Manifolds and Differential Forms

Reyer Sjamaar

DEPARTMENT OF MATHEMATICS, CORNELL UNIVERSITY, ITHACA, NEW YORK 14853-4201

E-mail address: sjamaar@math.cornell.edu

URL: http://www.math.cornell.edu/~sjamaar

Notes and exercises for Math 321

Last run: 23h51, 27 August 2002

Contents

Preface

These are the lecture notes for Math 321, Manifolds and Differential Forms, as taught at Cornell University in the Fall of 2001 and 2002. The course covers differential geometry for an audience of undergraduates who have taken a typical calculus sequence at a North American university, including basic linear algebra and multivariable calculus up to the integral theorems of Green, Gauß and Stokes. With a view to the fact that nowadays such students are familiar with vector spaces, the text is not restricted to curves and surfaces in $\mathbf{R}^3$, but covers manifolds of arbitrary dimension. Some prerequisites are briefly reviewed within the text. The selection of material is similar to Flanders' book [**Fla89**], but the treatment is at a more elementary and informal level appropriate for sophomores and juniors.

This is only the second iteration of these lecture notes and as such leaves very much to be desired. They are being rewritten on the fly and if you are a student taking this course, you will have to download a new instalment every week along with the homework assignments. For this I ask your forbearance. Corrections, suggestions and any comments short of abuse will be received gratefully.

Ithaca, NY, 27 August 2002

CHAPTER 1

Introduction

Manifolds

We start with an informal, intuitive introduction to manifolds and how they arise in mathematical nature. Most of this material will be examined more thoroughly in later chapters.

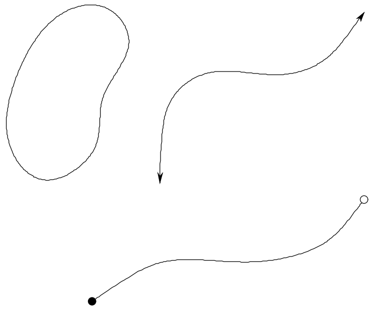
Recall that $\mathbf{R}^n$ is the set of all column vectors with n real entries

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

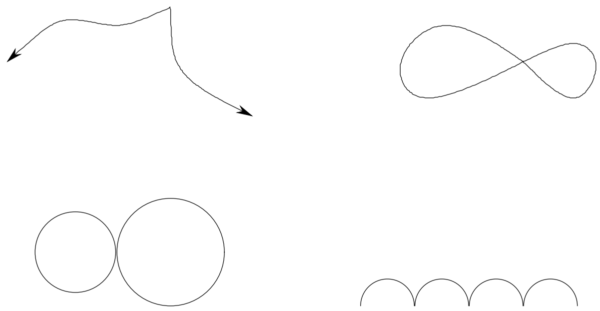
which we shall call *points* or sometimes n-*vectors*. A *manifold* is a certain type of subset of $\mathbf{R}^n$. A precise definition will follow later, but one important consequence of the definition is that a manifold has a well-defined tangent space at every point. This fact enables us to apply the methods of calculus and linear algebra to the study of manifolds. The *dimension* of a manifold is by definition the dimension of its tangent spaces. The dimension of a manifold in $\mathbf{R}^n$ can be no higher than n.

Dimension 1

A one-dimensional manifold is, loosely speaking, a curve without kinks or self-intersections. Instead of the tangent "space" at a point one usually speaks of the tangent line. A curve in $\mathbf{R}^2$ is called a *plane* curve and a curve in $\mathbf{R}^3$ is a *space* curve, but you can have curves in any $\mathbf{R}^n$. Curves can be closed (as in the first picture below), unbounded (as indicated by the arrows in the second picture), or have one or two endpoints (the third picture shows a curve with an endpoint on the left, indicated by the black dot; the white dot on the right indicates that that point does not belong to the curve; the curve "peters out" without coming to an endpoint). Endpoints are also called *boundary points*.



The four curves below are not manifolds. The first one has a kink; the second has a point of self-intersection with two distinct tangent lines instead of a single well-defined tangent line. The figure eight and the cycloid have well-defined tangent lines everywhere. Still they are not manifolds: the figure eight has a self-intersection and the cusps of the cycloid have a jagged appearance which is proscribed by the definition of a manifold (which we have not yet given). The points where these curves fail to be manifolds are called *singular*. The "good" points are called *regular* or *smooth*. Singularities are extremely interesting, but hard to study without a thorough understanding of the smooth points. We shall largely ignore singularities in this course.



