Introduction to Differential Geometry

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 600

Subsection 1

Differentiability for Functions of Several Variables

Partial Derivatives

- We consider real-valued functions of several variables, that is, functions whose domain is a subset $A \subseteq \mathbb{R}^n$ and whose range is $\mathbb{R}$.
- If $f : A \to \mathbb{R}$ is such a function, then $f(x) = f(x^1, \ldots, x^n)$ denotes its value at $x = (x^1, \ldots, x^n) \in A$.
- We assume now that f is a function on an open set $U \subseteq \mathbb{R}^n$.
- At each $a \in U$, the **partial derivative** $\left(\frac{\partial f}{\partial x^j}\right)_a$ **of** f **with respect to** x^j is the following limit, if it exists,

$$\left(\frac{\partial f}{\partial x^j}\right)_a = \lim_{h \to 0} \frac{f(a^1, \ldots, a^j + h, \ldots, a^n) - f(a^1 \ldots, a^j, \ldots, a^n)}{h}.$$

- If $\frac{\partial f}{\partial x^j}$ is defined, that is, the limit above exists, at each point of U, for $1 \leq j \leq n$, this defines n functions on U.
- Should these functions be continuous on U, for $1 \leq j \leq n$, f is said to be **continuously differentiable** on U, denoted by $f \in C^1(U)$.

Introducing Differentialbility

- Mere existence of partial derivatives is too weak a property for most purposes.

  Example: Consider the function defined on $\mathbb{R}^2$ by

$$f(x,y) = \begin{cases} 0, & \text{if } (x,y) = (0,0), \\ \frac{xy}{x^2+y^2}, & \text{otherwise.} \end{cases}$$

  f is not continuous at $(0,0)$.

  However, both derivatives are defined there.
- So generalize the notion of existence of the derivative as applies to functions of one variable.

Differentialbility

- We shall say that f is **differentiable at** $a \in U$ if there is a (homogeneous) linear expression $\sum_{i=1}^{n} b_i(x^i - a^i)$ such that the (inhomogeneous) linear function defined by

$$f(a) + \sum_{i=1}^{n} b_i(x^i - a^i)$$

approximates $f(x)$ near a in the sense that

$$\lim_{x \to a} \frac{f(x) - f(a) - \sum b_i(x^i - a^i)}{\|x - a\|} = 0.$$

- Equivalently, there exist constants $b_1, \ldots, b_n$ and a function $r(x, a)$, defined on a neighborhood V of $a \in U$, which satisfy:
  - $f(x) = f(a) + \sum b_i(x^i - a^i) + \|x - a\| r(x, a)$, for all $x \in V$;
  - $\lim_{x \to a} r(x, a) = 0$.

Differentiability on a Domain

- If f is differentiable for every $a \in U$, we say it is **differentiable on** U.
  - This is a technical definition from advanced calculus.
  - However, later, *differentiable* will be used rather loosely to mean differentiable of some order, usually infinitely differentiable (C^∞).
- Note that differentiability on U is a local concept.
- That is, if f is differentiable on a neighborhood of each point of U, then f is differentiable on U.
- By the Mean Value Theorem, for a function of one variable the existence of the derivative at $a \in U$ is equivalent to differentiability.
- For functions of several variables this is not the case.

Differentiability, Continuity and Partial Derivatives

- If f is differentiable at a, then it is continuous at a and all the partial derivatives $\left(\frac{\partial f}{\partial x^i}\right)_a$ exist.
- Moreover the b_i are uniquely determined for each a at which f is differentiable.
- In fact, we have

$$b_i = \left(\frac{\partial f}{\partial x^i}\right)_a.$$

Differentials

- By virtue of the preceding property, when f is differentiable at a, we have

$$f(x) - f(a) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x^i} \right)_a (x^i - a^i) + \|x - a\| r(x, a).$$

- We denote by $(df)_a$, or simply df, the homogeneous linear expression on the right:

$$df = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x^i} \right)_a (x^i - a^i).$$

- It is called the **differential of** f **at** a.

Existence and Continuity of Partial Derivatives

- Consider the partial derivatives

$$\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n}.$$

- Suppose that:
  - They are defined in a neighborhood of a;
  - They are continuous at a.
- Then f is differentiable at a.
- Thus, existence and continuity of the partial derivatives of f on an open set $U \subseteq \mathbb{R}^n$ implies differentiability of f at every point of U.

Continuous Differentiability

- We define the notion of an r-**fold continuously differentiable function** on an open set $U \subseteq \mathbb{R}^n$ (a function of class C^r) inductively.

  f is of class C^r on U if its first derivatives are of class C^{r-1}.
- Equivalently we may say that f has continuous derivatives of order $1, 2, \ldots, r$ on U.
- f is **smooth**, or of class C^∞, if f is of class C^r, for all r.
- As in the case of C^1, we denote these classes of functions on U by $C^r(U)$ and $C^\infty(U)$.

Chain Rule: First Version

- We state the first version of the chain rule.
- A more general version is given in the next section.
- Define a **differentiable** (C^r) **curve in** $\mathbb{R}^n$ to be a mapping

$$f : (a, b) \to \mathbb{R}^n$$

of an open interval $(a, b) = \{x \in \mathbb{R} : a < x < b\}$ of the real numbers into $\mathbb{R}^n$, with

$$f(t) = (x^1(t), \ldots, x^n(t)),$$

where the n coordinate functions $x^1(t), \ldots, x^n(t)$ are differentiable (resp. C^r) on the interval.

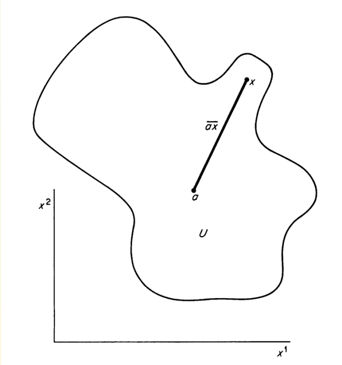
Chain Rule: First Version (Cont'd)

- Let $f : (a, b) \to \mathbb{R}^n$ be a differentiable curve.
- Assume f maps (a, b) into U, an open subset of $\mathbb{R}^n$.
- Let $a < t_0 < b$ and suppose that g is a function on U which is differentiable at $f(t_0) \in U$.
- The composite function $g \circ f$ is a real-valued function on (a, b).
- Moreover, $g \circ f$ is differentiable at t_0 and its derivative at t_0 is given by the **chain rule**

$$\frac{d}{dt}(g \circ f)_{t_0} = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x^i}\right)_{f(t_0)} \left(\frac{dx^i}{dt}\right)_{t_0}.$$

Starlike Domains

- We shall say that a domain U is **starlike with respect to** $a \in U$ provided that, whenever $x \in U$, the segment $\overline{ax}$ lies entirely in U.



- This is a somewhat weaker property than convexity of U.
- A convex set is starlike with respect to everyone of its points.

Mean Value Theorem

Theorem (Mean Value Theorem)

Let g be a differentiable function on an open set $U \subseteq \mathbb{R}^n$.
Let $a \in U$ and suppose that U is starlike with respect to a.
Then given $x \in U$, there exists $\theta \in \mathbb{R}$, $0 < \theta < 1$, such that

$$g(x) - g(a) = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial x^i} \right)_{a+\theta(x-a)} (x^i - a^i),$$

the derivatives $\frac{\partial g}{\partial x^1}, \ldots, \frac{\partial g}{\partial x^n}$ all being evaluated at the same point $a + \theta(x - a)$ on the segment $\overline{ax}$.

- Set $f(t) = a + t(x - a)$, that is

$$x^i(t) = a^i + t(x^i - a^i).$$

Mean Value Theorem (Cont'd)

- Then the corresponding curve is a line segment with:
  - $f(0) = a$;
  - $f(1) = x$.

  This curve is differentiable, in fact C^∞.

  So $g \circ f$ maps $[0, 1]$ into U and is differentiable on $(0, 1)$.

  Apply the standard Mean Value Theorem for functions of one variable (as in elementary differential calculus).

  Then use the Chain Rule to compute the derivatives.

  This gives the formula.

Mean Value Theorem (Cont'd)

Corollary

Let g be a differentiable function on an open set $U \subseteq \mathbb{R}^n$.
Let $a \in U$ and suppose that U is starlike with respect to a.
If $\left|\frac{\partial g}{\partial x^i}\right| < K$ on U, $i = 1, 2, \ldots, n$, then for any $x \in U$, we have

$$|g(x) - g(a)| < K\sqrt{n}\|x - a\|.$$

- Take absolute values in the formula of the Mean Value Theorem and use the Schwarz Inequality,

$$\begin{aligned} |g(x) - g(a)| &= \left|\sum_{i=1}^n \left(\frac{\partial g}{\partial x^i}\right)(x^i - a^i)\right| \\ &\leq \left[\sum_i \left(\frac{\partial g}{\partial x^i}\right)^2\right]^{1/2} \left[\sum_i (x^i - a^i)^2\right]^{1/2}. \end{aligned}$$

Therefore $|g(x) - g(a)| < K\sqrt{n}\|x - a\|$.

The Order of Differentiation

Corollary

If f is of class C^r on U, then at any point of U the value of the derivatives of order k, $1 < k \leq r$ is independent of the order of differentiation.
That is, if $(j_1, \ldots, j_k)$ is a permutation of $(i_1, \ldots, i_k)$, then

$$\frac{\partial^k}{\partial x^{i_1} \cdots \partial x^{i_k}} = \frac{\partial^k f}{\partial x^{j_1} \cdots \partial x^{j_k}}.$$

Analyticity

- Using the technique of the Mean Value Theorem, one may extend to functions of several variables:
  - Taylor's Theorem on polynomial approximation, with its formula for the remainder R_{N+1}, or error, of the approximation of degree N;
  - The corollary theorem on power series expansions.
- A necessary, but not sufficient, condition that a function be (real) analytic, that is, can be expanded in a power series at each $a \in U$, an open set of $\mathbb{R}^n$, is that it be in $C^\infty(U)$.
- We write $f \in C^\omega(U)$ if f is real analytic on U.

