

Dr M. Dörrzapf
Lent Term 2006

Vector Calculus
Mathematical Tripos, Part IA
Lecture Notes

I INTRODUCTION AND NOTATION

This course is about functions of more than one variable and develops the calculus of scalar and vector quantities in two and three dimensions. A sound knowledge of these topics is a vital prerequisite for almost all the later courses in applied mathematics and theoretical physics. It is an applied course, meaning that you are expected to be able to apply techniques, but we will not necessarily prove rigorously that they work, this will be done in future Analysis courses. In the first part of the course, the idea of integration is extended from \mathbb{R} to \mathbb{R}^2 and \mathbb{R}^3 (with an obvious extension to higher dimensions): integrals along the x-axis are replaced by integrals over curves, surfaces and volumes. Then the idea of differentiation is extended to vectors (div, grad and curl), which is a basic tool in many areas of theoretical physics (such as Electromagnetism and Fluid Dynamics). Two important theorems are introduced, namely the divergence theorem and Stokes' theorem; in both cases, an integral over a region (in \mathbb{R}^3 and in \mathbb{R}^2 , respectively) is converted to an integral over the boundary of the region. All the previous ideas are then applied to Laplace's equation $\nabla^2\phi = 0$, which is one of the most important equations in the whole of mathematics and physics. Finally, the notion of a vector is generalised to that of a tensor. A vector can be thought of as a 3×1 matrix that carries physical information: namely, magnitude and direction. This information is preserved when the axes are rotated only if the components change according to a certain rule. Very often, it is necessary to describe physical quantities using a 3×3 matrix (or even a $3 \times 3 \times 3 \dots$ matrix). Such a quantity is called a tensor if its components transform according to a certain rule when the axes are rotated. This rule means that the physical information in the tensor (essentially the eigenvalues) is preserved.

By the end of this course, you should be able to manipulate, and solve problems using vector operators; be able to calculate line, surface and volume integrals in \mathbb{R}^3 , using Stokes' theorem and the divergence theorem; be able to solve Laplace's equation in simple cases, and be able to prove standard uniqueness theorems for Laplace's and related equations; understand the notion of a tensor and the general properties of tensors in simple cases.

The course is structured in the following way:

I INTRODUCTION AND NOTATION

Introduction, books, notation, important results from Part IA Differential Equations.

II CURVES IN \mathbb{R}^3

Parameterised curves and arc length, tangents and normals to curves in \mathbb{R}^3 , the radius of curvature. [1 hour]

III INTEGRATION IN \mathbb{R}^2 AND \mathbb{R}^3

Line integrals. Surface and volume integrals: definitions, examples using Cartesian, cylindrical and spherical coordinates; change of variables. [4 hours]

IV VECTOR OPERATORS

Directional derivatives. The gradient of a real-valued function: definition; interpretation as normal to level surfaces; examples including the use of cylindrical, spherical and general orthogonal curvilinear coordinates; conservative fields. Divergence, curl and ∇^2 in Cartesian coordinates, examples; formulae for these operators (statement only) in cylindrical, spherical and general orthogonal curvilinear coordinates. Solenoidal fields and irrotational fields. Vector derivative identities. [5 hours]

V INTEGRATION THEOREMS

Green's theorem in the plane, Divergence theorem, Stokes' theorem, Green's first and second theorem: statements; informal proofs; examples; application to fluid dynamics and electromagnetism. [5 hours]

VI LAPLACE'S EQUATION

Laplace's equation in \mathbb{R}^2 and \mathbb{R}^3 : uniqueness theorem and maximum principle. Solution of Poisson's equation by Gauss' method (for spherical and cylindrical symmetry) and as an integral. [4 hours]

VII CARTESIAN TENSORS IN \mathbb{R}^3

Tensor transformation laws, addition, multiplication, contraction, with emphasis on tensors of second rank. Isotropic second and third rank tensors. Symmetric and antisymmetric tensors. Revision of principal axes and diagonalization. Quotient theorem. Examples including inertia and conductivity. [5 hours]

You will find in the present lecture notes all relevant definitions, theorems and important equations discussed in the lectures. However the notes are not meant to give a full presentation of the course material neither are they intended to replace attendance of the lectures. The notes also contain worked examples at the end of each chapter. These notes and supporting material can be found at http://www.damtp.cam.ac.uk/user/md131/vector_calculus. Please send corrections and comments for improvements to md131@cam.ac.uk.

Books

- S. J. Colley. *Vector Calculus*. Prentice Hall 2005.
 P.C. Matthews. *Vector Calculus*. Springer 1998.
 K.F. Riley, M.P. Hobson and S.J. Bence. *Mathematical Methods for Physicists and Engineers*. Cambridge University Press 1998.
 J.E. Marsden and A.J. Tromba. *Vector Calculus*. Freeman 1996.
 D.E. Bourne and P.C. Kendall. *Vector Analysis and Cartesian Tensors*. Chapman and Hall 1992.
 T.M. Apostol. *Calculus*. Wiley 1975.
 H.M. Schey. *Div, grad, curl and all that: an informal text on vector calculus*. Norton 1996.
 H. Anton. *Calculus*. Wiley 1995.
 M.L. Boas. *Mathematical Methods in the Physical Sciences*. Wiley 1983.
 E. Kreyszig. *Advanced Engineering Mathematics*. Wiley 1999.
 M.R. Spiegel. *Schaum's outline of Vector Analysis*. McGraw Hill 1974.

Notation

The following is a table of the notation used in this course.

\forall	for all	\mathbb{R}	real numbers
\exists	there exists	\mathbb{C}	complex numbers
\exists_1	there exists a unique	\mathbb{Z}	integers
\nexists	there is no	\mathbb{N}	positive integers, $\mathbb{N} := \{1, 2, 3, \dots\}$
$:=$	defined equal	\mathbb{N}_0	non-negative integers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$
\Rightarrow	implication	$[a, b]$	closed interval $x \in \mathbb{R}$: $a \leq x \leq b$
\therefore	therefore	(a, b)	open interval $x \in \mathbb{R}$: $a < x < b$
\because	because	\in	contained in
\Leftrightarrow	equivalence	\notin	not contained in
$:\Leftrightarrow$	defined equivalent	\wedge	and
$\stackrel{!}{=}$	should be equal to	\vee	or
$:$	such that	\perp	perpendicular
\equiv	identical	\square	end of proof
\approx	approximately equal	ζ	contradiction

Examples

$$A := \{x \in \mathbb{R} : x < 2\}$$

$$A \subset B \quad :\Leftrightarrow \quad x \in A \Rightarrow x \in B$$

$$A = B \quad :\Leftrightarrow \quad A \subset B \wedge B \subset A$$

$$\{x \in \mathbb{R} : 1 \leq x \leq 2\} =: [1, 2]$$

A is defined to be the set of all real x such that $x < 2$.

A is a subset of B is defined to be equivalent to 'x is contained in A implies that x is also contained in B'. the set A equals the set B is defined to be equivalent to 'A is a subset of B and B is a subset of A'.

the closed interval [1, 2] is defined to be the set of real numbers between and including 1 and 2.

Labels

Equations are referred to by 'EQ. number', e.g. EQ. 1, theorems and definitions by [number] e.g. [1] and sections by 'SEC. number' e.g. SEC. 1. Finally, figures are referred to by FIG. iii. For this course non-examinable sections are indicated with a * behind the number.

