{n} h_j \left(\partial_j f_i\right)(x) u_i \right\| \leq \|h\| \varepsilon \,.$$

This says that f is differentiable at x. The matrix $[f_i'(x)]$ consists of the row $(\partial_1 f_i)(x), \ldots, (\partial_n f_i)(x)$; and since $\partial_1 f_i, \ldots, \partial_n f_i$ are continuous functions on E, it's easy to see that $f \in C^1(E, \mathbb{R}^m)$. $\square$

Exercise 1.1. Define f and g on $\mathbb{R}^2$ by: $f(0,0) = g(0,0) = 0, f(x,y) = xy^2/\left(x^2+y^4\right), g(x,y) = xy^2/\left(x^2+y^6\right)$ if $(x,y) \neq (0,0)$. Prove that f is bounded on $\mathbb{R}^2$, that g is unbounded in every neighborhood of $(0,0)$, and that f is not continuous at (0,0); nevertheless, the restrictions of both f and g to every straight line in $\mathbb{R}^2$ are continuous!

Exercise 1.2. Define $f(0,0) = 0$ and

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

(i) Prove that $\partial_1 f$ and $\partial_2 f$ are bounded functions in $\mathbb{R}^2$. (Hence f is continuous.)

(ii) Let u be any unit vector in $\mathbb{R}^2$. Show that the directional derivative $(\partial_u f)(0,0)$ exists, and that its absolute value is at most 1.

(iii) Let γ be a differentiable mapping of $\mathbb{R}^1$ into $\mathbb{R}^2$, with $\gamma(0) = (0,0)$ and $\|\gamma'(0)\| > 0$. Put $g(t) = f(\gamma(t))$ and prove that g is differentiable for every $t \in \mathbb{R}^1$. If $\gamma \in C^1$, prove that $g \in C^1$.

(iv) In spite of this, prove that f is not differentiable at $(0,0)$.

1.2.3. The Basic Laws of Differentiation Next, we will discuss several basic laws of differentiation. First of all, the operation of differentiation is linear, which is easy to prove and hence we omit the proof.

Proposition 1.4 (Linearity). If the functions $f: E \to \mathbb{R}$ and $g: E \to \mathbb{R}$, defined on a set $E \subset \mathbb{R}^m$ are differentiable at the point $x_0 \in E$, then a linear combination of them $(\alpha f + \beta g): E \to \mathbb{R}^m$ is also differentiable at that point, and the following equality holds:

$$(\alpha f + \beta g)'(x_0) = \left(\alpha f' + \beta g'\right)(x_0).$$

If the functions in question are real-valued, the operations of multiplication and division (when the denominator is not zero) can also be performed. We have then the following theorem. The proof of this theorem is the same as the proof in the case that $E \subset \mathbb{R}$ is an interval, so we omit the details.

Proposition 1.5. If the functions $f : E \to \mathbb{R}$ and $g : E \to \mathbb{R}$, defined on a set $E \subset \mathbb{R}^n$ are differentiable at the point $x_0 \in E$, then the following statements hold.

(i) Their product fg is differentiable at x_0 and

$$(f \cdot g)'(x_0) = g(x_0)f'(x_0) + f(x_0)g'(x_0) .$$

(ii) Their quotient f/g is differentiable at x_0 if $g(x_0) \neq 0$, and

$$\left(\frac{f}{g}\right)'(x_0) = \frac{1}{g^2(x_0)} \left[g(x_0)f'(x_0) - f(x_0)g'(x_0)\right] .$$

The next proposition asserts that mutually inverse differentiable mappings have mutually inverse derivative mappings at corresponding points.

Proposition 1.6 (Differentiation of an Inverse Mapping). Let $f : U(x_0) \to V(y_0)$ be a mapping of a open neighborhood $U(x_0) \subset \mathbb{R}^n$ of the point x_0 onto a open neighborhood $V(y_0) \subset \mathbb{R}^n$ of the point $y_0 = f(x_0)$. Assume that f is continuous at the point x_0 and has an inverse mapping $f^{-1} : V(y_0) \to U(x_0)$ that is continuous at the point y_0. If the mapping f is differentiable at x_0 and the derivative mapping $f'(x_0) \in L(\mathbb{R}^n)$ has an inverse $f'(x_0)^{-1}$, then the mapping $f^{-1} : V(y_0) \to U(x_0)$ is differentiable at the point $y_0 = f(x_0)$, and the following equality holds:

$$\left(f^{-1}\right)'(y_0) = f'(x_0)^{-1} .$$

Proof. Take $k \in \mathbb{R}^n$ for which $y_0 + k \in V(y_0)$ and put

$$\Delta f^{-1}(y_0; k) = f^{-1}(y_0 + k) - f^{-1}(y_0) = f^{-1}(y_0 + k) - x_0 .$$

Observe that

$$\begin{aligned} k = y_0 + k - y_0 &= f(f^{-1}(y_0 + k)) - f(f^{-1}(y_0)) \\ &= f(x_0 + \Delta f^{-1}(y_0; k)) - f(x_0) \\ &= f'(x_0)\Delta f^{-1}(y_0; k) + r(\Delta f^{-1}(y_0; k))\,, \end{aligned}$$

where $r(h) \coloneqq f(x_0 + h) - f(x_0) - f'(x_0)h = o(h)$ as $h \to 0$. Then we get

$$\Delta f^{-1}(y_0; k) = f'(x_0)^{-1}k + f'(x_0)^{-1}r(\Delta f^{-1}(y_0; k))\,. \tag{1.12}$$

We have only to show that

$$r(\Delta f^{-1}(y_0; k)) = o(k) \quad \text{as} \quad k \to 0\,.$$

To this end, it suffices to show that $\Delta f^{-1}(y_0; k) = O(k)$ as $k \to 0$, that is

$$\limsup_{k \to 0} \frac{\|\Delta f^{-1}(y_0; k)\|}{\|k\|} < \infty\,.$$

By assumption, f^{-1} is continuous at y_0 so $\Delta f^{-1}(y_0; k) \to 0$ as $k \to 0$. By (1.12), since $r(h) = o(h)$, for k with sufficiently small norm, we have

$$\|f'(x_0)^{-1}r(\Delta f^{-1}(y_0; k))\| \leq \frac{1}{2}\|\Delta f^{-1}(y_0; k)\|\,,$$

and hence

$$\begin{aligned} \|\Delta f^{-1}(y_0; k)\| &\leq \|f'(x_0)^{-1}\|\|k\| + \|f'(x_0)^{-1}r(\Delta f^{-1}(y_0; k))\| \\ &\leq \|f'(x_0)^{-1}\|\|k\| + \frac{1}{2}\|\Delta f^{-1}(y_0; k)\|\,. \end{aligned}$$

So $\|\Delta f^{-1}(y_0; k)\| \leq 2\|f'(x_0)^{-1}\|\|k\|$ for k with sufficiently small norm, as desired. We now complete the proof. □

Example 1.2 (Polar coordinates in $\mathbb{R}^2$). Let $f : (0, \infty) \times (0, 2\pi) \to \mathbb{R}^2 \backslash \{0\}$ be given by $f(r, \varphi) = (r\cos\varphi, r\sin\varphi) =: (x, y)$. Let g be the inverse function to f. From

$$f'(r, \varphi) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

we deduce that $\det[f'(r,\varphi)] = r > 0$. Then

$$g'(x,y) = f'(r,\varphi)^{-1} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix}$$
$$= \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix}.$$

We now extend the *chain rule* to the present situation.

PROPOSITION 1.7 (The Chain Rule). Suppose f maps $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$ and f is differentiable at $x_0 \in E$. Suppose g maps a set $F \subset \mathbb{R}^m$ containing $f(E)$ into $\mathbb{R}^k$, and g is differentiable at $f(x_0)$. Then the mapping $g \circ f$ of E into $\mathbb{R}^k$ is differentiable at x_0, and

$$(g \circ f)'(x_0) = g'(f(x_0))\, f'(x_0)\,. \tag{1.13}$$

On the right side of (1.13) we have the product (composition) of two linear operators.

Proof. Put $y_0 = f(x_0)$, and define the error term $u(h)$, $v(k)$ by

$$f(x_0+h) - f(x_0) = f'(x_0)h + u(h)\,,$$
$$g(y_0+k) - g(y_0) = g'(y_0)k + v(k)\,,$$

for all $h \in \mathbb{R}^n$ and $k \in \mathbb{R}^m$ with $x_0 + h \in E$ and $y_0 + k \in F$, respectively. We put $\Delta f(x_0;h) = f(x_0+h) - f(x_0)$. Then given h with $x_0 + h \in E$,

$$\begin{aligned}
&(g \circ f)(x_0+h) - (g \circ f)(x_0) = g(f(x_0+h)) - g(f(x_0)) \\
&= g'(f(x_0))[f(x_0+h) - f(x_0)] + v(f(x_0+h) - f(x_0)) \\
&= g'(y_0)\left[f'(x_0)h + u(h)\right] + v(\Delta f(x_0;h)) \\
&= g'(y_0)f'(x_0)h + g'(y_0)u(h) + v(\Delta f(x_0;h))\,.
\end{aligned}$$

So it suffices to show that $g'(y_0)u(h) + v(\Delta f(x_0;h)) = o(h)$ as $h \to 0$, $x_0 + h \in E$. Firstly, since f is differentiable at x_0, $u(h) = o(h)$ so

$$\limsup_{\substack{h\to 0\\ x_0+h\in E}} \frac{\|g'(y_0)u(h)\|}{\|h\|} \le \lim_{\substack{h\to 0\\ x_0+h\in E}} \|g'(y_0)\| \frac{\|u(h)\|}{\|h\|} = 0\,.$$

To show this for another term, note that as $h \to 0$, $x_0 + h \in E$ we have $\Delta f(x_0; h) \to 0$, $y_0 + \Delta f(x_0; h) = f(x_0 + h) \subset F$, then we have

$$\lim_{\substack{h\to 0\\ x_0+h\in E}} \frac{\|v(\Delta f(x_0;h))\|}{\|h\|} = \lim_{\substack{h\to 0\\ x_0+h\in E}} \frac{\|v(\Delta f(x_0;h))\|}{\|\Delta f(x_0;h)\|} \frac{\|\Delta f(x_0;h)\|}{\|h\|}$$
$$= \|f'(x_0)\| \lim_{\substack{k\to 0\\ y_0+k\in E}} \frac{\|v(k)\|}{\|k\|} = 0\,,$$

as desired. Thus $g \circ f$ is differentiable at x_0 and (1.13) holds. □

Remark 1.6. The chian rule can be rewritten in coordinate form, as following. As we know,

$$f'(x) = \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{pmatrix},$$

and $y = f(x)$, then

$$g'(y) = \begin{pmatrix} \partial_1 g_1(y) & \cdots & \partial_m g_1(y) \\ \vdots & \ddots & \vdots \\ \partial_1 g_k(y) & \cdots & \partial_m g_k(y) \end{pmatrix}.$$

The chian rule asserts that

$$(g\circ f)'(x) = \begin{pmatrix} \partial_1 (g_1\circ f)(x) & \cdots & \partial_n (g_1\circ f)(x) \\ \vdots & \ddots & \vdots \\ \partial_1 (g_k\circ f)(x) & \cdots & \partial_n (g_k\circ f)(x) \end{pmatrix}$$
$$= \begin{pmatrix} \partial_1 g_1(y) & \cdots & \partial_m g_1(y) \\ \vdots & \ddots & \vdots \\ \partial_1 g_k(y) & \cdots & \partial_m g_k(y) \end{pmatrix} \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{pmatrix}.$$

In particular, if g is a real function, i.e., $k = 1$, then

$$\partial_j (g\circ f)(x) = \sum_{j=1}^{n} \partial_i g(f(x)) \cdot \partial_j f_i(x)\,.$$

EXAMPLE 1.3. Let γ be a differentiable mapping of the segment $(a, b) \subset \mathbb{R}^1$ into an open set $E \subset \mathbb{R}^n$, in other words, γ is a differentiable curve in E. Let f be a real-valued differentiable function with domain E. Define

$$g(t) = f(\gamma(t)) \quad (a < t < b)\,.$$

The chain rule asserts then that

$$g'(t) = f'(\gamma(t))\gamma'(t) \quad (a < t < b)\,.$$

In particular, let $u \in \mathbb{R}^n$ be a unit vector (i.e., $\|u\| = 1$), and specialize γ so that

$$\gamma(t) = x + tu \quad (-\infty < t < \infty)\,.$$

Then $\gamma'(t) = u$ for every $t \in \mathbb{R}$. Hence $g'(0) = f'(x)u$.

Some of these ideas will play a role in the following theorem.

THEOREM 1.8 (The Finite-Increment Theorem). Suppose f maps a convex open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, f is differentiable in E, and there is a real number M such that

$$\big\|f'(x)\big\| \le M \;\text{ for all }\; x \in E\,.$$

Then for all a, b in E,

$$\|f(b) - f(a)\| \le M\|b - a\|\,.$$

Proof. Fix $a \in E, b \in E$. Define

$$\gamma(t) = (1 - t)a + tb \;\text{ for }\; t \in [0, 1]\,.$$

Since E is convex, $\gamma(t) \in E$. Put

$$g(t) = f(\gamma(t)) \;\text{ for }\; t \in [0, 1]\,.$$

Then g is continuous on $[0, 1]$ and, for $t \in (0, 1)$, by the chain rule we have

$$g'(t) = f'(\gamma(t))\gamma'(t) = f'(\gamma(t))(b - a)\,,$$

and hence

$$\left\|g'(t)\right\| \le \left\|f'(\gamma(t))\right\| \|b-a\| \le M\|b-a\| \quad \text{for all } t \in (0,1)\,.$$

Then by the mean value theorem, there exist $\xi \in (0,1)$ with

$$|g(1) - g(0)| = |g'(\xi)| \le M\|b-a\|\,.$$

But $g(0) = f(a)$ and $g(1) = f(b)$. This completes the proof. □

Corollary 1.9. If f maps a region $G \subset \mathbb{R}^n$ into $\mathbb{R}^m$ and f is differentiable in G with $f'(x) = 0$ for all $x \in G$, then f is constant.

1.3 The Basic Facts of Differential Calculus of Real-Valued Functions of Several Variables

Let $\Omega \subset \mathbb{R}^n$ be a *region* (sometimes also called a *domain*), that is Ω is an open connected set in $\mathbb{R}^n$. In this section we will always suppose that $f : \Omega \mapsto \mathbb{R}$ is a real-valued differentiable function.

1.3.1. Gradient and Directional Derivatives Take any $x \in \Omega$. Then $f'(x)$ is a linear functional on (the Hilbert space) $\mathbb{R}^n$, by the preceding remark, for $h \in \mathbb{R}^n$

$$f'(x)h = \sum_{i=1}^{n} \partial_i f(x) h_i = \langle h, \nabla f(x) \rangle\,,$$

where $\nabla f(x) \in \mathbb{R}^n$ is the so-called *gradient* of f at x defined by

$$\nabla f(x) = \sum_{i=1}^{n} \partial_i f(x) e_i = \begin{bmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{bmatrix}.$$

Hence the gradient ∇f is a mapping form $\Omega \subset \mathbb{R}^n$ into $\mathbb{R}^n$.

Remark 1.7. The gradient is perpendicular to level sets. In fact, let γ be a differentiable curve $[0,1] \to \Omega$, which lies in a level set of f, that is $f(\gamma(t)) = c$ for all $t \in [0,1]$. Then we have for all $t \in [0,1]$ that

$$\left\langle \nabla f(\gamma(t)), \gamma'(t) \right\rangle = 0 \,.$$

Let $u \in \mathbb{R}^n$ be a unit vector (i.e., $\|u\| = 1$), then we define the *directional derivative* of f at x, in the direction of the unit vector u by

$$\left(\partial_u f\right)(x) \coloneqq \lim_{t \to 0} \frac{f(x+tu) - f(x)}{t} = f'(x)u = \langle u, \nabla f(x) \rangle \,. \tag{1.14}$$

If f and x are fixed, but u varies, then (1.14) shows that $\partial_u f(x)$ attains its maximum when u is a positive scalar multiple of $\nabla f(x)$. (The case $(\nabla f)(x) = 0$ should be excluded here.) If $u = \sum_1^n u_i e_i$, then (1.14) shows that $\partial_u f(x)$ can be expressed in terms of the partial derivatives of f at x by the formula

$$\partial_u f(x) = \sum_{i=1}^{n} \partial_i f(x) u_i \,.$$

In other words, the direction of $\nabla f(x)$ is the direction of steepest ascent at x, $-\nabla f(x)$ is the direction of steepest descent.

1.3.2. The Mean-Value Theorem We know the following mean-value theorem for a differentiable function f with single variable: $f(x) - f(y) = f'(\xi)(x-y)$ for some $\xi \in (x,y)$. We cannot generalize this, however, for vector-valued functions with single variable, since in general we get a different ξ for every component. The fundamental theorem of calculus does not have this disadvantage: $f(y) - f(x) = \int_x^y f'(\xi) \mathrm{d}\xi$ is also true for vector-valued functions with single variable, but requires f' to be integrable.

We are now going to prove some versions of the mean-value theorem for functions of several variables. Let Ω be a region in $\mathbb{R}^n$. For any point a, b in $\mathbb{R}^n$, denote by $[a,b]$ and (a,b) the closed and open line segment with endpoints a and b respectively, i.e.,

$$[a,b] \coloneqq \{(1-t)a + tb : t \in [0,1]\} \;;$$
$$(a,b) \coloneqq \{(1-t)a + tb : t \in (0,1)\} \;.$$

Now we can state the mean-value theorem.