Dimension 2

A two-dimensional manifold in $\mathbf{R}^3$ is a smooth surface embedded in three-dimensional space, such as a sphere, a torus, a paraboloid, or a Möbius band. A two-dimensional manifold can also have a boundary, which is always a one-dimensional manifold. The boundary of a hemisphere is a great circle. What is the boundary of a Möbius band? (Sorry, no pictures. My drawing software is too primitive.)

There exist two-dimensional manifolds that cannot be depicted accurately as subsets of $\mathbf{R}^3$. Examples that you may have heard of are the Klein bottle and the projective plane. These can be represented as surfaces in $\mathbf{R}^3$ with self-intersections (the technical term is "immersed"), but to represent them without self-intersections (technically, to "embed" them) you need four dimensions.

You can also have two-dimensional manifolds in $\mathbf{R}^2$, but they are relatively boring. Examples are: arbitrary open subsets of $\mathbf{R}^2$, or a closed subset with a smooth boundary, such as a closed disc, or a closed annulus. A closed square is not a manifold, because the corners are not smooth.

Equations

Very often manifolds are given "implicitly", namely as the solution set of an equation $f(\mathbf{x}) = \mathbf{c}$. Here $f\colon \mathbf{R}^n \to \mathbf{R}^m$ is a function and $\mathbf{c}$ is a constant vector in $\mathbf{R}^m$. Manifolds defined by linear equations (i.e. where f is a matrix) are usually called *affine subspaces* of $\mathbf{R}^n$ and are studied in linear algebra. More interesting manifolds arise from nonlinear equations. As a simple example, the *sphere of radius* r in $\mathbf{R}^n$ is the set of all $\mathbf{x}$ in $\mathbf{R}^n$ satisfying the single equation $\|\mathbf{x}\| = r$. Here

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}\cdot\mathbf{x}} = \sqrt{x_1^2 + \cdots + x_n^2}$$

is the norm or length of $\mathbf{x}$. The tangent space to the sphere at a point $\mathbf{x}$ is the set of all vectors $\mathbf{y}$ such that $(\mathbf{y} - \mathbf{x}) \cdot \mathbf{x} =$, i.e. $y_1 x_1 + \cdots + y_n x_n = r^2$. It is useful to introduce a notation for this: if we call the sphere X and if $\mathbf{x}$ is a point on it, then the tangent space at $\mathbf{x}$ is denoted $T_{\mathbf{x}}X$. We can now write in symbols

$$T_{\mathbf{x}}X = \{\, \mathbf{y} \in \mathbf{R}^n : \mathbf{y} \cdot \mathbf{x} = r^2 \,\}.$$

$T_{\mathbf{x}}X$ is an affine subspace of $\mathbf{R}^n$ given by a single equation. Its dimension is therefore $n - 1$. Thus the unit sphere in $\mathbf{R}^n$ is $n - 1$-dimensional. What is a one-dimensional sphere? And a zero-dimensional sphere?

The solution set of a system of equations may have singularities and is therefore not necessarily a manifold. A simple example is $xy = 0$, the union of the two coordinate axes in the plane. Where is the singularity?

Parametrizations

A dual method for defining manifolds is the "explicit" way, namely by parametrizations. For instance,

$$x = \cos\theta \cos\phi, \quad y = \sin\theta \cos\phi, \quad z = \sin\phi$$

parametrizes the unit sphere in $\mathbf{R}^3$. A parametrization of a manifold is often harder to come by than a system of equations, but sometimes a parametrization is more useful, for example when one wants to integrate over the manifold. Also, it is usually impossible to parametrize a manifold in such a way that every point is covered exactly once. Such is the case for the two-sphere. One commonly restricts the variables (θ, ϕ) to the rectangle $(0, 2\pi) \times (-\pi/2, \pi/2)$ to avoid counting points twice. One then misses the meridian $\theta = 0$, but this does not matter for some purposes, such as integrating a continuous function.

A parametrization does not always describe a manifold. An example is the plane curve parametrized by

$$x = t - \sin t, \quad y = 1 - \cos t,$$

which is studied in Exercise 1.3.

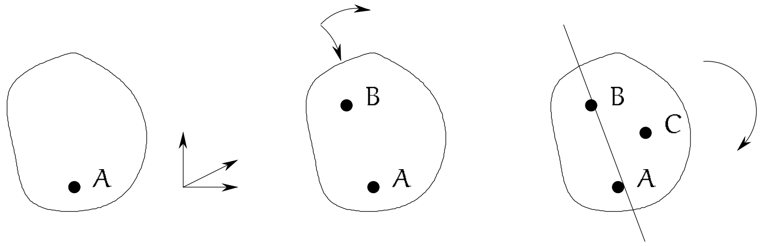
Configuration spaces

Sometimes manifolds come about in ways that are less readily visualized. A good example are the configuration spaces of physics and engineering. The *configuration* of a mechanical system (such as a pendulum, a spinning top, the solar system, a fluid, or a gas etc.) is its state or position at any given time. (The configuration ignores any motions that the system may be undergoing. So when the system moves, its configuration changes.) In practice one usually describes a configuration by specifying the coordinates of suitably chosen parts of the system. The configuration *space* of the system is an abstract space, the points of which are in one-to-one correspondence to all physically possible configurations of the system. Very often the configuration space turns out to be a manifold. Its dimension is called the *number of degrees of freedom* of the system. The configuration space of even a fairly small system can be quite complicated.