Polynomials and Rational Functions

- We do not require knowledge of analytic functions.
- However, C^ω implies C^∞.
- So it is helpful to know that any linear function $f(x) = \sum a_i x^i$, or any polynomial $P(x^1, \ldots, x^n)$ of n variables, is an analytic function on $U = \mathbb{R}^n$.
- The same is true for any quotient of polynomials (rational function) if we exclude from the domain the points at which the denominator is zero.

  Example: A determinant is an analytic function of its entries.

  Suppose that an $n \times n$ matrix A has nonzero determinant.

  Then each entry in the inverse A^{-1} of A is an analytic (and hence C^∞) function of the entries in the matrix A.

Subsection 2

Differentiability of Mappings and Jacobians

Maps and Coordinate Functions

- We generalize the preceding ideas to the case of functions defined on subsets of $\mathbb{R}^n$ but whose range is in $\mathbb{R}^m$ rather than $\mathbb{R}$.
- We will refer to them as **mappings** or **maps**.
- We try to reserve the term **function** for real-valued functions.
- Let $\pi^i : \mathbb{R}^m \to \mathbb{R}$ denote the projection to the i-th coordinate,

$$\pi^i(x^1, \ldots, x^i, \ldots, x^m) = x^i.$$

- Let $F : A \to \mathbb{R}^m$ be a mapping defined on $A \subseteq \mathbb{R}^n$.
- F is determined by its **coordinate functions** $f^i = \pi^i \circ F$,

$$F(x) = (f^1(x), \ldots, f^m(x)), \quad x \in A.$$

- Conversely, any set of m functions $f^1, \ldots, f^m$ on A with values in $\mathbb{R}$ determines a mapping $F : A \to \mathbb{R}^m$ with the coordinates of $F(x)$ given by $f^1(x), \ldots, f^m(x)$ as above.

Differentiability and Smoothness

- Let $A = U$ be an open set of $\mathbb{R}^n$, possibly all of $\mathbb{R}^n$.
- Since $\mathbb{R}^m$ and $\boldsymbol{V}^m$ may be identified, maps are sometimes referred to as **vector-valued functions on** $\mathbb{R}^n$.
- From general topology we know that F is continuous if and only if its coordinate functions are.
- We shall say that F is **differentiable**, **of class** C^r, C^∞, C^ω, and so on, **at** $a \in U$ or **on** U if each of its coordinate functions has the corresponding property.
- We may sometimes call a C^∞ mapping F a **smooth mapping**.
- If F is smooth, then:
  - Each coordinate function f^i possesses continuous partial derivatives of all orders;
  - Each such derivative is independent of the order of differentiation.

The Jacobean

- If F is differentiable on U, we know that the $m \times n$ **Jacobian matrix**

$$\frac{\partial(f^1,\ldots,f^m)}{\partial(x^1,\ldots,x^n)} = \begin{pmatrix} \frac{\partial f^1}{\partial x^1} & \cdots & \frac{\partial f^1}{\partial x^n} \\ \vdots & & \vdots \\ \frac{\partial f^m}{\partial x^1} & \cdots & \frac{\partial f^m}{\partial x^n} \end{pmatrix}$$

  is defined at each point of U, its mn entries being functions on U.
- These functions need not be continuous on U.
- They are so if and only if F is of class C^1.

Characterization of Differentiability of Maps

- We state, without proof, a characterization of differentiability.

Proposition

A mapping $F : U \to \mathbb{R}^m$, U an open subset of $\mathbb{R}^n$, is differentiable at $a \in U$ (or on U) if and only if there exist:

- An $m \times n$ matrix A of constants (respectively, functions on U);
- An m-tuple $R(x, a) = (r^1(x, a), \ldots, r^m(x, a))$ of functions defined on U (on $U \times U$),

such that:

- $\|R(x, a)\| \to 0$ as $x \to a$;
- For each $x \in U$,

$$F(x) = F(a) + A(x - a) + \|x - a\| R(x, a).$$

If such $R(x, a)$ and A exist, then A is unique and is the Jacobian matrix.

Comments

- In the expression

$$F(x) = F(a) + A(x-a) + \|x-a\| R(x,a),$$

  $A(x-a)$ denotes a matrix product.
- So we must write $(x-a)$ as an $n \times 1$ (column) matrix.
- Moreover, this is read as an equation in $m \times 1$ matrices.
- The last term means that each component of the m-tuple $R(x,a)$ is multiplied by $\|x-a\|$.

Starlike Domains

Theorem

Let $a \in U$ be an open subset of $\mathbb{R}^n$ which is starlike with respect to a. Let $F : U \to \mathbb{R}^m$ be differentiable on U, such that, for all $1 \leq i, j \leq k$,

$$\left|\frac{\partial f^i}{\partial x^j}\right| \leq K, \quad \text{at every point of } U.$$

Then the following inequality holds for all $x \in U$:

$$\|F(x) - F(a)\| \leq \sqrt{nm} K \|x - a\|.$$

Comments

- Let F be a differentiable mapping.
- We use DF to denote the Jacobian matrix of F.
- We use $DF(x)$ to denote its value at x.
- If F is differentiable on U, then, for $a \in U$, the previous expression becomes

$$F(x) = F(a) + DF(a)(x - a) + \|x - a\| R(x, a).$$

- We have $F \in C^1(U)$ iff $DF(x)$ varies continuously with x.
- Equivalently, $x \in DF(x)$ is a continuous map of U into the space $M(m, n)$ of $m \times n$ matrices, identified with $\mathbb{R}^{mn}$ and given the corresponding topology.

Composition of Mappings

- Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$, be open sets.
- Let $F : U \to V \subseteq \mathbb{R}^m$ and $G : V \to \mathbb{R}^p$.

$$U \xrightarrow{\quad F \quad} V \xrightarrow{\quad G \quad} \mathbb{R}^p$$

- Then $H = G \circ F$ is defined on U, which it maps into $\mathbb{R}^P$.
- Write the coordinate functions of H using those of F and G,

$$h^i(x) = g^i \circ F(x) = g^i(f^1(x), \ldots, f^m(x)), \quad i = 1, \ldots, p.$$

Chain Rule for Composition of Mappings

Theorem (Chain Rule)

Let F, G, H be as above.

$$U \xrightarrow{\quad F \quad} V \xrightarrow{\quad G \quad} \mathbb{R}^p$$

Suppose that:

- F is differentiable at $a \in U$
- G is differentiable at $b = F(a)$.

Then $H = G \circ F$ is differentiable at $x = a$ and we have

$$DH(a) = DG(F(a)) \cdot DF(a)$$

(where $\cdot$ indicates matrix multiplication).
If F is differentiable on U and G on V, then this holds for every $a \in U$.

Chain Rule for Composition of Mappings (Cont'd)

- According to the characterization above it is enough to show that the p-tuple $R_H(x,a)$ defined by
$$H(x) - H(a) - DG(F(a)) \cdot DF(a) \cdot (x-a) = \|x-a\| R_H(x,a)$$
approaches 0 as x approaches a.
Let $y = F(x)$ and $b = F(a)$.
Using the differentiability of F at a, we get
$$y - b = F(x) - F(a) = DF(a) \cdot (x-a) + \|x-a\| R_F(x,a).$$
Using the differentiability of G at b, we get
$$H(x) - H(a) = G(y) - G(b) = DG(b) \cdot (y-b) + \|y-b\| R_G(y,b).$$

Chain Rule for Composition of Mappings (Cont'd)

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

$$H(x) - H(a) = DG(a) \cdot DF(a) \cdot (x - a) \\ + \|x - a\| \left\{ DG(F(a)) \cdot R_F(x, a) + \tfrac{\|F(x) - F(a)\|}{\|x - a\|} R_G(F(x), F(a)) \right\}.$$

Using the continuity of F, by differentiability, and the properties of $R_F(x, a)$ and $R_G(y, b)$, we see that as $x \to a$ the expression

$$R_H(x, a) = \left\{ DG(F(a)) \cdot R_F(x, a) + \frac{\|F(x) - F(a)\|}{\|x - a\|} R_G(F(x), F(a)) \right\}$$

goes to zero.

This completes the proof.

The Class of the Composition

Corollary

If F and G are of class C^r (or smooth) on U and V, respectively, then $H = G \circ F$ is of class C^r (or smooth) on u.

- We prove only the statement for C^1.

  The proof of the general case uses mathematical induction.

  If F and G are C^1, then they are certainly differentiable.

  Moreover, DF and DG are continuous functions on U and V.

  Since F is C^1, it is continuous.

  So $DG(F(x))$ is continuous on U.

The Class of the Composition (Cont'd)

- The entries in the $m \times p$ product matrix are polynomials in the entries of the factors.

  So the product is a continuous, in fact C^{ω}, mapping of $\mathbb{R}^{mn} \times \mathbb{R}^{np}$.

  Thus, the chain rule formula gives $DH(x)$ as a composite of functions which are at least continuous.

  So it must be continuous.

  This is equivalent to its entries being continuous.

  I.e., the coordinate functions of H, thus H itself, are of class C^1.

Subsection 3

The Space of Tangent Vectors at a Point of $\mathbb{R}^n$

Tangent Space at a Point in $\boldsymbol{E}^3$

- We shall presently restrict our attention to $\mathbb{R}^n$.
- But we first consider $\boldsymbol{E}^n$, or $\boldsymbol{E}^3$ at least, for the sake of intuition.
- Our purpose is to attach to each point a of $\mathbb{R}^n$ an n-dimensional vector space $T_a(\mathbb{R}^n)$.
- We know how to do this in Euclidean space.
- If $a \in \boldsymbol{E}^3$, we let $T_a(\boldsymbol{E}^3)$ be the vector space whose elements are directed line segments X_a with a as initial point.
  - These are added using the Parallelogram Law;
  - $-X_a$ is the oppositely directed segment;
  - 0 is the segment consisting of the point a alone.
- We have supposed that a unit of length was chosen in $\boldsymbol{E}^3$.
- We may denote by $\|X_a\|$ the length of the segment.