THEOREM 1.10 (The Mean-Value Theorem). Let f be a real-valued function defined in a region $\Omega \subset \mathbb{R}^n$ and let the closed line segment $[x, x+h]$ be contained in Ω. If the function f is continuous at the points of the closed line segment $[x, x+h]$ and differentiable at points of the open interval $(x, x+h)$, then there exists a point $\xi \in (x, x+h)$ such that

$$f(x+h) - f(x) = f'(\xi)h = \langle h, \nabla f(\xi) \rangle .$$

Proof. Consider the auxiliary function

$$F(t) = f(x+th)$$

defined on the closed interval $0 \leq t \leq 1$. This function satisfies all the hypotheses of Lagrange mean-value theorem: it is continuous on $[0,1]$, being the composition of continuous mappings, and differentiable on the open interval $(0,1)$, being the composition of differentiable mappings. Consequently, there exists a point $\theta \in (0,1)$ such that

$$F(1) - F(0) = F'(\theta) .$$

But $F(1) = f(x+h), F(0) = f(x), F'(\theta) = f'(x+\theta h)h$, and hence the equality just written is the same as the assertion of the theorem. □

REMARK 1.8. The theorem is called the mean-value theorem because there exists a certain "average" point $\xi \in (x, x+h)$ at which Eq. (8.53) holds. We have already noted in our discussion of Lagrange's theorem (Sect. 5.3.1) that the mean-value theorem is specific to real-valued functions. A general finite-increment theorem for mappings has been given in THEOREM 1.8.

1.3.3. HIGHER-ORDER PARTIAL DERIVATIVES Let $f : \Omega \to \mathbb{R}$ be a function defined in a region $\Omega \subset \mathbb{R}^n$. Suppose f has the partial derivative $\partial_i f = \frac{\partial f}{\partial x_i}$ with respect to the variables x_i in Ω. Then this partial derivative $\partial_i f : \Omega \to \mathbb{R}$ is also a function which in turn may have a partial derivative $\partial_j (\partial_i f)$ with respect to a variable x_j at some point $x \in \Omega$. In this case,

$\partial_j(\partial_i f)(x)$ is called the *second partial derivative* of f with respect to the variables x_i and x_j at x and is denoted by one of the following symbols:

$$\partial_{ji} f(x) \quad \text{or} \quad \frac{\partial^2 f}{\partial x_j \partial x_i}(x)\,.$$

The order of the indices indicates the order in which the differentiation is carried out with respect to the corresponding variables.

We have now defined partial derivatives of second order. If a partial derivative of order k

$$\partial_{i_1 \cdots i_k} f = \frac{\partial^k f}{\partial x_{i_1} \ldots \partial x_{i_k}}$$

has been defined, we define by induction the partial derivative of order $k+1$ by the relation

$$\partial_{j i_1 \cdots i_k} f(x) \coloneqq \partial_j\left(\partial_{i_1 \cdots i_k} f\right)(x)\,.$$

At this point a question arises that is specific for functions of several variables: Does the order of differentiation affect the partial derivative computed?

Lemma 1.11. Let $f : \Omega \to \mathbb{R}$ be a real function having partial derivatives $\partial_i f$ and $\partial_j f$ in a region $\Omega \subset \mathbb{R}^n$. If the second order partial derivative $\partial_{ij} f$ exists in Ω and is continuous some point $x \in \Omega$, then $\partial_{ji} f$ exists at this point x. Moreover,

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\,.$$

Proof. Note that by definition,

$$\begin{aligned}\frac{\partial^2 f}{\partial x_j \partial x_i}(x) &= \lim_{t \to 0} \frac{\frac{\partial f}{\partial x_i}(x + te_j) - \frac{\partial f}{\partial x_i}(x)}{t} \\ &= \lim_{t \to 0} \lim_{s \to 0} \frac{f(x + te_j + se_i) - f(x + te_j) - f(x + se_i) + f(x)}{st}\,.\end{aligned}$$

To show this repeated limit exists, we show that the double limit exists and thie repeated limit equals to the double limit. Thus, we set, for s, t in $\mathbb{R}$,

$$\phi(s,t) \coloneqq f(x + te_j + se_i) - f(x + te_j) - f(x + se_i) + f(x)$$

Let $F(x) = f(x + se_i) - f(x)$, then by the mean-value theorem,

$$\begin{aligned}\phi(s,t) = F(x + te_j) - F(x) &= t\,\frac{\partial F}{\partial x_j}(x + \theta_1 te_j) \\ &= t\left[\frac{\partial f}{\partial x_j}(x + se_i + \theta te_j) - \frac{\partial f}{\partial x_j}(x + \theta_1 te_j)\right] \\ &= st\,\frac{\partial^2 f}{\partial x_i \partial x_j}(x + \theta_2 se_i + \theta_1 te_j)\,,\end{aligned}$$

where θ_1, θ_2 in $(0,1)$. Thus, by the continuity of $\partial_{ij} f$ at x,

$$\lim_{s,t\to 0} \frac{\phi(s,t)}{st} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\,.$$

Observe that for any fixed $t \neq 0$,

$$\lim_{s\to 0} \frac{\phi(s,t)}{st} = \frac{\frac{\partial f}{\partial x_i}(x + te_j) - \frac{\partial f}{\partial x_i}(x)}{t}\,.$$

Since the double limits exists, we get

$$\lim_{t\to 0} \frac{\frac{\partial f}{\partial x_i}(x + te_j) - \frac{\partial f}{\partial x_i}(x)}{t} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\,,$$

as desired. □

The following example shows that the condition in Lemma 1.11 that the second partial derivatives must be continuous is indeed necessary.

Example 1.4. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2}\,, & \text{for } (x,y) \neq (0,0)\,; \\ 0\,, & \text{for } (x,y) = (0,0)\,. \end{cases}$$

One can show that $f \in C^1\left(\mathbb{R}^2\right)$, but $\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1$ and $\frac{\partial^2 f}{\partial y \partial x} f(0,0) = -1$.

Let us agree to denote the set of functions $f : \Omega \to \mathbb{R}$ all of whose partial derivatives up to order k inclusive are defined and continuous in the domain $\Omega \subset \mathbb{R}^m$ by the symbol $C^{(k)}(\Omega;\mathbb{R})$ or $C^{(k)}(\Omega)$. Then using the preceding lemma, we obtain the following.

THEOREM 1.12. If $f \in C^{(k)}(\Omega)$, the value $\partial_{i_1 \dots i_k} f(x)$ of the partial derivative is independent of the order $i_1, \dots, i_k$ of differentiation, that is, remains the same for any permutation of the indices $i_1, \dots, i_k$.

Proof. In the case $k = 2$ this theorem is contained in LEMMA 1.11. Let us assume that the theorem holds up to order n inclusive. We shall show that then it also holds for order $n + 1$.

But $\partial_{i_1 i_2 \cdots i_{n+1}} f(x) = \partial_{i_1} \left(\partial_{i_2 \cdots i_{n+1}} f \right)(x)$. By the induction assumption the indices $i_2, \dots, i_{n+1}$ can be permuted without changing the function $\partial_{i_2 \cdots i_{n+1}} f(x)$, and hence without changing $\partial_{i_1 \cdots i_{n+1}} f(x)$. For that reason it suffices to verify that one can also permute, for example, the indices i_1 and i_2 without changing the value of the derivative $\partial_{i_1 i_2 \cdots i_{n+1}} f(x)$. Since

$$\partial_{i_1 i_2 \cdots i_{n+1}} f(x) = \partial_{i_1 i_2} \left(\partial_{i_3 \cdots i_{n+1}} f \right)(x)$$

the possibility of this permutation follows immediately from LEMMA 1.11. By the induction we complete the proof. □

EXAMPLE 1.5. If $f(x) = f(x_1, \dots, x_n)$ and $f \in C^{(k)}(\Omega)$, then, under the assumption that $[x, x+h] \subset \Omega$, for the function $\varphi(t) = f(x + th)$ defined on the closed interval $[0, 1]$ we obtain $\varphi \in C^{(k)}[0, 1]$ with

$$\varphi^{(k)}(t) = \sum_{1 \le i_1, \cdots, i_k \le n} h_{i_1} \cdots h_{i_k} \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x + th) \,. \tag{1.15}$$

We can also write formula (1.17) as

$$\varphi^{(k)}(t) = \left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n} \right)^k f(x + th) \,.$$

1.3.4. TAYLOR'S FORMULA In this subsection we generalize the Taylor's formula to the functions of several variables. To this end, we shall use the the polynomials in n-variables to approximate functions in n-variables with appropriate differentiability.

Recall that for $k \geq 1$,

$$(x_1 + \cdots + x_n)^k = \sum_{\alpha_1+\cdots+\alpha_n=k} \frac{k!}{\alpha_1! \cdots \alpha_n!} x_1^{\alpha_1} \cdots x_n^{\alpha_n} , \tag{1.16}$$

where $\alpha_1, \cdots, \alpha_n$ are nonnegative integers. To keep things simple, we introduce the following notations. We say $\alpha = (\alpha_1, \cdots, \alpha_n) \in \mathbb{N}_0^n$ is a *multi-index*, and put

$$|\alpha| := \alpha_1 + \cdots + \alpha_n \ ; \ \alpha! := \alpha_1! \cdots \alpha_n! \, .$$

For $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, we put

$$x^\alpha := x^{\alpha_1} \cdots x^{\alpha_n} ,$$

then (1.16) becomes

$$(x_1 + \cdots + x_n)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha \, . \tag{1.17}$$

Moreover, for multi-index $\alpha \in \mathbb{N}_0^n$, we define the differential operator ∂^α by

$$\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}} \, .$$

Now we can state the Taylor's formula.

Theorem 1.13. If the function $f : \Omega \to \mathbb{R}$ is defined and belongs to class $C^{(k+1)}(\Omega)$ in a region $\Omega \subset \mathbb{R}^n$. Suppose the closed interval $[x, x+h]$ is completely contained in Ω, then there exists $\theta \in (0,1)$ so that

$$f(x+h) = \sum_{0 \leq |\alpha| \leq k} \frac{1}{\alpha!} \partial^\alpha f(x) \, h^\alpha + r_k(f, x; h) \tag{1.18}$$

where

$$r_k(f, x; h) = \sum_{|\alpha|=k+1} \frac{1}{\alpha!} \partial^\alpha f(x + \theta h) \, h^\alpha$$

is called the *Lagrange form of the remainder term.*

Proof. Taylor's formula follows immediately from the corresponding Taylor formula for a function of one variable. In fact, consider the auxiliary function

$$\varphi(t) = f(x + th) \quad \text{for} \quad t \in [0,1] \, .$$

Then by EXAMPLE 1.5, φ belongs to the class $C^{(k+1)}[0,1]$. By Taylor's formula for functions of one variable, we can write that

$$\varphi(1) = \sum_{j=0}^{k} \frac{\varphi^{(j)}(0)}{j!} + \frac{1}{(k+1)!}\varphi^{(k+1)}(\theta)$$

for some $\theta \in (0,1)$. By EXAMPLE 1.5 and (1.15) we obtain that for $0 \le j \le k+1$ and $t \in [0,1]$,

$$\begin{aligned}\varphi^{(j)}(t) &= (h_1\partial_1 + \cdots + h_n\partial_n)^j\, f(x+th) \\ &= \sum_{|\alpha|=j} \frac{j!}{\alpha!}\,\partial^\alpha f(x+th)\, h^\alpha\,.\end{aligned}$$

Combine this with $\varphi(1) = f(x+h)$ and $\varphi(0) = f(x)$, we get the desired result. □

If we write the remainder term in relation (1.3.4) in the Lagrange form rather than the integral form, then the equality

$$\varphi(1) = \sum_{j=0}^{k} \frac{\varphi^{(j)}(0)}{j!} + \frac{1}{k!}\int_0^1 \varphi^{(k+1)}(t)(1-t)^k\,\mathrm{d}t\,.$$

implies Taylor's formula (1.18) with remainder term

$$r_k(f,x;h) = \sum_{|\alpha|=k+1} \frac{1}{k!}\int_0^1 \partial^\alpha f(x+th)\, h^\alpha\,(1-t)^k\,\mathrm{d}t\,.$$

This form of the remainder term, as in the case of functions of one variable, is called the *integral form of the remainder term* in Taylor's formula.

We end this subsection with the Taylor's formula with the remainder term in Peano form.

THEOREM 1.14. If the function $f:\Omega\to\mathbb{R}$ is defined and belongs to class $C^{(k)}(\Omega)$ in a region $\Omega\subset\mathbb{R}^n$. Suppose the closed interval $[x, x+h]$ is completely contained in Ω, then

$$f(x+h) = \sum_{0\le|\alpha|\le k} \frac{1}{\alpha!}\,\partial^\alpha f(x)\, h^\alpha + r_k(f,x;h)$$

where

$$r_k(f,x;h) = o(\|h\|^k) \quad \text{as} \quad h \to 0\,.$$

Proof. By THEOREM 1.13

$$\begin{aligned} f(x+h) &= \sum_{0\le|\alpha|\le k-1} \frac{1}{\alpha!}\,\partial^\alpha f(x)\,h^\alpha + \sum_{|\alpha|=k} \frac{1}{\alpha!}\,\partial^\alpha f(x+\theta h)\,h^\alpha \\ &= \sum_{0\le|\alpha|\le k} \frac{1}{\alpha!}\,\partial^\alpha f(x)\,h^\alpha + \sum_{|\alpha|=k} \frac{1}{\alpha!}\,\left[\partial^\alpha f(x+\theta h) - \partial^\alpha f(x)\right]\,h^\alpha \end{aligned}$$

Thus

$$r_k(f,x;h) = \sum_{|\alpha|=k} \frac{1}{\alpha!}\,\left[\partial^\alpha f(x+\theta h) - \partial^\alpha f(x)\right]\,h^\alpha\,.$$

Since for $|\alpha| = \alpha_1 + \cdots + \alpha_n = k$

$$|h^\alpha| = |h_1^{\alpha_1}|\cdots|h_n^{\alpha_n}| \le \|h\|^{\alpha_1+\cdots+\alpha_n} = \|h\|^k\,,$$

and $f \in C^{(k)}(\Omega)$, it's easy to see that

$$\frac{|r_k(f,x;h)|}{\|h\|^k} \to 0 \quad \text{as} \quad h \to 0\,,$$

and thus the desired result follows. □

1.3.5. EXTREMA OF FUNCTIONS OF SEVERAL VARIABLES One of the most important applications of differential calculus is its use in finding extrema of functions. Recall that a function $f : \Omega \to \mathbb{R}$ has a *local maximum* (resp. *local minimum*) at a point $a \in \Omega$ if there exists a neighborhood U of the point a such that $f(x) \le f(a)$ (resp. $f(x) \ge f(a)$) for all $x \in U$.

If the strict inequality $f(x) < f(a)$ holds for $x \in U\backslash\{a\}$ (or, respectively, $f(x) > f(a)$), the function has a *strict local maximum* (resp. *strict local minimum*) at a. The local minima and maxima of a function are called its *local extrema*.

LEMMA 1.15. $f : \Omega \to \mathbb{R}$ has partial derivatives with respect to each of the variables $x_1, \ldots, x_n$ at the point $a \in \Omega$. Then a necessary condition for the function to have a local extremum at a is $\nabla f(a) = 0$, in other words,

$$\frac{\partial f}{\partial x_1}(a) = 0, \cdots, \frac{\partial f}{\partial x_n}(a) = 0\,.$$

Proof. Consider the function $\varphi(x_1) = f(x_1, a_2, \ldots, a_m)$ of one variable defined, according to the hypotheses of the theorem, in some neighborhood of the point a_1 on the real line. At a_1 the function $\varphi(x_1)$ has a local extremum, and since

$$\varphi'(a_1) = \frac{\partial f}{\partial x_1}(a)$$

it follows that $\frac{\partial f}{\partial x_1}(a) = 0$ The other equalities are proved similarly. □

LEMMA 1.15 shows that the local extrema of $f : \Omega \to \mathbb{R}$ are found either among the points at which f is not differentiable or at the points where the differential $f'(a)$ vanishes.

If $a \in \Omega$ satisfying $\nabla f(a) = 0$, we say that a is a *critical point* (or *stationary point*) of f. A critical point may not be a extremum point. An example that confirms this is any example constructed for this purpose for functions of one variable. Thus, where previously we spoke of the function $x \mapsto x^3$, whose derivative is zero at zero, but has no extremum there, we can now consider the function

$$f(x_1, \ldots, x_n) = (x_1)^3$$

all of whose partial derivatives are zero at $a = (0, \ldots, 0)$, while the function obviously has no extremum at that point.

REMARK 1.9. Generally, the point a is a *critical* point of the mapping $f : \Omega \to \mathbb{R}^m$ if the rank of $f'(a)$ is less than $\min\{m, n\}$, that is, smaller than the maximum possible value it can have. In particular, if $m = 1$, the point a is critical if $\nabla f(a) = 0$.

After the critical points of a function have been found by solving the system $\nabla f(a) = 0$, the subsequent analysis to determine whether they are extrema or not can often be carried out using Taylor's formula and the following sufficient conditions for the presence or absence of an extremum provided by that formula.

THEOREM 1.16. Let $f : \Omega \to \mathbb{R}$ be a function of class $C^{(2)}(\Omega)$. Let $a \in \Omega$ be a critical point of f. If If, in the Taylor expansion of the function at the point a

$$f(a+h)-f(a)+\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\, h_i h_j + o\left(\|h\|^2\right) \tag{1.19}$$

the quadratic form

$$h \mapsto \frac{\partial^2 f}{\partial x_i \partial x_j}(a)\, h_i h_j \tag{1.20}$$

(i) is positive-definite or negative-definite, then f has a local extremum at a, which is a strict local minimum if (1.20) is positive-definite and a strict local maximum if it is negative-definite;

(ii) assumes both positive and negative values, then the function does not have an extremum at a.

Proof. Let $h \neq 0$ and $x_0 + h \in \Omega$. Let us represent (1.19) in the form

$$f\,(a+h)-f\,(a)=\frac{1}{2}\|h\|^2\left[\sum_{i,j=1}^{m}\frac{\partial^2 f}{\partial x^i \partial x^j}(a)\,\frac{h^i}{\|h\|}\frac{h^j}{\|h\|}+r(h)\right] \tag{1.21}$$

where $r(h) \to 0$ as $h \to 0$. It is clear that the sign of the difference $f\,(a+h)-f\,(a)$ is completely determined by the sign of the quantity in brackets. We now undertake to study this quantity.