1.1. EXAMPLE. A spherical pendulum is a weight or bob attached to a fixed centre by a rigid rod, free to swing in any direction in three-space. The state of the pendulum is therefore entirely determined by the position of the bob. The bob can move between any two points at a fixed distance (equal to the length of the rod) from the centre. The configuration space is therefore a two-dimensional sphere.

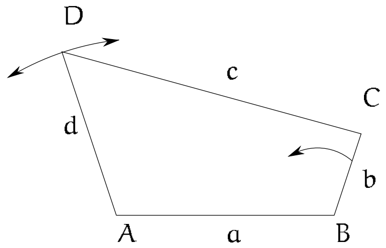
Some believe that only spaces of dimension ≤ 3 (or 4, for those who have heard of relativity) can have a basis in physical reality. The following example shows that this is not true.

1.2. EXAMPLE. What is the number of degrees of freedom of a rigid body moving in $\mathbf{R}^3$? Select any triple of points A, B, C in the solid that do not lie on one line. The point A can move about freely and is determined by three coordinates, and so it has three degrees of freedom. But the position of A alone does not determine the position of the whole solid. If A is kept fixed, the point B can perform two independent swivelling motions. In other words, it moves on a sphere centred at A, which gives two more degrees of freedom. If A and B are both kept fixed, the point C can rotate about the axis AB, which gives one further degree of freedom.



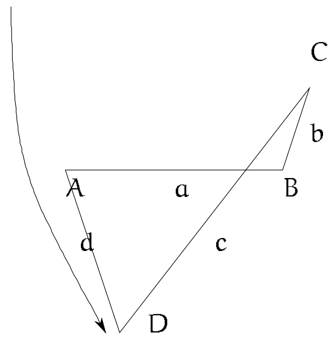
The positions of A, B and C determine the position of the solid uniquely, so the total number of degrees of freedom is $3 + 2 + 1 = 6$. Thus the configuration space of a rigid body is a six-dimensional manifold.

1.3. EXAMPLE (the space of quadrilaterals). Consider all quadrilaterals $ABCD$ in the plane with fixed sidelengths a, b, c, d, respectively. (Think of four rigid rods attached by hinges.) What are all the possibilities? For simplicity let us disregard translations by keeping the first edge AB fixed in one place.

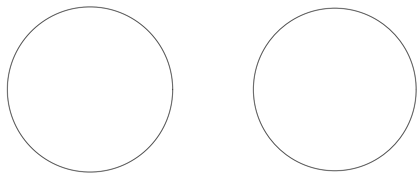


The short edge BC can spin full circle about the point B. During this motion the point D moves back and forth on a circle of radius d centred at A. (So we have built a little machine for converting a circular motion to a pumping motion, or vice versa. Arrangements such as this are of course quite common in engines etc.) The position of D, however, is wholly determined by that of C. This means that the configurations are in one-to-one correspondence with the points on the circle of radius b about the point B, i.e. the configuration space is a circle.

Actually, this is not completely accurate: for every choice of C, there are *two* choices for D! The second one can be obtained by reflecting D in the diagonal AC.

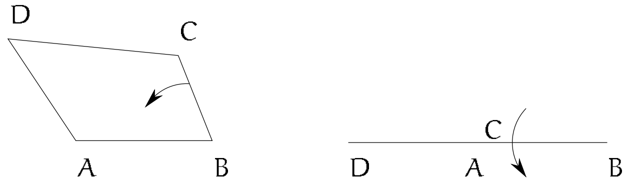


So there is in fact another circle's worth of possible configurations. It is not possible to move continuously from the first set of configurations to the second; in fact they are each other's mirror images. Thus the configuration space is a disjoint union of two circles.

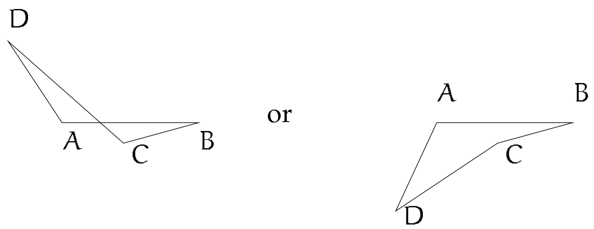


This is an example of a *disconnected* manifold consisting of two *connected components*.

