

1 Week 7 Lecture 1: Curves and Their Parametrisations

Introduction

We want to describe curves in space so that we can later do calculus on curves. One approach to describing a curve in the plane would be to consider the graph of a function $f(x)$. The graph is the set of points (x, y) where $y = f(x)$. For the graph to be what we think of as a curve, we need f to be continuous and the domain of f to be an interval of real numbers. There are problems with this approach. For example, it can be awkward even for relatively simple curves, for example a circle, and it is unnatural to treat x and y are treated differently. This brings us to vector functions.

1.1 Vector Functions

You know what a vector is. (A basic review is given at the end of the chapter.) The position of a point can be represented as a vector from the origin to the point. We typically use the vector \mathbf{r} to denote position and write

$$\mathbf{r} = (x, y) \in \mathbb{R}^2 \quad \text{for a point in the plane}$$

$$\mathbf{r} = (x, y, z) \in \mathbb{R}^3 \quad \text{for a point in space}$$

so x , y , and possibly z are the components of the position vector. These are also the usual **Cartesian coordinates** for a point. At times we consider the general case where the dimension is some unspecified n . In these cases we label the components of position by $\mathbf{r} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

You know what a real valued function of a real variable is: $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$. To each real number in U , a subset of \mathbb{R} , f assigns a value in \mathbb{R} .

A vector function
$$\mathbf{r} : U \subseteq \mathbb{R} \rightarrow \mathbb{R}^n$$

assigns a point in \mathbb{R}^n , that is a vector, to each real number in U .

We usually denote the independent variable by t .

Vector function can be viewed in terms of component functions

$$\mathbf{r}(t) = (x_1(t), x_2(t), \dots, x_j(t), \dots, x_n(t)),$$

where each component function $x_j(t)$ is a real valued function. You can think of a vector function as just a collection real-valued functions arranged as components of a vector.

We are primarily interested in either the plane ($n = 2$) or in space ($n = 3$), and in these cases we usually to use the notation $x(t)$, $y(t)$, and possibly $z(t)$ for component functions rather than $x_1(t)$, $x_2(t)$, and $x_3(t)$. For example, in space we would typically write

$$\mathbf{r}(t) = (x(t), y(t), z(t)).$$

You should also be familiar with the unit vectors \mathbf{i} , \mathbf{j} , and \mathbf{k} . Using these we can avoid outer parentheses and write

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$$

Things you know about vectors and functions can now be applied to vector functions is the obvious way, e.g., given vector functions

$$\mathbf{f}(t) = f_1(t)\mathbf{i} + f_2(t)\mathbf{j}, \quad \mathbf{g}(t) = g_1(t)\mathbf{i} + g_2(t)\mathbf{j},$$

and scalar α , we could define the vector function \mathbf{r}

$$\begin{aligned} \mathbf{r}(t) &= \mathbf{f}(t) + \alpha\mathbf{g}(t) \\ &= (f_1(t) + \alpha g_1(t))\mathbf{i} + (f_2(t) + \alpha g_2(t))\mathbf{j} \end{aligned}$$

Example: Linear vector function. Let $\mathbf{r}(t) = \mathbf{a}t + \mathbf{b}$, where $\mathbf{a} \neq \mathbf{0}$ and \mathbf{b} are constant vectors. The function values trace a straight line as t varies. This is superior to the graph $y = f(x) = mx + b$. With a vector function we can represent a line in any dimension and in any direction.

1.2 Curves Defined

There is a close connection between vector functions and curves. Essentially a curve is the set of points traced out by a vector function as the independent variable ranges over an interval $I = [a, b]$.

Let $\mathbf{r}(t): I \rightarrow \mathbb{R}^n$ be a continuous vector function, where $I \subseteq \mathbb{R}$ is an interval. Then the set of points

$$\mathcal{C} = \{\mathbf{r}(t) | t \in I\}$$

is a curve. The function $\mathbf{r}(t)$ is called a **parametrisation** of the curve \mathcal{C} .

Remarks

- The definition requires $\mathbf{r}(t)$ to be a *continuous* vector function. Continuous means what you think it does: it guarantees that there are no breaks in the curve \mathcal{C} .
- We will generically write the interval as $I = [a, b]$, but it is understood that I could be infinite or semi-infinite. For example: $I = (-\infty, \infty)$, $I = (-\infty, 0]$, or $I = [1, \infty)$.
- You should be aware that there is some variation in the definition of a curve. In some texts a curve is defined as \mathbf{r} itself and not the set of points \mathcal{C} . Some authors require I to be a closed interval, others do not.
- There are infinitely many possible parametrisations for a given curve \mathcal{C} .
- The definition in terms of parametrisations gives us something in addition to the set of points \mathcal{C} , it gives an orientation corresponding to increasing t . Different parametrisations can give opposite orientations, but given a parametrisation, there is an orientation attached to the curve beyond its geometrical shape.
- From the way we motivated the subject, one sees a close connection between curves (purely geometrical objects) and trajectories associated with motion in time, similar to what we have seen for ODEs. This is further reinforced by using t for the independent variable.

1.3 Working with Parametrisations

Examples

We will primarily be concerned with the following examples.

- $\mathbf{r}(t) = (R \cos t, R \sin t)$, $t \in [0, 2\pi]$ is the parametrisation for a circle of radius R centred on the origin.
- $\mathbf{r}(t) = (a + R \cos t, b + R \sin t)$, $t \in [0, 2\pi]$ is the parametrisation for a circle of radius R centred on (a, b) .
- $\mathbf{r}(t) = (0, b + R \cos t, c + R \sin t)$, $t \in [0, 2\pi]$ is the parametrisation for a circle in space. It has radius R , lies in the (y, z) plane, and is centred on $(0, b, c)$.
- If $\mathbf{r}(t) = (x(t), y(t))$ is a parametrisation for some curve, then $(ax(t), y(t))$ will stretch the x -coordinate if $a > 1$ and compress the x -coordinate stretched if $0 < a < 1$. For example $\mathbf{r}(t) = (a \cos t, b \sin t)$, $t \in [0, 2\pi]$ is the parametrisation for an ellipse if $a \neq b$.
- $\mathbf{r}(t) = (at \cos t, at \sin t)$, $t \geq 0$ is the parametrisation of an *Archimedean spiral*, (a spiral given in polar coordinates by $r = a\theta$). $\mathbf{r}(t) = (e^{at} \cos t, e^{at} \sin t)$, $t \in \mathbb{R}$ is the parametrisation of a *Logarithmic spiral*, (in polar coordinates $r = e^{a\theta}$).