The vector $s = \frac{h}{\|h\|}$ obviously has norm 1. The quadratic form (1.20) is continuous as a function $h \in \mathbb{R}^m$, and therefore its restriction to the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ is also continuous on S^{n-1}. But the sphere S^{n-1} is compact. Consequently, the form (1.20) has both a minimum point and a maximum point on S, at which it assumes respectively the values m and M.

If the form (1.20) is positive-definite, then $0 < m \leq M$, and there is a number $\delta > 0$ such that $|r(h)| < m$ for $\|h\| < \delta$.. Then for $\|h\| < \delta$ the bracket on the right-hand side of (1.21) is positive, and consequently $f\,(a+h)-f\,(a) > 0$ for $0 < \|h\| < \delta$. Thus, in this case the point a is a strict local minimum of the function.One can verify similarly that when the form (1.20) is negative-definite, the function has a strict local maximum at the point a.

Thus (i) is now proved. We now prove (ii).

Let s_m and s_M be points of the unit sphere at which the form (1.20) assumes the values m and M respectively, and let $m < 0 < M$.

Setting $h = ts_m$, where t is a sufficiently small positive number so that $a + ts_m \in \Omega$, then

$$f\left(a + ts_m\right) - f\left(a\right) = \frac{1}{2}\, t^2(m + r(ts_m))$$

where $r(ts_m) \to 0$ as $t \to 0$. Starting at some time (that is, for all sufficiently small values of t), the quantity $m + o(1)$ on the right-hand side of this equality will have the sign of m, that is, it will be negative. Consequently, the left-hand side will also be negative. Similarly, setting $h = ts_M$, we obtain

$$f\left(a + ts_M\right) - f\left(a\right) = \frac{1}{2!} t^2(M + r(ts_M))$$

and consequently for all sufficiently small t the difference $f\left(a + ts_M\right) - f\left(a\right)$ is positive.

Thus, if the quadratic form (1.20) assumes both positive and negative values on the unit sphere, or, what is obviously equivalent, in $\mathbb{R}^n$, then in any neighborhood of the point a there are both points where the value of the function is larger than $f\left(a\right)$ and points where the value is smaller than $f\left(a\right)$. Hence, in that case a is not a local extremum of the function. □

We now make a number of remarks in connection with this theorem.

Remark 1.10. Theorem 1.16 says nothing about the case when the form (1.20) is semi-definite, that is, non-positive or non-negative. It turns out that in this case the point may be an extremum, or it may not. This can be seen, in particular from the follow- ing example.

Example 1.6. Let us find the extrema of the function $f(x, y) = x^4 + y^4 - 2x^2$, which is defined in $\mathbb{R}^2$.

In accordance with the necessary conditions $\nabla f(x, y) = 0$ we write the system of equations

$$\begin{cases} \frac{\partial f}{\partial x}(x,y) = 4x^3 - 4x = 0 \\ \frac{\partial f}{\partial y}(x,y) = 4y^3 = 0 \end{cases}$$

from which we find three critical points: $(-1,0)$, $(0,0)$, $(1,0)$. Since

$$\frac{\partial^2 f}{\partial x^2}(x,y) = 12x^2 - 4, \quad \frac{\partial^2 f}{\partial x \partial y}(x,y) \equiv 0, \quad \frac{\partial^2 f}{\partial y^2}(x,y) = 12y^2\,,$$

at the three critical points the quadratic form (1.20) has respectively the form

$$8\,(h_1)^2\,, \quad -4\,(h_1)^2\,, \quad 8\,(h_1)^2\ .$$

That is, in all cases it is positive semi-definite or negative semi-definite. THEOREM 1.16 is not applicable, but since $f(x,y) = \left(x^2-1\right)^2 + y^4 - 1$, it is obvious that the function $f(x,y)$ has a strict minimum -1 (even a global minimum) at the points $(-1,0)$, and $(1,0)$, while there is no extremum at $(0,0)$, since for $x=0$ and $y \neq 0$ we have $f(0,y) = y^4 > 0$, and for $y=0$ and sufficiently small $x \neq 0$ we have $f(x,0) = x^4 - 2x^2 < 0$.

REMARK 1.11. It should be kept in mind that we have given necessary conditions (and sufficient conditions for an extremum of a function only at an interior point of its domain of definition. Thus in seeking the absolute maximum or minimum of a function, it is necessary to examine the boundary points of the domain of definition along with the critical interior points, since the function may assume its maximal or minimal value at one of these boundary points.

1.3.6. SOME GEOMETRIC IMAGES CONNECTED WITH FUNCTIONS OF SEVERAL VARIABLES

(a). The Graph of a Function and Curvilinear Coordinates

Let x, y, and z be Cartesian coordinates of a point in $\mathbb{R}^3$ and let $z = f(x,y)$ be a continuous function defined in some domain Ω of the plane $\mathbb{R}^2$ of the variables x and y.

By the general definition of the graph of a function, the graph of the function $f : \Omega \to \mathbb{R}$ in our case is the set

$$S = \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in \Omega, z = f(x, y)\}$$

in the space $\mathbb{R}^3$. It is obvious that the mapping $F : \Omega \to S$ defined by the relation $(x, y) \mapsto (x, y, f(x, y))$ is a continuous one-to-one mapping of Ω onto S, by which one can determine every point of S by exhibiting the point of Ω corresponding to it, or, what is the same, giving the coordinates (x, y) of this point of Ω.

Thus the pairs of numbers $(x, y) \in \Omega$ can be regarded as certain coordinates of the points of a set $S-$ the graph of the function $z = f(x, y)$. Since the points of S are given by pairs of numbers, we shall conditionally agree to call S a two-dimensional surface in $\mathbb{R}^3$.

If we define a path $\Gamma : I \to \Omega$ in Ω, then a path $F \circ \Gamma : I \to S$ automatically appears on the surface S. If $x = x(t)$ and $y = y(t)$ is a parametric definition of the path Γ, then the path $F \circ \Gamma$ on S is given by the three functions $x = x(t)$ $y = y(t), z = z(t) = f(x(t), y(t))$. In particular, if we set $x = x_0 + t, y = y_0$, we obtain a curve $x = x_0 + t, y = y_0, z = f\,(x_0 + t, y_0)$ on the surface S along which the coordinate $y = y_0$ of the points of S does not change. Similarly one can exhibit a curve $x = x_0, y = y_0 + t, z = f\,(x_0, y_0 + t)$ on S along which the first coordinate x_0 of the points of S does not change. By analogy with the planar case these curves on S are naturally called *coordinate lines* on the surface S. However, in contrast to the coordinate lines in $\Omega \subset \mathbb{R}^2$, which are pieces of straight lines, the coordinate lines on S are in general curves in $\mathbb{R}^3$. For that reason, the coordinates (x, y) of points of the surface S are often called *curvilinear coordinates* on S.

Thus the graph of a continuous function $z = f(x, y)$, defined in a domain $\Omega \subset R^2$ is a two-dimensional surface S in $\mathbb{R}^3$ whose points can be defined by curvilinear coordinates $(x, y) \in \Omega$. At this point we shall not go into detail on the general definition of a surface, since we are interested only in a special case of a surface - the graph of a function.

(b). The Tangent Plane to the Graph of a Function

Differentiability of a function $z = f(x,y)$ at the point $(x_0, y_0) \in \Omega$ means that

$$\begin{aligned} f(x,y) =& f(x_0,y_0) + A(x-x_0) + B(y-y_0) + \\ &+ o\left(\sqrt{(x-x_0)^2+(y-y_0)^2}\right) \quad \text{as } (x,y) \to (x_0,y_0) . \end{aligned} \tag{1.22}$$

where A and B are certain constants.

In $\mathbb{R}^3$ let us consider the plane

$$z = z_0 + A(x-x_0) + B(y-y_0) \tag{1.23}$$

where $z_0 = f(x_0,y_0)$. Comparing equalities (1.22) and (1.23), we see that the graph of the function is well approximated by the plane (1.23) in a neighborhood of the point (x_0, y_0, z_0). More precisely, the point $(x, y, f(x,y))$ of the graph of the function differs from the point $(x, y, z(x,y))$ of the plane (1.23) by an amount that is infinitesimal in comparison with the magnitude $\sqrt{(x-x_0)^2+(y-y_0)^2}$ of the displacement of its curvilinear coordinates (x,y) from the coordinates (x_0,y_0) of the point (x_0,y_0,z_0).

By the uniqueness of the differential of a function, the plane (1.23) possessing this property is unique and has the form

$$z = f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)(y-y_0) . \tag{1.24}$$

This plane is called the *tangent plane* to the graph of the function $z = f(x,y)$ at the point $(x_0, y_0, f(x_0,y_0))$.

Thus, the differentiability of a function $z = f(x,y)$ at the point (x_0,y_0) and the existence of a tangent plane to the graph of this function at the point $(x_0, y_0, f(x_0,y_0))$ are equivalent conditions.

(c). The Normal Vector

Writing (1.24) for the tangent plane in the canonical form

$$\frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)(y-y_0) - (z - f(x_0,y_0)) = 0$$

we conclude that the vector

$$\left(\frac{\partial f}{\partial x}\left(x_0, y_0\right), \frac{\partial f}{\partial y}\left(x_0, y_0\right), -1\right) \tag{1.25}$$

is the *normal vector* to the tangent plane. Its direction is considered to be the direction normal or orthogonal to the surface S (the graph of the function) at the point $(x_0, y_0, f(x_0, y_0))$.

In particular, if (x_0, y_0) is a critical point of the function $f(x, y)$, then the normal vector to the graph at the point $(x_0, y_0, f(x_0, y_0))$ has the form (0,0,-1) and consequently, the tangent plane to the graph of the function at such a point is horizontal (parallel to the xy-plane). The three graphs in

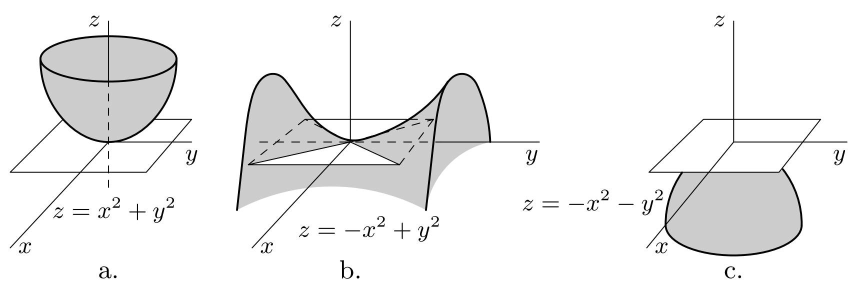

Figure 1: Tangent plane

Figture 1 illustrate what has just been said. Figture 1 a and c depict the location of the graph of a function with respect to the tangent plane in a neighborhood of a local extremum (minimum and maximum respectively), while Figture 1 b shows the graph in the neighborhood of a so-called *saddle point*.

(d). Tangent Planes and Tangent Vectors

We know that if a path $\Gamma : I \to \mathbb{R}^3$ in $\mathbb{R}^3$ is given by differentiable functions $x = x(t), y = y(t), z = z(t)$, then the vector $(x'(0), y'(0), z'(0))$ is the velocity vector at time $t = 0$. It is a direction vector of the tangent at the point $x_0 = x(0), y_0 = y(0)$ $z_0 = z(0)$ to the curve in $\mathbb{R}^3$ that is the support of the path Γ.

Now let us consider a path $\Gamma : I \to S$ on the graph of a function

$z = f(x, y)$ given in the form $x = x(t), y = y(t), z = f(x(t), y(t))$. In this particular case we find that

$$(x'(0), y'(0), z'(0)) = \left(x'(0), y'(0), \frac{\partial f}{\partial x}(x_0, y_0)\, x'(0) + \frac{\partial f}{\partial y}(x_0, y_0)\, y'(0)\right)$$

from which it can be seen that this vector is orthogonal to the vector (1.25) normal to the graph S of the function at the point $(x_0, y_0, f(x_0, y_0))$. Thus we have shown that if a vector (ξ, η, ζ) is tangent to a curve on the surface S at the point $(x_0, y_0, f(x_0, y_0))$, then it is orthogonal to the vector (1.25) and (in this sense) lies in the plane (1.24) tangent to the surface S at the point in question. More precisely we could say that the whole line $x = x_0 + \xi t, y = y_0 + \eta t, z = f(x_0, y_0) + \zeta t$ lies in the tangent plane (1.24).

Let us now show that the converse is also true, that is, if a line $x = x_0 + \xi t$ $y = y_0 + \eta t, z = f(x_0, y_0) + \zeta t$, or what is the same, the vector (ξ, η, ζ), lies in the plane (1.24), then there is a path on S for which the vector (ξ, η, ζ) is the velocity vector at the point $(x_0, y_0, f(x_0, y_0))$. The path can be taken, for example, to be

$$x = x_0 + \xi t, \quad y = y_0 + \eta t, \quad z = f(x_0 + \xi t, y_0 + \eta t)\,.$$

In fact, for this path,

$$x'(0) = \xi, \quad y'(0) = \eta, \quad z'(0) = \frac{\partial f}{\partial x}(x_0, y_0)\, \xi + \frac{\partial f}{\partial y}(x_0, y_0)\, \eta\,.$$

In view of the equality

$$\frac{\partial f}{\partial x}(x_0, y_0)\, x'(0) + \frac{\partial f}{\partial y}(x_0, y_0)\, y'(0) - z'(0) = 0$$

and the hypothesis that

$$\frac{\partial f}{\partial x}(x_0, y_0)\, \xi + \frac{\partial f}{\partial y}(x_0, y_0)\, \eta - \zeta = 0$$

We conclude that

$$(x'(0), y'(0), z'(0)) = (\xi, \eta, \zeta)$$

Hence the tangent plane to the surface S at the point (x_0, y_0, z_0) is formed by the vectors that are tangents at the point (x_0, y_0, z_0) to curves on the

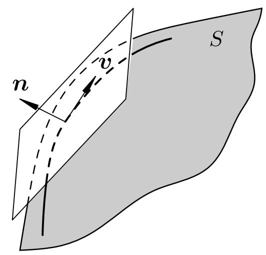

Figure 2: Vectors in the tangent plane

surface S passing through the point (see Figture 2). This is a more geometric description of the tangent plane. In any case, one can see from it that if the tangent to a curve is invariantly defined (with respect to the choice of coordinates), then the tangent plane is also invariantly defined.

(d). The General Case

We have been considering functions of two variables for the sake of visualizability, but everything that was said obviously carries over to the general case of a function

$$y = f(x_1, \ldots, x_n) \tag{1.26}$$

of n variables. At the point $(a_1, \ldots, a_n, f(a_1, \ldots, a_n))$ the plane tangent to the graph of such a function can be written in the form

$$y = f(a_1, \ldots, a_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(a_1, \ldots, a_n)(x_i - a_i) \tag{1.27}$$

the vector

$$\left(\frac{\partial f}{\partial x^1}(a), \ldots, \frac{\partial f}{\partial x^n}(a), -1\right)$$

is the normal vector to the tangent plane (1.27). This plane itself, like the graph of the function (1.26) has dimension n, that is, any point is now given by a set $(x_1, \ldots, x_n)$ of n coordinates.

Thus, (1.27) defines a hyperplane in $\mathbb{R}^{n+1}$.

Repeating verbatim the reasoning above, one can verify that the tangent plane (1.27) consists of vectors that are tangent to curves passing through the point $(a_1, \ldots, a_n, f(a_1, \ldots, a_n))$ and lying on the n-dimensional surface $S-$ the graph of the function (1.26).

1.4 The Implicit Function Theorem

In this section we shall prove the implicit function theorem, which is important both intrinsically and because of its numerous applications.

1.4.1. Introduction Let us begin by explaining the problem. Suppose, for example, we have the relation

$$x^2 + y^2 - 1 = 0 \tag{1.28}$$

between the coordinates x, y of points in the plane $\mathbb{R}^2$. The set of all points of $\mathbb{R}^2$ satisfying this condition is the unit circle.

The presence of the relation (1.28) shows that after fixing one of the coordinates, for example, x, we can no longer choose the second coordinate arbitrarily. Thus relation (1.28) determines the dependence of y on x. We are interested in the question of the conditions under which the implicit relation (1.28) can be solved as an explicit functional dependence $y = y(x)$ Solving Eq. (1.28) with respect to y, we find that

$$y = \pm\sqrt{1 - x^2} \tag{1.29}$$

that is, to each value of x such that $|x| < 1$, there are actually two admissible values of y. In forming a functional relation $y = y(x)$ satisfying relation (1.28) one cannot give preference to either of the values (1.29) without invoking additional requirements. For example, the function $y(x)$ that assumes the value $+\sqrt{1 - x^2}$ at rational points of the closed interval [-1,1] and the value $-\sqrt{1 - x^2}$ at irrational points obviously satisfies (1.28).

It is clear that one can create infinitely many functional relations satisfying (1.28) by varying this example.

The question whether the set defined in $\mathbb{R}^2$ by (1.28) is the graph of a function $y = y(x)$ obviously has a negative answer, since from the geometric point of view it is equivalent to the question whether it is possible to establish a one-to-one direct projection of a circle into a line.

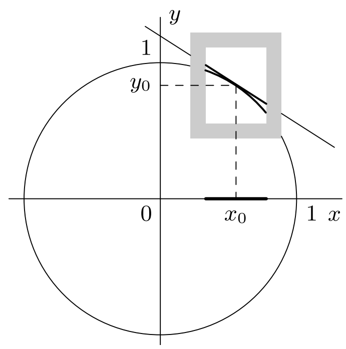

Figure 3: $x^2 + y^2 - 1 = 0$

But observation (see Figture 3) suggests that nevertheless, in a neighborhood of a particular point (x_0, y_0) the arc projects in a one-to-one manner into the x -axis, and that it can be represented uniquely as $y = y(x)$, where $x \mapsto y(x)$ is a continuous function defined in a neighborhood of the point x_0 and assuming the value y_0 at x_0. In this aspect, the only bad points are $(-1, 0)$ and $(1, 0)$, since no arc of the circle having them as interior points projects in a one-to-one manner into the x-axis. Even so, neighborhoods of these points on the circle are well situated relative to the y axis, and can be represented as the graph of a function $x = x(y)$ that is continuous in a neighborhood of the point 0 and assumes the value -1 or 1 according as the arc in question contains the point $(-1, 0)$ or $(1, 0)$.