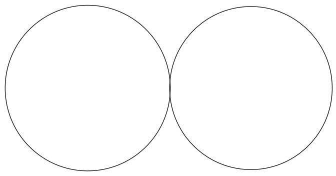
1.4. EXAMPLE (quadrilaterals, continued). Even this is not the full story: something strange happens when $b+c = a+d$ (and also when $a+b = c+d$). In this case, when BC points straight to the left, the quadrilateral collapses to a line segment:



and when C moves further down, there are two possible directions for D to go, up or down:



This means that when $b+c = a+d$ the two components of the configuration space are merged at a point.



The juncture represents the collapsed quadrilateral. This configuration space is not a manifold, but most configuration spaces occurring in nature are (and an engineer designing an engine wouldn't want to use this quadrilateral to make a piston drive a flywheel). More singularities appear in the case of a parallelogram (i.e. $a = c$ and $b = d$) and in the equilateral case ($a = b = c = d$).

Exercises

1.1. Draw a picture of the surface in $\mathbf{R}^3$ defined by $z^2 = x^2y$. Determine all the points where the surface does not have a well-defined tangent plane. (Computer output is OK, but bear in mind that few drawing programs do an adequate job of graphing this surface, so you may be better off drawing it by hand. As a preliminary step, determine the intersections of the surface with the coordinate planes $x =$ const, $y =$ const and $z =$ const.)

1.2. Let X be the sphere of radius $\sqrt{n}$ about the origin in $\mathbf{R}^n$ and let $\mathbf{a}$ be the point $(1, 1, \dots, 1)$ on X. Determine the tangent space $T_{\mathbf{a}}X$ explicitly in the following sense: using linear algebra find vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1}$ such that every element $\mathbf{x}$ of $T_{\mathbf{a}}X$ is of the form

$$\mathbf{x} = \mathbf{a} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_{n-1}\mathbf{v}_{n-1}$$

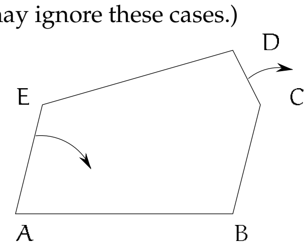
with real coefficients $t_1, t_2, \dots, t_{n-1}$.

1.3. The formulas $x = t - \sin t$, $y = 1 - \cos t$ ($t \in \mathbf{R}$) parametrize a plane curve X. Draw a careful picture of X. Also include a few tangent lines at judiciously chosen points. (E.g. find all tangent lines with slope 0, ± 1, and ∞.) Identify all points where X is not a manifold.

1.4. Same questions as in Exercise 1.3 for the curve $x = 3at/(1+t^3)$, $y = 3at^2/(1+t^3)$.

1.5. What is the number of degrees of freedom of a bicycle? (Imagine that it moves freely through empty space and is not constrained to the surface of the earth.)

1.6. What is the configuration space of all pentagons $ABCDE$ in the plane with fixed sidelengths a, b, c, d, e? (As in the case of quadrilaterals, for certain choices of sidelengths singularities may occur. You may ignore these cases.)



1.7. Choose two distinct positive real numbers a and b. What is the configuration space of all parallelograms $ABCD$ such that AB and CD have length a and BC and AD have length b? What happens if $a = b$? (As in Examples 1.3 and 1.4 assume that the edge AB is kept fixed in place so as to rule out translations.)

CHAPTER 2

Prerequisites

Derivatives

This is a brief review of some multivariable calculus needed in the study of manifolds. Let $f\colon \mathbf{R}^n \to \mathbf{R}^m$ be a map or function. (The word "map" is often used for a vector-valued function of several variables.) We can write f as

$$f(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix},$$

where each of the component functions f_i is a scalar-valued function $f_i\colon \mathbf{R}^n \to \mathbf{R}$. We say that f is *continuously differentiable* if the partial derivatives

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}) = \lim_{h\to 0} \frac{f_i(\mathbf{x}+h\mathbf{e}_j) - f_i(\mathbf{x})}{h} \tag{2.1}$$

are well-defined and continuous functions of $\mathbf{x}$ for all $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Here

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ldots, \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

are the standard basis vectors of $\mathbf{R}^n$. The *(total) derivative* or *Jacobi matrix* of f at $\mathbf{x}$ is then the $m \times n$-matrix

$$df(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \ldots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \ldots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

If $\mathbf{v}$ is any vector in $\mathbf{R}^n$, the *directional derivative* of f along $\mathbf{v}$ is defined to be the vector $df(\mathbf{x})\mathbf{v}$ in $\mathbf{R}^m$, obtained by multiplying the matrix $df(\mathbf{x})$ by the vector $\mathbf{v}$.