Vectors

We will denote vectors using underlining e.g. \underline{x} in handwritten formulae and bold face type e.g. \mathbf{x} in print. In this course we will mainly operate in \mathbb{R}^2 or \mathbb{R}^3 . The i th component of the vector \mathbf{x} , i.e. $(\mathbf{x})_i$, is always denoted x_i . The notation \mathbf{x}_i would mean the i th of a set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$. Unless otherwise stated, summation convention applies whenever there is a repeated suffix. The standard scalar product is defined by $\mathbf{x} \cdot \mathbf{y} := x_i y_i$. The norm $\|\mathbf{x}\|$ is defined to be the *Euclidean*^a norm $\|\mathbf{x}\| := \sqrt{\sum_{i=1}^n x_i^2}$. Vectors \mathbf{x} are usually assumed to be column vectors

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

but in some instances we will need \mathbf{x} to be a row vector. If it is required for clarity then may explicitly write the *transpose* of the vector: $\mathbf{x}^T = (x_1, \dots, x_n)$ but otherwise we may label row vectors also just by $\mathbf{x} = (x_1, \dots, x_n)$.

^aEuclid of Alexandria, about 325BC-265BC.

The vectors of an orthonormal triad are denoted by \mathbf{e}_i , where

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} .$$

The vector product between two vectors is defined by

$$\mathbf{x} \wedge \mathbf{y} \equiv \mathbf{x} \times \mathbf{y} = -\mathbf{y} \times \mathbf{x} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} ; \quad \text{and} \quad (\mathbf{x} \times \mathbf{y})_i = \epsilon_{ijk} x_j y_k .$$

The identity

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$$

can be obtained using the identity

$$\epsilon_{kij} \epsilon_{klm} = \delta_{il} \delta_{jm} - \delta_{im} \delta_{jl} .$$

The position vector is denoted \mathbf{r} or also \mathbf{x} , and in Cartesians $\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ or $(\mathbf{r})_i = x_i$. The length of the position vector is r , where $r^2 = \mathbf{r} \cdot \mathbf{r} = \mathbf{x} \cdot \mathbf{x}$.

Cylindrical polar co-ordinates (ρ, ϕ, z) in \mathbb{R}^3 .

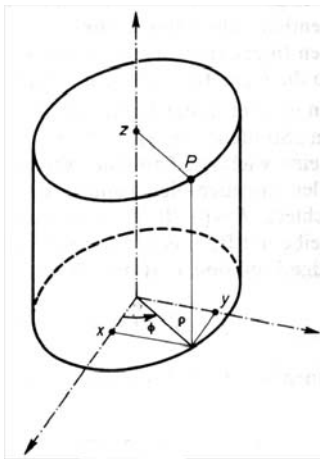


FIG. i Cylindrical polar co-ordinates

In cylindrical polar co-ordinates the position vector \mathbf{x} is given in terms of a radial distance ρ from an axis \mathbf{k} , a polar angle ϕ , and the distance z along the axis. With respect to Cartesian axes, the position vector is

$$\mathbf{x} = (\rho \cos \phi, \rho \sin \phi, z), \quad (1)$$

where $0 \leq \rho < \infty$, $0 \leq \phi \leq 2\pi$ and $-\infty < z < \infty$.

In terms of the orthonormal basis related to the cylindrical coordinates, \mathbf{e}_ρ , \mathbf{e}_ϕ , \mathbf{e}_z , the position vector is given by

$$\mathbf{r} = \rho \mathbf{e}_\rho + z \mathbf{e}_z .$$

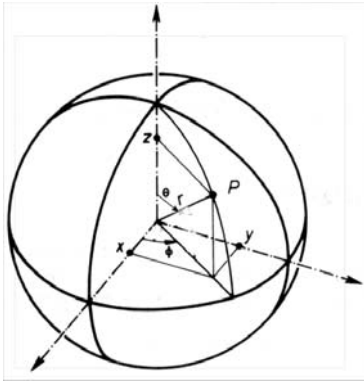
Spherical polar co-ordinates (r, θ, ϕ) in \mathbb{R}^3 .

FIG. ii Spherical polar co-ordinates

In spherical polar co-ordinates the position vector \mathbf{x} is given in terms of a radial distance r from the origin, a 'latitude' angle θ , and a 'longitude' angle ϕ . With respect to Cartesian axes, the position vector is

$$\mathbf{x} = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta), \quad (2)$$

where $0 \leq r < \infty$, $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$.

In terms of the orthonormal basis $\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi$, the position vector is given by

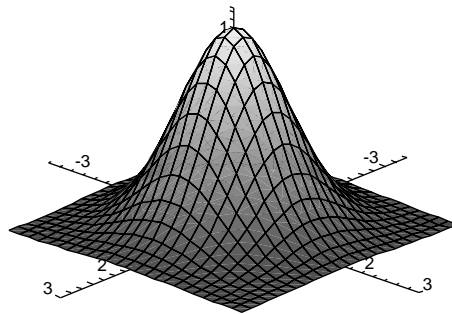
$$\mathbf{x} = r \mathbf{e}_r.$$

Functions

The functions we consider in this course map $\mathbf{x} \in \mathbb{R}^n$ to a vector in \mathbb{R}^m which we denote by $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Usually m and n will either be 1, 2 or 3. The domain of a function f is denoted by $\text{dom} f$. For $m = 1$ we simply say a *real valued* function and for $m > 1$ we call the function *vector valued*. Most of the time the domain and the range of a function is self-understood. If a function is defined by simply giving an expression for $f(\mathbf{x})$ then we mean that $\text{dom} f$ is the maximal possible domain and the range is self-understood. For example, the function

$$f = \begin{cases} \mathbb{R}^2 \rightarrow \mathbb{R} \\ \begin{pmatrix} x \\ y \end{pmatrix} \mapsto f(x, y) = e^{-(x^2+y^2)} \end{cases} \quad (3)$$

can simply be written as $f(x, y) = e^{-(x^2+y^2)}$. Since $\text{dom} f = \mathbb{R}^2$ and the range is \mathbb{R} we can sketch the graph of such a function in a three dimensional co-ordinate system. The graph of this particular function is given in the following figure.

FIG. iii Graph of $f(x, y) = e^{-(x^2+y^2)}$

Sometimes it may be necessary to define functions by sticking different branches together or by defining certain points separately. The notation we use in this case becomes clear in the following example.

$$f(x, y) = \begin{cases} \frac{xy^2}{x^2+y^4} & , (x, y) \neq (0, 0) \\ 0 & , (x, y) = (0, 0) \end{cases} \quad (4)$$

A function $f(x, y)$ as simple as the one in Eq. 4 can in fact have a very complicated graph as the following figure shows.

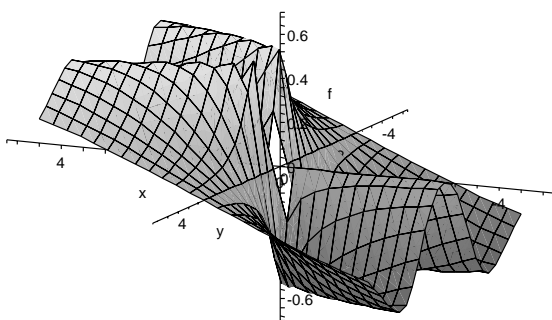


FIG. iv $\frac{xy^2}{x^2+y^4}$

Partial derivatives

In the Part IA course on Differential Equations the concept of differentiation has been generalised to functions $f(\mathbf{x})$ of more than one variable leading to the concept of partial differentiation. A sound knowledge of these results is required for this course. We will therefore give the following summary of the main results.