How is it possible to find out analytically when a geometric locus of points defined by a relation of the type (1.28) can be represented in the form of an explicit function $y = y(x)$ or $x = x(y)$ in a neighborhood of a point (x_0, y_0) on the locus?

We shall discuss this question using the following, now familiar, method. We have a function $F(x, y) = x^2 + y^2 - 1$. The local behavior of this function in a neighborhood of a point (x_0, y_0) is well described by its differential

$$F'_x(x_0, y_0)(x - x_0) + F'_y(x_0, y_0)(y - y_0)$$

since

$$F(x, y) = F(x_0, y_0) + F'_x(x_0, y_0)(x - x_0) + \\ + F'_y(x_0, y_0)(y - y_0) + o(|x - x_0| + |y - y_0|)$$

as $(x, y) \to (x_0, y_0)$.

If $F(x_0, y_0) = 0$ and we are interested in the behavior of the level curve

$$F(x, y) = 0$$

of the function in a neighborhood of the point (x_0, y_0), we can judge that behavior from the position of the (tangent) line

$$F'_x(x_0, y_0)(x - x_0) + F'_y(x_0, y_0)(y - y_0) = 0\,. \tag{1.30}$$

If this line is situated so that its equation can be solved with respect to y, then, since the curve $F(x, y) = 0$ differs very little from this line in a neighborhood of the point (x_0, y_0), we may hope that it also can be written in the form $y = y(x)$ in some neighborhood of the point (x_0, y_0). The same can be said about local solvability of $F(x, y) = 0$ with respect to x.

Writing (1.30) for the specific relation (1.28) we obtain the following equation for the tangent line:

$$x_0(x - x_0) + y_0(y - y_0) = 0\,.$$

This equation can always be solved for y when $y_0 \neq 0$, that is, at all points of the circle except $(-1, 0)$ and $(1, 0)$. It is solvable with respect to x at all points of the circle except $(0, -1)$ and $(0, 1)$.

1.4.2. An Elementary Version of the Implicit Function Theorem

In this section we shall obtain the implicit function theorem by a very intuitive, but not very constructive method, one that is adapted only to the case of real-valued functions of real variables. The reader can become familiar with another method of obtaining this theorem, one that is in many ways preferable, and with a more detailed analysis of its structure in the next section.

The following proposition is an elementary version of the implicit function theorem.

Proposition 1.17. $F:\Omega\to\mathbb{R}$ is a function defined in a region $\Omega\subset\mathbb{R}^2$ and $F\in C^{(p)}(U;\mathbb{R})$, where $p\geq 1$. Suppose $(x_0,y_0)\in\Omega$ satisfies that

(i) $F\left(x_0,y_0\right)=0$,

(ii) $F'_y\left(x_0,y_0\right)\neq 0$.

Then there exist open intervals U and V with $x_0\in U$, $y_0\in V$, $U\times V\subset\Omega$ and a function $f\in C^{(p)}\left(U;V\right)$ such that

$$F(x,y)=0 \ \text{ if and only if } \ y=f(x) \text{ for } \ (x,y)\in U\times V\,. \tag{1.31}$$

Moreover, the derivative of the function f at the points $x\in U$ can be computed from the formula

$$f'(x)=-\frac{F'_x(x,f(x))}{F'_y(x,f(x))}\,. \tag{1.32}$$

Before taking up the proof, we shall give some possible reformulations of (1.31), which should bring out the meaning of the relation itself.

Remark 1.12. Proposition 1.17 says that the portion of the set defined by the relation $F(x,y)=0$ that belongs to the neighborhood $U\times V$ of the point (x_0,y_0) is the graph of a function $f:U\to V$ of class $C^{(p)}\left(U;V\right)$. In other words, one can say that inside the neighborhood $U\times V$ of the point (x_0,y_0) the equation $F(x,y)=0$ has a unique solution for y, and the function $y=f(x)$ is that solution, that is, $F(x,f(x))\equiv 0$ on U.

It follows in turn from this that if $y=\tilde{f}(x)$ is a function defined on U that is known to satisfy the relation $F(x,f(x))\equiv 0$ on $U, \tilde{f}\left(x_0\right)=y_0$, and this function is continuous at the point $x_0\in U$, then there exists a neighborhood $W\subset U$ of x_0 such that $\tilde{f}(W)\subset V$, and then $\tilde{f}(x)\equiv f(x)$ for $x\in W$. Without the assumption that $\tilde{f}$ is continuous at the point x_0 and the condition $\tilde{f}\left(x_0\right)=y_0$, this last conclusion could turn out to be incorrect, as can be seen from the example of the circle already studied.

Let us now prove Proposition 1.17.

Proof. Step 1. We shall show that there exists open intervals U and V with $x_0 \in U$, $y_0 \in V$, $U \times V \subset \Omega$ and a function $f : U \to V$ satisfying (1.31).

Suppose for definiteness that $F'_y(x_0, y_0) > 0$. Since $F \in C^{(1)}(\Omega; \mathbb{R})$, there exists a closed disk $D = D(x_0, y_0; r)$ of radius $r = 2\beta$ centering at (x_0, y_0) that $F'_y(x, y) > \frac{1}{2} F'_y(x_0, y_0) > 0$ for all $(x, y) \in D$. The function $F(x_0, y)$ is defined and strictly increasing as a function of y on the closed interval $y_0 - \beta \leq y \leq y_0 + \beta$. Consequently, $F(x_0, y_0 - \beta) < F(x_0, y_0) = 0 < F(x_0, y_0 + \beta)$. By the continuity of F, there exists a positive number $\alpha < \beta$ such that for all x with $|x - x_0| \leq \alpha$ there holds

$$F(x, y_0 - \beta) < 0 < F(x, y_0 + \beta) \ .$$

We shall now show that the open intervals

$$U = \{x \in \mathbb{R} : |x - x_0| \alpha\} \ , \ V = \{y \in \mathbb{R} : |y - y_0| < \beta\}$$

is the required one in which relation (1.31) holds. For each fixed $x \in U$ we fix the vertical closed interval with endpoints $(x, y_0 - \beta)$ and $(x, y_0 + \beta)$. Regarding $F(x, y)$ as a function of y on that closed interval, we obtain a strictly increasing continuous function that assumes values of opposite sign at the endpoints of the interval. Consequently, for each $x \in U$, there is a unique point $y_x \in V$ such that $F(x, y_x) = 0$. Setting $f : U \to V\,; x \mapsto y_x$, we arrive at relation (1.31).

Step 2. We now establish that $f \in C(U; V)$.

Take any $x \in U$, it suffices to show that

$$\Delta f(x; h) \coloneqq f(x + h) - f(x) \to 0 \quad \text{as} \quad h \to 0\,.$$

Note that we have

$$\begin{aligned}
&F(x + h, f(x + h)) - F(x, f(x)) \\
&= F(x + h, f(x + h)) - F(x + h, f(x)) + F(x + h, f(x)) - F(x, f(x)) \\
&= F'_y(x + h, f(x) + \theta_1 \Delta f(x; h)) \Delta f(x; h) + F'_x(x + \theta_2 h, f(x)) h = 0
\end{aligned}$$

where we used the Lagrange's mean-value theorem and $\theta_1, \theta_2 \in (0, 1)$. Thus

$$\Delta f(x; h) = \frac{F'_x(x + \theta_2 h, f(x))}{F'_y(x + h, f(x) + \theta_1 \Delta f(x; h))} h\,. \tag{1.33}$$

Since $(x + \theta_2 h, f(x))$ and $(x + h, f(x) + \theta_1 \Delta f(x; h))$ belongs to the closed disk D, we have

$$|\Delta f(x;h)| \leq \frac{2 \sup_{(\xi,\eta)\in D} |F'_x(\xi,\eta)|}{F'_y(x_0,y_0)} |h| \, .$$

Thus $\Delta f(x;h) \to 0$ as $h \to 0$.

Step 3. We now establish that $f \in C^{(1)}(U;V)$ and (1.32) holds.

By (1.32), we have

$$\frac{f(x+h)-f(x)}{h} = \frac{F'_x(x+\theta_2 h, f(x))}{F'_y(x+h, f(x)+\theta_1 \Delta f(x;h))} \, .$$

Since $F \in C^1(\Omega)$ and $f \in C(U;V)$, we get that

$$\begin{aligned} &\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0} \frac{F'_x(x+\theta_2 h, f(x))}{F'_y(x+h, f(x)+\theta_1 \Delta f(x;h))} \\ &= \frac{F'_x(x,f(x))}{F'_y(x,f(x))} \, . \end{aligned}$$

Thus f is differentiable in U and (1.32) holds. By the theorem on continuity of composite functions, it follows from formula (1.33) that f' is continuous, i.e., $f \in C^{(1)}(U;V)$.

Step 4. We now establish that $f \in C^{(p)}(U;V)$.

If $F \in C^{(2)}(U)$, the right-hand side of formula (1.32) can be differentiated with respect to x, and we find that

$$f''(x) = -\frac{\left[F''_{xx} + F''_{xy} \cdot f'(x)\right] F'_y - F'_x \left[F''_{xy} + F''_{yy} \cdot f'(x)\right]}{\left(F'_y\right)^2} \tag{1.34}$$

where $F'_x, F'_y, F''_{xx}, F''_{xy}$, and F''_{yy} are all computed at the point $(x, f(x))$. Thus $f \in C^{(2)}(U;V)$ if $F \in C^{(2)}(\Omega)$.

Since the order of the derivatives of f on the right-hand side of (1.32), (1.34) and so forth, is one less than the order on the left-hand side of the equality, we find by induction that $f \in C^{(p)}(U;V)$ if $F \in C^{(p)}(\Omega)$. □

Example 1.7. Let us return to relation (1.28) studied above, which defines a circle in $\mathbb{R}^2$, and verify Proposition 1.17 on this example. In this case

$$F(x, y) = x^2 + y^2 - 1$$

and it is obvious that $F \in C^{(\infty)}\left(\mathbb{R}^2\right)$. Next,

$$F_x'(x, y) = 2x, \quad F_y'(x, y) = 2y$$

so that $F_y'(x, y) \neq 0$ if $y \neq 0$. Thus, for any point (x_0, y_0) of this circle different from the points (-1,0) and (1,0) there is a neighborhood such that the arc of the circle contained in that neighborhood can be written in the form $y = f(x)$. Direct computation confirms this, and either $f(x) = \sqrt{1 - x^2}$ or $f(x) = -\sqrt{1 - x^2}$. Next, by Proposition 1.17

$$f'\left(x_0\right) = -\frac{F_x'\left(x_0, y_0\right)}{F_y'\left(x_0, y_0\right)} = -\frac{x_0}{y_0} \tag{1.35}$$

Direct computation yields

$$f'(x) = \begin{cases} -\frac{x}{\sqrt{1-x^2}}, & \text{if } f(x) = \sqrt{1 - x^2} \\ \frac{x}{\sqrt{1-x^2}}, & \text{if } f(x) = -\sqrt{1 - x^2} \end{cases}$$

which can be written as the single expression

$$f'(x) = -\frac{x}{f(x)} = -\frac{x}{y}$$

and computation with it leads to the same result,

$$f'\left(x_0\right) = -\frac{x_0}{y_0}$$

as computation from formula (1.35) obtained from Proposition 1.17.

It is important to note that formula (1.32) or (1.35) makes it possible to compute $f'(x)$ without even having an explicit expression for the relation $y = f(x)$, if only we know that $f\left(x_0\right) = y_0$. The condition $y_0 = f\left(x_0\right)$ must be prescribed, however, in order to distinguish the portion of the level curve $F(x, y) = 0$ that we intend to describe in the form $y = f(x)$.

It is clear from the example of the circle that giving only the coordinate x_0 does not determine an arc of the circle, and only after fixing y_0 have we distinguished one of the two possible arcs in this case.

The following proposition is a simple generalization of PROPOSITION 1.17 to the case of a relation $F(x_1, \ldots, x_n, y) = 0$.

PROPOSITION 1.18. Suppose $F : \Omega \to \mathbb{R}$ is a function defined in a region $\Omega \subset \mathbb{R}^{n+1}$ and $F \in C^{(p)}(\Omega)$, where $p \geq 1$. Suppose $(x_0, y_0) \in \Omega$ where $x_0 \in \mathbb{R}^n$ and $y_0 \in \mathbb{R}$, satisfies that

(i) $F(x_0, y_0) = 0$,

(ii) $F'_y(x_0, y_0) \neq 0$.

Then there exist n-dimension open rectangle U and V with $x_0 \in U$, $y_0 \in V$, $U \times V \subset \Omega$ and a function $f \in C^{(p)}(U; V)$ such that

$$F(x, y) = 0 \text{ if and only if } y = f(x) \text{ for } (x, y) \in U \times V .$$

Moreover, the partial derivatives of the function f at the points $x \in U$ can be computed from the formula

$$\frac{\partial f}{\partial x_j}(x) = -\frac{\frac{\partial F}{\partial x_j}(x, f(x))}{\frac{\partial F}{\partial y}(x, f(x))}, \text{ for } 1 \leq j \leq n . \tag{1.36}$$

Proof. The proof of the existence of U, V and the existence of the function $f : U \to V$ and its continuity in U is a verbatim repetition of the corresponding part of the proof of PROPOSITION 1.17 with only a single change, which reduces to the fact that the symbol x must now be interpreted as $(x_1, \ldots, x_n)$.

If we now fix all the variables in the functions $F(x_1, \ldots, x_n, y)$ and $f(x_1, \ldots, x_n)$ except x_j and y, we have the hypotheses of PROPOSITION 1.17 where now the role of x is played by the variable x_j. Formula (1.36) follows from this. It is clear from this formula that $\frac{\partial f}{\partial x_j} \in C(U)$ $(j = 1, \ldots, n)$, that is, $f \in C^{(1)}(U; V)$. Reasoning as in the proof of PROPOSITION 1.17 we establish by induction that $f \in C^{(p)}(U; V)$ when $F \in C^{(p)}(\Omega)$. □

EXAMPLE 1.8. Assume that the function $F : \Omega \to \mathbb{R}$ is defined in a domain $\Omega \subset \mathbb{R}^n$ and belongs to the class $C^{(1)}(\Omega)$; $a = (a_1, \ldots, a_n) \in \Omega$ and $F(a) =$

$F\left(a_1, \ldots, a_n\right)=0$. If a is not a critical point of F, then at least one of the partial derivatives of F at a is nonzero. Suppose, for example, that

$$\frac{\partial F}{\partial x_n}(a) \neq 0 .$$

Then, by PROPOSITION 1.18 in some neighborhood of a the subset of Ω defined by the equation $F\left(x_1, \ldots, x_n\right)=0$ can be defined as the graph of a function $f\left(x_1, \ldots, x_{n-1}\right)$, defined in a neighborhood of the point $\left(a_1, \cdots, a_n\right) \in \mathbb{R}^{n-1}$ that is continuously differentiable in this neighborhood and such that $f\left(a_1, \ldots, a_{n-1}\right)=a_n$. Thus, in a neighborhood of a noncritical point a of F the equation

$$F\left(x_1, \ldots, x_n\right)=0$$

defines an $(n-1)$-dimensional surface.

In particular, in the case of $\mathbb{R}^3$ the equation

$$F(x, y, z)=0$$

defines a two-dimensional surface in a neighborhood of a noncritical point $\left(x_0, y_0, z_0\right)$ satisfying the equation, which, when $F_z^{\prime}\left(x_0, y_0, z_0\right) \neq 0$ holds, can be locally written in the form

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

As we know, the equation of the plane tangent to the graph of this function at the point $\left(x_0, y_0, z_0\right)$ has the form

$$z-z_0=\frac{\partial f}{\partial x}\left(x_0, y_0\right)\left(x-x_0\right)+\frac{\partial f}{\partial y}\left(x_0, y_0\right)\left(y-y_0\right) .$$

But by formula

$$\frac{\partial f}{\partial x}\left(x_0, y_0\right)=-\frac{F_x^{\prime}\left(x_0, y_0, z_0\right)}{F_z^{\prime}\left(x_0, y_0, z_0\right)}, \quad \frac{\partial f}{\partial y}\left(x_0, y_0\right)=-\frac{F_y^{\prime}\left(x_0, y_0, z_0\right)}{F_z^{\prime}\left(x_0, y_0, z_0\right)}$$

and therefore the equation of the tangent plane can be rewritten as

$$F_x^{\prime}\left(x_0, y_0, z_0\right)\left(x-x_0\right)+F_y^{\prime}\left(x_0, y_0, z_0\right)\left(y-y_0\right)+F_z^{\prime}\left(x_0, y_0, z_0\right)\left(z-z_0\right)=0$$

which is symmetric in the variables x, y, z.

Similarly, in the general case we obtain the equation

$$\sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(a)(x_i - a_i) = 0$$

of the hyperplane in $\mathbb{R}^m$ tangent at the point $a = (a_1, \ldots, a_n)$ to the surface given by the equation $F(x_1, \ldots, x_n) = 0$ (naturally, under the assumptions that $F(a) = 0$ and that a is a noncritical point of F).

It can be seen from these equations that, given the Euclidean structure on $\mathbb{R}^n$, one can assert that the vector

$$\nabla F(a) = \left(\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial x_n}\right)(a)$$

is orthogonal to the r-level surface $F(x) = r$ of the function F at a corresponding point $a \in \mathbb{R}^n$.