The Tangent Space at a Point in $\boldsymbol{E}^3$ (Cont'd)

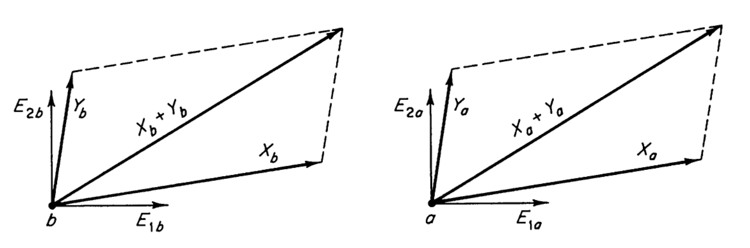
- Multiplication by positive (negative) real numbers leaves the direction unchanged (reversed) and multiplies the length by the absolute value of the number.
- This does indeed give a vector space of dimension 3 over $\mathbb{R}$.
- Thus we attach to each point of $\boldsymbol{E}^3$ a three-dimensional vector space called the **tangent space** at that point.

Euclidean Space and Naturality

- We shall ultimately attach vector spaces at each point of more complicated spaces, namely manifolds.
- There is a unique feature of the tangent spaces of Euclidean space which is not shared by the tangent spaces at points of manifolds.
- The tangent spaces at any two points of Euclidean space are naturally isomorphic.
- This means that there is an isomorphism determined in some unique fashion by the geometry of the space and not chosen by us.
- Without the restriction of naturality the statement would be trivial, since any two vector spaces of the same dimension over the real numbers are isomorphic.
- In general, there is no unique isomorphism singled out.
- We must, rather, choose one arbitrarily from a very large collection.

Naturality Explained

- If a, b are points of $\boldsymbol{E}^3$, then there is exactly one translation of the space taking a to b.
- This translation moves each line segment issuing from a to a line segment from b.
- Thus it carries $T_a(\boldsymbol{E}^3)$ to $T_b(\boldsymbol{E}^3)$.
- Since parallelograms go to congruent parallelograms and lengths are preserved, this correspondence is an isomorphism.
- It is uniquely determined by the geometry.



Naturality Explained (Cont'd)

- Choose a fixed point a as origin.
- Choose at a three linearly independent vectors E_{1a}, E_{2a}, E_{3a}.
- E.g., E_{1a}, E_{2a}, E_{3a} could be three mutually perpendicular unit vectors.
- This automatically determines a basis of $T_a(\boldsymbol{E}^3)$.
- It also determines (by parallel translation) a basis of $T_b(\boldsymbol{E}^3)$, for every $b \in \boldsymbol{E}^3$.
- We used geometric intuition, but we have not really proved our claims.
- We turn to $\mathbb{R}^n$ to be more precise and rigorous.

Tangent Spaces in $\mathbb{R}^n$

- Let $a = (a^1, \ldots, a^n)$ be any point of $\mathbb{R}^n$.
- We define $T_a(\mathbb{R}^n)$, the **tangent (vector) space attached to** a, as follows.
- As a set it consists of all pairs of points (a, x), or $\overrightarrow{ax}$, with
$$a = (a^1, \ldots, a^n) \quad \text{and} \quad x = (x^1, \ldots, x^n)$$
corresponding to initial and terminal points of a segment.
- We also denote such a pair by X_a using upper case letters for vectors.
- We establish a one-to-one correspondence $\varphi_a : T_a(\mathbb{R}^n) \to \boldsymbol{V}^n$ between this set and the vector space of n-tuples of real numbers.
- If $X_a = \overrightarrow{ax}$, then
$$\varphi_a(X_a) = (x^1 - a^1, \ldots, x^n - a^n).$$

Tangent Spaces in $\mathbb{R}^n$ (Cont'd)

- Finally the vector space operations (addition and multiplication by scalars) are defined in the one way possible so that φ_a is an isomorphism.
- This requires that

$$\begin{array}{rcl} X_a + Y_a & = & \varphi_a^{-1}(\varphi_a(X_a) + \varphi_a(Y_a)), \\ \alpha X_a & = & \varphi_a^{-1}(\alpha \varphi_a(X_a)), \quad \alpha \in \mathbb{R}, \end{array}$$

  the right-hand side being used to define the operations on the left.
- The guiding principle is that $\mathbb{R}^n$ and $\boldsymbol{E}^n$ may be identified if we choose rectangular Cartesian coordinates in $\boldsymbol{E}^n$.
- This is equivalent to choosing an origin 0 and n mutually orthogonal unit vectors there, $(E_1)_0, \ldots, (E_n)_0$, lying on each (positive) coordinate axis, like $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ in the usual model for $\boldsymbol{E}^3$.

Tangent Spaces in $\mathbb{R}^n$ (Cont'd)

- Then vectors at any point a are uniquely determined by their components relative to the basis $E_{1a}, \ldots, E_{na}$.
- These are given by subtracting from the coordinates of the terminal point of each vector, the coordinates of its initial point a.
- Note that $\boldsymbol{V}^n$ has a canonical basis

$$\boldsymbol{e}^1 = (1, 0, \ldots, 0), \ldots, \boldsymbol{e}^n = (0, \ldots, 1).$$

- This gives at each $a \in \mathbb{R}^n$ a natural or canonical basis of $T_a(\mathbb{R}^n)$,

$$E_{1a} = \varphi^{-1}(\boldsymbol{e}^1), \ldots, E_{na} = \varphi^{-1}(\boldsymbol{e}^n).$$

Example

- Consider again the canonical isomorphism given by translation in the case of $\boldsymbol{E}^n$.
- It is
$$\varphi_b^{-1} \circ \varphi_a : T_a(\mathbb{R}^n) \to T_b(\mathbb{R}^n).$$
- We have $X_a = \overrightarrow{ax}$ corresponds to $Y_b = \overrightarrow{by}$ if and only if
$$x^i - a^i = y^i - b^i, \quad i = 1, 2, \ldots, n.$$

Remarks

- We never identify the tangent spaces into a single vector space, as is often done in discussions of vectors on Euclidean space.
- That is, we never equate vectors with different initial points.
- In particular, we cannot add a vector in $T_a(\mathbb{R}^n)$ and one in $T_b(\mathbb{R}^n)$, where $a \neq b$.
- The reason becomes apparent when we learn how to attach a tangent space $T_p(M)$ to each point p of a manifold in general.
- Then, we have nothing corresponding to the natural isomorphisms of $T_a(\boldsymbol{E}^3)$ and $T_b(\boldsymbol{E}^3)$ given by the translations of $\boldsymbol{E}^3$.
- Our method of defining $T_a(\mathbb{R}^n)$ at each a, which depended on such an isomorphism, is not suitable for generalization in its present form.

C^1 Curves Through a Point

- Consider a C^1 curve in $\mathbb{R}^n$,

$$x(t), \quad -\varepsilon < t < \varepsilon.$$

- Suppose $x(t)$ passes through $a \in \mathbb{R}^n$ when $t = 0$.
- That is, assume $x(t) = (x^1(t), \ldots, x^n(t))$, where:
  - $x^i(t)$ is C^1;
  - $x^i(0) = a^i$, $i = 1, \ldots, n$.
- Let

$$I_\varepsilon = \{t \in \mathbb{R} : |t| < \varepsilon\}.$$

- Then each such curve is a C^1 map of

$$I_\varepsilon \to \mathbb{R}^n,$$

where $\varepsilon > 0$ and may vary from curve to curve.

Equivalence and Tangent Spaces

- Consider two C^1 curves $x(t)$ and $y(t)$ through a.
- The two curves are **equivalent**, written

  $$x(t) \sim y(t),$$

  if at $t = 0$, the derivatives with respect to t of their coordinate functions are equal,

  $$\dot{x}^i(0) = \dot{y}^i(0), \quad i = 1, \ldots, n.$$

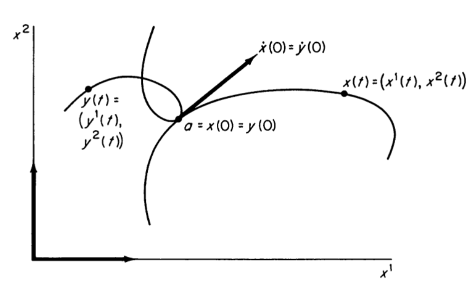
- Let $[x(t)]$ denote the equivalence class of $x(t)$.
- To each $[x(t)]$ corresponds an n-tuple of numbers

  $$\dot{x}(0) = (\dot{x}^1(0), \ldots, \dot{x}^n(0)),$$

  that is, an element of $\boldsymbol{V}^n$.
- Using this map we obtain a vector space structure on the collection of equivalence classes which we denote by $T_a(\mathbb{R}^n)$.

Intuition Behind Definition

- Suppose we use the identification of $\mathbb{R}^n$ with $\boldsymbol{E}^n$ plus a rectangular Cartesian coordinate system.
- Then $\dot{x}^i(0)$ is the ith component of the velocity vector of the particle whose motion is given by $x(t) = (x^1(t), \ldots, x^n(t))$ at the instant it passes through a.
- Two curves are equivalent if they represent two motions with the same velocity at this instant.



Subsection 4

Another Definition of $T_a(\mathbb{R}^n)$

The Space $C^\infty(a)$

- Let us denote by $C^\infty(a)$ the collection of all C^∞ functions whose domain includes a.
- Here we are only interested in the derivatives at a.
- So in $C^\infty(a)$, we identify functions which agree on an open set containing a.

Directional Derivative

- Let a vector X_a of $T_a(\mathbb{R}^n)$ be expressed in the canonical basis as
$$X_a = \sum_{i=1}^{n} \alpha^i E_{ia}.$$
- We define the **directional derivative** Δf of f at a in the "direction of X_a" by
$$\Delta f = \sum_{i=1}^{n} \alpha^i \frac{\partial f}{\partial x^i}, \quad \frac{\partial f}{\partial x^i} \text{ evaluated at } a = (a^1, \ldots, a^n).$$
- This is a slight extension of the usual definition in that we do not require X_a to be a unit vector.
- Since Δf depends on f, a, and X_a we shall write it as $X_a^* f$.
- Thus
$$X_a^* f = \sum_{i=1}^{n} \alpha^i \left(\frac{\partial f}{\partial x^i} \right)_a.$$

The Mapping X_a^*

- We may take the directional derivative in the "direction of X_a" of any C^∞ function defined in a neighborhood of a.
- We get a mapping assigning to each $f \in C^\infty(a)$ a real number,

$$X_a^* : \quad \begin{array}{rcl} C^\infty(a) & \to & \mathbb{R}; \\ f & \to & X_a^* f. \end{array}$$

- It is reasonable to denote this mapping by

$$X_a^* = \sum_{i=1}^n \alpha_i \frac{\partial}{\partial x^i},$$

  where the derivatives are to be evaluated at a.
- We remark that $X_a^* x^i = \alpha^i$, $i = 1, \ldots, n$.
- So the vector X_a is completely determined if the value of X_a^* on every C^∞ function at a, or even on $f^i(x) = x^i$, is known.