You may have previously studied curves given as radius as a function of angle: $r = f(\theta)$. We shall call such curves **polar graphs** or **graphs of a polar function** $r = f(\theta)$. The Archimedean and Logarithmic spirals are simply expressed as polar functions. (A basic review of polar coordinate is given at the end of this weeks' notes, but we will not consider polar representation of curves.)

- In space: $\mathbf{r}(t) = (R \cos t, R \sin t, kt)$, $t \in \mathbb{R}$, $k \neq 0$, is the parametrisation of a helix.
- In space: $\mathbf{r}(t) = (e^t \cos bt, e^t \sin bt, e^t)$, $t \in \mathbb{R}$, $b \neq 0$, is the parametrisation of a curve winding around a cone (circular conical surface).

Sketching and finding parametrisations

You should learn how to sketch the examples in the previous section. With a computer, curve sketching is relatively easy, so you should learn to use Python or other software to plot curves and verify your sketches.

Finding a parametrisation is somewhat the opposite of sketching. Given some description or specification of a curve, you need to produce a parametrisation for it, generally for the purpose of doing some further calculation involving the curve. This can be quite challenging in general, but you will not be asked to do anything too complicated.

You need to be able to parametrise curves made of straight segments. You need to be able to parametrise circles and circular arcs. You need to be able to parametrise ellipses, spirals, and helices and combinations of these. (See next section on curves with multiple segments).

1.4 Curves in Multiple Segments

We carefully defined curves and parametrisations so that curves are continuous images of a single interval I . Yet, in many applications this is annoying and unnecessary. Consider parametrising the plane curve \mathcal{C} which starts at $(1, 0)$ and proceed to $(-1, 0)$ along the upper unit semicircle and then returns to $(1, 0)$ along the x -axis. The annoyance is that curve is easily described as the union of two segments, each of which is easily parametrised:

$$\begin{aligned}\mathbf{r}_1(t) &= \cos(t)\mathbf{i} + \sin(t)\mathbf{j}, & t \in [0, \pi] \\ \mathbf{r}_2(t) &= \mathbf{i}, & t \in [-1, 1]\end{aligned}$$

One can adjust things to produce a mapping over a single interval I , (probably you would adjust the second parametrisation). However, for all things we care about in this moduel, this is simply unnecessary. It is sufficient to work with a set of parametrisations, $\mathbf{r}_1(t), \dots, \mathbf{r}_k(t)$ defined on intervals $I_1(t), \dots, I_k(t)$, if these naturally describe a curve composed of a union of k segments. *You may work with multiple parametrisations when it is natural to do so.*

1.5 Derivative of a Vector Function

The derivative of a vector function is itself a vector and is computed by differentiating component by component. We often denote the derivative with a $'$, e.g., $\mathbf{r}'(t)$, so

$$\mathbf{r}'(t) = \frac{d\mathbf{r}}{dt}(t) = \left(\frac{dx_1}{dt}(t), \frac{dx_2}{dt}(t), \dots, \frac{dx_n}{dt}(t) \right)$$

For example, in space we can write the derivative as

$$\mathbf{r}'(t) = x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k}$$

If $\mathbf{r}(t)$ is a parametrisation of a curve \mathcal{C} and if the derivative exists and is nonzero ($\mathbf{r}'(t) \neq \mathbf{0}$), then the vector $\mathbf{r}'(t)$ is tangent to \mathcal{C} at the point $\mathbf{r}(t)$. (We return to this next lecture.)

1.6 Differentiation Rules

Suppose $\mathbf{r}(t)$ and $\mathbf{s}(t)$ are differentiable vector functions, and $f(t)$ is a differentiable real function, and a is a real number. Then the following hold:

- $\frac{d}{dt} a\mathbf{r}(t) = a\mathbf{r}'(t)$
- $\frac{d}{dt} (\mathbf{r}(t) + \mathbf{s}(t)) = \mathbf{r}'(t) + \mathbf{s}'(t)$
- $\frac{d}{dt} f(t)\mathbf{r}(t) = f(t)\mathbf{r}'(t) + f'(t)\mathbf{r}(t)$
- $\frac{d}{dt} (\mathbf{r}(t) \cdot \mathbf{s}(t)) = \mathbf{r}'(t) \cdot \mathbf{s}(t) + \mathbf{r}(t) \cdot \mathbf{s}'(t)$
- $\frac{d}{dt} \mathbf{r}(f(t)) = \mathbf{r}'(f(t))f'(t)$

These are obvious extensions of what you know for functions from $\mathbb{R} \rightarrow \mathbb{R}$. You should learn and be able to verify each of these.

1.7 Further Concepts

There are a few further concepts related to curves which are worth defining. Only the first concept is essential in this module.

A curve is **closed** if $I = [a, b]$ and $\mathbf{r}(a) = \mathbf{r}(b)$. A closed curve is also called a **loop**.

A curve is **embedded** if it does not self intersect. In terms of parametrisations, a curve is embedded if it can be parametrised by a mapping $\mathbf{r} : I \rightarrow \mathbb{R}^n$ that is injective except possibly at the end points of $I = [a, b]$. We need to allow the possibility that $\mathbf{r}(a) = \mathbf{r}(b)$ since this just means that the curve is closed and does not correspond to an intersection. From the definition of injective, a curve is embedded if it can be parametrised by $\mathbf{r}(t)$ where $t_1 \neq t_2$ implies $\mathbf{r}(t_1) \neq \mathbf{r}(t_2)$, except if $t_1 = a$ and $t_2 = b$.

A curve is **regular** if there exists a parametrisation such that its derivative \mathbf{r}' is defined and nonzero at all points. In other words, regular curves have a tangent vector at each point. Regular curves are nice in that they do not have corners or cusps and these are mainly the ones we want to do calculus on. We will commonly assume, without explicitly stating so, that curves we consider are regular, or at least piecewise regular (made up of pieces that are each regular).