For $n = 1$ f is a vector-valued function of one variable x, often called a *path* or *curve* in $\mathbf{R}^m$. In this case the matrix $df(x)$ consists of a single column vector and is usually denoted simply by $f'(x)$.

For $m = 1$ f is a scalar-valued function of n variables and $df(\mathbf{x})$ is a single row vector. The transpose of $df(\mathbf{x})$ is therefore a column vector, usually called the *gradient* of f: $df(\mathbf{x})^T = \operatorname{grad} f(\mathbf{x})$. The directional derivative of f along $\mathbf{v}$ is can then be written as an inner product $df(\mathbf{x})\mathbf{v} = \operatorname{grad} f(\mathbf{x}) \cdot \mathbf{v}$.

Frequently a function is not defined on all of $\mathbf{R}^n$, but only on a subset U. We must be a little careful in defining the derivative of such a function. Recall that the *open ball* of radius ε about a point $\mathbf{x} \in \mathbf{R}^n$ is

$$B_\varepsilon(\mathbf{x}) = \{\, \mathbf{y} \in \mathbf{R}^n : \|\mathbf{y} - \mathbf{x}\| < \varepsilon \,\}.$$

A subset U of $\mathbf{R}^n$ is *open* if for every $\mathbf{x} \in U$ there exists an $\varepsilon > 0$ such that $B_\varepsilon(x)$ is contained in U. Intuitively this means that at every point in U there is a little bit of room inside U to move around in any direction you like. An *open neighbourhood* of $\mathbf{x} \in \mathbf{R}^n$ is any open set containing $\mathbf{x}$.

Now let $f\colon U \to \mathbf{R}^m$ be a function defined on an open subset U of $\mathbf{R}^n$. Let $\mathbf{x} \in U$. Then there exists $\varepsilon > 0$ such that the points $\mathbf{x} + t\mathbf{e}_j$ are contained in U for $-\varepsilon < t < \varepsilon$. Therefore $f_i(\mathbf{x} + t\mathbf{e}_j)$ is well-defined for $-\varepsilon < t < \varepsilon$ and thus it makes sense to ask whether the partial derivatives (2.1) exist. If they do, for all $\mathbf{x} \in U$ and all i and j, and if they are continuous, the function f is called *continuously differentiable* or C^1.

If the second partial derivatives

$$\frac{\partial^2 f_i}{\partial x_j \partial x_k}(\mathbf{x})$$

exist and are continuous for all $\mathbf{x} \in U$ and for all $i = 1, 2, \dots, n$ and $j, k = 1, 2, \dots, m$, then f is called *twice continuously differentiable* or C^2. Likewise, if all k-fold partial derivatives

$$\frac{\partial^k f_i}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_k}}(\mathbf{x})$$

exist and are continuous, then f is k *times continuously differentiable* or C^k. If f is C^k for all $k \geq 1$, then we say that f is *infinitely many times differentiable*, C^∞, or *smooth*.

Let us now review some of the most important facts concerning derivatives.

Chain rule

Recall that if A, B and C are sets and $f\colon A \to B$ and $g\colon B \to C$ are functions, we can apply g after f to obtain the composite function $(g \circ f)(x) = g(f(x))$.

2.1. THEOREM (chain rule). *Let $U \subseteq \mathbf{R}^n$ and $V \subseteq \mathbf{R}^m$ be open and let $f\colon U \to V$ and $g\colon V \to \mathbf{R}^k$ be continuously differentiable. Then $g \circ f$ is continuously differentiable and*

$$\boxed{d(g \circ f)(\mathbf{x}) = dg(f(\mathbf{x}))df(\mathbf{x})}$$

for all $\mathbf{x} \in U$.

Here $dg(f(\mathbf{x}))df(\mathbf{x})$ denotes the composition or the product of the two matrices $dg(f(\mathbf{x}))$ and $df(\mathbf{x})$.