The Algebra $C^\infty(a)$

- We have agreed not to distinguish between C^∞ functions f, g in $C^\infty(a)$ if they agree on some open set containing a.
- Two functions of $C^\infty(a)$ may be added or multiplied to give another element of $C^\infty(a)$, whose domain is the intersection of their domains.
- If $\alpha \in \mathbb{R}$, then αf is a C^∞ function with the same domain as f.
- So $f \in C^\infty(a)$ implies $\alpha f \in C^\infty(a)$.
- The same result would be obtained by multiplying f by a C^∞ function whose value is α on some open set about a.
- Thus, $C^\infty(a)$ is an algebra over $\mathbb{R}$ containing $\mathbb{R}$ as a subalgebra.

Linearity and the Leibniz Rule

- Recall the fundamental properties of derivatives.
- If α, β are real numbers and f, g are C^∞ functions, defined in open sets containing a, then we have:
  (i) $X_a^*(\alpha f + \beta g) = \alpha(X_a^* f) + \beta(X_a^* g)$ (**linearity**);
  (ii) $X_a^*(fg) = (X_a^* f)g(a) + f(a)(X_a^* g)$ (**Leibniz rule**).

Space of Derivations

- Let $\mathcal{D}(a)$ denote all mappings of $C^\infty(a)$ to $\mathbb{R}$ satisfying linearity and the Leibniz Rule.
- We may call the elements of $\mathcal{D}(a)$ "**derivations**" on $C^\infty(a)$ into $\mathbb{R}$.
- Suppose $D_1, D_2 : C^\infty(a) \to \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}$.
- Define
$$(\alpha D_1 + \beta D_2)f = \alpha(D_1 f) + \beta(D_2 f),$$
where the operations on the right are in $\mathbb{R}$.
- This defines in $\mathcal{D}(a)$ both addition and multiplication by real numbers α, β.

Space of Derivations (Cont'd)

- We see that $\mathcal{D}(a)$ is a vector space over $\mathbb{R}$.
- One must check that the vector space axioms are indeed satisfied.
- In particular, it must be verified that:
  - If $D \in \mathcal{D}(a)$, then $\alpha D \in \mathcal{D}(a)$;
  - If $D_1, D_2 \in \mathcal{D}(a)$, then so also are $D_1 + D_2$.
- This means:
  - Checking the linearity of $\alpha D : C^\infty(a) \to \mathbb{R}$ and $D_1 + D_2 : C^\infty(a) \to \mathbb{R}$;
  - Checking that the Leibniz rule is satisfied.

Space of Derivations (Cont'd)

- We do this for γD only.
- Suppose then $\gamma, \alpha, \beta \in \mathbb{R}$, $D \in \mathcal{D}(a)$, and $f, g \in C^\infty(a)$.
- Then

$$\begin{array}{rcll}(\gamma D)(\alpha f + \beta g) & = & \gamma[D(\alpha f + \beta g)] & \text{(by definition of } \gamma D\text{)} \\ & = & \gamma[\alpha(Df) + \beta(Dg)] & \text{(by Property (i))} \\ & = & \gamma\alpha(Df) + \gamma\beta(Dg) & \text{(by the dist. law of } \mathbb{R}\text{)} \\ & = & \alpha(\gamma D)f + \beta(\gamma D)g & \text{(by our definition of } \gamma D\text{).}\end{array}$$

- It follows that the map $\gamma D : C^\infty \to \mathbb{R}$ is linear.

Space of Derivations (Cont'd)

- γD also satisfies the Leibniz rule for differentiation of products:

$$\begin{aligned}(\gamma D)(fg) &= \gamma[D(fg)] && \text{(by definition of } \gamma D\text{)}\\ &= \gamma[(Df)g(a)+f(a)(Dg)] && \text{(by Property (ii))}\\ &= \gamma(Df)g(a)+f(a)\gamma(Dg) && \text{(real numbers)}\\ &= ((\gamma D)f)g(a)+f(a)((\gamma D)g) && \text{(by definition of } \gamma D\text{).}\end{aligned}$$

- A similar verification shows that $D_1 + D_2$ is a derivation into $\mathbb{R}$.

A Vector Space Isomorphism

Theorem

The vector space $T_a(\mathbb{R}^n)$ is isomorphic to the vector space $\mathcal{D}(a)$ of all derivations of $C^\infty(a)$ into $\mathbb{R}$. This isomorphism is given by making each X_a correspond to the directional derivative X_a^* in the direction of X_a.

- Consider the mapping

$$T_a(\mathbb{R}^n) \to \mathcal{D}(a); \quad X_a \mapsto X_a^*.$$

  This mapping is one-to-one.
  Suppose $X_a^* = Y_a^*$. Then $X_a^* f = Y_a^* f$, for every $f \in C^\infty(a)$.
  The ith component of X_a relative to the natural basis is just $X_a^* x^i$.
  So we get

$$X_a = \sum_{i=1}^n (X_a^* x^i) E_{ia} = \sum_{i=1}^n (Y_a^* x^i) E_{ia} = Y_a.$$

A Vector Space Isomorphism (Cont'd)

- It is easy to see that this mapping is linear.

  Suppose $Z_a = \alpha X_a + \beta Y_a \in T_a(\mathbb{R}^n)$.

  For the directional derivatives, we have, for any $f \in C^\infty(a)$,

  $$Z_a^* f = \alpha(X_a^* f) + \beta(Y_a^* f).$$

  If interpreted in terms of the operations in $\mathcal{D}(a)$, this means exactly that the mapping $T_a(\mathbb{R}^n) \to \mathcal{D}(a)$ is linear.

  In summary, $X_a \to X_a^*$ defines an isomorphism of the vector space $T_a(\mathbb{R}^n)$ into the vector space $\mathcal{D}(a)$, which allows us to identify $T_a(\mathbb{R}^n)$ with a subspace of $\mathcal{D}(a)$.

  It only remains to show that $X_a \to X_a^*$ is a map onto $\mathcal{D}(a)$.

  That is, every derivation of $C^\infty(a)$ into $\mathbb{R}$ is a directional derivative.

  This will result from two lemmas.

Lemma 1: Constant Functions

Lemma

Let D be an arbitrary element of $\mathcal{D}(a)$. Then D is zero on any function $f \in C^\infty(a)$ which is constant in a neighborhood of a.

- The map D is linear. So it suffices to show $D1 = 0$, where 1 denotes the constant function of value 1.
  We have
  $$D1 = D(1 \cdot 1) = (D1)1 + 1(D1) = D1 + D1 = 2D1.$$
  So $D1 = 0$.
- We have identified $\mathbb{R}$ with the subalgebra of $C^\infty(a)$ of functions whose value is constant in some open set (possibly $\mathbb{R}^n$) containing a.
- So, in interpreting the equalities, multiplying $f \in C^\infty(a)$ by a real number α gives exactly the same result as multiplying by the C^∞ function whose value is α in some open set containing a.

Lemma 2: Decomposing a Function

Lemma

Let $f(x^1, \ldots, x^n)$ be defined and C^∞ on some open set U. If $a \in U$, then there is a spherical neighborhood B of a, $B \subseteq U$, and C^∞-functions $g^1, \ldots, g^n$ defined on B such that:

(i) $g^i(a) = \left(\frac{\partial f}{\partial x^i}\right)_{x=a}$;

(ii) $f(x^1, \ldots, x^n) = f(a) + \sum_{i=1}^n (x^i - a^i) g^i(x)$.

- Let $B \subseteq U$ be a spherical neighborhood of a.

  Note that, for $x \in B$,

$$f(x) = f(a) + \int_0^1 \frac{\partial}{\partial t} f(a + t(x - a)) dt.$$

Lemma 2: Decomposing a Function (Cont'd)

- Hence,

$$f(x) = f(a) + \sum_{i=1}^{n} (x^i - a^i) \int_0^1 \left[\frac{\partial f}{\partial x^i}\right]_{a+t(x-a)} dt.$$

Let

$$g^i(x) = \int_0^1 \left[\frac{\partial f}{\partial x^i}\right]_{a+t(x-a)} dt, \quad i = 1, \ldots, n.$$

These are C^∞-functions and satisfy the two conditions.

Proof of the Isomorphism Theorem

- Suppose D is any derivation on $C^\infty(a)$.
  We wish to show that, given $D \in \mathcal{D}(a)$, there is a vector $X_a \in T_a(\mathbb{R}^n)$ such that, for all $f \in C^\infty(a)$, $X_a^* f = Df$.
  If this is the case, then $X_a^* = D$.
  So every derivation of $C^\infty(a)$ into $\mathbb{R}$ is a directional derivative.
  Thus, the map $X_a \mapsto X_a^*$ of $T_a(\mathbb{R}^n)$ to $\mathcal{D}(a)$ is an isomorphism onto.
  Suppose
  $$h^i(x^1, \ldots, x^n) = x^i.$$
  Denote by α^i the value of Dh^i, that is, $\alpha^i = Dh^i$.
  Consider $X_a = \sum_{i=1}^n \alpha^i E_{ia}$.
  As an operator on $C^\infty(a)$, it gives
  $$X_a^* f = \sum_{i=1}^n \alpha^i \left(\frac{\partial f}{\partial x^i} \right)_a.$$

Proof of the Isomorphism Theorem (Cont'd)

- By Lemma 2, on some $B_\varepsilon(a)$ in the domain of f,

$$f(x) = f(a) + \sum_{i=1}^{n} (x^i - a^i) g^i(x).$$

Restricting to $B_\varepsilon(a)$ and using the properties of D, we may write

$$Df = D(f(a)) + \sum_{i=1}^{n} \{(D(x^i - a^i)) g^i(a) + 0 \cdot Dg^i\}.$$

By Lemma 1, $D(f(a)) = 0$ and $D(x^i - a^i) = Dx^i = \alpha^i$.
By Lemma 2, $g^i(a) = \left(\frac{\partial f}{\partial x^i}\right)_a$.
Therefore,

$$Df = \sum_{i=1}^{n} \alpha^i \left(\frac{\partial f}{\partial x^i}\right)_a = X_a^* f.$$

Since f is an arbitrary element of $C^\infty(a)$, we have $D = X_a^*$.