Review Material

Basics of vectors

This is just a quick reminder of some basic things you already be familiar with. In Linear Algebra you will be spending a lot of time discussing vectors.

Notation. Vectors will be denoted by boldface and sometime over arrow, e.g., \mathbf{v} or \vec{v} . In lectures, underline will be used instead of bold face, e.g., \underline{v} . We only consider real vectors in this course, so $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$. The real numbers v_1, v_2, \dots, v_n are call the **components** of \mathbf{v} .

Scalars. We shall use the term **scalar** to refer to a real number, to contrast with vectors. You should know things such as how to multiply or divide a vector \mathbf{v} by a scalar α :

$$\alpha\mathbf{v} = (\alpha v_1, \alpha v_2, \dots, \alpha v_n), \quad \frac{\mathbf{v}}{\alpha} = \frac{1}{\alpha}(v_1, v_2, \dots, v_n) = \left(\frac{v_1}{\alpha}, \frac{v_2}{\alpha}, \dots, \frac{v_n}{\alpha}\right)$$

Norms. The **Euclidean norm**, **length**, or **magnitude** of a vector \mathbf{v} is defined as: $\|\mathbf{v}\| = \left(\sum_{i=1}^n v_i^2\right)^{1/2}$.

Dot Product. You should know properties of the dot product and how to compute the dot product, either as

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta$$

where θ is the angle between \mathbf{u} and \mathbf{v} , or as

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u_i v_i,$$

where u_i and v_i are the components of \mathbf{u} and \mathbf{v} . The dot product of two vectors is a scalar, and $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$. Notice that $\|\mathbf{v}\| = (\mathbf{v} \cdot \mathbf{v})^{1/2}$.

Unit Vectors. A **unit vector** is a vector with length one. Often these are denoted with hats, for example $\hat{\mathbf{v}}$ where $\|\hat{\mathbf{v}}\| = 1$. We say $\hat{\mathbf{v}}$ has “unit norm”,

Important unit vectors are the standard basis vectors. For example, the basis vectors for Cartesian coordinates in three dimensions are (hats are not used for these unit vectors)

$$\mathbf{i} = (1, 0, 0) \quad \mathbf{j} = (0, 1, 0) \quad \mathbf{k} = (0, 0, 1)$$

Then we can write the position vector \mathbf{r} as a sum of components times basis vectors:

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = (x, y, z)$$

One sometimes uses the following notation for these bases vectors:

$$\mathbf{e}_1 = (1, 0, 0) \quad \mathbf{e}_2 = (0, 1, 0) \quad \mathbf{e}_3 = (0, 0, 1)$$

(See Linear Algebra for the details on basis vectors. In this course we will only be concerned with standard basis vectors for typical coordinate systems such as Cartesian and polar coordinates.)

Given a nonzero vector \mathbf{v} , that is $\mathbf{v} \neq (0, 0, \dots, 0) = \mathbf{0}$, the **direction** of \mathbf{v} is: $\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$

So each nonzero vector can be written $\mathbf{v} = \|\mathbf{v}\| \hat{\mathbf{v}} = (\text{magnitude}) \times (\text{direction}) = (\text{scalar}) \times (\text{unit vector})$

Polar coordinates

You should be familiar with polar coordinates, typically denoted (r, θ) . The relationship between polar and Cartesian coordinate is:

$$\begin{aligned} x &= r \cos \theta & r^2 &= x^2 + y^2 \\ y &= r \sin \theta & \tan \theta &= \frac{y}{x} \end{aligned}$$

You should know how to go back and forth between Cartesian and polar coordinates and be able to use polar coordinates as needed. There is a small issue here with non-uniqueness of polar coordinates. You should know that coordinates (r, θ) , $(r, \theta + 2\pi)$, and $(-r, \theta + \pi)$ all correspond to the same point in the plane. One sometimes works with $r \geq 0$ and θ restricted to a fixed range to avoid the non-uniqueness. In this case I generally prefer to work with $\theta \in [0, 2\pi)$, but this depends on the problem. Sometimes it is more convenient to work with $\theta \in (-\pi, \pi]$. Sometimes it is best to live with the non-uniqueness. For $r = 0$ the value of θ is irrelevant. (This is due to the coordinate singularity at $r = 0$, but we will not concern ourselves with this.)

Conic sections

The following curves are conic sections: circle, ellipse, parabola, and hyperbola. These curves all come from intersecting a circular conical surface with a plane, hence the name conic section. The curves also come about as sets of points (x, y) satisfying the general quadratic equation:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

where A, B, C, D, E, F are constants with not all three of A, B , and C are equal to zero. See the *Wikipedia* page on *Conic Sections*.

These curves arise frequently and you should know standard forms for the equations for each case:

- *Circle*: Points satisfying $x^2 + y^2 = a^2$.
- *Ellipse*: Points satisfying $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$
- *Hyperbola*: Points satisfying $\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 1$
- *Parabola*: Points satisfying $y = ax^2$, or sometimes for consistency with the other expressions $y^2 = 4ax$. (In the first case the parabola opens upward or downward depending on the sign of a , while in the second case it opens to the left or right.)

2 Week 7 Lecture 2: Arc Length and Differential Geometry of Curves

2.1 Arc Length

We are going to want to perform integration along curves, now and later in this module. To begin, let us ask the question: “given a parametrised curve \mathcal{C} , what is its length?” The answer will require integration along \mathcal{C} .

Our method for deriving the integral will be to take \mathcal{C} and divide it up into “small” pieces, or arcs, and sum up the lengths of these pieces. We then take the limit of the sum as the size of the pieces goes to zero while the number of pieces going to infinity.

Let $\mathbf{r} : I = [a, b] \rightarrow \mathbb{R}^n$ be a parametrisation of \mathcal{C} . To divide \mathcal{C} into small arcs, all we need do is divide I into small segments of size Δt . Let

$$t_0 = a, t_1 = a + \Delta t, \dots, t_j = a + j\Delta t, \dots, t_N = b.$$

So there are N segments of length $\Delta t = (b - a)/N$.