For example, for the function

$$F(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}$$

defined in $\mathbb{R}^3$, the r-level is the empty set if $r < 0$, a single point if $r = 0$, and the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = r$$

if $r > 0$. If (x_0, y_0, z_0) is a point on this ellipsoid, then by what has been proved, the vector

$$\nabla F(x_0, y_0, z_0) = \left(\frac{2x_0}{a^2}, \frac{2y_0}{b^2}, \frac{2z_0}{c^2}\right)$$

is orthogonal to this ellipsoid at the point (x_0, y_0, z_0), and the tangent plane to it at this point has the equation

$$\frac{x_0(x - x_0)}{a^2} + \frac{y_0(y - y_0)}{b^2} + \frac{z_0(z - z_0)}{c^2} = 0$$

which, when we take account of the fact that the point (x_0, y_0, z_0) lies on the ellipsoid, can be rewritten as

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} + \frac{z_0 z}{c^2} = r\,.$$

1.4.3. The Implicit Function Theorem We now turn to the general case of a system of equations

$$\begin{cases} F_1(x_1,\dots,x_n,y_1,\dots y_m)=0 \\ \vdots \\ F_m(x_1,\dots,x_n,y_1,\dots,y_m)=0 \end{cases} \tag{1.37}$$

which we shall solve with respect to $y_1,\dots,y_m$, that is, find a system of functional relations

$$\begin{cases} y_1=f_1(x_1,\dots,x_n) \\ \vdots \\ y_m=f_m(x_1,\dots,x_n) \end{cases} \tag{1.38}$$

locally equivalent to the system (1.37).

For the sake of brevity, convenience in writing, and clarity of statement, let us agree that $x=(x_1,\dots,x_n)$, $y=(y_1,\dots,y_m)$. We shall write the left-hand side of the system (1.37) as $F(x,y)$, the system of equations (1.37) as $F(x,y)=0$, and the mapping (1.38) as $y=f(x)$.

As we know,

$$F'_x(x,y)=\begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{pmatrix}(x,y);$$

$$F'_y(x,y)=\begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix}(x,y).$$

We remark that the matrix $F'_y(x,y)$ is square and hence invertible if and only if its determinant is nonzero. In the case $n=1$, it reduces to a single element, and in that case the invertibility of $F'_y(x,y)$ is equivalent to the condition that that single element is nonzero. As usual, we shall denote the matrix inverse to $F'_y(x,y)$ by $\left[F'_y(x,y)\right]^{-1}$.

We now state the main result of the present section.

THEOREM 1.19 (The Implicit Function Theorem). Suppose $F : \Omega \to \mathbb{R}^m$ is a vector-valued function defined in a region $\Omega \subset \mathbb{R}^{n+m}$ and $F \in C^{(p)}(\Omega)$, where $p \geq 1$. Suppose $(x_0, y_0) \in \Omega$ where $x_0 \in \mathbb{R}^n$ and $y_0 \in \mathbb{R}^m$, satisfies that

(i) $F(x_0, y_0) = 0$,

(ii) $F'_y(x_0, y_0)$ is invertible.

Then there exist n-dimensional open rectangle U and m-dimensional open rectangle V with $x_0 \in U$, $y_0 \in V$, $U \times V \subset \Omega$ and a function $f \in C^{(p)}(U; V)$ such that

$$F(x, y) = 0 \ \text{ if and only if } \ y = f(x) \text{ for } \ (x, y) \in U \times V \, .$$

Moreover, the derivative mapping of the function f at the points $x \in U$ can be computed from the formula

$$f'(x) = -\left[F'_y(x, f(x))\right]^{-1} \left[F'_x(x, f(x))\right] \, . \tag{1.39}$$

Proof. The proof of the theorem will rely on PROPOSITION 1.18 and the elementary properties of determinants. We shall break it into stages, reasoning by induction. For $m = 1$, the theorem is the same as PROPOSITION 1.18 and is therefore true. Suppose the theorem is true for dimension $m - 1$. We shall show that it is then valid for dimension m.

Step 1. By hypothesis (ii), the determinant of $F'_y(x_0, y_0)$ is nonzero at the point $(x_0, y_0) \in \mathbb{R}^{n+m}$. Consequently at least one element of the last row of this matrix is nonzero. Up to a change in the notation, we may assume that the element $\frac{\partial F_m}{\partial y_m}(x_0, y_0)$ is nonzero. Since $F \in C^1(\Omega)$, $\frac{\partial F_m}{\partial y_m}$ is nonzero in some neighborhood of the point (x_0, y_0).

Step 2. Then applying PROPOSITION 1.18 to the relation

$$F_m(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0 \, ,$$

we find a $(n + m - 1)$-dimensional open rectangle W and an open interval V^1 with $(x_0, (y_0)_1, \ldots, (y_0)_{m-1}) \in W$ and $(y_0)_m \in V^1$ and a function $\tilde{f} \in$

$C^{(p)}\left(W;V^1\right)$ such that

$$F_m\left(x_1,\dots,x_n,y_1,\dots,y_m\right)=0 \quad \text{if and only if}$$
$$y_m=\tilde f\left(x_1,\dots,x_n,y_1,\dots,y_{m-1}\right) \qquad \text{for all } (x,y)\in W\times V^1\,.$$

Step 3. Substituting the resulting expression $y_n = \tilde f\left(x,y_1,\dots,y_{m-1}\right)$ for the variable y_n in the first $m-1$ equations of (1.37) we obtain $m-1$ relations

$$\begin{cases} \Phi_1\left(x,y_1,\dots,y_{m-1}\right) \\ \quad := F_1\Big(x,y_1,\dots,y_{m-1},\tilde f\left(x,y_1,\dots,y_{m-1}\right)\Big)=0\,; \\ \vdots \\ \Phi_{m-1}\left(x,y_1,\dots,y_{m-1}\right) \\ \quad := F_{n-1}\Big(x,y_1,\dots,y_{m-1},\tilde f\left(x,y_1,\dots,y_{m-1}\right)\Big)=0\,. \end{cases} \tag{1.40}$$

It is clear that $\Phi_i\in C^{(p)}\left(W\right)$ $(i=1,\dots,m-1)$, and

$$\Phi_i\left(x_0;(y_0)_1,\dots,(y_0)_{m-1}\right)=0 \quad (i,\dots,m-1)\,.$$

By definition of the functions Φ_i, for $1\le i,j\le m-1$,

$$\begin{aligned} &\frac{\partial\Phi_i}{\partial y_j}(x,y_1,\dots,y_{m-1})=\frac{\partial F_i}{\partial y_j}(x,y_1,\dots,y_{m-1},\tilde f\left(x,y_1,\dots,y_{m-1}\right)) \\ &+\frac{\partial F_i}{\partial y_m}(x,y_1,\dots,y_{m-1},\tilde f\left(x,y_1,\dots,y_{m-1}\right))\frac{\partial \tilde f}{\partial y_i}(x,y_1,\dots,y_{m-1})\,. \end{aligned} \tag{1.41}$$

Since

$$F_m\Big(x,y_1,\dots,y_{m-1},\tilde f\left(x,y_1,\dots,y_{n-1}\right)\Big)\equiv 0$$

we have for $1\le j\le m-1$,

$$\begin{aligned} &\frac{\partial F_m}{\partial y_j}(x,y_1,\dots,y_{m-1},\tilde f\left(x,y_1,\dots,y_{m-1}\right)) \\ &+\frac{\partial F_m}{\partial y_m}(x,y_1,\dots,y_{m-1},\tilde f\left(x,y_1,\dots,y_{m-1}\right))\frac{\partial \tilde f}{\partial y_j}(x,y_1,\dots,y_{m-1})\equiv 0\,. \end{aligned} \tag{1.42}$$

Taking account of relations (1.41) and (1.42) and the properties of determinants, we can now observe that

$$\begin{aligned}
&\det\begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix} \\
&= \det\begin{pmatrix} \frac{\partial F_1}{\partial y_1} + \frac{\partial F_1}{\partial y_n}\cdot\frac{\partial \tilde{f}}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_{m-1}} + \frac{\partial F_1}{\partial y_m}\cdot\frac{\partial \tilde{f}}{\partial y_{m-1}} & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial F_m}{\partial y_1} + \frac{\partial F_m}{\partial y_m}\cdot\frac{\partial \tilde{f}}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_{m-1}} + \frac{\partial F_m}{\partial y_m}\cdot\frac{\partial \tilde{f}}{\partial y_{m-1}} & \frac{\partial F_m}{\partial y_m} \end{pmatrix} \\
&= \det\begin{pmatrix} \frac{\partial \Phi_1}{\partial y_1} & \cdots & \frac{\partial \Phi_1}{\partial y_{m-1}} & \frac{\partial F_m}{\partial y_m} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial \Phi_{m-1}}{\partial y_1} & \cdots & \frac{\partial \Phi_{m-1}}{\partial y_{m-1}} & \frac{\partial F_{m-1}}{\partial y_m} \\ 0 & \cdots & 0 & \frac{\partial F_m}{\partial y_m} \end{pmatrix} \neq 0\,.
\end{aligned}$$

Since $\frac{\partial F_m}{\partial y_m} \neq 0$, thus

$$\det\begin{pmatrix} \frac{\partial \Phi_1}{\partial y_1} & \cdots & \frac{\partial \Phi_1}{\partial y_{m-1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Phi_{m-1}}{\partial y_1} & \cdots & \frac{\partial \Phi_{m-1}}{\partial y_{m-1}} \end{pmatrix}(x_0, (y_0)_1, \dots (y_0)_{m-1}) \neq 0\,.$$

Then by the induction hypothesis there exist a n-dimensional open rectangle U, a $(m-1)$-dimensional open rectangle V^{m-1} so that $x_0 \in U$ and $((y_0)_1, \dots, (y_0)_{m-1}) \in V$ and $U \times V^{m-1} \subset W$; and a mapping $g \in C^{(p)}\left(U; V^{m-1}\right)$ such that for $(x, y_1, \dots, y_{m-1}) \in U \times V^{m-1}$

$$\Phi(x, y_1, \dots, y_{m-1}) = 0 \;\text{ if and only if }\; y_i = g_i(x) \text{ for each } 1 \le i \le m-1\,.$$

Step 4. Let $V = V^{m-1} \times V^1$, then V is a m-dimensional open rectangle with $y_0 \in V$. Clearly $U \times V \subset W \times V^1 \subset \Omega$ and $(x_0, y_0) \in U \times V$. We define $f : U \to V$ by

$$\begin{aligned}
f_i(x) &= g_i(x) \;\text{ for }\; x \in U, 1 \le i \le m-1; \\
f_m(x) &= \tilde{f}(x, g_1(x), \cdots, g_{m-1}(x)) \;\text{ for } x \in U\,.
\end{aligned}$$

First of all, f is well-defined since $g : U \to V^{m-1}$, $(x, g_1(x), \dots, g_{m-1}(x)) \in U \times V^{m-1} \subset W$; since $\tilde{f} : W \mapsto V^1$,

$$f(x) = (f_1(x), \cdots, f_{m-1}(x), f_m(x)) \in V^{m-1} \times V^1 = V\,.$$

Moreover, $f \in C^{(p)}(U; V)$, since $g \in C^{(p)}(U; V^{m-1})$ and $\tilde{f} \in C^{(p)}(W; V^1)$. It's easy to see that for $(x, y) \in U \times V$,

$$F(x, y) = 0 \ \text{ if and only if } \ y = f(x)\,.$$

To complete the proof of the theorem it remains only to verify formula (1.39). Since

$$F(x, f(x)) \equiv 0, \qquad \text{for } x \in U \tag{1.43}$$

and $f \in C^{(p)}(U; V)$ and $F \in C^{(p)}(\Omega)$, where $p \geq 1$, it follows that $F(\cdot, f(\cdot)) \in C^{(p)}(U; \mathbb{R}^n)$ and, differentiating the identity (1.43) we obtain

$$F'_x(x, f(x)) + F'_y(x, f(x)) \cdot f'(x) = 0\,.$$

Taking account of the invertibility of the matrix $F'_y(x, y)$ in a neighborhood of the point (x_0, y_0), we find by this equality that

$$f'(x) = -\left[F'_y(x, f(x))\right]^{-1}\left[F'_x(x, f(x))\right]\,,$$

and the theorem is completely proved. □

1.5 Some Corollaries of the Implicit Function Theorem

1.5.1. The Inverse Function Theorem A mapping $f : U \to V$, where U and V are open subsets of $\mathbb{R}^n$, is called a $C^{(p)}$-*diffeomorphism*, where $p \in \mathbb{N}_0 \cup \{\infty\}$, if f is a bijection; $f \in C^{(p)}(U; V)$ and $f^{-1} \in C^{(p)}(V; U)$. As we know, a $C^{(0)}$-diffeomorphism is a *homeomorphism*.

The basic idea of the following frequently used theorem is that if the differential of a mapping is invertible at a point, then the mapping itself is invertible in some neighborhood of the point.

Theorem 1.20 (Inverse Function Theorem). Let $f : \Omega \to \mathbb{R}^n$ be a mapping of a region $\Omega \subset \mathbb{R}^n$ such that $f \in C^{(p)}(\Omega; \mathbb{R}^n)$ for $p \geq 1$. If $f'(a)$ is invertible for some $a \in \Omega$, then there exists an open neighborhood $U \subset \Omega$ of a and an open neighborhood V of $f(a)$ such that $f : U \to V$ is a $C^{(p)}$-diffeomorphism. Moreover, if $x \in U$ and $y = f(x) \in V$, then

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

Proof. We define

$$F(x, y) = f(x) - y \quad \text{ for } x \in \Omega \text{ and } y \in \mathbb{R}^n .$$

Then F is defined in the neighborhood $\Omega \times \mathbb{R}^n$ of the point $(a, f(a)) \in \mathbb{R}^{2n}$ and

$$F'_x(x, y) = f'(x), \quad F'_y(x, y) = -I$$

where I is the identity mapping on $\mathbb{R}^n$. By hypotheses of the theorem the mapping $F(x, y)$ has the property that

$$\begin{aligned}
&F \in C^{(p)}(\Omega \times \mathbb{R}^n; \mathbb{R}^n), \; p \geq 1 ; \\
&F(a, f(a)) = 0 ; \\
&F'_x(a, f(a)) = f'(a) \quad \text{is invertible} .
\end{aligned}$$

By the implicit function theorem there exist open neighborhoods W, V of the points a, $f(a)$, respectively; and a mapping $g \in C^{(p)}(V; W)$ such that for all $(x, y) \in W \times V$,

$$f(x) - y = 0 \;\text{ if and only if }\; x = g(y) ; \tag{1.44}$$

and

$$g'(y) = -\left[F'_x(g(y), y)\right]^{-1}\left[F'_y(g(y), y)\right] = \left[f'(g(y))\right]^{-1} .$$

Note that $g : V \to W$ is injective. Indeed if $g(y_1) = g(y_2)$ for $y_1, y_2 \in V$, by (1.44), we $y_1 = f(g(y_1)) = g(g(y_2)) = y_2$. So let $U = g(V)$, then $g : V \to U$ is bijective and it follows form (1.44) that

$$g^{-1} = f \;\text{ i.e., } g = f^{-1} .$$

However, our proof is not completed since we have to show that U is an open neighborhood of a.

Since $f(a) \in V$, $a = g(f(a)) \in g(V) = U$. It suffices to show that U is open. Note that

$$U = g(V) = W \cap f^{-1}(V)$$

and f is continuous, so U is open. □

The following is an immediate consequence of the inverse function theorem.

COROLLARY 1.21 (Open Mapping). Let f be a $C^{(1)}$-mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^n$. If $f'(x)$ is invertible for every $x \in E$, then f is an open mapping, i.e., $f(U)$ is open for every open $U \subset E$.

Proof. Since for any open $U \subset E$ and any $x \in U$, there exists a open neighborhood $V_x \subset U$ of x so that f is a $C^(1)$-diffeomorphism from V_x onto $f(V_x)$, and hence $f(V_x)$ is open. So $f(U) = \cup_{x \in U} f(V_x)$ is open. □

REMARK 1.13. The hypotheses made in this theorem ensure that each point $x \in E$ has a neighborhood in which f is $1-1$. This may be expressed by saying that f is locally one-to-one in E. But f need not be $1-1$ in E under these circumstances.

We shall now give several examples that illustrate the inverse function theorem. The inverse function theorem is very often used in converting from one coordinate system to another. The simplest version of such a change of coordinates was studied in analytic geometry and linear algebra and has the form

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

This linear transformation $A : \mathbb{R}^n_x \to \mathbb{R}^n_y$ has an inverse $A^{-1} : \mathbb{R}^n_y \to \mathbb{R}^n_x$ defined on the entire space $\mathbb{R}^m_y$ if and only if the matrix (a_{ij}) is invertible, that is, $\det(a_{ij}) \neq 0$ The inverse function theorem is a local version of this proposition, based on the fact that in a neighborhood of a point a smooth mapping behaves approximately like its differential at the point.

EXAMPLE 1.9. (Polar Coordinates) The mapping $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}^2$ of the half-plane $\mathbb{R}_+ \times \mathbb{R} = \{(\rho, \varphi) \in \mathbb{R}^2 \mid \rho \geq 0\}$ onto the plane $\mathbb{R}^2$ defined by the formula

$$\begin{aligned} x &= \rho \cos\varphi \\ y &= \rho \sin\varphi \end{aligned} \tag{1.45}$$

is illustrated in Figture 4.

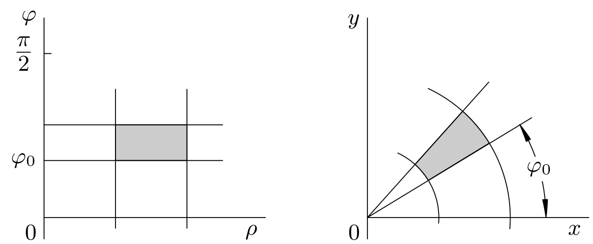

Figure 4: Polar coordinates

The Jacobian of this mapping, as can be easily computed, is ρ, that is, it is nonzero in a neighborhood of any point (ρ, φ), where $\rho > 0$. Therefore formulas (1.45) are locally invertible and hence locally the numbers ρ and φ can be taken as new coordinates of the point previously determined by the Cartesian coordinates x and y.