Identification and Notation

- The theorem allows us to identify the vector space $T_a(\mathbb{R}^n)$ with the space $\mathcal{D}(a)$ of linear operators on functions of $C^\infty(a)$ into $\mathbb{R}$ which satisfy the product rule of Leibniz, that is, the "**derivations into** $\mathbb{R}$".
- Note that, under this identification, the following are identified:
  - The canonical basis vectors $E_{1a}, \ldots, E_{na}$ of $T_a(\mathbb{R}^n)$;
  - The directional derivatives $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ (evaluated at $x = a$) in the directions of the coordinate axes.
- That is, we have

$$E_{ia} \to E_{ia}^* f = \left(\frac{\partial f}{\partial x^i}\right)_a.$$

- We will make this identification from now on for vectors in $T_a(\mathbb{R}^n)$.
- For this reason we will drop the asterisk * which distinguishes the vector X_a, as a segment or point pair, from the directional derivative, i.e.,

$$X_a^* f \text{ will be written } X_a f.$$

Vectors and Linear Operators

- In $\mathbb{R}^n$ we may use either E_{ia} or $\frac{\partial}{\partial x^i}$ to denote the unit vector parallel to the ith coordinate axis.
- This characterization of $T_a(\mathbb{R}^n)$ requires C^∞ functions.
- Although $C^r(a)$ is an algebra, it is known to have other derivations than directional derivatives.
- Now we have a dual view of tangent spaces.
  - We shall rely on Euclidean space for our geometric intuition of the space of tangent vectors at a point.
  - However, in formal definitions and proofs, we view a vector at a point as a linear operator, satisfying the product rule for derivatives, on the C^∞ functions at the point.

Subsection 5

Vector Fields on Open Subsets of $\mathbb{R}^n$

Vector Fields

- A **vector field** on an open subset $U \subseteq \mathbb{R}^n$ is a function which assigns to each point $p \in U$ a vector $X_p \in T_p(\mathbb{R}^n)$.
- A similar definition applies to Euclidean space $\boldsymbol{E}^n$.

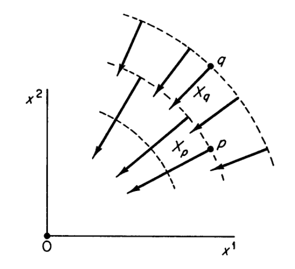
Example (**Gravitational Field**):

Let an object of mass μ be located at a point 0.

Then to each point p in $U = \boldsymbol{E}^n - \{0\}$, there is assigned a vector denoting the force of attraction on a particle of unit mass placed at the point.

This vector is represented by an arrow from p (as initial point) directed toward 0 and having length $\frac{k\mu}{r^2}$, where:

- r denotes the distance $d(0, p)$;
- k a constant determined by the units.



Example (Cont'd)

- We now introduce Cartesian coordinates with 0 as origin.
- Consider the point p with coordinates (x^1, x^2, x^3).
- The components of X_p in the canonical basis are

$$\frac{-x^1}{r^3}, \frac{-x^2}{r^3}, \frac{-x^3}{r^3}, \text{ with } r = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}.$$

- That is

$$\begin{array}{rcl} X_p & = & \frac{-1}{r^3}(x^1 E_{1p} + x^2 E_{2p} + x^3 E_{3p}) \\ & = & \frac{-1}{r^3}\left(x^1 \frac{\partial}{\partial x^1} + x^2 \frac{\partial}{\partial x^2} + x^3 \frac{\partial}{\partial x^3}\right). \end{array}$$

- We note that the components of X_p are C^∞ functions of the coordinates.

Smooth Vector Fields

- We shall say that a vector field on $\mathbb{R}^n$ is C^∞ or **smooth** if its components relative to the canonical basis are C^∞ functions on U.
- All vector fields considered will be assumed to have this property.
- However, it is quite possible to define continuous, C^1, and so on, vector fields also.
- When dealing with vector fields, as with functions, the independent variable will be omitted from the notation.
- Thus we write X rather than X_p, just as we customarily use f rather than $f(p)$ for a function.
- Then X_p is the value at p of X, that is, the vector of the field which is attached to p, and which lies in $T_p(\mathbb{R}^n)$.

Field of Frames

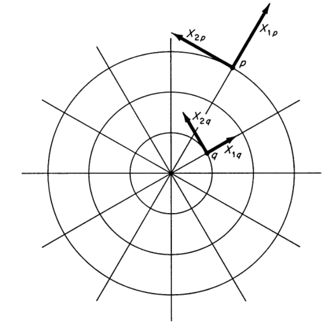
- Further examples of vector fields are given, for each $i = 1, \ldots, n$, by the fields
$$E_i = \frac{\partial}{\partial x^i}.$$
- These assign to every $p \in \mathbb{R}^n$ the naturally defined basis vector E_i at that point.
- The vector fields $E_1, \ldots, E_n$ are independent, even orthogonal unit vectors, at each point p.
- So they form a basis of $T_p(\mathbb{R}^n)$ at p.
- Such a set of fields is called a **field of frames**.

Example

- Consider the vector fields X_1, X_2 on $U = \mathbb{R}^2 - \{0\}$ defined by

$$\begin{aligned} X_1 &= x^1 E_1 + x^2 E_2; \\ X_2 &= x^2 E_1 - x^1 E_2. \end{aligned}$$

- They also define a field of frames.
- Geometrically:
  - X_{1p} is a vector along a ray from 0 to p;
  - X_{2p} is a vector perpendicular to it, that is, tangent to the circle through p with center at 0.



The C^∞-Function Xf

- Let X be a C^∞-vector field on U.
- Let f be a C^∞ function on U.
- Then Xf is the C^∞-function on U defined by

$$(Xf)(p) = X_p f.$$

- Suppose the components of X are the functions $\alpha^1(p), \ldots, \alpha^n(p)$.
- Then $X = \sum_{i=1}^n \alpha^i E_i$.
- Moreover,

$$(Xf)(p) = \sum_{i=1}^n \alpha^i(p) \left(\frac{\partial f}{\partial x^i}\right)_p.$$

- By hypothesis, $\alpha^i(p) \in C^\infty(U)$ and $\frac{\partial f}{\partial x^i} \in C^\infty(U)$.
- So the right-hand side shows that Xf is a C^∞ function of p on U.
- Thus $f \mapsto Xf$ maps $C^\infty(U) \mapsto C^\infty(U)$.

X as a Derivation on $C^\infty(U)$

- $C^\infty(U)$ is an algebra over $\mathbb{R}$ with unit, where:
  - $\mathbb{R}$ is identified with the constant functions;
  - In particular, the constant function 1 plays the role of the unit.
- We ask whether X is a linear map of $C^\infty(U)$ to $C^\infty(U)$.
- More generally, we ask whether it is a derivation.
- This is the case, since

$$\begin{aligned}[X(\alpha f+\beta g)](p) &= X_p(\alpha f+\beta g)\\ &= \alpha(X_p f)+\beta(X_p g)\\ &= \alpha(Xf)(p)+\beta(Xg)(p).\end{aligned}$$

X as a Derivation on $C^\infty(U)$ (Cont'd)

- Moreover

$$\begin{aligned}[X(fg)](p) &= (X_p f)g(p) + f(p)(X_p g) \\ &= [(Xf)(p)]g(p) + f(p)[(Xg)(p)].\end{aligned}$$

- Since the functions on the right and left agree for each $p \in U$, they are equal as functions.
- Thus $X : C^\infty(U) \to C^\infty(U)$ is a derivation which maps $C^\infty(U)$ into itself, a slight variation from the previous case.
- Note that this, in fact, is the customary use of the term "derivation" of an algebra.
- If A is an algebra over $\mathbb{R}$, then a **derivation** is a map $D : A \to A$ which is linear and satisfies the product rule of Leibniz.
- For example, $\frac{\partial}{\partial x}$ is a derivation on the algebra of all polynomials in two variables x and y.

A Separation Theorem

- We close by proving an important property of C^∞-functions.
- Together with the corollary given, it is used very often in discussions of vector fields.
- It is a "separation theorem" and contrasts strongly the behavior of C^∞ and C^ω functions on $\mathbb{R}^n$.

Theorem

Let $F \subseteq \mathbb{R}^n$ be a closed set and $K \subseteq \mathbb{R}^n$ compact, such that

$$F \cap K = \emptyset.$$

Then, there is a C^∞ function $\sigma(x)$, whose domain is all of $\mathbb{R}^n$ and whose range of values is the closed interval $[0,1]$, such that:

- $\sigma(x) \equiv 1$ on K;
- $\sigma(x) \equiv 0$ on F.

Proof of the Theorem (Step (a))

- We prove the theorem in two steps.

(a) Let $B_\varepsilon(a)$ be an open ball of center a and radius ε.

We show that there is a C^∞ function $g(x)$ on $\mathbb{R}^n$, such that:

- g is positive on $B_\varepsilon(a)$;
- g is identically 1 on $\overline{B}_{\varepsilon/2}$;
- g is 0 outside $B_\varepsilon(a)$.

Consider the function $h(t)$ defined by

$$h(t) = \begin{cases} 0, & \text{if } t \leq 0, \\ e^{-1/t}, & \text{if } t > 0. \end{cases}$$

By calculus, all of its derivatives exist and are zero at $t = 0$.

Moreover, it is analytic for other values of t.

We conclude that g is C^∞.

Proof of the Theorem (Step (a) Cont'd)

- We now define

$$\overline{g}(x) = \frac{h(\varepsilon - \|x\|)}{h(\varepsilon - \|x\|) + h(\|x\| - \frac{1}{2}\varepsilon)}.$$

  Note that, in at least one of $\varepsilon - \|x\|$ or $\|x\| - \frac{1}{2}\varepsilon$, h is positive.

  As the denominator is never zero, $\overline{g}$ is a C^∞ function.

  When $\|x\| \geq \varepsilon$, the numerator is zero.

  Otherwise it is positive.

  When $0 \leq \|x\| \leq \frac{1}{2}\varepsilon$, the value of $\overline{g}(x)$ is identically 1.