Now, the j^{th} segment $[t_j, t_{j+1}]$ will get mapped by \mathbf{r} to a small arc from $\mathbf{r}(t_j)$ to $\mathbf{r}(t_{j+1})$ of the curve \mathcal{C} . The length of the arc, Δs_j , will be approximately the length of the chord from $\mathbf{r}(t_j)$ to $\mathbf{r}(t_{j+1})$

$$\Delta s_j \approx \|\mathbf{r}(t_{j+1}) - \mathbf{r}(t_j)\|$$

or, multiplying and dividing by Δt

$$\Delta s_j \approx \left\| \frac{\mathbf{r}(t_j + \Delta t) - \mathbf{r}(t_j)}{\Delta t} \right\| \Delta t \quad (1)$$

The total length s is then approximately

$$s = \sum_{j=0}^{N-1} \Delta s_j \approx \sum_{j=0}^{N-1} \left\| \frac{\mathbf{r}(t_j + \Delta t) - \mathbf{r}(t_j)}{\Delta t} \right\| \Delta t$$

Taking the limit $N \rightarrow \infty$ with $\Delta t \rightarrow 0$ we obtain the following

The length s of a curve \mathcal{C} , denoted $\ell(\mathcal{C})$, is given by

$$s = \ell(\mathcal{C}) = \int_a^b \left\| \frac{d\mathbf{r}}{dt}(t) \right\| dt$$

where $\mathbf{r}(t)$, $t \in [a, b]$ is a parametrisation of \mathcal{C} .

Discussion

- We have implicitly assumed that the parametrisation $r(t)$ does not repeat points of \mathcal{C} . In practice this is always obvious. For example, to compute the length of a circle (i.e. its circumference) one would not use $r(t) = (R \cos t, R \sin t)$, with $t \in [0, 4\pi]$.
- Independence of parametrisation. While we defined the length of a curve based on a parametrisation, the length of a curve is independent of the parametrisation. If $\mathbf{r}_1 : [a, b] \rightarrow \mathbb{R}^n$ and $\mathbf{r}_2 : [c, d] \rightarrow \mathbb{R}^n$ are two parametrisations of the same curve \mathcal{C} , the lengths will be the same.
- Arc length is never negative. The limits of integration are *always* from the left endpoint (smallest value) of I to the right endpoint (largest value) of I .
- An infinitesimal increment dt in the parameter t corresponds to an infinitesimal increment of arc length ds along the curve \mathcal{C} . These are related by

$$ds = \|\mathbf{r}'(t)\| dt$$

This is the infinitesimal version of Eq. (1). It can be equivalently written as

$$\frac{ds}{dt} = \|\mathbf{r}'(t)\| \quad (2)$$

- Eq. (2) and the formula for length have simple physical interpretations. Let $\mathbf{r}(t)$ represent a particle path. Then Eq. (2) states that the change in distance per change in time, ds/dt , i.e., the speed of the particle, is $\|\mathbf{r}'(t)\|$. The equation for arc length states

$$\text{Distance} = \int_a^b \text{Speed } dt$$

i.e., distance travelled is just the integral of speed over time a interval.

- We learnt that to differentiate a vector-valued function one simply differentiates component by component, e.g., given

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$$

the derivative is

$$\mathbf{r}'(t) = x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k}$$

You might have therefore expected to see a similar start to the integration of vector-valued functions, e.g.,

$$\int_a^b \mathbf{r}(t)dt = \int_a^b x(t)dt \mathbf{i} + \int_a^b y(t)dt \mathbf{j} + \int_a^b z(t)dt \mathbf{k}$$

This formula is correct and it is the correct meaning of $\int_a^b \mathbf{r}(t)dt$. The reason we did not start with this is that such integration does not often arise, and in any case it is easy. Integration along a curve, such as in computing arc length, is more important and we will see it again later in this module when we do line integrals. Be sure you understand the difference between

$$\int_a^b \mathbf{r}(t)dt \quad \text{and} \quad \int_a^b \|\mathbf{r}'(t)\|dt.$$

2.2 Curves in Multiple Segments

We already noted that we often want to work with curves composed of multiple segments. Let \mathcal{C} be the union of k segments which meet end to end

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_k,$$

Assuming each segment \mathcal{C}_j is regular, then the curve \mathcal{C} is **piecewise regular**. Such curves may fail to be regular where the curves join. For example if \mathcal{C} is square then it is not regular because it has 4 corners. However it can clearly be composed of 4 regular curves and hence \mathcal{C} is piecewise regular.

The length of \mathcal{C} is given by

$$\ell(\mathcal{C}) = \ell(\mathcal{C}_1) + \ell(\mathcal{C}_2) + \cdots + \ell(\mathcal{C}_k)$$

In practice one would typically use whatever parametrisation is most convenient for each segment to compute the length of that segment, and then add up the lengths to obtain the length of \mathcal{C} .

Differential Geometry of Curves

We now know how to differentiate and integrate along curves. We now explore some of the geometrical properties of curves that can be addressed using differential calculus. We will be particularly interested in the bending of curves (curvature).

Unit tangent vector

Given a parametrisation $\mathbf{r}(t)$ of a curve \mathcal{C} , $\mathbf{r}'(t)$ is tangent to \mathcal{C} at $\mathbf{r}(t)$, assuming $\mathbf{r}'(t) \neq \mathbf{0}$. However the length of this tangent vector will depend on the particular parametrisation. We define the **unit tangent vector** to be

$$\mathbf{T} = \frac{\mathbf{r}'}{\|\mathbf{r}'\|}$$

Principal normal vector

Start with a little calculation. We know that by definition $\|\mathbf{T}(t)\| = 1 = \text{const}$ for all t . So

$$\|\mathbf{T}\|^2 = \mathbf{T} \cdot \mathbf{T} = \text{const}$$

Differentiating gives

$$\frac{d}{dt}(\mathbf{T} \cdot \mathbf{T}) = \mathbf{T}' \cdot \mathbf{T} + \mathbf{T} \cdot \mathbf{T}' = 2 \mathbf{T}' \cdot \mathbf{T} = 0$$

so

$$\mathbf{T}' \cdot \mathbf{T} = 0$$

Thus either $\mathbf{T}' = \mathbf{0}$ or \mathbf{T}' is perpendicular to \mathbf{T} . One can also see this graphically. We define the **principal normal vector** to be the unit vector in the direction of \mathbf{T}' :

$$\mathbf{N} = \frac{\mathbf{T}'}{\|\mathbf{T}'\|}$$

If $\mathbf{T}' = \mathbf{0}$ the principal normal vector is not defined.