Simple partial derivatives (e.g. $\frac{\partial f(x,y,z)}{\partial x}$) are effectively ordinary one dimensional derivatives where we simply treat all variables but one as constants. As a result all differentiation rules we know from one dimensional derivatives can easily be applied to partial derivatives. For example:

$$\begin{aligned} \frac{\partial x^2 \sin(ye^z)}{\partial x} &= 2x \sin(ye^z) , \\ \frac{\partial x^2 \sin(ye^z)}{\partial y} &= x^2 \cos(ye^z) e^z , \\ \frac{\partial \frac{\cos(x+y)}{x+y}}{\partial x} &= -\frac{\sin(x+y)}{x+y} - \frac{\cos(x+y)}{(x+y)^2} . \end{aligned}$$

Instead of $\frac{\partial f(x,y,z)}{\partial x}$ we may also write f_x or simply f_1 where the latter should not be confused with the first component of a vector \mathbf{f} . In case we only want to keep some of the other

variables constant but not all of them then we have to say explicitly which variables are being kept constant e.g. $\left(\frac{\partial f(x,y,z)}{\partial x}\right)_y$ keeps y invariant but not z and therefore takes account of any changes in z if x is being varied. The standard partial derivative $\frac{\partial f(x,y,z)}{\partial x}$ can in this notation also be written as $\left(\frac{\partial f(x,y,z)}{\partial x}\right)_{y,z}$. If none of the variables are being kept constant then we call it the total derivative e.g. $\frac{df(x,y,z)}{dx}$. The chain rule relates these derivatives to simple partial derivatives:

$$\frac{d}{dt}f(x_1(t), x_2(t), \dots, x_n(t)) = \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dt} + \dots + \frac{\partial f}{\partial x_n} \frac{dx_n}{dt}, \quad (5)$$

and more generally for $f(x_1, \dots, x_n)$:

$$\left(\frac{\partial f}{\partial t}\right)_{x_{i_1}, \dots, x_{i_k}} = \frac{\partial f}{\partial x_1} \left(\frac{\partial x_1}{\partial t}\right)_{x_{i_1}, \dots, x_{i_k}} + \dots + \frac{\partial f}{\partial x_n} \left(\frac{\partial x_n}{\partial t}\right)_{x_{i_1}, \dots, x_{i_k}}. \quad (6)$$

A particular case of EQ. 6 for the function $f(x, y, z)$ is for example

$$\begin{aligned} \left(\frac{\partial f}{\partial x}\right)_z &= \frac{\partial f}{\partial x} \left(\frac{\partial x}{\partial x}\right)_z + \frac{\partial f}{\partial y} \left(\frac{\partial y}{\partial x}\right)_z + \frac{\partial f}{\partial z} \left(\frac{\partial z}{\partial x}\right)_z \\ &= \frac{\partial f}{\partial x} 1 + \frac{\partial f}{\partial y} \left(\frac{\partial y}{\partial x}\right)_z + \frac{\partial f}{\partial z} 0 = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \left(\frac{\partial y}{\partial x}\right)_z. \end{aligned} \quad (7)$$

The Analysis of the rate of change of a multi dimensional function can easily become very difficult due to the fact that in multi dimensions we can approach points not only along straight lines but also along complicated curves such as spirals etc.. If we assume that the function is sufficiently smooth then these difficulties will not occur. We will leave it to future Analysis courses to study this in more detail. For this course we shall always assume that all our functions are continuous and sufficiently smooth and well-behaved (which basically means that the functions are continuous and all the partial derivatives we need exist and are again continuous). In this case a range of nice theorems hold:

1 Schwarz's^b theorem

Mixed partial derivatives are independent of the order of differentiation e.g.:

$$\frac{\partial^2 f(x, y)}{\partial x \partial y} = \frac{\partial^2 f(x, y)}{\partial y \partial x}. \quad (8)$$

2 Increments of real valued functions:

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1} h_1 + \dots + \frac{\partial f(\mathbf{x})}{\partial x_n} h_n + r(\mathbf{h}) \quad (9)$$

with $\frac{r(\mathbf{h})}{\|\mathbf{h}\|} \rightarrow 0$ as $\mathbf{h} \rightarrow \mathbf{0}$.

^b Hermann Amandus Schwarz, 1843-1921.

Alternatively, we can write Eq. 9 as a matrix product or as a scalar product:

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right) \mathbf{h} + r(\mathbf{h}) = \nabla f(\mathbf{x}) \cdot \mathbf{h} + r(\mathbf{h}), \quad (10)$$

with the *gradient* of a real valued function defined as

$$\nabla f(\mathbf{x}) := \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

Alternatively to the notation ∇f we may also use $grad(f)$.

The amazing fact of [2] is that $r(\mathbf{h})$ does not only tend to 0 as $\mathbf{h} \rightarrow \mathbf{0}$, it still tends to 0 even if divided by $\|\mathbf{h}\|$. Therefore $f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{h}$ approximates f to first order, in other words, $f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{h} + \mathcal{O}(\|\mathbf{h}\|)$. Analogously to the one dimensional case we say that the function f can be *approximated linearly*^c.

For vector valued functions $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$

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

the gradient in Eq. 10 becomes a matrix, the *Jacobi matrix* $J_{\mathbf{f}}(\mathbf{x})$:

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = J_{\mathbf{f}}(\mathbf{x})\mathbf{h} + \mathbf{r}(\mathbf{h}), \quad (11)$$

with

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

We then find $\mathbf{r}(\mathbf{h}) = (r_1(\mathbf{h}), \dots, r_m(\mathbf{h}))^T$ such that each $\frac{r_i(\mathbf{h})}{\|\mathbf{h}\|} \rightarrow 0$ as $\mathbf{h} \rightarrow \mathbf{0}$, in other words $\frac{\mathbf{r}(\mathbf{h})}{\|\mathbf{h}\|} \rightarrow \mathbf{0}$ as $\mathbf{h} \rightarrow \mathbf{0}$. We therefore say $\mathbf{r}(\mathbf{h}) = \mathcal{O}(\|\mathbf{h}\|)$.

Finally, Taylor's theorem^d generalises Eq. 11 to higher order:

3 Taylor's theorem

$$\begin{aligned} \mathbf{f}(\mathbf{x} + \mathbf{h}) &= \mathbf{f}(\mathbf{x}) + \sum_i \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_i} h_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial x_i \partial x_j} h_i h_j + \dots \\ &\dots + \frac{1}{k!} \sum_{i_1, \dots, i_k} \frac{\partial^k \mathbf{f}(\mathbf{x})}{\partial x_{i_1} \dots \partial x_{i_k}} h_{i_1} \dots h_{i_k} + \mathbf{r}_k(\mathbf{h}), \end{aligned} \quad (12)$$

^cIt should be noted that in the multi dimensional case the linear approximation of differentiable functions is valid in whichever way \mathbf{h} approaches $\mathbf{0}$. Whether on a straight line, on a spiral or on any complicated curve, the linear expression always approximates the increment. Compared to the one dimensional case $f(x+h) = f(x) + f'(x)h + \mathcal{O}(h)$ it is now the gradient $\nabla f(\mathbf{x})$ that plays the rôle of the derivative. The amazing property of smooth functions in higher dimensions is that the function can be approximated linearly in the multidimensional sense by a vector of partial derivatives. Considering that partial derivatives are just limits along straight lines it is remarkable that the set of these straight line limits describes a differentiable function in whichever way we approach the point \mathbf{x} , even if it is not on a straight line. On the other hand, as shall be shown in Analysis, for a function which is not differentiable the information contained in the partial derivatives is usually not enough to describe the behaviour of the function if we approach \mathbf{x} along a path which is not parallel to a co-ordinate axis not even along a straight path.

^dBrook Taylor, 1685-1731, St John's College (1703-1709, LL.B. 1709).

with $\frac{\mathbf{r}_k(\mathbf{h})}{\|\mathbf{h}\|^k} \rightarrow \mathbf{0}$ as $\mathbf{h} \rightarrow \mathbf{0}$.

The linear term of the Taylor expansion Eq. 12 is of course just the Jacobi^e matrix acting on \mathbf{h} . For many of our applications the linear term will be sufficient:

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + J_{\mathbf{f}}(\mathbf{x})\mathbf{h} + \dots \quad (13)$$

The real valued case $m = 1$ of [3] is particularly important for us. This case can also be written in a simple way using the matrix of second order partial derivatives, the so-called *Hessian*^f $H_{\mathbf{f}}(\mathbf{x})$.