  So $\overline{g}(x)$ is C^∞ satisfies the following:
  - It vanishes outside $B_\varepsilon(0)$;
  - It is positive on the interior of $B_\varepsilon(0)$;
  - $\overline{g}(x) = 1$ for $x \in \overline{B}_{\varepsilon/2}(0)$.

  Hence $g(x) = \overline{g}(x - a)$ has the desired properties.

Proof of the Theorem (Step b)

(b) By hypothesis, K is compact and $K \cap F = \emptyset$.

So we can find a finite collection of n-balls $B_\varepsilon(a_i)$, $i = 1, \ldots, k$, such that

$$\bigcup_{i=1}^{k} B_{\varepsilon/2}(a_i) \supseteq K.$$

For $B_\varepsilon(a_i)$, let $g_i(x)$ have the properties of Step (a).

Define $\sigma(x)$ by

$$\sigma(x) = 1 - \prod_{i=1}^{k}(1 - g_i).$$

On each $x \in K$ at least one g_i has the value 1. So $\sigma \equiv 1$ on K.

Outside $\bigcup_{i=1}^{k} B_\varepsilon(a_i)$ each g_i vanishes. So $\sigma(x) = 0$.

Since F lies outside this union, $\sigma \equiv 0$ on F.

Function Vanishing Outside a Neighborhood

Corollary

Let $f(x^1, \ldots, x^n)$ be C^∞ on an open set $U \subseteq \mathbb{R}^n$ and let $a \in U$.
Then there exists an open set $V \subseteq U$, which is a neighborhood of a, and a C^∞ function $f^*(x^1, \ldots, x^n)$ defined on all of $\mathbb{R}^n$, such that:

- $f^*(x) = f(x)$, for all $x \in V$;
- $f^*(x) = 0$, for x outside U.

- Choose any neighborhoods V_1, V_2 of a, such that:
  - $\overline{V}_1 \subseteq V_2$;
  - $\overline{V}_2 \subseteq U$;
  - $\overline{V}_1$ is compact.

Function Vanishing Outside a Neighborhood (Cont'd)

- Let $\overline{V}_1 = K$ and $F = \mathbb{R}^n - V_2$ in the theorem.

  Then take $\sigma(x)$, a C^∞ function, such that:
  - $\sigma(x)$ is 1 on $\overline{V}_1$;
  - $\sigma(x)$ is 0 outside V_2, i.e., on F.

  Define

  $$f^*(x) = \begin{cases} \sigma(x)f(x), & \text{if } x \in U, \\ 0, & \text{if } x \in \mathbb{R}^n - V_2. \end{cases}$$

  f^*, thus defined, is:
  - C^∞ on U, where it is equal to σf;
  - C^∞ on $\mathbb{R}^n - V_2$, where it is identically zero.

  Further, on the (open) intersection $U - \overline{V}_2$ of these two sets, both definitions agree.

  It follows that f^* is C^∞ on $\mathbb{R}^n$ and has the properties needed.

Subsection 6

The Inverse Function Theorem

Diffeomorphisms

- Suppose that $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^n$ are open sets.
- We then shall say that a mapping $F : U \to V$ is a C^r-**diffeomorphism** if:
  - (i) F is a homeomorphism;
  - (ii) Both F and F^{-1} are of class C^r, $r \geq 1$.
- When $r = \infty$, we simply say **diffeomorphism**.
- We need to require both F and F^{-1} to be of class C^r, because we wish the relation to be symmetric.

Diffeomorphisms and Inverses

- Even when F is a homeomorphism, the differentiability of F^{-1} is not a consequence of that of F.

  Example: Let $U = \mathbb{R}$ and $V = \mathbb{R}$.

  Consider
  $$F : t \mapsto s = t^3.$$

  This is a homeomorphism.

  Moreover, F is analytic.

  Now consider
  $$F^{-1} : s \mapsto t = s^{1/3}.$$

  It has no derivative at $s = 0$.

  So it is not C^1 on V.

Redundancies

- The definition contains some redundant requirements.
  - First, it would not be possible to have a diffeomorphism between open subspaces of Euclidean spaces of different dimensions.
    Brouwer's Invariance of Domain Theorem asserts that even a homeomorphism between open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, $m \neq n$, is impossible.
  - Second, in the example given above the derivative of F vanishes at $t = 0$, thus behaving atypically.
    - If it vanished everywhere, then F could not be a homeomorphism of $\mathbb{R}$ to $\mathbb{R}$.
    - If it vanished at no point, then F^{-1} would indeed be a differentiable map.

C^1-Homeomorphisms and Derivatives

- Suppose $F : U \to V$ is a homeomorphism.
- Assume, also, that both F and F^{-1} are of class C^1 at least.
- Then $DF(x)$ is nonsingular, i.e., has nonvanishing determinant at each $x \in U$.
- We have $F^{-1} \circ F = I$, the identity map of U to U.
- By the chain rule

$$DI(x) = DF^{-1}(F(x)) \cdot DF(x).$$

- Now $DI(x)$ is just the identity matrix for every $x \in U$.
- So $DF(x)$ is nonsingular and its determinant is never zero.
- This includes the assertion that, if $F : \mathbb{R} \to \mathbb{R}$ is a diffeomorphism, then its derivative can never be zero.

Example

- Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be the translation taking

$$a = (a^1, \ldots, a^n) \mapsto b = (b^1, \ldots, b^n).$$

- Then F is given by

$$F(x^1, \ldots, x^n) = (x^1 + (b^1 - a^1), \ldots, x^n + (b^n - a^n)),$$

  or $F(x) = x + (b - a)$.
- The coordinate functions $f^i(x) = x^i + (b^i - a^i)$ are analytic, and hence C^∞.
- The translation $G(x) = x + (a - b)$ is F^{-1}.
- This is then also C^∞.
- Since F, F^{-1} are defined and continuous, F is a homeomorphism.
- Thus, F is a diffeomorphism.

Example

- Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation

$$F(x^1, \ldots, x^n) = \left(\sum_{j=1}^n \alpha_j^1 x^j, \ldots, \sum_{j=1}^n \alpha_j^n x^j \right).$$

- Using matrix notation with x as an $n \times 1$ (column matrix) and $A = (\alpha_j^i)$,

$$F(x) = Ax.$$

- Computation shows that $DF(x)$ is the constant matrix A,

$$DF(x) = A.$$

- If $\det A \neq 0$, then A has an inverse B.
- Moreover, F^{-1} is the homogeneous linear transformation

$$G(x) = Bx.$$

Example (Cont'd)

- Suppose, on the other hand, that $\det A = 0$,.
- Then F is not one-to-one.
- In fact it maps at least a line through the origin onto the single point

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

- Obviously F is analytic and C^∞ in either case.
- So F is a diffeomorphism if and only if $DF(x) = A$ is nonsingular.

Equivalence Property of Diffeomorphism

- Diffeomorphism is an equivalence relation on the open subsets of $\mathbb{R}^n$.
- Symmetry and reflexivity are part of the definition.
- The following lemma, stated without proof, gives the transitivity property.

Lemma

Let U, V, W be open subsets of $\mathbb{R}^n$.
Let $F : U \to V$, $G : V \to W$ be mappings onto.
Let $H = G \circ F : U \to W$ be their composition.
If any two of these maps is a diffeomorphism, then the third is also.

Contracting Mapping Theorem

Theorem (Contracting Mapping Theorem)

Let M be a complete metric space with metric $d(x, y)$.
Let $T : M \to M$ be a mapping of M into itself.
Assume there is a constant λ, $0 \leq \lambda < 1$, such that for all $x, y \in M$,

$$d(T(x), T(y)) \leq \lambda d(x, y).$$

Then T has a unique fixed point a in M.

- Applying T repeatedly we see that

$$d(T^n(x), T^n(y)) \leq \lambda^n d(x, y).$$

Suppose we choose arbitrarily $x_0 \in M$ and let $x_n = T^n(x_0)$.
We assert that

$$d(x_n, x_{n+m}) \leq \lambda^n K,$$

where $K \geq 0$, a constant independent of n, m.

Contracting Mapping Theorem (Cont'd)

- We have
$$T^{n+m}(x_0) = T^n(T^m(x_0)).$$
We write
$$d(x_n, x_{n+m}) = d(T^n(x_0), T^n(T^m(x_0))) \leq \lambda^n d(x_0, T^m(x_0)).$$
By the triangle inequality
$$\begin{array}{rcl} d(x_0, T^m(x_0)) & \leq & d(x_0, T(x_0)) + d(T(x_0), T^2(x_0)) \\ & & + \cdots + d(T^{m-1}(x_0), T^m(x_0)) \\ & \leq & (1 + \lambda + \lambda^2 + \cdots + \lambda^{m-1}) d(x_0, T(x_0)) \\ & \leq & \frac{1}{1-\lambda} d(x_0, T(x_0)). \end{array}$$
This shows that we may take $K = \frac{1}{1-\lambda} d(x_0, T(x_0))$.
The assertion has been proved.

Contracting Mapping Theorem (Cont'd)

- It follows that $\{x_n\}$ is a Cauchy sequence.

  So it has a limit point a.

  Obviously, $T(x_n) = x_{n+1}$ has the same limit.

  We see that

  $$d(T(a), a) = \lim d(T(x_n), x_n) = \lim d(x_{n+1}, x_n) = 0.$$

  So $T(a) = a$ and a is a fixed point of T.

  Suppose there were two fixed points a, b.

  Then

  $$d(a, b) = d(T(a), T(b)) \leq \lambda d(a, b).$$

  This contradicts the fact that $\lambda < 1$.

Inverse Function Theorem

Theorem (Inverse Function Theorem)

Let W be an open subset of $\mathbb{R}^n$.
Let $F : W \to \mathbb{R}^n$ a C^r mapping, $r = 1, 2, \ldots,$ or ∞.
Suppose $a \in W$ and $DF(a)$ is nonsingular.
Then there exists an open neighborhood U of a in W, such that $V = F(U)$ is open and $F : U \to V$ is a C^r diffeomorphism.
If $x \in U$ and $y = F(x)$, then we have the following formula for the derivatives of F^{-1} at y,

$$DF^{-1}(y) = (DF(x))^{-1},$$

where the term on the right denotes the inverse matrix to $DF(x)$.