Curvature

Intuitively curves can have different amounts of bending and this is seen in how quickly the unit tangent vector \mathbf{T} changes as one moves along the curve. Unfortunately, \mathbf{T}' itself will depend on the parametrisation used. To remove this dependence, one defines the **curvature** κ to be the magnitude of instantaneous change in \mathbf{T} with respect to arc length

$$\kappa = \left\| \frac{d\mathbf{T}}{ds} \right\|$$

To obtain a useful expression, we use the chain rule

$$\frac{d\mathbf{T}}{ds} = \frac{d\mathbf{T}}{dt} \frac{dt}{ds} = \mathbf{T}' \frac{1}{ds/dt} = \mathbf{T}' \frac{1}{\|\mathbf{r}'\|}$$

Thus, given a parametrisation $\mathbf{r}(t)$ of a curve, the curvature is computed using

$$\kappa = \frac{\|\mathbf{T}'\|}{\|\mathbf{r}'\|}$$

The **radius of curvature** ρ is defined to be

$$\rho = \frac{1}{\kappa}$$

Example: Consider a circle of radius R parametrised by $\mathbf{r} = (R \cos t, R \sin t)$. Then $\mathbf{T} = (-\sin t, \cos t)$ and $\mathbf{N} = (-\cos t, -\sin t)$. The curvature is $\kappa = 1/R$ and the radius of curvature is $\rho = R$. Note any regular parametrisation of the circle would give the same \mathbf{T} , within a \pm sign. Flipping the orientation flips \mathbf{T} but not \mathbf{N} . You should verify this.

2.3 Parametrisation by Arc Length (optional)

Imagine that the interval I is a straight piece of wire and that the mapping \mathbf{r} corresponds to taking the wire and bending it around into a curve \mathcal{C} *without any stretching or compression*. Then there is an exact correspondence between distances along the interval I and arc length along \mathcal{C} . For example, let $I = [0, 10]$ (think of a 10 centimetre long straight wire). Each subinterval $[0, 1], [2, 3], \dots, [9, 10]$ of I gets mapped by \mathbf{r} to a segment of \mathcal{C} with arc length one. (Each centimetre of straight wire gets bent around, but maintains its length of one centimetre.)

Such a parametrisation of a curve \mathcal{C} is called an **arc-length parametrisation** and the curve is said to be parametrised by arc length. This is also called the **natural parametrisation**. Denote such a parametrisation by \mathbf{r}_{arc} . It has the property that the arc length between $\mathbf{r}_{\text{arc}}(t_1)$ and $\mathbf{r}_{\text{arc}}(t_2)$ is precisely $|t_2 - t_1|$, the distance between t_1 and t_2 .

We shall simplify our discussion by assuming \mathcal{C} has finite length $\ell(\mathcal{C}) < \infty$ and when parametrised by arc-length the interval I is $I = [0, \ell(\mathcal{C})]$. (One can in fact parametrise infinitely long curves by arc-length.)

Arc-length parametrisations are important for two reasons. The first is conceptual. We know that there are infinitely many ways to parametrise a curve, yet these all give the same arc-length. By choosing arc length along the curve *as the parameter* we select a parametrisation that is intrinsic to the curve rather than arbitrary. Apart from orientation, the parametrisation by arc length is unique.

The second reason is more important in practice. Let $\mathbf{r}_{\text{arc}}(t)$ be an arc-length parametrisation of \mathcal{C} . Then by definition

$$\int_0^s \|\mathbf{r}'_{\text{arc}}(t)\| dt = s$$

for all $s \in [0, \ell(\mathcal{C})]$. Differentiating this with respect to s gives (see Additional Material),

$$\begin{aligned} \frac{d}{ds} \int_0^s \|\mathbf{r}'_{\text{arc}}(t)\| dt &= \frac{d}{ds} s \\ \|\mathbf{r}'_{\text{arc}}(s)\| &= 1 \end{aligned}$$

Hence the derivative \mathbf{r}'_{arc} has unit length for an arc-length parametrisation. One often says that an arc-length parametrisation has speed one, whether or not one is thinking of particle paths. Many calculations for curves, such as those we will see next week, greatly simplify for arc-length parametrisations.

To obtain an arc-length parametrisation of a curve \mathcal{C} , start with some parametrisation $\mathbf{r}(t) : [a, b] \rightarrow \mathbb{R}^n$ and compute the arc length from $\mathbf{r}(a)$ to $\mathbf{r}(t)$ for $t \in [a, b]$

$$s = \ell(t) = \int_a^t \|\mathbf{r}'(\tau)\| d\tau$$

Inverting this relationship, $t = \ell^{-1}(s)$, gives us the parameter value t corresponding to arc length s from the end point $\mathbf{r}(a)$. Using this to eliminate t in favour of s , we can in principle obtain an arc-length parametrisation

$$\mathbf{r}_{\text{arc}}(s) = \mathbf{r}(\ell^{-1}(s))$$

Be warned, however, even though this parametrisation exists in principle, obtaining an arc-length parametrisation is generally impossible because no formula exists for $\ell^{-1}(s)$.

It is common practice to use s rather than t for the independent variable in an arc-length parametrisation, and so to write $\mathbf{r}_{\text{arc}}(s)$ as opposed to $\mathbf{r}_{\text{arc}}(t)$. We will sometimes do this if we want to emphasise arc-length parametrisation.

3 Week 7 Lecture 3: Functions of Several Variables and Partial Derivatives

3.1 Introduction

A functions of several variables

$$f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$$

is a rule that assigns a real number to each point in U , a subset of \mathbb{R}^n ,

We sometimes call f a function of n variables, or say f is a function on \mathbb{R}^n (meaning perhaps a subset of \mathbb{R}^n). When we focus specifically on $n = 2$ or $n = 3$ we commonly write

$$w = f(x, y) \quad \text{or} \quad w = f(x, y, z)$$

In fact, when we consider graphs (see below) for $n = 2$ we frequently use z for the dependent variable, e.g., $z = f(x, y)$.