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^T H_{\mathbf{f}}(\mathbf{x}) \mathbf{h} + \dots, \quad (14)$$

with

$$H_{\mathbf{f}}(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}.$$

Examples

Example 1.1 Consider the function $f(x, y) = x^2 + yx$. Then

$$\left(\frac{\partial f}{\partial x}\right)_y = 2x + y, \quad \left(\frac{\partial f}{\partial y}\right)_x = x.$$

Note that it is not really necessary to indicate which variable is being held constant, since $f(x, y)$ is unambiguously a function of x and y . There would be no confusion in writing

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

Example 1.2 Let $f(x, y, z) = r$ where $r = |\mathbf{x}| = (x^2 + y^2 + z^2)^{\frac{1}{2}}$. Then

$$\left(\frac{\partial r}{\partial x}\right)_{y,z} = \frac{\partial}{\partial x}(x^2 + y^2 + z^2)^{\frac{1}{2}} = \frac{2x}{2(x^2 + y^2 + z^2)^{\frac{1}{2}}} = \frac{x}{r}, \quad (15)$$

and similarly

$$\left(\frac{\partial r}{\partial y}\right)_{x,z} = \frac{y}{r}, \quad \left(\frac{\partial r}{\partial z}\right)_{x,y} = \frac{z}{r}. \quad (16)$$

^e Carl Gustav Jacob Jacobi, 1804-1851.

^f Ludwig Otto Hesse, 1811-1874.

^gNote that Schwarz's theorem says that $H_{\mathbf{f}}(\mathbf{x})$ is a symmetric matrix and therefore we know from Part IA Algebra & Geometry that $H_{\mathbf{f}}(\mathbf{x})$ can always be diagonalised with real eigenvalues and using an orthonormal basis transformation.

Example 1.3 We will find the partial derivatives of the Cartesian coordinate x_1 . We have

$$\left(\frac{\partial x_1}{\partial x_1}\right)_{x_2, x_3} = 1, \quad \left(\frac{\partial x_1}{\partial x_2}\right)_{x_1, x_3} = 0, \quad \text{etc.}$$

Thus

$$\frac{\partial x_1}{\partial x_j} = \begin{cases} 1 & \text{if } j = 1 \\ 0 & \text{if } j \neq 1 \end{cases} = \delta_{j1},$$

where δ_{jk} is the Kronecker delta.

A similar result holds for any of the coordinates. For the i th coordinate, we have

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij}. \quad (17)$$

We can now use this result to obtain (15) and (16) more neatly. Starting with that $r^2 = x_i x_i$ (don't forget we are using summation convention), we see that

$$2r \frac{\partial r}{\partial x_k} = \frac{\partial(r^2)}{\partial x_k} = \frac{\partial(x_i x_i)}{\partial x_k} = 2x_i \frac{\partial x_i}{\partial x_k} = 2x_i \delta_{ik} = 2x_k.$$

Dividing through by $2r$ gives

$$\frac{\partial r}{\partial x_k} = \frac{x_k}{r}, \quad (18)$$

which, in terms of suffix notation, is the same as (15) and (16).

Example 1.4 Let

$$f(x, y) = x^2 + y, \quad \mathbf{x}(t) = (x, y) = (t^2, e^t),$$

and let $F(t) = f(\mathbf{x}(t))$. We will calculate $\frac{dF}{dt}$ in two ways.

1. Using the chain rule. We have

$$\left(\frac{\partial f}{\partial x}\right)_y = 2x, \quad \left(\frac{\partial f}{\partial y}\right)_x = 1, \quad \frac{dx}{dt} = 2t, \quad \frac{dy}{dt} = e^t.$$

Hence from using the chain rule

$$\frac{dF}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} = 2x \cdot 2t + e^t = 4t^3 + e^t.$$

2. By direct substitution. We have

$$F(t) = t^4 + e^t, \quad \text{and hence} \quad \frac{dF}{dt} = 4t^3 + e^t.$$

The second (direct) method may seem much easier, but the chain rule allows us to find a formula for the rate of change of a general function $f(x, y)$ along the curve $\mathbf{x}(t) = (t^2, e^t)$:

$$\frac{df(\mathbf{x}(t))}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} = 2t f_x(t^2, e^t) + e^t f_y(t^2, e^t).$$

Example 1.5 We will find the gradient at $x = 1, y = 1$ of the curve of the family given by $f(x, y) = c$, where $f(x, y) = \sin(\pi xy) + e^y$. We have

$$\frac{dy}{dx} = -\frac{f_x(1, 1)}{f_y(1, 1)} = -\frac{\pi y \cos(\pi xy)}{\pi x \cos(\pi xy) + e^y} = -\frac{-\pi}{-\pi + e}.$$

Example 1.6 Consider the coordinate transformation (actually a rotation of the co-ordinate axes through a fixed angle θ)

$$x = u \cos \theta - v \sin \theta, \quad y = u \sin \theta + v \cos \theta.$$

We will calculate how the derivatives of an arbitrary function $f(x, y)$ can be calculated in the new coordinates. Let $F(\mathbf{u}) = f(\mathbf{x}(\mathbf{u}))$. Then

$$\begin{aligned} \frac{\partial F}{\partial u} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial u} = \cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y}, \\ \frac{\partial F}{\partial v} &= \frac{\partial f}{\partial x} \frac{\partial x}{\partial v} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial v} = -\sin \theta \frac{\partial f}{\partial x} + \cos \theta \frac{\partial f}{\partial y}. \end{aligned}$$

Sometimes, indeed often, $F(u, v) = f(x(u, v), y(u, v))$ is written (incorrectly) by applied mathematicians (but never by pure mathematicians) as $f(u, v)$. Thus f refers to the value of the function, whatever variables are used. The reason for doing this is that in many cases the function referred to has physical significance, such as temperature, and it is very inconvenient to refer to it by different letters according to which variables it is regarded as a function of. Since all the other variables have standard names too, this abuse of notation never becomes confusing. Then the above becomes

$$\begin{aligned} \left(\frac{\partial f}{\partial u}\right)_v &= \cos \theta \left(\frac{\partial f}{\partial x}\right)_y + \sin \theta \left(\frac{\partial f}{\partial y}\right)_x, \\ \left(\frac{\partial f}{\partial v}\right)_u &= -\sin \theta \left(\frac{\partial f}{\partial x}\right)_y + \cos \theta \left(\frac{\partial f}{\partial y}\right)_x, \end{aligned}$$

Using this notation, it is of course essential to state explicitly the variable(s) that are held constant in each partial differentiation.

Example 1.7 Let $f(x, y)$ be an arbitrary function, and let $F(\rho, \phi) = f(\rho \cos \phi, \rho \sin \phi)$ (i.e. F has the same value as f but is calculated in plane polars rather than Cartesians). We will transform the equation

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

into polar coordinates. First, note that

$$\frac{\partial f}{\partial x} = \frac{\partial \rho}{\partial x} \frac{\partial F}{\partial \rho} + \frac{\partial \phi}{\partial x} \frac{\partial F}{\partial \phi} = \cos \phi \frac{\partial F}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial F}{\partial \phi},$$

or, omitting f and F ,

$$\frac{\partial}{\partial x} = \cos \phi \frac{\partial}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial}{\partial \phi}.$$

Hence

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x} \frac{\partial f}{\partial x} = \left(\cos \phi \frac{\partial}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial}{\partial \phi} \right) \left(\cos \phi \frac{\partial F}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial F}{\partial \phi} \right) \\ &= \cos \phi \frac{\partial}{\partial \rho} \left(\cos \phi \frac{\partial F}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial F}{\partial \phi} \right) - \frac{\sin \phi}{\rho} \frac{\partial}{\partial \phi} \left(\cos \phi \frac{\partial F}{\partial \rho} - \frac{\sin \phi}{\rho} \frac{\partial F}{\partial \phi} \right) \end{aligned}$$

$$\begin{aligned}
&= \cos^2 \phi \frac{\partial^2 F}{\partial \rho^2} + \frac{\sin \phi \cos \phi}{\rho^2} \frac{\partial F}{\partial \phi} - \frac{\sin \phi \cos \phi}{\rho} \frac{\partial}{\partial \rho} \frac{\partial F}{\partial \phi} + \frac{\sin^2 \phi}{\rho} \frac{\partial F}{\partial \rho} \\
&\quad - \frac{\sin \phi \cos \phi}{\rho} \frac{\partial}{\partial \phi} \frac{\partial F}{\partial \rho} + \frac{\sin \phi \cos \phi}{\rho^2} \frac{\partial F}{\partial \phi} + \frac{\sin^2 \phi}{\rho^2} \frac{\partial^2 F}{\partial \phi^2}.
\end{aligned}$$