- We shall organize the proof in several steps in order to make it somewhat easier to follow.

Inverse Function Theorem (i)

(i) We assume $F(0) = 0$ and $DF(0) = I$, the identity matrix.

This may be done without loss of generality by virtue of the preceding lemma, combined with the use of the preceding examples.

Next we define the mapping G on the same domain by

$$G(x) = x - F(x).$$

Then, obviously:

- $G(0) = 0$;
- $DG(0) = 0$ (the right-hand side is the 0 matrix).

Inverse Function Theorem (ii)

(ii) There exists a real number $r > 0$, such that:

- DF is nonsingular on the closed ball $\overline{B}_{2r}(0) \subseteq W$;
- For $x_1, x_2 \in \overline{B}_r(0)$, we have

$$\begin{array}{rcl} \|G(x_1) - G(x_2)\| & \leq & \frac{1}{2}\|x_1 - x_2\|; \\ \|x_1 - x_2\| & \leq & 2\|F(x_1) - F(x_2)\|. \end{array}$$

To verify these statements we choose r so that:

- $\overline{B}_{2r}(0) \subseteq W$;
- $\det(DF(x))$, which is a continuous function of x and not zero at 0, does not vanish on $\overline{B}_{2r}(0)$;
- The derivatives of the coordinate functions of G, all of which are zero at 0, are bounded in absolute value by $\frac{1}{2n}$ on $\overline{B}_{2r}(0)$.

With these assumptions, $x_1, x_2 \in \overline{B}_r(0)$ implies $\|x_1 - x_2\| < 2r$.

By a previous theorem, we get

$$\|G(x_1) - G(x_2)\| \leq \sqrt{n \cdot n}\frac{1}{2n}\|x_1 - x_2\| = \frac{1}{2}\|x_1 - x_2\|.$$

Inverse Function Theorem (ii)

- The second inequality results by replacing $G(x_i)$ by $x_i - F(x_i)$, $i = 1, 2$, in the first inequality and using a standard property of norms.
  First, by the first inequality, we get
  $$\|x_1 - F(x_1) - x_2 + F(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|.$$
  But, we have
  $$\|x_1 - x_2\| - \|F(x_2) - F(x_1)\| \leq \|(x_1 - x_2) + F(x_2) - F(x_1)\|.$$
  Combining, we get
  $$\begin{aligned} \|x_1 - x_2\| &\leq \|(x_1 - x_2) + F(x_2) - F(x_1)\| + \|F(x_1) - F(x_2)\| \\ &\leq \tfrac{1}{2}\|x_1 - x_2\| + \|F(x_1) - F(x_2)\|. \end{aligned}$$
  This gives $\|x_1 - x_2\| \leq 2\|F(x_1) - F(x_2)\|$.

Inverse Function Theorem (iii)

(iii) If $\|x\| \leq r$, then $\|G(x)\| \leq \frac{r}{2}$, that is,

$$G(\overline{B}_r(0)) \subseteq \overline{B}_{r/2}(0).$$

Moreover for each $y \in \overline{B}_{r/2}(0)$, there is a unique $x \in \overline{B}_r(0)$, such that

$$F(x) = y.$$

The first inequality in (ii) with $x_1 = x$, $x_2 = 0$, yields the first claim.
The second uses the preceding lemma.
Suppose $y \in \overline{B}_{r/2}(0)$ and $x \in \overline{B}_r(0)$.
Then

$$\|y + G(x)\| \leq \|y\| + \|G(x)\| \leq \frac{1}{2}r + \frac{1}{2}r = r.$$

For $y \in \overline{B}_{r/2}(0)$, define a mapping $T_y : \overline{B}_r(0) \to \overline{B}_r(0)$, by

$$T_y(x) = y + G(x).$$

Then $T_y(x) = x$ iff $y = x - G(x)$ iff $F(x) = y$.

Inverse Function Theorem (iii)

- Now rewrite the first inequality in (ii) in the form

$$\|T_y(x_1) - T_y(x_2)\| = \|G(x_1) - G(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|,$$

valid for $x_1, x_2 \in \overline{B}_r(0)$.
So $T_y(x)$ is a contracting map of the compact set $\overline{B}_r(0)$ into itself.
Therefore, by the preceding theorem, there is a unique x, such that

$$y = F(x).$$

This is valid for any $y \in \overline{B}_{r/2}(0)$.
So we see that F^{-1} is defined on that set.
In particular, F being continuous, $U = F^{-1}(B_{r/2}(0))$ is an open subset of $\overline{B}_r(0)$. Let $V = B_{r/2}(0)$.
Since $\overline{B}_r(0) \subseteq W$ we see that

(iv) F is a homeomorphism of the open set $U \subseteq W$ onto the open set V.

Inverse Function Theorem (iv)

(iv) F is a homeomorphism of the open set $U \subseteq W$ onto the open set V.

It remains only to prove continuity of F^{-1}.

This is a consequence of the second inequality in (ii).

Whenever $x_1, x_2 \in U$, we have $y_1 = F(x_1)$ and $y_2 = F(x_2)$.

So (ii) becomes

$$\|F^{-1}(y_1) - F^{-1}(y_2)\| \leq 2\|y_1 - y_2\|.$$

This implies that $F^{-1} : V \to U$ is continuous.

Inverse Function Theorem (v)

(v) Let $b = F(a)$ be in V. Then F^{-1} is differentiable at b and

$$DF^{-1}(b) = [DF(a)]^{-1},$$

the matrix inverse to $DF(a)$.

Since F is of class C^r, $r \geq 1$, on W it is differentiable on all of U.
In particular, it is differentiable at $a = F^{-1}(b)$.
Thus, by definition,

$$F(x) - F(a) = DF(a) \cdot (x - a) + \|x - a\| r(x, a),$$

where $r(x, a) \to 0$ as $x \to a$.
By (ii), $DF(a)$ is nonsingular, and we let A be its inverse matrix.
Multiplying the above expression by A and using $y = F(x)$, $x = F^{-1}(y)$, and $a = F^{-1}(b)$, and so on, we obtain

$$\begin{aligned} A \cdot (y - b) = {} & F^{-1}(y) - F^{-1}(b) \\ & + \|F^{-1}(y) - F^{-1}(b)\| A \cdot r(F^{-1}(y), F^{-1}(b)). \end{aligned}$$

Inverse Function Theorem (v)

- This, in turn, gives

$$F^{-1}(y) = F^{-1}(b) + A \cdot (y - b) + \|y - b\|\widetilde{r}(y, b)$$

if we suppose $y \neq b$ and define

$$\widetilde{r}(y, b) = -\frac{\|F^{-1}(y) - F^{-1}(b)\|}{\|y - b\|} A \cdot r(F^{-1}(y), F^{-1}(b)).$$

Inequality (ii) shows that the initial fraction is bounded by 2

Furthermore, A is a matrix of constants and $F^{-1}(y)$ is continuous.

So it is clear that $\lim_{y \to b} \widetilde{r}(y, b) = 0$.

This proves the differentiability of F^{-1} at any $b \in V$.

Hence $DF^{-1}(b) = A = [DF(a)]^{-1}$, as claimed.

Inverse Function Theorem (vi)

(vi) If F is of class C^r on U, then F^{-1} is of class C^r on V.

For $y \in V$, we have just seen that $DF^{-1}(y) = [DF(F^{-1}(y))]^{-1}$.

We know that:

- $F^{-1}(y)$ is continuous as a function of y on V and its range is U;
- DF is of class C^r and nonsingular on U;
- The entries in the inverse of a nonsingular matrix are C^∞ functions of the entries of the matrix.

It follows that DF^{-1} is continuous on V.

Thus, F^{-1} is of class C^1 at least.

In fact, if F^{-1} is of class $k < r$, the entries of DF^{-1} are of class $k-1$ at least on V.

However, the formula above for them shows these entries to be given by composition of functions of class C^k or greater.

Hence they are of class C^k at least. This implies F^{-1} is of class C^{k+1}.

By induction, F^{-1} is of class C^r.

Consequences

- The following two corollaries are immediate consequences of the theorem.
- We use the notation of the theorem:
  - W is an open subset of $\mathbb{R}^n$;
  - $F : W \to \mathbb{R}^n$.

Corollary

If DF is nonsingular at every point of W, then F is an open mapping of W. That is, it carries W and open subsets of $\mathbb{R}^n$ contained in W to open subsets of $\mathbb{R}^n$.

Corollary

A C^∞ map F is a diffeomorphism from W to $F(W)$ if and only if:

- It is one-to-one;
- DF is nonsingular at every point of W.

Subsection 7

The Rank of a Mapping

Rank of a Linear Transformation

- In linear algebra the **rank** of an $m \times n$ matrix A is defined in three equivalent ways:
  (i) The dimension of the subspace of $\boldsymbol{V}^n$ spanned by the rows;
  (ii) The dimension of the subspace of $\boldsymbol{V}^m$ spanned by the columns;
  (iii) The maximum order of any nonvanishing minor determinant.
- We see at once from (i) and (ii) that the $\text{rank} A \leq m, n$.
- The **rank** of a linear transformation is defined to be the dimension of the image.
- It can be shown that this is the rank of any matrix which represents the transformation.
- It follows that, if P and Q are nonsingular matrices, then

$$\text{rank}(PAQ) = \text{rank}(A).$$

Rank of the Derivative

- Let $F : U \to \mathbb{R}^m$ be a C^1 mapping of an open set $U \subseteq \mathbb{R}^n$.
- Then $DF(x)$ has a rank at each $x \in U$.
- The value of a determinant is a continuous function of its entries.
- From (iii), if $DF(a) = k$, then for some open neighborhood V of a, $\text{rank} DF(x) \geq k$.
- Further, if $k = \inf(m, n)$, then $\text{rank} DF(x) = k$ on this V.
- In general, the inequality is possible.

Example

- Let
$$F(x^1, x^2) = ((x^1)^2 + (x^2)^2, 2x^1x^2).$$
It has Jacobian
$$DF(x^1, x^2) = \begin{pmatrix} \frac{\partial f^1}{\partial x^1} & \frac{\partial f^1}{\partial x^2} \\ \frac{\partial f^2}{\partial x^1} & \frac{\partial f^2}{\partial x^2} \end{pmatrix} = \begin{pmatrix} 2x^1 & 2x^2 \\ 2x^2 & 2x^1 \end{pmatrix}.$$
Its rank is 2 on all of $\mathbb{R}^2$ except the lines $x^2 = \pm x^1$.
The rank is 1 on these lines except at $(0, 0)$ where it is zero.