2.2. EXAMPLE. In the one-variable case $n = m = k = 1$ the derivatives df and dg are 1×1-matrices $(f'(x))$ and $(g'(y))$, and matrix multiplication is ordinary multiplication, so we get the usual chain rule

$$(g \circ f)'(x) = g'(f(x))f'(x).$$

2.3. EXAMPLE. If $n = k = 1$, then $g \circ f$ is a real-valued function of one variable x, so $d(g \circ f)$ is a 1×1-matrix containing the single entry $(g \circ f)'$. Moreover,

$$df(x) = \begin{pmatrix} \frac{df_1}{dx}(x) \\ \vdots \\ \frac{df_m}{dx}(x) \end{pmatrix} \qquad \text{and} \qquad dg(\mathbf{y}) = \begin{pmatrix} \frac{\partial g}{\partial y_1}(\mathbf{y}) & \dots & \frac{\partial g}{\partial y_m}(\mathbf{y}) \end{pmatrix},$$

so by the chain rule

$$\frac{d(g \circ f)}{dx}(x) = dg(f(x))df(x) = \sum_{i=1}^{m} \frac{\partial g}{\partial y_i}(f(x))\frac{df_i}{dx}(x).$$

Sometimes we are sloppy and abbreviate this as

$$\frac{d(g \circ f)}{dx} = \sum_{i=1}^{m} \frac{\partial g}{\partial y_i}\frac{df_i}{dx}.$$

Implicit function theorem

Let $F \colon W \to \mathbf{R}^m$ be a continuously differentiable function defined on an open subset W of $\mathbf{R}^{n+m}$. Let us think of a vector in $\mathbf{R}^{n+m}$ as an ordered pair of vectors $(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{y} \in \mathbf{R}^m$. Consider the equation

$$F(\mathbf{x}, \mathbf{y}) = \mathbf{0}.$$

Under what circumstances is it possible to solve for $\mathbf{y}$ as a function of $\mathbf{x}$? The answer is given by the implicit function theorem. We form the Jacobi matrices of F with respect to the $\mathbf{x}$- and $\mathbf{y}$-variables separately,

$$d_{\mathbf{x}}f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}, \qquad d_{\mathbf{y}}f = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial y_1} & \dots & \frac{\partial f_m}{\partial y_m} \end{pmatrix}.$$

Observe that the matrix $d_{\mathbf{y}}f$ is square. We are in business if we have a point at which F is $\mathbf{0}$ and $d_{\mathbf{y}}F$ is invertible.

2.4. THEOREM (implicit function theorem). *Let $F \colon W \to \mathbf{R}^m$ be continuously differentiable, where W is open in $\mathbf{R}^{n+m}$. Suppose that $(\mathbf{x}_0, \mathbf{y}_0) \in W$ is a point such that $F(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$ and $d_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0)$ is invertible. Then there are open neighbourhoods $U \subseteq \mathbf{R}^n$ of $\mathbf{x}_0$ and $V \subseteq \mathbf{R}^m$ of $\mathbf{y}_0$ such that for each $\mathbf{x} \in U$ there exists a unique $\mathbf{y} = \phi(\mathbf{x}) \in V$ satisfying $F(\mathbf{x}, \phi(\mathbf{x})) = \mathbf{0}$. The function $\phi \colon U \to V$ is continuously differentiable with derivative given by implicit differentiation:*

$$\boxed{d\phi(\mathbf{x}) = -d_{\mathbf{y}}F(\mathbf{x}, \mathbf{y})^{-1}\, d_{\mathbf{x}}F(\mathbf{x}, \mathbf{y})\big|_{\mathbf{y}=\phi(\mathbf{x})}}$$

for all $\mathbf{x} \in U$.

This is well-known for $m = n = 1$, when F is a function of two real variables (x, y). If $\partial F/\partial y \neq 0$ at a certain point (x_0, y_0), then for x close to x_0 and y close to y_0 we can solve the equation $F(x, y) = 0$ for y as a function $y = \phi(x)$ of x, and

$$\phi' = -\frac{\partial F/\partial x}{\partial F/\partial y}.$$

Now let us take F to be of the form $F(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}) - \mathbf{x}$, where $f\colon W \to \mathbf{R}^n$ is a given function with W open in $\mathbf{R}^n$. Solving $F(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ here amounts to *inverting* the function f. Moreover, $d_{\mathbf{y}}F = df$, so the implicit function theorem yields the following result.

2.5. THEOREM (inverse function theorem). *Let $f\colon W \to \mathbf{R}^n$ be continuously differentiable, where W is open in $\mathbf{R}^n$. Suppose that $\mathbf{y}_0 \in W$ is a point such that $df(\mathbf{y}_0)$ is invertible. Then there is an open neighbourhood $U \subseteq \mathbf{R}^n$ of $\mathbf{y}_0$ such that $f(U)$ is an open neighbourhood of $f(\mathbf{y}_0)$ and the function $f\colon U \to f(U)$ is invertible. The inverse $f^{-1}\colon V \to U$ is continuously differentiable with derivative given by*

$$\boxed{df^{-1}(\mathbf{x}) = df(\mathbf{y})^{-1}\big|_{\mathbf{y}=f^{-1}(\mathbf{x})}}$$

for all $\mathbf{y} \in V$.