On assuming the equality of mixed derivatives we obtain

$$\frac{\partial^2 f}{\partial x^2} = \cos^2 \phi \frac{\partial^2 F}{\partial \rho^2} + \frac{\sin^2 \phi}{\rho} \frac{\partial F}{\partial \rho} - \frac{2 \sin \phi \cos \phi}{\rho} \frac{\partial^2 F}{\partial \rho \partial \phi} + \frac{2 \sin \phi \cos \phi}{\rho^2} \frac{\partial F}{\partial \phi} + \frac{\sin^2 \phi}{\rho^2} \frac{\partial^2 F}{\partial \phi^2}.$$

We can calculate $\frac{\partial^2 f}{\partial y^2}$ by observing that $x = \rho \sin(\phi + \frac{\pi}{2})$ and $y = -\rho \cos(\phi + \frac{\pi}{2})$. Hence after applying the transformations $x \rightarrow -y$, $y \rightarrow x$, $\phi \rightarrow (\phi + \frac{\pi}{2})$, we obtain

$$\frac{\partial^2 f}{\partial y^2} = \sin^2 \phi \frac{\partial^2 F}{\partial \rho^2} + \frac{\cos^2 \phi}{\rho} \frac{\partial F}{\partial \rho} + \frac{2 \sin \phi \cos \phi}{\rho} \frac{\partial^2 F}{\partial \rho \partial \phi} - \frac{2 \sin \phi \cos \phi}{\rho^2} \frac{\partial F}{\partial \phi} + \frac{\cos^2 \phi}{\rho^2} \frac{\partial^2 F}{\partial \phi^2}.$$

and hence that

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 F}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial F}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 F}{\partial \phi^2}.$$

II CURVES IN \mathbb{R}^3

In Part IA Differential Equations you studied how differentiation can be generalised to higher dimensions simply by transferring the same limit technique from one dimension to higher dimensions which ultimately led to the partial derivative. What about integration? The integral over a one dimensional function is a limit process which chops up the area underneath the graph of the function in lots of little rectangles and sums over their areas. Could this concept or a similar mechanism easily be applied to functions of more than one variable and what would this give us? Before we look at the process of dividing areas up into little rectangles in more detail and then applying this to volumes and surfaces which are not flat we will first study a much simpler structure that is length of a path.

Let us take the circle $x^2 + y^2 = 1r^2$ which is obviously described by the path $\gamma(t) = (r \sin t, r \cos t)$ as the time t runs from 0 to 2π . What is the length of this path (i.e. the circumference of the circle)? We could take the perimeter of the *inscribed* square which is $4\sqrt{2}r$ as an approximation of the circumference of the circle but it is obvious that the real circumference must be bigger. Obviously, the perimeter of the inscribed pentagon would be a better approximation and indeed the perimeter of the inscribed hexagon is an even better approximation. In fact, the perimeter of the inscribed regular n -gon becomes a better and better approximation as n gets bigger and should ultimately tend to the correct circumference in the limit $n \rightarrow \infty$. The perimeter of the n -gon is given by $\sum_{i=1}^n \left\| \begin{pmatrix} r \cos \frac{2\pi i}{n} \\ r \sin \frac{2\pi i}{n} \end{pmatrix} - \begin{pmatrix} r \cos \frac{2\pi(i-1)}{n} \\ r \sin \frac{2\pi(i-1)}{n} \end{pmatrix} \right\|$. Using the linear term of the Taylor expansion of \cos and \sin this is approximately equal to $\sum_{i=1}^n \left\| \begin{pmatrix} -r \sin \frac{2\pi(i-1)}{n} \\ r \cos \frac{2\pi(i-1)}{n} \end{pmatrix} \right\| \frac{2\pi}{n}$. For large n the higher order terms are negligible and therefore in the limit we expect this to give the circumference, and indeed

$$\sum_{i=1}^n \left\| \begin{pmatrix} -r \sin \frac{2\pi(i-1)}{n} \\ r \cos \frac{2\pi(i-1)}{n} \end{pmatrix} \right\| \frac{2\pi}{n} = \sum_{i=1}^n \frac{2\pi r}{n} = 2\pi r .$$

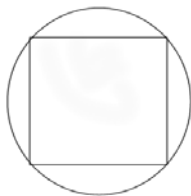


FIG. v Inscribed square

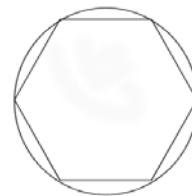
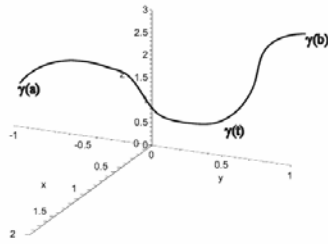


FIG. vi Inscribed hexagon

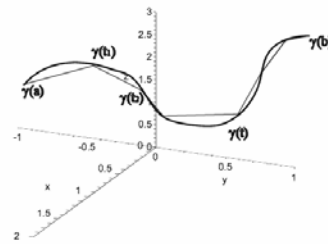
This limit process seems to rely on taking points on the path and replacing the arcs of the path connecting these points by straight lines. The sum of the Euclidean lengths of these lines would then approximate the length of the path. These ideas will form the basis of the following section which will lead to the definition of the length of a general curve.

1 Parametrised curves

We now want to formalise the ideas developed above for a general (but continuous and smooth) curve in \mathbb{R}^n . For our applications n would either be 2 or 3 but everything we are going to say would of course also work for curves in spaces of dimensions higher than 3.

FIG. vii A path $\gamma(t)$ in \mathbb{R}^3 .

Following our initial remarks, we take a partition \mathcal{Z} of the interval $[a, b]$ with $t_0 = a$ and $t_b = b$ and $t_0 < t_1 < \dots < t_n$. \mathcal{Z} could be an equal distance partition but does not need to be, therefore it is useful to define the partition length $\|\mathcal{Z}\|$ to be the maximum $\|\mathcal{Z}\| := \max |t_i - t_{i-1}|$. We can now connect the points $\gamma(t_i)$ with straight lines as in FIG.viii and adding up the lengths of these straight lines $L(\gamma, \mathcal{Z}) := \sum_{i=1}^n \|\gamma(t_i) - \gamma(t_{i-1})\|$ would obviously approximate the length of the path γ (and in fact would always be smaller than the real length of γ).

FIG. viii Partitioning a path $\gamma(t)$ in \mathbb{R}^3 .

If we consider a sequence of partitions \mathcal{Z} of $[a, b]$ with the partition lengths $\|\mathcal{Z}\|$ becoming smaller and smaller and ultimately tending to 0 then the sequence of the corresponding $L(\gamma, \mathcal{Z})$ which are approximations for the length of γ would obviously tend to what we should define as the real *length* of the path $\gamma(t)$. In technical terms mathematicians call a path for which this limit exists a *rectifiable path*. We denote such a limit by $\lim_{\|\mathcal{Z}\| \rightarrow 0}$.