The coordinates (ρ, φ) are a well known system of curvilinear coordinates on the plane-polar coordinates. Their geometric interpretation is shown in Figture 4. We note that by the periodicity of the functions $\cos\varphi$ and $\sin\varphi$ the mapping (1.45) is only locally a diffeomorphism when $\rho > 0$; it is not bijective on the entire plane. That is the reason that the change from Cartesian to polar coordinates always involves a choice of a branch of the argument φ (that is, an indication of its range of variation).

EXAMPLE 1.10 (Spherical Coordinates). Polar coordinates (ρ, ψ, φ) in three-dimensional space $\mathbb{R}^3$ are called *spherical coordinates*. They are connected with Cartesian coordinates by the formulas

$$\begin{aligned} z &= \rho\cos\psi\,; \\ y &= \rho\sin\psi\sin\varphi\,; \\ x &= \rho\sin\psi\cos\varphi\,. \end{aligned} \tag{1.46}$$

The geometric meaning of the parameters ρ, ψ, and φ is shown in Figture 5.

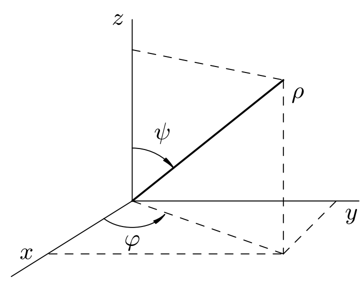

Figure 5: Spherical coordinates

The Jacobian of the mapping (1.46) is $\rho^2\sin\psi$, and so by Theorem 1 the mapping is invertible in a neighborhood of each point (ρ, ψ, φ) at which $\rho > 0$ and $\sin\psi \neq 0$.

The sets where $\rho = \text{const}$, $\varphi = \text{const}$, or $\psi = \text{const}$ in (x, y, z)-space obviously correspond to a spherical surface, a half-plane passing through the z-axis, and the surface of a cone whose axis is the z-axis respectively. Thus in passing from coordinates (x, y, z) to coordinates (ρ, ψ, φ), for example, the spherical surface and the conical surface are flattened; they correspond to pieces of the planes $\rho = \text{const}$ and $\psi = \text{const}$ respectively. We observed a similar phenomenon in the two-dimensional case, where an arc of a circle in the (x, y)-plane corresponded to a closed interval on the line in the plane with coordinates (ρ, φ) (see Figture 4). Please note that this is a local straightening.

In the n-dimensional case polar coordinates are introduced by the relations

$$\begin{aligned}
x_1 &= \rho\cos\varphi_1 \\
x_2 &= \rho\sin\varphi_1\cos\varphi_2 \\
&\vdots \\
x_{n-1} &= \rho\sin\varphi_1\sin\varphi_2\cdots\sin\varphi_{n-2}\cos\varphi_{n-1} \\
x_n &= \rho\sin\varphi_1\sin\varphi_2\cdots\sin\varphi_{n-2}\sin\varphi_{n-1}
\end{aligned}$$

The Jacobian of this transformation is

$$\rho^{n-1}\sin^{n-2}\varphi_1\sin^{n-3}\varphi_2\cdots\sin\varphi_{n-2}$$

and by the inverse function theorem it is also locally invertible everywhere where this Jacobian is nonzero.

EXAMPLE 1.11 (The General Idea of Local Rectification of Curves). New coordinates are usually introduced for the purpose of simplifying the analytic expression for the objects that occur in a problem and making them easier to visualize in the new notation. Suppose for example, a curve in the plane $\mathbb{R}^2$ is defined by the equation

$$F(x, y) = 0$$

Assume that F is a smooth function, that the point (x_0, y_0) lies on the curve, that is, $F(x_0, y_0) = 0$, and that this point is not a critical point of F. For example, suppose $F'_y(x, y) \neq 0$. Let us try to choose coordinates ξ, η so that in these coordinates a closed interval of a coordinate line, for example, the line $\eta = 0$, corresponds to an arc of this curve.

We set

$$\xi = x - x_0, \quad \eta = F(x, y)$$

The Jacobi matrix

$$\begin{pmatrix} 1 & 0 \\ F'_x & F'_y \end{pmatrix}(x, y)$$

of this transformation has as its determinant the number $F'_y(x, y)$, which by assumption is nonzero at (x_0, y_0). Then by the inverse function theorem, this mapping is a diffeomorphism of a neighborhood of (x_0, y_0) onto a neighborhood of the point $(\xi, \eta) = (0,0)$. Hence, inside this neighborhood, the numbers ξ and η can be taken as new coordinates of points lying in a neighborhood of (x_0, y_0). In the new coordinates, the curve obviously has the equation $\eta = 0$, and in this sense we have indeed achieved a local rectification of it (see Figture 6).

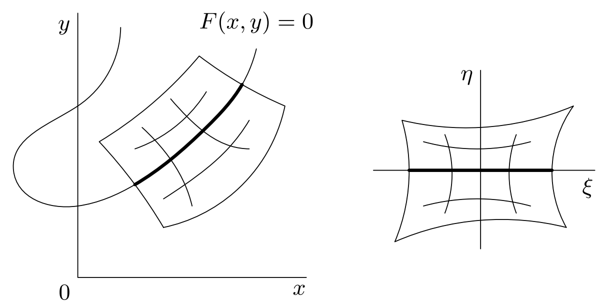

Figure 6: Local rectification of curves

1.5.2. Local Reduction of a Smooth Mapping to Canonical Form
In this subsection we shall consider only one question of this type. To be specific, we shall exhibit a canonical form to which one can locally reduce any smooth mapping of constant rank by means of a suitable choice of coordinates.

We recall that the rank of a mapping $f : \Omega \to \mathbb{R}^n$ of a domain $\Omega \subset \mathbb{R}^n$ at a point $x \in \Omega$ is the rank of the linear transformation tangent to it at the point, that is, the rank of the matrix $f'(x)$. The rank of a mapping at a point is usually denoted rank $f(x)$.

Theorem 1.22 (The Rank Theorem). Let $f : U \to \mathbb{R}^m$ be a mapping defined in an open neighborhood $U \subset \mathbb{R}^m$ of a point $x_0 \in \mathbb{R}^n$. If $f \in C^{(p)}(U; \mathbb{R}^m)$ for some $p \geq 1$, and the mapping f has the same rank k at every point $x \in U$, then there exist open neighborhoods $O(x_0)$ of x_0 and $O(y_0)$ of $y_0 = f(x_0)$ and diffeomorphisms $u = \varphi(x), v = \psi(y)$ of those neighborhoods, of class $C^{(p)}$, such that the mapping $v = \psi$ o $f \circ \varphi^{-1}(u)$ has the coordinate representation

$$(u_1, \ldots, u_k, \ldots, u_n) = u \mapsto v = (v_1, \ldots, v_m) = (u_1, \ldots, u_k, 0, \ldots, 0) \quad (1.47)$$

in the neighborhood $O(u_0) = \varphi(O(x_0))$ of $u_0 = \varphi(x_0)$.

In other words, the theorem asserts (see Figture 7) that one can choose coordinates $(u_1, \ldots, u_n)$ in place of $(x_1, \ldots, x_n)$ and $(v_1, \ldots, v_m)$ in place of $(y_1, \ldots, y_m)$ in such a way that locally the mapping has the form (1.47) in the new coordinates, that is, the canonical form for a linear transformation of rank k.

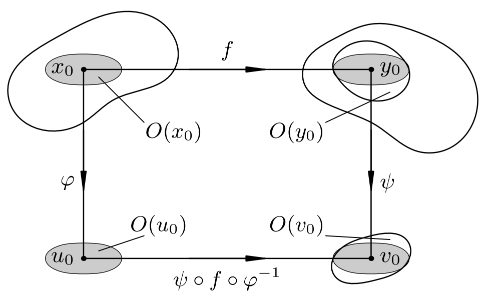

Figure 7: The rank theorem

Proof. In order to avoid relabeling the coordinates and the neighborhood U, we shall assume that at every point $x \in U$, the principal minor of order k in the upper left corner of the matrix $f'(x)$ is nonzero.

Let us consider the mapping φ defined in a neighborhood U of x_0 by the equalities

$$\begin{aligned} \varphi_1(x_1, \ldots, x_n) &= f_1(x_1, \ldots, x_n) \ , \\ &\vdots \\ \varphi_k(x_1, \ldots, x_n) &= f_k(x_1, \ldots, x_n) \ , \\ \varphi_{k+1}(x_1, \ldots, x_n) &= x_{k+1} \ , \\ &\vdots \\ \varphi_n(x_1, \ldots, x_n) &= x_n \ . \end{aligned} \quad (1.48)$$

Then

$$\varphi'(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_k} & \frac{\partial f_1}{\partial x_{k+1}} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \cdots & \frac{\partial f_k}{\partial x_k} & \frac{\partial f_k}{\partial x_{k+1}} & \cdots & \frac{\partial f_k}{\partial x_n} \\ & & & 1 & & 0 \\ & 0 & & & \ddots & \\ & & & 0 & & 1 \end{pmatrix} (x)\,,$$

and by assumption its determinant is nonzero in U. By the inverse function theorem, φ is a $C^{(p)}$-diffeomorphism of some open neighborhood $\tilde{O}\,(x_0) \subset U$ of x_0 onto an open neighborhood $\tilde{O}\,(u_0) = \varphi\left(\tilde{O}\,(x_0)\right)$ of $u_0 = \varphi\,(x_0)$.

We now consider the composite function

$$g \coloneqq f \circ \varphi^{-1} : \tilde{O}\,(u_0) \to \mathbb{R}^m\,.$$

I) Since $\varphi(\varphi^{-1}(u)) = u$ for all $u \in \tilde{O}\,(u_0)$ and $\varphi_j = f_j$ for $1 \leq j \leq k$, we see that

$$\begin{aligned} g_1\,(u_1, \ldots, u_n) &= u_1\,, \\ &\vdots \\ g_k\,(u_1, \ldots, u_n) &= u_k\,. \end{aligned} \tag{1.49}$$

II) Since the mapping $\varphi^{-1} : \tilde{O}\,(u_0) \to \tilde{O}\,(x_0)$ has maximal rank n at each point $u \in \tilde{O}\,(u_0)$, and the mapping $f : \tilde{O}\,(x_0) \to \mathbb{R}^n_y$ has rank k at every point $x \in \tilde{O}\,(x_0)$, it follows, as is known from linear algebra, that the matrix

$$g'(u) = f'\left(\varphi^{-1}(u)\right)\left(\varphi^{-1}\right)'(u)$$

has rank k at every point $u \in \tilde{O}\,(u_0)$.

Direct computation of the Jacobi matrix of the mapping (1.49) yields

$$\begin{pmatrix} 1 & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & 1 & & & \\ \frac{\partial g_{k+1}}{\partial u_1} & \cdots & \frac{\partial g_{k+1}}{\partial u_k} & \frac{\partial g_{k+1}}{\partial u_{k+1}} & \cdots & \frac{\partial g_{k+1}}{\partial u_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial u_1} & \cdots & \frac{\partial g_m}{\partial u_k} & \frac{\partial g_m}{\partial u_{k+1}} & \cdots & \frac{\partial g_m}{\partial u_n} \end{pmatrix},$$

hence at each point $u \in \tilde{O}\,(u_0)$ we obtain

$$\frac{\partial g_j}{\partial u_i}(u) = 0 \quad \text{for } k+1 \leq i \leq n\,, k+1 \leq j \leq m\,.$$

Thus, assuming that $\tilde{O}\,(u_0)$ is convex (which can be achieved by shrinking $\tilde{O}\,(u_0)$ to a ball with center at u_0, for example), we can conclude from this that the functions $g_{k+1}, \ldots, g_m$ really are independent of the variables $u_{k+1}, \ldots, u_n$. So for $k+1 \leq j \leq m$ and $u \in \tilde{O}\,(u_0)$, we can write $g_j(u_1, \cdots, u_k)$ instead of $g_j(u_1, \cdots, u_n)$.

At this point we can exhibit the mapping ψ. We set

$$\begin{aligned} \psi_1(y_1, \cdots, y_m) &= y_1 \\ &\vdots \\ \psi_k(y_1, \cdots, y_m) &= y_k \\ \psi_{k+1}(y_1, \cdots, y_m) &= y_{k+1} - g_{k+1}\,(y_1, \ldots, y_k) \\ &\vdots \\ \psi_m(y_1, \cdots, y_m) &= y_m - g_m\,(y_1, \ldots, y_k) \end{aligned}$$

Since $y_0 = f(x_0)$, $u_0 = \varphi(x_0)$ and $\varphi_j = f_j$ for $1 \leq j \leq k$, $g_{k+1}, \cdots, g_m$ are well defined in an open neighborhood of y_0. Thus the mapping ψ is defined in an open neighborhood of y_0 and belongs to class $C^{(p)}$ in that neighborhood. Moreover,

$$\psi'(y) = \begin{pmatrix} 1 & & 0 & & \\ & \ddots & & & 0 \\ 0 & & 1 & & \\ -\frac{\partial g_{k+1}}{\partial y_1} & \cdots & -\frac{\partial g_{k+1}}{\partial y_k} & 1 & & 0 \\ \vdots & \ddots & \vdots & & \ddots & \\ -\frac{\partial g_m}{\partial y_1} & \cdots & -\frac{\partial g_m}{\partial y_k} & 0 & & 1 \end{pmatrix} .$$

Its determinant equals 1 and so by the inverse function theorem the mapping ψ is a $C^{(p)}$-diffeomorphism of some neighborhood $\tilde{O}\,(y_0)$ of $y_0 \in \mathbb{R}^n_y$ onto a neighborhood $\tilde{O}\,(v_0) = \psi\left(\tilde{O}\,(y_0)\right)$ of $v_0 \in \mathbb{R}^n_v$.

Note that $y_0 = f(x_0) = g(u_0)$, in an open neighborhood $O\,(u_0) \subset \tilde{O}\,(u_0)$ of u_0 so small that $g\,(O\,(u_0)) \subset \tilde{O}\,(y_0)$, the mapping

$$\psi \circ f \circ \varphi^{-1} : O\,(u_0) \to \tilde{O}\,(v_0)$$

is a mapping of smoothness p from this neighborhood onto some open neighborhood $O\,(v_0) \subset \tilde{O}\,(v_0)$ of $v_0 \in \mathbb{R}^n_v$ and that it has the canonical form:

$$\begin{aligned} v_1 &:= (\psi_1 \circ f \circ \varphi^{-1})(u) = \psi_1(g(u)) = u_1\,, \\ &\vdots \\ v_k &:= (\psi_k \circ f \circ \varphi^{-1})(u) = \psi_k(g(u)) = u_k\,, \\ v_{k+1} &:= (\psi_{k+1} \circ f \circ \varphi^{-1})(u) = \psi_{k+1}(g(u)) = 0\,, \\ &\vdots \\ v_m &:= (\psi_m \circ f \circ \varphi^{-1})(u) = \psi_m(g(u)) = 0\,. \end{aligned}$$

Setting $\varphi^{-1}\,(O\,(u_0)) := O\,(x_0)$ and $\psi^{-1}\,(O\,(v_0)) := O\,(y_0)$, we obtain the neighborhoods of x_0 and y_0 whose existence is asserted in the theorem. □

1.6 The Inverse Function Theorem

The inverse function theorem states, roughly speaking, that a continuously differentiable mapping f is invertible in a neighborhood of any point x at which the linear transformation $f'(x)$ is invertible.

We will introduce a fixed point theorem that is valid in arbitrary complete metric spaces. It will be used in the proof of the inverse function theorem. Let X be a metric space, with metric d. If φ maps X into X and if there is a number $\lambda < 1$ such that

$$d(\varphi(x), \varphi(y)) \leq \lambda d(x, y)\,,$$

for all $x, y \in X$, then φ is said to be a *contraction* of X into X.

LEMMA 1.23 (Contraction Principle). If X is a complete metric space, and if φ is a contraction of X into X, then there exists one and only one $x \in X$ such that $\varphi(x) = x$. In other words, φ has a unique fixed point.

Now we are prepared to prove the inverse function theorem.

THEOREM 1.24 (Inverse Function Theorem). Suppose f is a $C^{(1)}$-mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^n$, $f'(a)$ is invertible for some $a \in E$. Then

(i) there exist open sets U and V in $\mathbb{R}^n$ such that $a \in U$, $f(a) \in V$, f is one-to-one on U, and $f(U) = V$;

(ii) if g is the inverse of f (which exists, by (i)), defined in V, by

$$g(f(x)) = x \quad \text{for} \quad x \in U\,,$$

then $g \in C^{(1)}(V)$.

REMARK 1.14. Since the linear map $f'(a)$ is the best linear approximation to f at a, it is plausible that f is invertible in a neighborhood of a if and only if $f'(a)$ is also.

REMARK 1.15. Although the inverse function theorem apparently reduces the invertibility of f on an open set to a single number at a, because f is continuously differentiable, the invertibility of the derivative at a is equivalent to its invertibility in a neighborhood of a.

REMARK 1.16. Writing the equation $y = f(x)$ in component form, we arrive at the following interpretation of the conclusion of the theorem: The system of n equations

$$y_i = f_i\left(x_1, \ldots, x_n\right) \quad (1 \leq i \leq n)$$

can be solved for $x_1, \ldots, x_n$ in terms of $y_1, \ldots, y_n$, if we restrict x and y to small enough neighborhoods of a and b; the solutions are unique and continuously differentiable.

Without loss of generality, suppose that $f(a) = 0$.

Proof of (i). Since f' is continuous at a, there is an open ball U with center at a so that $\overline{U} \subset E$, and

$$\left\|f'(x) - f'(a)\right\| < \frac{1}{2\|f'(a)^{-1}\|} \quad (x \in U)\,. \tag{1.50}$$

We associate to each fixed $y \in \mathbb{R}^n$ a function φ, defined by

$$\varphi_y(x) = x + f'(a)^{-1}(y - f(x)) \quad (x \in E)\,.$$

Note that x is a fixed point of φ_y if and only if $f(x) = y$.