Many physical systems are expressed as functions of several variables and the governing laws are expressed in the calculus of such functions. Consider for example the temperature in a room. Temperature is a real number that will be a function of both position and time. Call this function T , so $T(x, y, z, t)$ is the temperature at position (x, y, z) and time t . Under certain assumptions the physical law governing the evolution of temperature is:

$$\frac{\partial T}{\partial t}(x, y, z, t) = \frac{\partial^2 T}{\partial x^2}(x, y, z, t) + \frac{\partial^2 T}{\partial y^2}(x, y, z, t) + \frac{\partial^2 T}{\partial z^2}(x, y, z, t)$$

You are familiar with ordinary differential equations. This is a **partial differential equation**. By the end of this week you will be able to verify whether $T(x, y, z, t)$ satisfies the equation. Finding solutions to partial differential equations will come in later years.

Visualising functions on \mathbb{R}^n

There are two primary ways to visualise functions of several variables: graphs for $n = 2$ and level set for $n = 2$ and $n = 3$. One can also make movies of graphs or level sets, and thereby visualise functions of up to four variables. For larger n visualisation is very difficult.

Graphs

For $n = 2$, f can be visualised as the graph

$$G_f = \{(x, y, z) \mid (x, y) \in U, z = f(x, y)\}$$

The function is seen as a sheet of height $f(x, y)$ above or below each point (x, y) .

Level sets

Level sets, also called contours in \mathbb{R}^2 or isosurfaces in \mathbb{R}^3 , are subsets of U which are all mapped to the same value by f . Formally, the definition is

The **level sets** of a function $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ are sets of points

$$\mathcal{L}_k = \{\mathbf{x} \in U \mid f(\mathbf{x}) = k\}$$

for each constant k in the range of f .

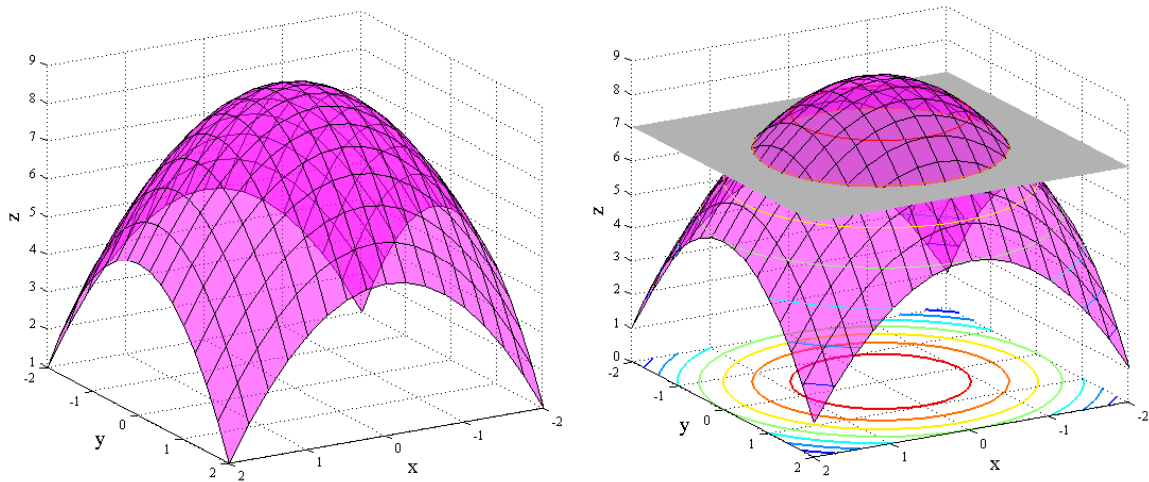


Figure 1: The function $f(x, y) = 9 - x^2 - y^2$ visualised as a graph (left) and as a contour plot (right) by slicing the graph at constant heights.

The intuition is easy for function on \mathbb{R}^2 . Plot the graph $z = f(x, y)$ then intersect the graph with plane $z = k$ for some constant k . Project the intersection points down onto \mathbb{R}^2 and these will make up the contour for this value of k . Typically contours will be curves in \mathbb{R}^2 . To represent a function using contours one typically plots several contours with the corresponding values of k labelled in some way. These are called **contour maps** or **contour plots**. The concept is familiar from topographic maps, weather maps, and such.

The level sets of a function of three variables $f(x, y, z)$ are typically surfaces in \mathbb{R}^3 called **isosurfaces** (iso meaning equal, so a surface of equal, i.e., constant, value of f). Using transparency or clipping, computers can often make several isosurfaces visible simultaneously, allowing for good understanding of the underlying function f .

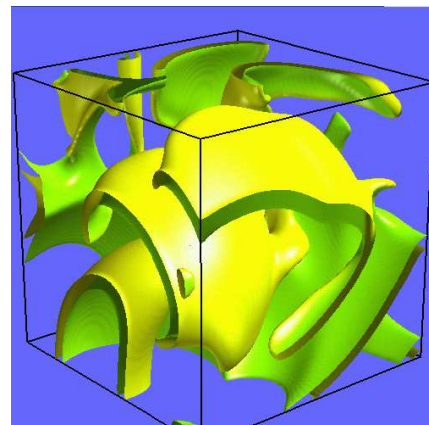
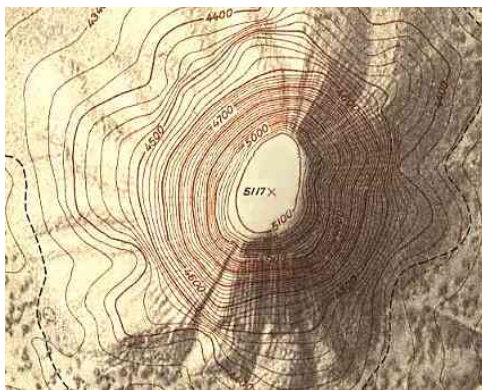


Figure 2: Example of a topographic map (contour plot) and an isosurface. (Map reproduced from http://mail.colonial.net/~hkaiter/topographic_maps)

3.2 Partial Derivatives

Partial derivatives are easy. For simplicity we initially restrict to the case $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ and define

The **partial derivatives** of $f(x, y)$ with respect to x and y at the point (a, b) are

$$\frac{\partial f}{\partial x}(a, b) = \lim_{h \rightarrow 0} \frac{f(a+h, b) - f(a, b)}{h}$$

$$\frac{\partial f}{\partial y}(a, b) = \lim_{h \rightarrow 0} \frac{f(a, b+h) - f(a, b)}{h}$$

Assuming the limits exist.