4 Length of a path

The limit $\lim_{\|\mathcal{Z}\| \rightarrow 0} L(\gamma, \mathcal{Z})$ of a (rectifiable) path $\gamma(t) : [a, b] \rightarrow \mathbb{R}^n$ is called the *length* $L(\gamma)$ of the path γ and is denoted by

$$\int_{\gamma} ds := \lim_{\|\mathcal{Z}\| \rightarrow 0} L(\gamma, \mathcal{Z}). \quad (19)$$

Let us now assume that each component function $\gamma_i(t)$ of the vector valued function $\gamma(t)$ can be differentiated, then Taylor's theorem tells us that the increment $\gamma(t_k) - \gamma(t_{k-1})$ can be approximated by $\gamma'(t_k)(t_k - t_{k-1})$ and therefore

$$L(\gamma, \mathcal{Z}) = \sum_k \|\gamma(t_k) - \gamma(t_{k-1})\| \quad (20)$$

can be approximated by

$$\sum_k \|\gamma'(t_k)\| (t_k - t_{k-1}). \quad (21)$$

You will show in Analysis that the higher order terms do not matter in a limit where the differences $t_k - t_{k-1}$ tend to 0 and therefore EQ. 20 and EQ. 21 tend to the same value under such a limit. We now of course recognise that the limit of [21] is just the standard integral $\int_a^b \|\gamma'(t)\| dt$. This leads us to the very significant and amazing result that this new and at first rather complicated looking limit we constructed above can in fact be converted into a standard integral provided we can differentiate the path $\gamma(t)$!

5 Length of a differentiable path

If the path $\gamma : [a, b] \rightarrow \mathbb{R}^n$ can be differentiated, then the length of γ is given by:

$$\int_{\gamma} ds = \int_a^b \|\gamma'(t)\| dt . \quad (22)$$

We call $\gamma'(t)$ the velocity, $\|\gamma'(t)\|$ the speed and $\gamma''(t)$ the acceleration of the path $\gamma(t)$.

Example 2.1 The path $\gamma(t) = (t \cos t, t \sin t)$ describes a spiral turning counterclockwise as t runs from 0 to $2\pi n$, $n \in \mathbb{N}$. What is the length of the spiral?

We find the velocity vector as $\gamma' = (\cos t - t \sin t, \sin t + t \cos t)$ and therefore its norm (the speed) is $\|\gamma'\| = \sqrt{1+t^2}$. Therefore $\int_{\gamma} ds = \int_0^{2\pi n} \sqrt{1+t^2} dt = \frac{1}{2}(t\sqrt{1+t^2} + \sinh^{-1} t) \Big|_0^{2\pi n} = \pi n \sqrt{1+4\pi^2 n^2} + \frac{1}{2} \sinh^{-1}(2\pi n)$.

If we have two paths $\gamma_1 : [a, b] \rightarrow \mathbb{R}^n$ and $\gamma_2 : [b, c] \rightarrow \mathbb{R}^n$ with $a < b < c$ and $\gamma_1(b) = \gamma_2(b)$ then we define the sum of the two paths $\gamma_1 \oplus \gamma_2$ as the path $[a, c] \rightarrow \mathbb{R}^n$ with

$$\gamma_1 \oplus \gamma_2(t) := \begin{cases} \gamma_1(t) & , t \in [a, b] \\ \gamma_2(t) & , t \in [b, c] \end{cases} . \quad (23)$$

For the length of $\gamma_1 \oplus \gamma_2$ we obviously find that $L(\gamma_1 \oplus \gamma_2) = L(\gamma_1) + L(\gamma_2)$. In the same way we can restrict a path $\gamma : [a, b] \rightarrow \mathbb{R}^n$ to a shorter interval: $\gamma_t : [a, t] \rightarrow \mathbb{R}^n$ with $\gamma_t(u) = \gamma(u)$ for $a \leq u \leq t \leq b$. The length of the restricted path γ_t will then define a function of t which we want to call the arclength function $s(t)$ of γ with basepoint $\gamma(a)$ (or simply the length function):

$$s(t) = L(\gamma_t) = \int_{\gamma_t} ds . \quad (24)$$

$s(t)$ obviously describes the length of the part of the path γ a particle would travel if time went from time a until time t . Using [5] we immediately find that $s(t) = \int_a^t \|\gamma'(u)\| du$ and therefore $s'(t) = \|\gamma'(t)\|$ which means that $s'(t)$ is simply the speed of $\gamma(t)$.

6 The sum of paths and the length function

If $s(t)$ describes the length of the path γ restricted to the interval $[a, t]$ then $s'(t) = \|\gamma'(t)\|$. If γ is the sum of finitely many paths $\gamma = \gamma_1 \oplus \dots \oplus \gamma_k$ then

$$\int_{\gamma} ds = \int_{\gamma_1} ds + \dots + \int_{\gamma_k} ds . \quad (25)$$

Let Γ be some continuous injective curve in \mathbb{R}^3 . In order to measure the length of the curve Γ we want to describe the curve with a path γ which is not going back and forth on parts of the curve, otherwise the length of the path would obviously not describe the length of the curve. Therefore we require a path to be injective in order to describe the length of

the curve described by the path. It is obvious that two injective paths describing the same curve have the same length. If $\gamma_1(t)$ and $\gamma_2(t)$ are two injective paths both describing Γ with the same starting points then $s_2^{-1}(s_1(t))$ gives the time at which the path γ_2 reaches the point $\gamma_1(t)$. Therefore we find $\gamma_2(s_2^{-1}(s_1(t))) = \gamma_1(t)$. Obviously $s_2^{-1}(s_1(t))$ is a strictly monotonous function^a. Whenever two paths are related to each other by a strictly increasing function $\kappa(t)$: $\gamma_2(\kappa(t)) = \gamma_1(t)$, then we call γ_2 a reparametrisation of γ_1 . Therefore two injective paths describing the same curve Γ are reparametrisations of each other.

An injective path γ is called a *Jordan path*^b and the corresponding curve is called a *simple curve* or a *Jordan arc*. A Jordan path γ describing a Jordan arc Γ is called a *Jordan description* of Γ . In summary, the length function of a Jordan path is always strictly increasing and therefore invertible. If two Jordan paths describe the same Jordan curve then they are reparametrisations of one another with some continuous and strictly monotonous function κ such that $\gamma_2 = \gamma_1 \circ \kappa$. Conversely, any continuous and strictly monotonous function κ generates another Jordan description $\gamma_2 = \gamma_1 \circ \kappa$ from a Jordan description γ_1 of the Jordan arc Γ . It is therefore obvious that we can easily generate all Jordan descriptions of a Jordan arc Γ out of just one of its Jordan descriptions simply by applying suitable reparametrisations. It is obvious from the limit definition that a reparametrisation should lead to the same length of the Jordan arc. Using [5] we easily see for $\gamma_2 = \gamma_1 \circ \kappa$ that $\int_{\gamma_2} ds = \int_a^b \|\gamma_2'(t)\| dt = \int_a^b \|\gamma_1'(\kappa(t))\| \kappa'(t) dt = \int_{\kappa(a)}^{\kappa(b)} \|\gamma_1'(t)\| dt = \int_{\gamma_1} ds$.

7 Arc length of a Jordan arc

The arc length of a Jordan arc Γ is the length of a Jordan path, an injective and rectifiable path γ describing Γ . Two such Jordan paths describing Γ have the same length and are reparametrisations of one another. We denote the length of a Jordan arc by $\int_{\Gamma} ds$

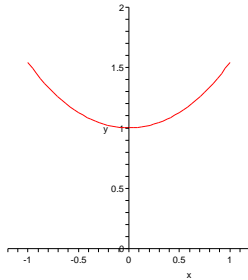


FIG. ix Hyperbolic cosine $\cosh x$.