Since $\varphi_y'(x) = I - f'(a)^{-1}f'(x) = f'(a)^{-1}\left(f'(a) - f'(x)\right)$, so we have

$$\left\|\varphi_y'(x)\right\| < \frac{1}{2} \quad (x \in U)$$

Hence by Theorem 1.8,

$$\left\|\varphi_y\left(x_1\right) - \varphi_y\left(x_2\right)\right\| \leq \frac{1}{2}\left\|x_1 - x_2\right\| \quad \left(x_1, x_2 \in U\right)\,, \tag{1.51}$$

which implies that φ_y has at most one fixed point in U. So $f(x) = y$ for at most one $x \in U$. Thus f is $1-1$ in U. In fact, it follows from (1.51) that

$$\|x_1 - x_2\| \leq 2\|f'(a)^{-1}\|\|f(x_1) - f(x_2)\| \quad \left(x_1, x_2 \in U\right)\,, \tag{1.52}$$

which implies that the inverse of f is continuous, as we will see later.

Next, put $V = f(U)$. We shall show that V is open. Let $y_0 \in V$, i.e., $y_0 = f\left(x_0\right)$ for some $x_0 \in U$. We will show that $y \in V$ whenever $\|y - y_0\|$ is sufficiently small. To this end, for the given y we will use contraction theorem to show that there esists a fixed point of φ_y in U. Take a closed ball $B = \overline{B}(x_0, r) \subset U$, we shall show that φ_y maps B into B if $\|y - y_0\|$

is small. Then it follows that φ_y has a fixed point $x \in B$ and for this x, $f(x) = y$. So $V = f(U)$ is open.

Observe that, for $x \in \bar{B}$,

$$\begin{aligned} \|\varphi_y(x) - x_0\| &\le \|\varphi_y(x) - \varphi_y(x_0)\| + \|\varphi_y(x_0) - x_0\| \\ &\le \frac{1}{2}\|x - x_0\| + \left\|f'(a)^{-1}\right\| \|y - y_0\| \\ &\le \frac{r}{2} + \left\|f'(a)^{-1}\right\| \|y - y_0\|, \end{aligned}$$

hence $\varphi_y(x) \in B$ if $\left\|f'(a)^{-1}\right\| \|y - y_0\| \le \frac{r}{2}$. This proves part (i) of the theorem. □

Proof of (ii). Pick y and $y+k \in V$. Let $x = g(y) \in U$, and $x+h = g(y+k) \in U$. Hence $y = f(x), y+k = f(x+h)$. It follows from (1.52) that

$$\begin{aligned} \|g(y+k) - g(y)\| = \|x + h - x\| &= \|h\| \\ &\le 2\|f'(a)^{-1}\|\|f(x+h) - f(x)\| \\ &= 2\|f'(a)^{-1}\|\|k\|. \end{aligned} \tag{1.53}$$

so g is continuous at y. Moreover, it follows from (1.50) that $f'(x)$ is invertible. Then

$$\begin{aligned} g(y+k) - g(y) - f'(x)^{-1}k &= x + h - x - f'(x)^{-1}k \\ &= f'(x)^{-1}\{f'(x)h - [f(x+h) - f(x)]\}, \end{aligned}$$

and hence

$$\begin{aligned} &\frac{\|g(y+k) - g(y) - f'(x)^{-1}k\|}{\|k\|} \\ &\le \|f'(x)^{-1}\| \frac{\|f(x+h) - f(x) - f'(x)h\|}{\|h\|} \frac{\|h\|}{\|k\|} \\ &\le 2\|f'(a)^{-1}\|\|f'(x)^{-1}\| \frac{\|f(x+h) - f(x) - f'(x)h\|}{\|h\|}. \end{aligned}$$

As $k \to 0$, (1.53) shows that $h \to 0$. The right side of the last inequality thus tends to 0. Hence the same is true of the left. We have thus proved that $g'(y) = f'(x)^{-1}$; i.e.,

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

Finally, note that g is a continuous mapping of V onto U, that f' is a continuous mapping of U into the set $GL_n(\mathbb{R})$ of all invertible elements of $L(\mathbb{R}^n)$, and that inversion is a continuous mapping of $GL_n(\mathbb{R})$ onto $GL_n(\mathbb{R})$, by Theorem 1.1. If we combine these facts, we see that $g \in \mathcal{C}'(V)$. This completes the proof. $\square$

A The Basic Theorems of Differential Calculus

A.1 Mean-Value Theorems

A.1.1. Fermat's Lemma and Rolle's Theorem A point $x_0 \in E \subset \mathbb{R}$ is called a *local maximum point* (resp. *local minimum point*) and the value of a function $f : E \to \mathbb{R}$ at that point a *local maximum value* (resp. *local minimum value*), if there exists a neighborhood $U_E(x_0)$ of x_0 in E such that

$$f(x) \leq f(x_0) \ \ (\text{ resp. } f(x) \geq f(x_0)) \quad \text{for all } x \in U_E(x_0) .$$

If the strict inequality $f(x) < f(x_0)$ (resp. $f(x) > f(x_0)$) holds at every point $x \in U_E(x_0) \setminus \{x_0\}$, the point x_0 is called *strict local maximum point*(resp. *strict local minimum point*) and the value of the function $f : E \to \mathbb{R}$ a *strict local maximum value* (resp. *strict local minimum value*).

The local maximum and minimum points are called *local extremum points* and the values of the function at these points *local extreme values* of the function. We say an extremum point $x_0 \in E$ of the function $f : E \to \mathbb{R}$ is *interior*, if x_0 is a limit point of both sets $\{x \in E : x < x_0\}$ and $\{x \in E : x > x_0\}$.

Lemma A.1 (Fermat). If a function $f : E \to \mathbb{R}$ is differentiable at an interior extremum point $x_0 \in E$, then its derivative at x_0 is 0: $f'(x_0) = 0$.

Proof. By definition of differentiability at x_0 we have

$$f(x_0 + h) - f(x_0) = f'(x_0)\, h + \alpha(x_0; h)\, h$$

where $\alpha(x_0; h) \to 0$ as $h \to x, x_0 + h \in E$. Let us rewrite this relation as follows:

$$f(x_0 + h) - f(x_0) = \left[f'(x_0) + \alpha(x_0; h)\right] h \tag{A.1}$$

Since x_0 is an extremum point, the left-hand side of (A.1) is either non-negative or non-positive for all values of h sufficiently close to 0 and for which $x_0 + h \in E$. If $f'(x_0) \neq 0$, then for h sufficiently close to 0 the quantity $f'(x_0) + \alpha(x_0; h)$ would have the same sign as $f'(x_0)$, since $\alpha(x_0; h) \to 0$ as $h \to 0, x_0 + h \in E$. But the value of h can be both positive or negative, given that x_0 is an interior extremum point.

Thus, assuming that $f'(x_0) \neq 0$, we find that the right-hand side of (A.1) changes sign when h does (for h sufficiently close to 0), while the left-hand side cannot change sign when h is sufficiently close to 0. This contradiction completes the proof. □

Remark A.1. Geometrically Fermat's lemma is obvious, since it asserts that at an extremum of a differentiable function the tangent to its graph is horizontal. (After all, $f'(x_0)$ is the tangent of the angle the tangent line makes with the x -axis.) Physically this lemma means that in motion along a line the velocity must be zero at the instant when the direction reverses (which is an extremum point!).

Remark A.2. Fermat's lemma gives a necessary condition for an interior extremum of a differentiable function. For non-interior extremum points, it is generally not true that $f'(x_0) = 0$.

This lemma and the theorem on the maximum (or minimum) of a continuous function on a compact interval together imply the following proposition.

Theorem A.2 (Rolle's Theorem). *Let f be a real-valued function that is continuous on a compact interval $[a, b]$ and differentiable on the open interval (a, b). If $f(a) = f(b)$, then there exists a point $\xi \in (a, b)$ such that $f'(\xi) = 0$.*

Proof. Since the function f is continuous on $[a, b]$, there exist points x_m, x_M in $[a, b]$ at which it assumes its minimal and maximal values respectively.

If $f(x_m) = f(x_M)$, then the function is constant on $[a, b]$; and since in that case $f'(x) \equiv 0$ in (a, b), the assertion is obviously true.

If $f(x_m) < f(x_M)$, then, since $f(a) = f(b)$, one of the points x_m and x_M must lie in the open interval (a, b). We denote it by ξ. Fermat's lemma now implies that $f'(\xi) = 0$. □

The following theorem states that every function that results from the differentiation of another function has the *intermediate value property*: the image of an interval is also an interval.

Theorem A.3 (Darboux's Theorem). Let $I \subset \mathbb{R}$ be an interval. Let $f : I \to \mathbb{R}$ be a differentiable function. Then f' has the intermediate value property: If a and b are points in I with $a < b$, then for every t between $f'(a)$ and $f'(b)$, there exists an ξ in $[a, b]$ such that $f'(\xi) = t$.

Proof. If t equals $f'(a)$ or $f'(b)$, then setting x equal to a or b, respectively, gives the desired result. Now assume that t is strictly between $f'(a)$ and $f'(b)$. Without loss of generality, suppose that $f'(b) < t < f'(a)$.

Let $\varphi : I \to \mathbb{R}$ such that $\varphi(x) = f(x) - tx$. Since φ is continuous on the compact interval $[a, b]$, the maximum value of φ on $[a, b]$ is attained at some point in $[a, b]$.

Because $\varphi'(a) = f'(a) - t > 0$, we know φ cannot attain its maximum value at a. (If it did, then $(\varphi(t) - \varphi(a))/(t - a) \leq 0$ for all $t \in [a, b]$, which implies $\varphi'(a) \leq 0$.) Likewise, because $\varphi'(b) = f'(b) - t < 0$, we know φ cannot attain its maximum value at b. Therefore, φ must attain its maximum value at some point $\xi \in (a, b)$. Hence, by Fermat's lemma, $\varphi'(\xi) = 0$, i.e. $f'(\xi) = t$. □

A.1.2. The Lagrange Mean-Value Theorem The following theorem is one of the most frequently used and important methods of studying real-valued functions. Lagrange's theorem is important in that it connects the increment of a function over a finite interval with the derivative of the function on that interval. Up to now we have not had such a theorem on finite increments and have characterized only the local (infinitesimal) increment of a function in terms of the derivative at a given point.

THEOREM A.4 (Lagrange's Theorem). If a real-valued function f is continuous on a compact interval $[a, b]$ and differentiable on the open interval (a, b), there exists a point $\xi \in (a, b)$ such that

$$f(b) - f(a) = f'(\xi)(b - a)\,. \tag{A.2}$$

Proof. Consider the auxiliary function

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)\,,$$

which is obviously continuous on the compact interval $[a, b]$ and differentiable on the open interval (a, b) and has equal values at the endpoints:

$$F(a) = F(b) = 0\,.$$

Applying Rolle's theorem to $F(x)$, we find a point $\xi \in (a, b)$ at which

$$F'(\xi) = f'(\xi) - \frac{f(b) - f(a)}{b - a} = 0\,.$$

□

REMARK A.3. In geometric language Lagrange's theorem means that at some point $(\xi, f(\xi))$, where $\xi \in (a, b)$, the tangent to the graph of the function is parallel to the chord joining the points $(a, f(a))$ and $(b, f(b))$, since the slope of the chord equals $\frac{f(b)-f(a)}{b-a}$.

REMARK A.4. If x is interpreted as time and $f(b) - f(a)$ as the amount of displacement over the time $b - a$ of a particle moving along a line, Lagrange's theorem says that the velocity $f'(x)$ of the particle at some time $\xi \in (a, b)$ is such that if the particle had moved with the constant velocity $f'(\xi)$ over the whole time interval, it would have been displaced by the same amount $f(b) - f(a)$. It is natural to call $f'(\xi)$ the *average velocity* over the time interval $[a, b]$.

Example A.1. We note nevertheless that for motion that is not along a straight line there may be no average speed in the sense of Remark A.4. Indeed, suppose the particle is moving around a circle of unit radius at constant angular velocity $\omega = 1$. Its law of motion, as we know, can be written as

$$\mathbf{r}(t) = (\cos t, \sin t)$$

Then

$$\mathbf{r}'(t) = \mathbf{v}(t) = (-\sin t, \cos t)$$

and $\|\mathbf{v}(t)\| = \sqrt{\sin^2 t + \cos^2 t} = 1$. The particle is at the same point $\mathbf{r}(0) = \mathbf{r}(2\pi) = (1, 0)$ at times $t = 0$ and $t = 2\pi$ and the equality

$$\mathbf{r}(2\pi) - \mathbf{r}(0) = \mathbf{v}(\xi)(2\pi - 0)$$

would mean that $\mathbf{v}(\xi) = \mathbf{0}$. But this is impossible.

Even so, we shall learn that there is still a relation between the displacement over a time interval and the velocity. It consists of the fact that the full length L of the path traversed cannot exceed the maximal absolute value of the velocity multiplied by the time interval of the displacement. What has just been said can be written in the following more precise form:

$$\|\mathbf{r}(b) - \mathbf{r}(a)\| \leq \sup_{t \in (a,b)} \|\mathbf{r}'(t)\| |b - a| \,.$$

As will be shown later, this natural inequality does indeed always hold. It is also called Lagrange's finite-increment theorem, while relation (A.2), which is valid only for real-valued functions, is often called the *Lagrange mean-value theorem* (the role of the mean in this case is played by both the value $f'(\xi)$ of the velocity and by the point ξ between a and b).

Now we give some applications of the Lagrange mean-value theorem.

Proposition A.5 (Criterion for Monotonicity). Let f be a differentiable real-valued function on an open interval $I \subset \mathbb{R}$.

(i) If f' is nonnegative (resp. positive) at every point of I, then f is increasing (resp. strictly increasing) on I.

(ii) If f' is nonzero at every point of I, then f is strictly monotone on I.

Proof. To show part (i), indeed, if x_1 and x_2 are two points of the interval and $x_1 < x_2$, that is, $x_2 - x_1 > 0$, then by formula (A.2)

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1), \qquad \text{where } x_1 < \xi < x_2;$$

and therefore, the sign of the difference on the left-hand side of this equality is the same as the sign of $f'(\xi)$.

To show part (ii), note that by the intermediate value property of f', either $f'(x) > 0$ for all $x \in I$ or $f'(x) < 0$ for all $x \in I$, then the desired result follows form (i). □

Proposition A.6 (Criterion for a Function to be Constant). A real-valued function that is continuous on a compact interval $[a, b]$ is constant on it if and only if its derivative equals zero at every point of the open interval (a, b).

Proof. Only the fact that $f'(x) \equiv 0$ on (a, b) implies that $f(x_1) = f(x_2)$ for all $x_1, x_2, \in [a, b]$ is of interest. But this follows from Lagrange's formula, according to which

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) = 0$$

since ξ lies between x_1 and x_2, that is, $\xi \in (a, b)$, and so $f'(\xi) = 0$. □

From this we can draw the following conclusion (which as we shall see, is very important for integral calculus): If the derivatives $F_1'(x)$ and $F_2'(x)$ of two functions $F_1(x)$ and $F_2(x)$ are equal on some interval, that is, $F_1'(x) = F_2'(x)$ on the interval, then the difference $F_1(x) - F_2(x)$ is constant.

Proposition A.7. Let f be a differentiable real-valued function defined on a compact interval $[a, b]$. Then $f' : [a, b] \to \mathbb{R}$ has no discontinuity point of the first kind.

Proof. Suppose for contradiction that $c \in [a, b]$ is a discontinuity point of the first kind for f'. Without loss of generality we suppose $c \in (a, b)$. Then by assumption the limits

$$\lim_{x\downarrow c} f'(x) =: f'(c+) \ , \ \lim_{x\uparrow c} f'(x) =: f'(c-)$$

exists and are finite. However, note that

$$f'(c) = \lim_{x\downarrow c} \frac{f(x) - f(c)}{x - c} = \lim_{x\downarrow c} f'(\xi_x) \, ,$$

where $c < \xi_x < x$, since $\xi_x \downarrow c$ as $x \downarrow c$, we get $f'(c) = f'(c^+)$. Similarly, $f'(c) = f'(c^-)$, and hence f' is continuous at c, which is absurd. □

EXERCISE A.1. Let $f : [a, b] \to \mathbb{R}$ is continuos. If f is differentiable in (a, b) and the limits

$$\lim_{x\downarrow a} f'(x) =: A$$

exists and is finite. Then show that f is differentiable at a and $f'(a) = A$.

A.1.3. THE CAUCHY MEAN-VALUE THEOREM The following proposition is a useful generalization of Lagrange's theorem, and is also based on Rolle's theorem.

THEOREM A.8 (Cauchy's Theorem). Let $x = x(t)$ and $y = y(t)$ be functions that are continuous on a compact interval $[a, b]$ and differentiable on the open interval (a, b). Then there exists a point $\xi \in (a, b)$ such that

$$x'(\xi)(y(b) - y(a)) = y'(\xi)(x(b) - x(a)) \, .$$

If in addition $x'(t) \neq 0$ for each $t \in (a, b)$, then $x(a) \neq x(b)$ and we have

$$\frac{y(b) - y(a)}{x(b) - x(a)} = \frac{y'(\xi)}{x'(\xi)} \, . \tag{A.3}$$

Proof. The function

$$F(t) = x(t)[y(b) - y(a)] - y(t)[x(b) - x(a)]$$

satisfies the hypotheses of Rolle's theorem on the compact interval $[a, b]$. Therefore there exists a point $\xi \in (a, b)$ at which $F'(\xi) = 0$, which is equivalent to the equality to be proved.

To obtain relation (A.3) from it, it remains only to observe that if $x'(t) \neq 0$ on (a, b), then $x(a) \neq x(b)$, again by Rolle's theorem. $\square$

Clearly, Lagrange's theorem can be obtained from Cauchy's by setting $x(t) = t$ and $y(t) = f(t)$.