The Rank of a Function

- We shall refer to the rank of $DF(x)$ as the **rank of** F **at** x.
- Diffeomorphisms have nonsingular Jacobians.
- So, if we compose F with diffeomorphisms, the facts cited and the chain rule imply that the rank of the composition is the rank of F.
- We say F **has rank** k **on a set** A, if it has rank k for each $x \in A$.

The Rank Theorem

Theorem (Rank Theorem)

Let $A_0 \subseteq \mathbb{R}^n$, $B_0 \subseteq \mathbb{R}^m$ be open sets.
Let $F : A_0 \to B_0$ be a C^r mapping.
Suppose the rank of F on A_0 is equal to k.
If $a \in A_0$ and $b = F(a)$, then, there exist open sets $A \subseteq A_0$ and $B \subseteq B_0$ with $a \in A$ and $b \in B$, and C^r diffeomorphisms

$$\begin{array}{llll} G: & A & \to & U \text{ (open) } \subseteq \mathbb{R}^n, \\ H: & B & \to & V \text{ (open) } \subseteq \mathbb{R}^m, \end{array}$$

such that

$$H \circ F \circ G^{-1}(U) \subseteq V$$

and such that this map has the simple form

$$H \circ F \circ G^{-1}(x^1, \ldots, x^n) = (x^1, \ldots, x^k, 0, \ldots, 0).$$

The Rank Theorem (Comments)

- The Rank Theorem tells us that a mapping of constant rank k behaves locally like:
  - Projection of $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$ to $\mathbb{R}^k$;
  - Followed by injection of $\mathbb{R}^k$ onto $\mathbb{R}^k \times \{0\} \subseteq \mathbb{R}^k \times \mathbb{R}^{m-k} = \mathbb{R}^m$.
- This is an important tool and we shall use it frequently.
- Later it will be rephrased in terms of local coordinates.
- It implies the Inverse Function Theorem as a special case.

The Rank Theorem (Proof)

- To begin with, we may suppose:
  - $a = 0$, the origin of $\mathbb{R}^n$;
  - $b = 0$, the origin of $\mathbb{R}^m$.
- Composition of F with two translations gives
$$\widetilde{F}(u) = F(u + a) - b.$$
- This has the property that $\widetilde{F}(0) = 0$.
- So, if the theorem holds for this case, then it holds in general.
- Similar, we may use linear maps which permute the coordinates.
- So we may suppose that a $k \times k$ minor of nonzero determinant in $DF(a)$ is the upper left $k \times k$ minor,
$$\frac{\partial(f^1, \ldots, f^k)}{\partial(u^1, \ldots, u^k)} = \left(\begin{array}{ccc} \frac{\partial f^1}{\partial u^1} & \cdots & \frac{\partial f^1}{\partial u^k} \\ \vdots & & \vdots \\ \frac{\partial f^k}{\partial u^1} & \cdots & \frac{\partial f^k}{\partial u^k} \end{array} \right)_{u=a}.$$

The Rank Theorem (Proof)

- We define the C^r-mapping $G : A_0 \to \mathbb{R}^n$ by

$$G(u^1,\ldots,u^n) = (f^1(u^1,\ldots,u^n),\ldots,f^k(u^1,\ldots,u^n),u^{k+1},\ldots,u^n).$$

Then

$$DG = \left(\begin{array}{ccc|c} \frac{\partial f^1}{\partial u^1} & \cdots & \frac{\partial f^1}{\partial u^k} & \\ \vdots & & \vdots & * \\ \frac{\partial f^k}{\partial u^1} & \cdots & \frac{\partial f^k}{\partial u^n} & \\ \hline & 0 & & I_{n-k} \end{array}\right)$$

where I_{n-k} is the $(n-k) \times (n-k)$ identity, the terms in the lower left block are zero, and those in the upper right do not interest us.

This matrix is nonsingular at $u = a$.

Hence, there is in A_0 an open set A_1, containing a, on which G is a diffeomorphism onto an open subset $U_1 = G(A_1)$.

The Rank Theorem (Proof)

- From the expression for G and the definition of U_1 we have:
  - $F \circ G^{-1}(0) = 0$;
  - $F \circ G^{-1}(U_1) \subseteq B_0$;
  - Setting $\overline{f}^{k+j}(x) = f^{k+j} \circ G^{-1}(x)$,

$$F \circ G^{-1}(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n) = (x^1, \ldots, x^k, \overline{f}^{k+1}(x), \ldots, \overline{f}^m(x)).$$

For the latter, recall that G^{-1} is one-to-one on U_1.

Moreover, for $\ell = 1, \ldots, k$ the values of x^ℓ are $f^\ell(u)$, $u \in G^{-1}(U_1)$.

So $f^\ell \circ G^{-1}(x) = f^\ell(u) = x^\ell$.

So far we have used only the fact that the rank of DF at a (hence in a neighborhood of a) is at least k.

We have not yet used the fact that it is identically k on A_0.

This is needed in the next step, requiring that the rank be at most k.

The Rank Theorem (Proof)

- We compute $D(F \circ G^{-1})$ from the formula above for $F \circ G^{-1}$, giving

$$D(F \circ G^{-1})(x) = \left(\begin{array}{c|ccc} I_k & & 0 & \\ \hline & \frac{\partial \overline{f}^{k+1}}{\partial x^{k+1}} & \cdots & \frac{\partial \overline{f}^{k+1}}{\partial x^n} \\ * & \vdots & & \vdots \\ & \frac{\partial \overline{f}^{m}}{\partial x^{k+1}} & \cdots & \frac{\partial \overline{f}^{m}}{\partial x^n} \end{array}\right).$$

  This is valid on U_1 where $F \circ G^{-1}$ is defined.

  Now, DG^{-1} is nonsingular on U_1 and $G^{-1}(U_1) = A_1 \subseteq A_0$.

  Therefore, $\text{rank} D(F \circ G^{-1}) = \text{rank}(DF \circ DG^{-1}) \equiv k$ on U_1.

  This implies that all terms in the lower right-hand block of the matrix are zero on U_1.

  That is, the functions $\overline{f}^{k+1}, \ldots, \overline{f}^m$ depend on $x^1, \ldots, x^k$ only.

The Rank Theorem (Proof)

- Now we define a function T from a neighborhood V_1 of 0 in $\mathbb{R}^m$ into $B_0 \subseteq \mathbb{R}^m$ by the formula

$$T(y^1, \ldots, y^k, y^{k+1}, \ldots, y^m) = (y^1, \ldots, y^k, \\ y^{k+1} + \overline{f}^{k+1}(y^1, \ldots, y^k), \ldots, y^m + \overline{f}^m(y^1, \ldots, y^k)).$$

  The domain V_1 is subject to two restrictions:
  - First it must be small enough so that for $y = (y^1, \ldots, y^m) \in V$, the functions $\overline{f}^{k+j}(y^1, \ldots, y^k)$ are defined;
  - Second, it must be small enough so that $T(V_1) \subseteq B_0$.

  It is clear that $T(0) = 0$.

  If we compute DT, we see that it is nonsingular everywhere on V_1, since it takes the form

$$DT(y) = \left(\begin{array}{c|c} I_k & 0 \\ \hline * & I_{m-k} \end{array}\right).$$

The Rank Theorem (Proof)

- Therefore T is a C^r diffeomorphism of a neighborhood V of 0 in V_1 onto an open set $B \subseteq \mathbb{R}^m$.

  The origin of $\mathbb{R}^m$ is in B and $B \subseteq B_1$.

  Choose a neighborhood $U \subseteq U_1$ of the origin in $\mathbb{R}^n$, such that

  $$F \circ G^{-1}(U) \subseteq B.$$

  Let $A = G^{-1}(U)$ and let $H = T^{-1}$.

  Then

  $$U \xrightarrow{G^{-1}} A \xrightarrow{F} B \xrightarrow{H} V$$

  are C^r maps of these open sets and G^{-1}, H are C^r diffeomorphisms onto A and V, respectively.

The Rank Theorem (Proof)

- We have

$$F \circ G^{-1}(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n) = (x^1, \ldots, x^k, \overline{f}^{k+1}(x^1, \ldots, x^k), \ldots, \overline{f}^m(x^1, \ldots, x^k)).$$

It follows that

$$H \circ F \circ G^{-1}(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n) = (x^1, \ldots, x^k, 0, \ldots, 0) \in \mathbb{R}^m.$$

On the other hand, suppose we set:
  - $y^i = x^i$, for $i = 1, \ldots, k$;
  - $y^i = 0$, for $i = k+1, \ldots, m$.

Then, according to its definition above, T must take the value $(x^1, \ldots, x^k, \overline{f}^{k+1}(x^1, \ldots, x^k), \ldots, \overline{f}^m(x^1, \ldots, x^k))$.
But T is one-to-one.
Hence, T^{-1} takes $(x^1, \ldots, x^k, \overline{f}^{k+1}(x^1, \ldots, x^m), \ldots, \overline{f}^m(x^1, \ldots, x^m))$ to $(x^1, \ldots, x^k, 0, \ldots, 0)$, as claimed.

Corollary

Corollary

We may choose the neighborhoods U and V in either of the following ways:

(i) $U = B_\epsilon^n(0)$ and $V = B_\epsilon^m(0)$ or

(ii) $U = C_\epsilon^n(0)$ and $V = C_\epsilon^m(0)$

with the same $\epsilon > 0$ for both U and V.

Suppose that:

- π denotes the projection of $\mathbb{R}^m = \mathbb{R}^k \times \mathbb{R}^{m-k}$ to $\mathbb{R}^k$;
- $i : \mathbb{R}^k \to \mathbb{R}^k \times \mathbb{R}^{n-k}$ is the injection to the subset $\mathbb{R}^k \times \{0\}$.

Then we have

$$\pi \circ H \circ F \circ G^{-1} \circ i$$

is the identity on $B_\epsilon^k(0)$ in Case (i) or on $C_\epsilon^k(0)$ in Case (ii).