Sometimes partial derivatives are indicated by subscripts, e.g., $f_x(a, b)$ and $f_y(a, b)$, or $f_1(a, b)$ and $f_2(a, b)$. Sometimes upper case D is used, e.g., $D_x(a, b)$ and $D_y(a, b)$. We will not use any of these notations. We will on occasion use the following. Letting $z = f(x, y)$ we will denote partial derivatives of f by $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$.

Interpretation

What this definition says is that *the partial derivative of f with respect to x is just the ordinary one-dimensional derivative treating y as a fixed constant*. Concretely, define $g(x) = f(x, b)$ and then compute the ordinary derivative dg/dx at a . This is the partial derivative of f with respect to x . The partial derivative with respect to y is analogous. So the pedantic view of partial differentiation is:

Given $f(x, y)$,

$$\left\{ \begin{array}{l} \text{let } g(x) = f(x, b), \text{ then } \frac{\partial f}{\partial x}(a, b) = \frac{dg}{dx}(a) \\ \text{let } h(y) = f(a, y), \text{ then } \frac{\partial f}{\partial y}(a, b) = \frac{dh}{dy}(b) \end{array} \right.$$

This is illustrated in the following pictures. The function $g(x)$ is obtained by slicing f with a plane $y = b$, and similarly for $h(y)$.

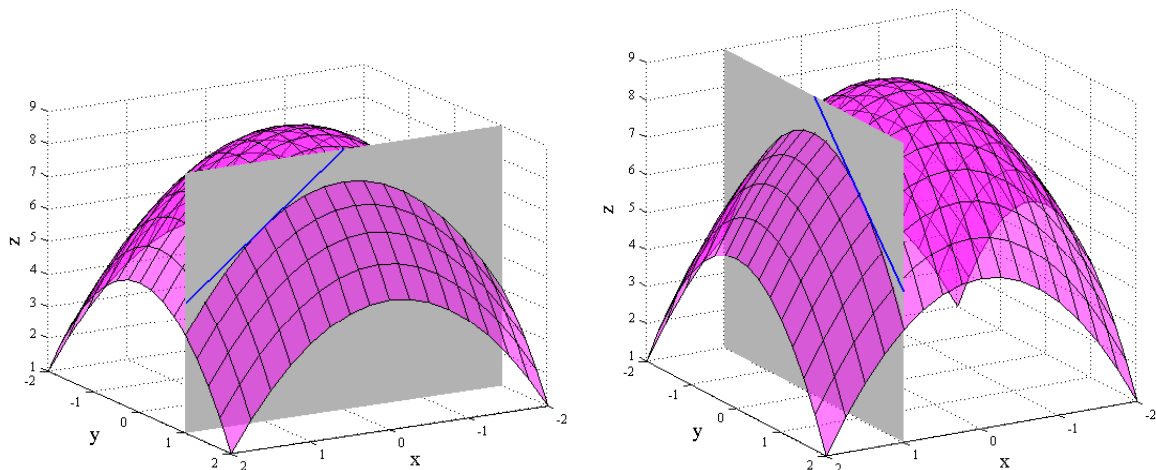


Figure 3: The x partial derivative (left) and y partial derivative (right) of the function $f(x, y) = 9 - x^2 - y^2$.

While defining the auxiliary functions $g(x)$ and $h(y)$ is pedagogically useful for explaining partial derivatives, in practice it is unnecessary to explicitly form these functions. You will quickly master computing partial derivatives by doing examples.

Partial derivatives are functions

In the above definition we defined the partial derivatives $\frac{\partial f}{\partial x}(a, b)$ and $\frac{\partial f}{\partial y}(a, b)$ at a point (a, b) . If we allow this point to vary, then each partial derivatives will itself be a function of (x, y) . In which case we would write

$$\frac{\partial f}{\partial x}(x, y), \quad \frac{\partial f}{\partial y}(x, y)$$

or with the arguments understood,

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}$$

In practice, given a function $f(x, y)$ one first computes the partial derivatives as functions of (x, y) and then evaluates them at particular points (a, b) as necessary. Sometimes vertical bars to indicate this evaluation, e.g.,

$$\frac{\partial f}{\partial x}(a, b) = \left. \frac{\partial f}{\partial x} \right|_{(a, b)}$$

You are of course already familiar with everything just stated from functions of one variable. The derivative, $f'(x)$, is itself a function of x . One often suppresses the argument x by writing just $\frac{df}{dx}$ or f' . To compute the derivative at a point one differentiates and then evaluates the derivative function at the required point, e.g., $f(x) = \sin(x)$, gives $f'(x) = \cos(x)$, from which $f'(0) = 1$.

3.3 Gradient

The gradient plays a fundamental role in the differential calculus of functions of several variables. This week and next week we will discuss different uses and interpretations of the gradient. It will appear in many subsequent courses.

Let f be a functions of n variables. The **gradient vector**, denoted by ∇f , is the vector formed from the n partial derivatives

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_i}, \dots, \frac{\partial f}{\partial x_n} \right)$$

The gradient is vector quantity, it has n components, and it a function of coordinates (x_1, \dots, x_n) . We are particularly interested in functions of two and three variable, for which we can write explicitly

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y) \right) = \frac{\partial f}{\partial x}(x, y)\mathbf{i} + \frac{\partial f}{\partial y}(x, y)\mathbf{j}$$

and

$$\nabla f(x, y, z) = \frac{\partial f}{\partial x}(x, y, z)\mathbf{i} + \frac{\partial f}{\partial y}(x, y, z)\mathbf{j} + \frac{\partial f}{\partial z}(x, y, z)\mathbf{k}$$

3.4 Chain Rule

Almost all of the differentiation rules you know for functions of one variable go over to rules for partial derivative exactly as you expect. In fact, one usually does not even state them as rules for partial differentiation. For example, given $f(x, y)$ and $g(x, y)$ a partial derivative of their product is

$$\frac{\partial fg}{\partial x} = g \frac{\partial f}{\partial x} + f \frac{\partial g}{\partial x}$$

but this is obvious (or soon will be to you) since taking the x partial derivative means treating y as a constant and so the product rule really is just the product rule from ordinary differentiation.