REMARK A.5. If we regard the pair $(x(t), y(t))$ as the law of motion of a particle, then $(x'(t), y'(t))$ is its velocity vector at time t, and $(x(b) - x(a), y(b) - y(a))$ is its displacement vector over the time interval $[a, b]$. The theorem then asserts that at some instant of time $\xi \in [a, b]$ these two vectors are collinear. However, this fact, which applies to motion in a plane, is the same kind of pleasant exception as the mean-velocity theorem in the case of motion along a line. Indeed, imagine a particle moving at uniform speed along a *helix*:

$$\mathbf{r}(t) = (\cos t, \sin t, t)$$

Its velocity makes a constant nonzero angle with the vertical, while the displacement vector can be purely vertical (after one complete turn).

We give some applications of Cauchy's mean-value theorem.

THEOREM A.9 (L'Hôpital's Rule I). Let f and g be two differentiable real-valued function on (a, b) such that $g'(x) \neq 0$ for all $x \in (a, b)$. If one of the following statements hold

(i) $\lim_{x \downarrow a} f(x) = \lim_{x \downarrow a} g(x) = 0$,

(ii) $\lim_{x \downarrow a} g(x) = \infty$,

and in addition the limit $\lim_{x \downarrow a} \frac{f'(x)}{g'(x)}$ exists in $[-\infty, \infty]$, then we have

$$\lim_{x \downarrow a} \frac{f(x)}{g(x)} = \lim_{x \downarrow a} \frac{f'(x)}{g'(x)} .$$

Proof. If (i) holds, put $f(a) = g(a) = 0$, then f and g is continuous on $[a, x]$, for each $x \in (a, b)$. By Cauchy's mean-value theorem, there exists ξ_x in (a, x) so that

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(\xi_x)}{g'(\xi_x)} .$$

Thus $\xi_x \downarrow a$ as $x \downarrow a$ and hence

$$\lim_{x \downarrow a} \frac{f(x)}{g(x)} = \lim_{x \downarrow a} \frac{f'(\xi_x)}{g'(\xi_x)} = \lim_{x \downarrow a} \frac{f'(x)}{g'(x)} \,,$$

as desired.

If (ii) holds, put $l = \lim_{x \downarrow a} \frac{f'(x)}{g'(x)}$. We begin with assuming that l is finite. For any given $\epsilon > 0$, there exists $\delta = \delta_\epsilon > 0$ so that for all $x \in (a, a+\delta)$,

$$l - \epsilon < \frac{f'(x)}{g'(x)} < l + \epsilon \,.$$

Thus for every $x \in (a, a+\delta)$, by Cauchy's mean-value theorem, there exists $\xi \in (x, a+\delta) \subset (a, a+\delta)$ so that

$$l - \epsilon < \frac{f(x) - f(a+\delta)}{g(x) - g(a+\delta)} = \frac{f'(\xi)}{g'(\xi)} < l + \epsilon \,. \tag{A.4}$$

Observe that since $g(x) \to \infty$ as $x \downarrow a$, we have

$$\limsup_{x \downarrow a} \frac{f(x) - f(a+\delta)}{g(x) - g(a+\delta)} = \limsup_{x \downarrow a} \frac{f(x)}{g(x)} \,,$$

$$\liminf_{x \downarrow a} \frac{f(x) - f(a+\delta)}{g(x) - g(a+\delta)} = \liminf_{x \downarrow a} \frac{f(x)}{g(x)} \,.$$

Thus letting $x \downarrow a$ in (A.4) we get

$$l - \epsilon < \liminf_{x \downarrow a} \frac{f(x)}{g(x)} \le \limsup_{x \downarrow a} \frac{f(x)}{g(x)} < l + \epsilon \,.$$

Since ϵ is arbitrary, we get

$$\liminf_{x \downarrow a} \frac{f(x)}{g(x)} = \limsup_{x \downarrow a} \frac{f(x)}{g(x)} = l \,,$$

and the desired result follows.

If l is ∞ or $-\infty$, the theorem follows by the same argument. □

COROLLARY A.10 (L'Hôpital's Rule II). Let f and g be two differentiable real-valued function on (a, ∞) such that $g'(x) \neq 0$ for all $x \in (a, \infty)$. If one of the following statements hold

(i) $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = 0$,

(ii) $\lim_{x\to\infty} g(x) = \infty$,

and in addition the limit $\lim_{x\to\infty} \frac{f'(x)}{g'(x)}$ exists in $[-\infty, \infty]$, then we have

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)} .$$

Proof. Since $t \mapsto \frac{1}{t}$ is continuous on $(0, \infty)$, using change of variable we have

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{t\downarrow 0} \frac{f\left(\frac{1}{t}\right)}{g\left(\frac{1}{t}\right)} .$$

Thus we define F and G by

$$F(t) = f\left(\frac{1}{t}\right) \quad \text{and} \quad G(t) = g\left(\frac{1}{t}\right) \quad \text{for} \quad t \in (0, \frac{1}{a}) .$$

Then F and G are differentiable on $(0, a^{-1})$ with

$$F'(t) = -\frac{1}{t^2}\, f'\left(\frac{1}{t}\right) \quad \text{and} \quad G'(t) = -\frac{1}{t^2}\, g'\left(\frac{1}{t}\right)$$

It's easy to see that the conditions of THEOREM A.6 are satisfied, so

$$\begin{aligned} \lim_{x\to\infty} \frac{f(x)}{g(x)} &= \lim_{t\downarrow 0} \frac{F(t)}{G(t)} = \lim_{t\downarrow 0} \frac{F'(t)}{G'(t)} \\ &= \lim_{t\downarrow 0} \frac{f'\left(\frac{1}{t}\right)}{g'\left(\frac{1}{t}\right)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)} , \end{aligned}$$

as desired. ☐

A.2 Taylor's Formula

From the amount of differential calculus that has been explained up to this point one may obtain the correct impression that the more derivatives of two functions coincide (including the derivative of zeroth order) at a point, the better these functions approximate each other in a neighborhood of that point. We have mostly been interested in approximations of a function in the neighborhood of a point by a polynomial

$$P_n(x) = c_0 + c_1 (x - x_0) + \cdots + c_n (x - x_0)^n ,$$

and that will continue to be our main interest. We know that an algebraic polynomial can be represented as

$$P_n(x) = P_n(x_0) + \frac{P_n'(x_0)}{1!}(x - x_0) + \cdots + \frac{P_n^{(n)}(x_0)}{n!}(x - x_0)^n$$

that is, $c_k = \frac{P_n^{(k)}(x_0)}{k!}$ for $k = 0, 1, \ldots, n$. This can easily be verified directly. Thus, if we are given a function $f(x)$ having derivatives up to order n inclusive at x_0, we can immediately write the polynomial

$$\begin{aligned} T_n(f, x_0; x) &= \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \\ &= f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n , \end{aligned} \tag{A.5}$$

whose derivatives up to order n inclusive at the point x_0 are the same as the corresponding derivatives of $f(x)$ at that point. The algebraic polynomial given by is the *Taylor polynomial* of order n of f at x_0.

We shall be interested in the value of

$$f(x) - T_n(f, x_0; x) = r_n(f, x_0; x)$$

of the discrepancy between the polynomial $P_n(x)$ and the function $f(x)$, which is often called the *remainder*, more precisely, the nth remainder or the nth remainder term in Taylor's formula:

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + r_n(f, x_0; x) . \tag{A.6}$$

The equality (A.6) itself is of course of no interest if we know nothing more about the function $r_n(f, x_0; x)$ than its definition.

Theorem A.11 (The Peano Form of Remainder). Let f be a real-valued function on an interval $I \subset \mathbb{R}$ such that its nth derivative $f^{(n)}(x_0)$ at $x_0 \in I$ exists. Then

$$r_n(f, x_0; x) = o(|x - x_0|^n) \quad \text{as } x \to x_0\,, x \in I\,.$$

Proof. We shall use the induction. When $n = 1$, the theorems follows form the definition of derivative. Suppose now the theorems holds for $n - 1$. Observe that

$$T_n'(f, x_0; x) = T_{n-1}\left(f', x_0; x\right)\,,$$

then by L'Hospital's rule and by assumption

$$\lim_{\substack{x \to x_0 \\ x \in I}} \frac{r_n(f, x_0; x)}{(x - x_0)^n} = \lim_{\substack{x \to x_0 \\ x \in I}} \frac{f(x) - T_n(f, x_0; x)}{(x - x_0)^n}$$
$$= \lim_{\substack{x \to x_0 \\ x \in I}} \frac{f'(x) - T_{n-1}(f', x_0; x)}{n(x - x_0)^{n-1}} = \frac{1}{n} \lim_{\substack{x \to x_0 \\ x \in I}} \frac{r_{n-1}(f', x_0; x)}{(x - x_0)^{n-1}} = 0\,.$$

Thus the theorem also holds for n, as desired. □

Remark A.6. Taylor's formula with the Peano form of the remainder, is obviously a generalization of the definition of differentiability of a function at a point, to which it reduces when $n = 1$.

We return once again to the problem of the local representation of a function $f : O \to \mathbb{R}$ by a polynomial. We wish to choose the polynomial $P_n(x_0; x) = x_0 + c_1(x - x_0) + \cdots + c_n(x - x_0)^n$ so as to have

$$f(x) = P_n(x_0, x) + o\left(|x - x_0|^n\right) \quad \text{as } x \to x_0, x \in I \tag{A.7}$$

Clearly Taylor;s polynomial satisfies us if $f^{(n)}(x_0)$ exists. In fact, if the polynomial exists it must be unique.

Proposition A.12. If there exists a polynomial $P_n(x_0; x)$ satisfying condition (A.7), that polynomial must be unique.

Proof. Indeed, from relation (A.7) we obtain the coefficients of the polynomial successively and completely unambiguously

$$\begin{aligned}
c_0 &= \lim_{I \ni x \to x_0} f(x)\,; \\
c_1 &= \lim_{I \ni x \to x_0} \frac{f(x) - c_0}{x - x_0}\,; \\
&\vdots \\
c_n &= \lim_{I \ni x \to x_0} \frac{f(x) - \left[c_0 + \cdots + c_{n-1}\,(x - x_0)^{n-1}\right]}{(x - x_0)^n}\,.
\end{aligned}$$

Thus the polynomial is unique. □

We shall now use a highly artificial device to obtain information on the remainder term. A more natural route to this information will come from the integral calculus.

THEOREM A.13 (The Mean-Value Form of the Remainder). If the function f is continuous on the closed interval with end-points x_0 and x along with its first n derivatives, and it has a derivative of order $n+1$ at the interior points of this interval, then for any function φ that is continuous on this closed interval and has a nonzero derivative at its interior points, there exists a point ξ between x_0 and x such that

$$r_n\,(f, x_0; x) = \frac{f^{(n+1)}(\xi)}{n!\,\varphi'(\xi)}(x - \xi)^n\ \ [\varphi(x) - \varphi\,(x_0)]\,. \tag{A.8}$$

REMARK A.7 (The Largane Form and the Cauchy Form of the Remainder). A particularly elegant formula results if we set $\varphi(t) = (x-t)^{n+1}$ in (A.8), then we get the so-called the *Lagrange form of the remainder*:

$$r_n\,(x_0; x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\,(x - x_0)^{n+1}\,.$$

If we setting $\varphi(t) = x - t$ in (A.8) we obtain the *Cauchy's form of the remainder*:

$$r_n\,(x_0; x) = \frac{f^{(n+1)}(\xi)}{n!}(x - \xi)^n\,(x - x_0)\,.$$

Taylor's formula with the Lagrange form or the Cauchy form of the remainder, is obviously a generalization of Lagrange's mean-value theorem, to which it reduces when $n = 0$

Proof. On the closed interval I with endpoints x_0 and x we consider the auxiliary function

$$\begin{aligned} F(t) &:= r_n(f, t; x) = f(x) - T_n(f, t; x) \\ &= f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(t)}{k!}(x - t)^k \;, \quad \text{for } t \in I\,. \end{aligned}$$

Clearly $F(x_0) = r_n(f, x_0; x)$ and $F(x) = r_n(f, x; x) = 0$, and we see from the definition of F and the hypotheses of the theorem that F is continuous on the closed interval I and differentiable at its interior points, with

$$\begin{aligned} F'(t) &= -\left[\sum_{k=0}^{n} \frac{f^{(k+1)}(t)}{k!}(x - t)^k - \sum_{k=1}^{n} \frac{f^{(k)}(t)}{(k-1)!}(x - t)^{k-1}\right] \\ &= -\frac{f^{(n+1)}(t)}{n!}(x - t)^n\,. \end{aligned}$$

Applying Cauchy's theorem to the pair of functions $F(t), \varphi(t)$ on the closed interval I, we find a point ξ between x_0 and x at which

$$\frac{F(x) - F\left(x_0\right)}{\varphi(x) - \varphi\left(x_0\right)} = \frac{F'(\xi)}{\varphi'(\xi)}\,.$$

Substituting the expression for $F'(\xi)$ here and observing that

$$F(x) - F\left(x_0\right) = -r_n\left(f, x_0; x\right)\,,$$

we obtain the desired equality. □

If the function f has derivatives of all orders $n \in \mathbb{N}$ at a point x_0, the series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}\left(x_0\right)}{n!}\left(x - x_0\right)^n$$

is called the *Taylor series* of f at the point x_0.

It should not be thought that the Taylor series of an infinitely differentiable function converges in some neighborhood of x_0. Since for given any sequence $c_0, c_1, \ldots, c_n, \ldots$ of real numbers, one can construct (although this is not simple to do) a function f such that $f^{(n)}(x_0) = c_n$, for all $n \in \mathbb{N}_0$. As we know, the radius of convergence of the series is given by

$$R = \frac{1}{\limsup\limits_{n\to\infty} \left| \frac{f^{(n)}(x_0)}{n!} \right|^{\frac{1}{n}}} .$$

Obviously, $f(x) = \sum_0^\infty \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n$, i.e., $T_n(f, x_0; x) \to f(x)$ if and only if $r_n(f, x_0; x) \to 0$. It should also not be thought that if the Taylor series converges, it necessarily converges to the function that generated it. A Taylor series converges to the function that generated it only when the generating function belongs to the class of so-called *analytic functions*.

PROPOSITION A.14. Let f be a infinitely differentiable real-valued function in an open interval I, i.e., $f \in C^\infty(I)$. If there exists a constant $M > 0$ so that

$$|f^{(n)}(x)| \le M \quad \text{for all} \quad x \in I, n \in \mathbb{N}_0 ,$$

then for each $x_0 \in I$, the Taylor series of f at x_0 converges to f in I, i.e.,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n \quad \text{for all} \quad x \in I .$$

If in addition I is bounded, then the convergence is uniform in x.

Proof. It suffices to show that for each given x_0 and x in I, $r_n(f, x_0; x) \to 0$. By the Taylor formula with the Lagrange form of the remainder,

$$|r_n(f, x_0; x)| = \left| \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - x_0)^{n+1} \right| \le M \frac{|x - x_0|^{n+1}}{(n+1)!} .$$

Clearly $r_n(f, x_0; x) \to 0$ as $n \to \infty$. If I is bounded, then $|x - x_0| \le d :=$ diam$(I) < \infty$ and hence

$$|r_n(f, x_0; x)| \le M \frac{d^{n+1}}{(n+1)!} ,$$

so $r_n(f, x_0; x) \to 0$ uniformly in x as $n \to \infty$. □

Example A.2 (A Nonanalytic Function). Here is Cauchy's example of a nonanalytic function:

$$f(x) = \begin{cases} e^{-1/x^2}, & \text{if } x \neq 0\,; \\ 0, & \text{if } x = 0\,. \end{cases}$$

Starting from the definition of the derivative and the fact that for each fixed $k \in \mathbb{N}_0$

$$x^k e^{-1/x^2} \to 0 \quad \text{as } x \to 0$$

it's easy to verify that $f^{(n)}(0) = 0$ for $n = 0, 1, 2, \ldots$. Thus, the Taylor series in this case has all its terms equal to 0 and hence its sum is identically equal to 0, while $f(x) \neq 0$ if $x \neq 0$.

We end this section with the Taylor formula with integral form of the remainder, which gives a more natural route to the Lagrange form of the reminder.

Theorem A.15 (Integral Form of the Remainder). If the real-valued function f has continuous derivatives up to order $n+1$ inclusive on the closed interval with endpoints x_0 and x, then

$$r_n(f, x_0; x) = \frac{1}{n!} \int_{x_0}^{x} f^{(n+1)}(t)(x-t)^n \,\mathrm{d}t\,.$$

Proof. Using the Newton-Leibniz formula and integration by parts, we carry out the following chain of transformations, in which all differentiations and substitutions are carried out on the variable t:

$$\begin{aligned}
f(x) &= f(x_0) + \int_{x_0}^{x} f'(t)\,\mathrm{d}t = f(x_0) - \int_{x_0}^{x} f'(t)(x-t)'\,\mathrm{d}t \\
&= f(x_0) - f'(t)(x-t)\Big|_{x_0}^{x} + \int_{x_0}^{x} f''(t)(x-t)\,\mathrm{d}t \\
&= f(x_0) + f'(x_0)(x-x_0) + \int_{x_0}^{x} f''(t)(x-t)\,\mathrm{d}t \\
&= f(x_0) + f'(x_0)(x-x_0) - \frac{1}{2}\int_{x_0}^{x} f''(t)\left((x-t)^2\right)'\,\mathrm{d}t \\
&= f(x_0) + f'(x_0)(x-x_0) - \frac{1}{2} f''(t)(x-t)^2\Big|_{x_0}^{x} + \frac{1}{2}\int_{x_0}^{x} f'''(t)(x-t)^2\,\mathrm{d}t
\end{aligned}$$

$$= \cdots = T_n(f, x_0; x) + \frac{1}{n!} \int_{x_0}^{x} f^{(n+1)}(t)(x-t)^n \, \mathrm{d}t \, .$$

Thus $r_n(f, x_0; x) = f(x) - T_n(f, x_0; x)$ is the integral above. □

B Differential Calculus from a More General Point of View

B.1 Multilinear Transformations