Again let us spell out the one-variable case $n = 1$. Invertibility of $df(y_0)$ simply means that $f'(y_0) \neq 0$. This implies that near y_0 the function f is strictly monotone increasing (if $f'(y_0) > 0$) or decreasing (if $f'(y_0) < 0$). Therefore if I is a sufficiently small open interval around x_0, then $f(I)$ is an open interval around $f(x_0)$ and the restricted function $f\colon I \to f(I)$ is invertible. The inverse function has derivative

$$(f^{-1})'(x) = \frac{1}{f'(y)},$$

with $y = f^{-1}(x)$.

2.6. EXAMPLE (square roots). Let $f(y) = y^2$. Then $f'(y_0) \neq 0$ whenever $y_0 \neq 0$. For $y_0 > 0$ we can take $I = (0, \infty)$. Then $f(I) = (0, \infty)$, $f^{-1}(x) = \sqrt{x}$, and $(f^{-1})'(x) = 1/(2\sqrt{x})$. For $y_0 < 0$ we can take $I = (-\infty, 0)$. Then $f(I) = (0, \infty)$, $f^{-1}(x) = -\sqrt{x}$, and $(f^{-1})'(x) = -1/(2\sqrt{x})$. In a neighbourhood of 0 it is not possible to invert f.

Substitution formula for integrals

Let V be an open subset of $\mathbf{R}^n$ and let $f\colon V \to \mathbf{R}$ be a function. Suppose we want to change the variables in the integral $\int_V f(\mathbf{y})\, d\mathbf{y}$. (This is shorthand for an n-fold integral over $y_1, y_2, \dots, y_n$.) This means we substitute $\mathbf{y} = p(\mathbf{x})$, where $p\colon U \to V$ is a map from an open $U \subseteq \mathbf{R}^n$ to V. Under a suitable hypothesis we can change the integral over $\mathbf{y}$ to an integral over $\mathbf{x}$.

2.7. THEOREM (change of variables formula). *Let U and V be open subsets of $\mathbf{R}^n$ and let $p\colon U \to V$ be a map. Suppose that p is bijective and that p and its inverse are continuously differentiable. Then for any integrable function f we have*

$$\boxed{\int_V f(\mathbf{y})\, d\mathbf{y} = \int_U f(p(\mathbf{x}))|\det dp(\mathbf{x})|\, d\mathbf{x}.}$$

Again this should look familiar from one-variable calculus: if $p\colon (a, b) \to (c, d)$ is C^1 and has a C^1 inverse, then

$$\int_c^d f(y)dy = \begin{cases} \int_a^b f(p(x))p'(x)\, dx & \text{if } p \text{ is increasing} \\ \int_b^a f(p(x))p'(x)\, dx & \text{if } p \text{ is decreasing.} \end{cases}$$

This can be written succinctly as $\int_c^d f(y)dy = \int_a^b f(p(x))|p'(x)|\, dx$, which looks more similar to the multi-dimensional case.

Exercises

2.1. Prove that the open ball $B_\varepsilon(\mathbf{x}) = \{\,\mathbf{y} \in \mathbf{R}^n : \|\mathbf{y}-\mathbf{x}\| < \varepsilon\,\}$ is open in $\mathbf{R}^n$. State your reasons as precisely as you can, using the definition of openness. (You will need the triangle inequality $\|\mathbf{y}-\mathbf{x}\| \leq \|\mathbf{y}-\mathbf{z}\| + \|\mathbf{z}-\mathbf{x}\|$.)

2.2. Let $\mathbf{x}$ and $\mathbf{v}$ be constant vectors in $\mathbf{R}^n$. Define $\mathbf{c}(t) = \mathbf{x} + t\mathbf{v}$. Find $\mathbf{c}'(t)$.

2.3. According to Newton's law of gravitation, a particle of mass m_1 placed at the origin in $\mathbf{R}^3$ exerts a force on a particle of mass m_2 placed at $\mathbf{x} \in \mathbf{R}^3 - \{\mathbf{0}\}$ equal to

$$\mathbf{F} = -\frac{Gm_1m_2}{\|\mathbf{x}\|^3}\mathbf{x},$$

where G is a constant of nature. Show that $\mathbf{F}$ is the gradient of $f(\mathbf{x}) = Gm_1m_2/\|\mathbf{x}\|$.