Let us for example calculate the length of a hyperbolic cosine curve defined on the interval $[-1, 1]$. We choose the parametrisation

$$\gamma(t) = \begin{pmatrix} t \\ \cosh t \end{pmatrix}, t \in [-1, 1], \quad (26)$$

which is obviously a Jordan description of Γ . We find:

$$\int_{\Gamma} ds = \int_{-1}^1 \left\| \begin{pmatrix} 1 \\ \sinh t \end{pmatrix} \right\| dt = \int_{-1}^1 \sqrt{1 + \sinh^2 t} dt = 2 \sinh 1. \quad (27)$$

Generally, if we want to calculate the length of a curve Γ_f described by the graph of a function $f(x) : [a, b] \rightarrow \mathbb{R}$ we can find a Jordan description $\gamma(x) = (x, f(x))^T$ which leads to the following length:

^aIn case the paths go in opposite directions along Γ then the function is strictly decreasing rather than strictly increasing.

^bMarie Ennemond *Camille* Jordan, 1838-1922.

8 Length of a curve described by a function $f(x)$

The length of a continuous curve Γ_f described by a function $f(x) : [a, b] \rightarrow \mathbb{R}$ is

$$\int_{\Gamma_f} ds = \int_a^b \sqrt{1 + \left(\frac{df}{dx}\right)^2} dx. \quad (28)$$

Let us consider a Jordan path γ corresponding to a Jordan arc Γ . Its length function $s(t)$ is strictly increasing in t and let $t(s)$ denote its inverse function. $t(s)$ gives the time t at which the path will have reached length s from the starting point $\gamma(a)$. We can now choose a particular parametrisation of the time using $t(s)$: we define the path $\gamma_c(s) := \gamma(t(s))$ on the interval $[0, L(\gamma)]$ which is obviously just a reparametrisation of $\gamma(t)$. Using the chain rule and [5] we find that

$$\frac{d\gamma_c(s)}{ds} = \gamma'(t) \frac{dt(s)}{ds} = \frac{\gamma'(t)}{\frac{ds}{dt}} = \frac{\gamma'(t)}{\|\gamma'(t)\|}, \quad (29)$$

and therefore $\left\| \frac{d\gamma_c(s)}{ds} \right\| = 1$. The Jordan path $\gamma_c(s)$ describes the same Jordan arc Γ as $\gamma(t)$ but $\gamma_c(s)$ is the parametrisation of Γ with constant *speed* 1 everywhere along the path. This parametrisation is obviously unique (provided we do not change the direction along the path). We want to call it the *canonical parametrisation* of the Jordan arc Γ .

Let us for example consider the *circular helix*^c described by the Jordan path $\gamma(t) = (a \cos t, a \sin t, ct)^T$ where $a, c \in \mathbb{R}^+$ and t starts at time 0.

The length function is obviously $s(t) = \sqrt{a^2 + c^2} t$ with inverse function $t(s) = \frac{s}{\sqrt{a^2 + c^2}}$. The canonical parametrisation of the helix is therefore

$$\gamma_c(s) = \begin{pmatrix} a \cos \frac{s}{\sqrt{a^2 + c^2}} \\ a \sin \frac{s}{\sqrt{a^2 + c^2}} \\ \frac{cs}{\sqrt{a^2 + c^2}} \end{pmatrix}.$$

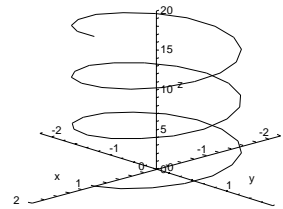


FIG. x Circular helix for $a = c = 1$.

Finally, on a technical note, [5] requires that the path γ is of course differentiable since we have to construct the derivative $\gamma'(t)$. In case the path is differentiable almost everywhere except at a finite number of points n , then we can split the path up into a sum of $n + 1$ paths γ_i , each of them being differentiable. For each γ_i we can use [5] in order to compute its length. Adding these lengths up leads to the length of γ . In this case mathematicians say that γ is *piecewise* differentiable. Most of what we said and what we are going to say about differentiable paths applies in the same way to piecewise differentiable paths following exactly this mechanism.

If a Jordan arc is closed we call it a *Jordan curve*. In this case a corresponding Jordan description can only be injective on $[a, b)$ since we obviously need $\gamma(a) = \gamma(b)$ in order to close the curve. But a Jordan curve does not allow any other points where injectivity is violated. For example, a circle is a Jordan curve whilst an '8' is not a Jordan curve. But we can of course generate an '8' using two Jordan curves.

^cA circular helix is found by winding a line around a cylinder. In contrast, a *conical helix* is found by winding a line around a cone.

Sometimes it may be useful to weight certain parts of a path more than other parts. For example, if the path represents an object which is heavier at one end than at the other end. Therefore we find a *line density* function $f(\mathbf{x})$ describing the mass distribution along the curve. In order to compute the total mass we obtain limits of the type $\lim_{\|\mathcal{Z}\| \rightarrow 0} \sum f(\gamma(t_k)) \|\gamma(t_k) - \gamma(t_{k-1})\|$. We denote such a limit by $\int_{\gamma} f(\mathbf{x}) ds$. It is clear that under the assumptions of [5] and for continuous functions $f(\gamma(t))$ we again find a simple integral just as in [5]:

9 Integrals over $f(\mathbf{x})$ along paths

If the path $\gamma : [a, b] \rightarrow \mathbb{R}^n$ can be differentiated then:

$$\int_{\gamma} f(\mathbf{x}) ds = \int_a^b f(\gamma(t)) \|\gamma'(t)\| dt . \quad (30)$$

We easily find the following integration rules.

10 Integration rules for integrals over $f(\mathbf{x})$ along paths

$$\int_{\gamma} (\alpha f(\mathbf{x}) + \beta g(\mathbf{x})) ds = \alpha \int_{\gamma} f(\mathbf{x}) ds + \beta \int_{\gamma} g(\mathbf{x}) ds , \quad (31)$$

$$\int_{\gamma_1 \oplus \gamma_2} f(\mathbf{x}) ds = \int_{\gamma_1} f(\mathbf{x}) ds + \int_{\gamma_2} f(\mathbf{x}) ds . \quad (32)$$

We can now also define integrals over vector valued functions $\mathbf{f}(x)$ along a path γ simply by defining the integral to be the vector of integrals over the component functions:

$$\int_{\gamma} \mathbf{f}(\mathbf{x}) ds := \begin{pmatrix} \int_{\gamma} f_1(\mathbf{x}) ds \\ \vdots \\ \int_{\gamma} f_m(\mathbf{x}) ds \end{pmatrix} . \quad (33)$$

2 Tangents and normals to curves in \mathbb{R}^3

In geometry you are mainly interested in two types of theorems, *classification theorems* and theorems relating *local* properties to *global* properties. For example, we know from simple geometry that two Euclidean triangles are congruent (can be mapped onto each other using translations, reflections and rotations) if and only if the lengths of their corresponding sides are equal. This is a classification theorem. It allows us to determine whether two mathematical objects are equivalent under some appropriate equivalence relation (here congruence) just by comparing a small (or at least finite) number of computable invariants. We also know from simple geometry that the sum of the interior angles of a Euclidean triangle is π . This theorem relates a local geometric property (angles) to a global property (that of being a triangle). Another example for a classification theorem is that two circles in the Euclidean plane are congruent if and only if they have the same radius. Whilst the theorem that the circumference of a Euclidean circle of radius r is $2\pi r$ is again a theorem that relates the local property of the angle 2π with the global property of being a circle. We could give numerous examples for these types of theorems in simple geometry, but if we want to continue the study of plane geometry beyond figures constructed from lines and circles we realise that an arbitrary curve cannot be

completely described by a few numbers such as lengths and radii. Instead we will find that the basic invariant is *curvature*. Curvature is a function of the position on the curve.

In the previous section we found that a Jordan arc in \mathbb{R}^2 can be uniquely described by its canonical Jordan path $\gamma_c(t)$ such that $\|\gamma'_c(s)\| = 1$ everywhere along the path. If we reparametrise the path then obviously the *speed* $\|\gamma'\|$ changes and therefore first order derivatives are not a suitable measure to classify the curve uniquely up to congruence. But it can be shown that the second order derivatives can be used instead.