2.4. A function $f\colon \mathbf{R}^n \to \mathbf{R}$ is *homogeneous* of degree k if $f(t\mathbf{x}) = t^k f(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{R}^n$ and $t > 0$. Show that if f is homogeneous of degree k and C^1, then

$$\sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}(x) = kf(\mathbf{x}).$$

(Differentiate the relation $f(t\mathbf{x}) = t^k f(\mathbf{x})$ with respect to t.)

2.5. Define a function $f\colon \mathbf{R} \to \mathbf{R}$ by $f(0) = 0$ and $f(x) = e^{-1/x^2}$ for $x \neq 0$.

(i) Show that f is differentiable at 0 and that $f'(0) = 0$.
(ii) Show that f is smooth (i.e. infinitely many times differentiable or C^∞) and that $f^{(n)}(0) = 0$ for all n.
(iii) Plot the function f over the interval $-5 \leq x \leq 5$. Using software or a graphing calculator is fine, but pay special attention to the behaviour near $x = 0$.

The following two calculus problems will be needed in a later chapter.

2.6. For $x > 0$ define

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\,dt$$

and prove the following assertions.

(i) $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$.
(ii) $\Gamma(n) = (n-1)!$ for positive integers n.
(iii) $\displaystyle\int_0^\infty e^{-u^2}u^a\,du = \frac{1}{2}\Gamma\Big(\frac{1+a}{2}\Big)$.

2.7. Let $\gamma = \Gamma(1/2)$ and prove these clever identities.

(i) $\displaystyle\gamma = \int_{-\infty}^\infty e^{-s^2}\,ds$.
(ii) $\displaystyle\gamma^2 = \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-x^2-y^2}\,dx\,dy$.
(iii) $\displaystyle\gamma^2 = \int_0^{2\pi} \int_0^\infty re^{-r^2}\,dr\,d\theta$.
(iv) $\gamma = \sqrt{\pi}$.

2.8. Define a map f from $\mathbf{R}^2$ to $\mathbf{R}^3$ by

$$f(u,v) = \Big(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, \frac{u^2+v^2-1}{u^2+v^2+1}\Big).$$

(i) Show that $(x,y,z) = f(u,v)$ is contained in the unit sphere about the origin.
(ii) Show that $f(u,v)$ is the intersection point of the unit sphere and the line through the points $(0,0,1)$ and $(u,v,0)$. (Draw a picture! The inverse map from the sphere to the plane $\mathbf{R}^2$ is called *stereographic projection* from the point $(1,0,0)$.)

(iii) Compute $df(u, v)$.

2.9. A map $f\colon \mathbf{R}^n \to \mathbf{R}^m$ is called *even* if $f(-\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x}$ in $\mathbf{R}^n$. Find $df(\mathbf{0})$ if f is even and C^1.

2.10 (spherical coordinates in n dimensions). Let us write a point in $\mathbf{R}^n$ as

$$(r, \theta_1, \theta_2, \dots, \theta_{n-1}).$$

Let P_1 be the function $P_1(r) = r$. For each $n \geq 1$ define a map $P_{n+1}\colon \mathbf{R}^{n+1} \to \mathbf{R}^{n+1}$ by

$$P_{n+1}(r, \theta_1, \theta_2, \dots, \theta_n) = \big((\cos\theta_n)P_n(r, \theta_1, \theta_2, \dots, \theta_{n-1}), r\sin\theta_n\big).$$

(This is an example of a *recursive* definition. If you know P_1, you can compute P_2, and then P_3, etc.)

(i) Show that P_2 and P_3 are the usual polar, resp. spherical coordinates on $\mathbf{R}^2$, resp. $\mathbf{R}^3$.
(ii) Give an explicit formula for P_4.
(iii) Show that the Jacobi matrix of P_{n+1} is the $(n+1) \times (n+1)$-matrix

$$dP_{n+1} = \begin{pmatrix} A & u \\ v & w \end{pmatrix},$$

where A is an $n \times n$-matrix, u is a column vector, v is a row vector and w is a function given respectively by

$$A = (\cos\theta_n)\,dP_n, \qquad u = -(\sin\theta_n)P_n,$$
$$v = \begin{pmatrix}\sin\theta_n & 0 & 0 & \dots & 0\end{pmatrix}, \qquad w = r\cos\theta_n.$$

(iv) Show that

$$\det dP_{n+1} = r\cos^{n-1}\theta_n \det dP_n$$

for $n \geq 1$. (Expand $\det dP_{n+1}$ with respect to the last row, using the formula in (iii). This is a little tricky. Perhaps you should begin by examining the cases $n = 1, 2$. But even if you can't do this part, you should be able to do the next two.)
(v) Using the formula in (iv) calculate $\det dP_n$ for $n = 1, 2, 3, 4$.
(vi) Find an explicit formula for $\det dP_n$ for general n.