11 Curvature of a plane curve

$\gamma_c(s)$ is the canonical Jordan path of the Jordan arc Γ . The curvature $\kappa(s)$ of Γ at the position s is defined as the magnitude of the acceleration vector

$$\kappa(s) := \|\gamma''(s)\|. \quad (34)$$

Along a straight path the velocity vector γ' never changes direction and therefore a straight path has curvature 0 everywhere. On the other hand a circle

$$\gamma_c(s) = \begin{pmatrix} r \cos \frac{s}{r} \\ r \sin \frac{s}{r} \end{pmatrix}, \quad s \in [0, 2\pi r], \quad (35)$$

has the constant curvature $\kappa = \frac{1}{r}$ at each point s on the circle. For a general curve Γ at some point s on Γ we find that among all circles which are tangent to Γ at this point there is only one circle whose acceleration also matches the canonical acceleration along the curve. This circle is called the *osculating circle* and the curvature of the curve at s is simply $\frac{1}{\text{radius}}$ of the osculating circle. If the radius is huge or even infinity for a straight line then obviously the curve is less curved at this point whilst a small radius means a big curvature. In summary, the bigger the curvature, the greater the acceleration and therefore the smaller the osculating circle.

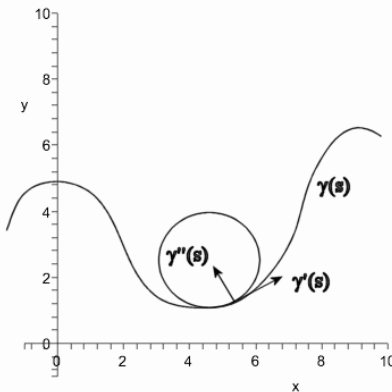


FIG. xi Osculating circle to a Jordan arc.

So far we have only defined curvature to be non-negative. Sometimes it is useful to introduce a (normal) direction and we would then define positive or negative curvature with respect to this direction. This is called the *signed curvature*. It can be shown that the signed curvature describes a curve uniquely up to congruence. Therefore we can obtain a classification theorem saying that two Jordan arcs are congruent if and only if the curvatures match at each point. This is the classification theorem we were looking for but one can also prove a theorem that

relates the local property of curvature to a global property of being a closed Jordan arc, meaning a Jordan curve: a Jordan arc is closed if and only if $\int_{\gamma} \kappa ds = \int_a^b \kappa(s) ds = 2\pi$ where $\kappa(s)$ is the signed curvature of γ .

We now want to apply similar ideas to curves in \mathbb{R}^3 . It is immediately clear that the radius of a plane osculating circle cannot be sufficient to describe curves in three dimensions. We will in addition need a way of describing the curvature of the curve coming out of the tangent plane. One finds that the following local properties define a curve in \mathbb{R}^3 up to congruence.

12 Curves in \mathbb{R}^3

Let Γ be a Jordan arc in \mathbb{R}^3 and let γ_c be the corresponding canonical Jordan path $\gamma_c : [a, b] \rightarrow \mathbb{R}^3$.

$$\mathbf{t} := \gamma'_c(s) \quad (36)$$

is called the tangent and it is clear that $\|\mathbf{t}(s)\| = 1 \quad \forall s$.

$$\mathbf{p} := \frac{\gamma''_c(s)}{\|\gamma''_c(s)\|} \quad (37)$$

is called the principal normal and $\kappa(s) := \|\gamma''_c(s)\|$ is the curvature. The binormal is given by

$$\mathbf{b} := \mathbf{t} \times \mathbf{p}, \quad (38)$$

and the torsion τ is the (negative) projection of the change in the binormal $\frac{\partial \mathbf{b}}{\partial s}$ onto the principal normal:

$$\tau := -\frac{\partial \mathbf{b}}{\partial s} \cdot \mathbf{p}. \quad (39)$$

It is worth mentioning that $\gamma'_c(s) \cdot \gamma''_c(s) = 0$ since $\gamma'_c(s) \cdot \gamma'_c(s) = 1$ and therefore \mathbf{t} , \mathbf{p} and \mathbf{b} form an orthonormal system. The tangent and the principal normal form a plane tangent to the curve which describes the movement of the curve to second order. In this plane there exists an osculating circle and the curvature describes the size of the osculating circle just like in the two dimensional case. The binormal is normal to this plane. If the curve describes a true three dimensional movement then it will obviously leave this plane. The torsion describes how strongly it leaves the plane. The minus sign in the definition is such that a movement outside this plane towards the direction of the binormal has a positive torsion, as in the case of the circular helix in FIG. x.

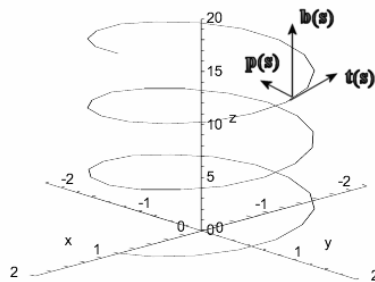


FIG. xii Tangent, principal normal and binormal for a circular helix.

Examples

Example 2.2 (i) The curve given parametrically by

$$\mathbf{x} = \boldsymbol{\psi}(t) = (at, bt^2, 0)$$

is the parabola $a^2y = bx^2$ in the x - y plane. The length of the curve from $t = 0$ to $t = 1$ is

$$\begin{aligned} \int ds &\equiv \int \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt \equiv \int \left\| \frac{d\mathbf{x}}{dt} \right\| dt \\ &= \int_0^1 \sqrt{a^2 + 4b^2t^2} dt = \text{whatever you get when you do this integral.} \end{aligned}$$

(ii) The curve

$$\mathbf{x} = \boldsymbol{\psi}(t) = (5 \cos t, 4 \sin t, 3 \sin t)$$

lies on the cylinder (with elliptic cross-section) $x^2/25 + y^2/16 = 1$. Setting $t = \phi$, we see that the ‘height’ z of the curve as it wraps around the cylinder is $3 \sin \phi$, so it looks similar to a sine curve wrapped around the cylinder. Clearly, the curve is closed: it joins back onto itself when t increases by 2π .

Note that it also lies on the cylinder $x^2/25 + z^2/9 = 1$, so it is the intersection of the two cylinders.

The (unit) tangent \mathbf{t} is the unit vector in the direction $\frac{d\boldsymbol{\psi}}{dt}$:

$$\mathbf{t} = (-5 \sin t, 4 \cos t, 3 \cos t)/5$$

The arc length, s , is given in terms of the parameter t by

$$ds = \left\| \frac{d\mathbf{t}}{dt} \right\| dt = 5dt,$$

so we could reparametrise the curve using s instead of t as follows:

$$\mathbf{x} = (5 \cos(s/5), 4 \sin(s/5), 3 \sin(s/5)).$$

Then

$$\frac{d\mathbf{t}}{ds} = -(5 \cos(s/5), 4 \sin(s/5), 3 \sin(s/5))/(25) \equiv \kappa \mathbf{p},$$

Using the fact that $\frac{d\mathbf{t}}{ds}$ and \mathbf{p} are unit vectors, we see that the curvature $\kappa = 1/5$ and that the principal normal $\mathbf{p} = -(5 \cos(s/5), 4 \sin(s/5), 3 \sin(s/5))/5$. Finally, the binormal $\mathbf{b} = \mathbf{t} \times \mathbf{p} = (0, -3, 4)/5$. Since $\frac{d\mathbf{b}}{ds} = 0$, the torsion $\tau = 0$. This is a little unexpected. It means that the curve lies in a plane with normal $(0, -3, 4)/5$ — which of course is now obvious: $\mathbf{x} \cdot (0, -3, 4) = 0$.

We can see this in another way. The curve lies on the sphere $r = 5$. If we use polar coordinates based on the x -axis instead of the z -axis, we see that $t = \theta$ and $\phi = \tan^{-1}(3/4)$; the curve is therefore a circle (a line of longitude).