The Chain Rule is different. It is also pervasive in the treatment of functions of several variables. Recall the Chain Rule for functions of one variable. It tells us how to differentiate functions of functions. Let $g(t) = f(h(t))$ then we have

$$g'(t) = f'(h(t)) h'(t)$$

In this chapter we consider the basic case of the multivariable Chain Rule where we have a real valued function of several variables, and each of these variables is a function of a single other variable. In later chapters this will be generalised.

For simplicity, consider a function of just two variables $f(x, y)$. Let both x and y be functions of a third variable t . We name these functions with the variable names and write $x(t)$ and $y(t)$. Using composition we can construct a function $g : \mathbb{R} \rightarrow \mathbb{R}$, $g(t) = f(x(t), y(t))$.

The chain Rule for this case is

$$\frac{dg}{dt}(t) = \frac{\partial f}{\partial x}(x(t), y(t)) \frac{dx}{dt}(t) + \frac{\partial f}{\partial y}(x(t), y(t)) \frac{dy}{dt}(t)$$

which is often written simply as

$$\frac{dg}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}$$

In the general case of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ where $f(x_1, \dots, x_n)$ and where each x_i is itself a function of a single variable t , we have

The **Chain Rule**. Let $g(t) = f(x_1(t), \dots, x_n(t))$, then

$$\frac{dg}{dt}(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_1(t), \dots, x_n(t)) \frac{dx_i}{dt}(t) \quad (3)$$

or

$$\frac{dg}{dt} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} \quad (4)$$

In words, the derivative is computed as follows: starting at the left, compute the partial derivative of f with respect to its first argument and multiply by the ordinary derivative of that argument function. Now do the same for the next argument of f and add that on. Continue until you get to the last component of argument of f .

Warning: *Most aspects of partial differentiation are straightforward, almost trivial extensions of what you know from functions of one variable. However, the Chain Rule has a tendency to cause trouble. The reason is compact notation that is used, as in Eq. (4). It is assumed you understand where the functions are being evaluated, so be sure you do understand. If necessary write out the arguments in full as in Eq. (3).*

3.5 Chain Rule (again)

Let us now re-approach the Chain Rule using vector notation. Given a function of n variables $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and a vector function $\mathbf{r} : \mathbb{R} \rightarrow \mathbb{R}^n$, into same n -dimensions, we can compose these to obtain

$$g = f \circ \mathbf{r} : \mathbb{R} \rightarrow \mathbb{R}, \quad \text{with} \quad g(t) = f(\mathbf{r}(t))$$

This is the same composition considered in previous section. We have just notationally replaced all of the component functions $x_i(t)$ with a single vector function $\mathbf{r}(t)$ and used the \circ notation for function composition.

Now re-write the Chain Rule using the gradient vector and the fact that $\frac{dx_i}{dt}$ are components of $\frac{d\mathbf{r}}{dt}$. Then

$$\frac{dg}{dt} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} = \nabla f \cdot \frac{d\mathbf{r}}{dt} = \nabla f \cdot \mathbf{r}'$$

The Chain Rule reduces to the dot product between the gradient vector and the derivative vector \mathbf{r}' .

The **Chain Rule** (again). Let $g(t) = f(\mathbf{r}(t))$, then

$$\frac{dg}{dt}(t) = \nabla f(\mathbf{r}(t)) \cdot \frac{d\mathbf{r}}{dt}(t) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

You should see clearly that this is simply the previous Chain Rule written using different notation.

3.6 Higher-Order Derivatives

Just as for functions of a single variable, it is generally possible to differentiate the derivative to obtain the second and higher derivatives. In the case of functions of several variables, there are potentially many second derivatives. For example, $f(x, y)$ has the following second derivatives:

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right) & \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) \\ \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) & \frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial y} \right) \end{aligned}$$

You can see that there are many possibilities for high derivatives of functions of several variables. One thing that you will learn is that the order of differentiation does not matter for **mixed partial derivatives**, e.g.,

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

in the case where $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ are themselves continuous functions. This will normally be the case for functions we consider in this course, but be careful about always assuming it to be true.

Additional Material (optional)

3.7 Directional Derivative

For simplicity we again initially restrict to the case $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Using vector notation, we can write the definitions of partial derivatives as

$$\begin{aligned}\frac{\partial f}{\partial x}(\mathbf{x}) &= \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{i}) - f(\mathbf{x})}{h} \\ \frac{\partial f}{\partial y}(\mathbf{x}) &= \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{j}) - f(\mathbf{x})}{h}\end{aligned}$$

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

As you may have guessed, there is nothing special about the unit vectors \mathbf{i} and \mathbf{j} and the derivative can be generalised to any direction \mathbf{u} , where $\mathbf{u} \in \mathbb{R}^2$ is a unit vector. This is called the **directional derivative** of $f(x, y)$ in the direction \mathbf{u} and is denoted by $D_{\mathbf{u}}f$. Specifically,

$$D_{\mathbf{u}}f(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h}$$

In the general case we have

The **directional derivative** of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ in the direction \mathbf{u} is

$$D_{\mathbf{u}}f(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h}$$

where $\mathbf{u} \in \mathbb{R}^n$ is a unit vector.

While one can compute the directional derivative from the definition, it is more common to use the gradient vector as follows. Let

$$g(t) = f(\mathbf{r}(t)), \quad \text{where } \mathbf{r}(t) = \mathbf{x} + t\mathbf{u}$$

with \mathbf{x} and \mathbf{u} fixed with \mathbf{u} a unit vector. These have the same meaning as above: \mathbf{x} will be the point where we evaluate the directional derivative and \mathbf{u} is the direction. Note

$$\mathbf{r}(0) = \mathbf{x} \quad \mathbf{r}'(t) = \mathbf{u}$$

Now we compute $\frac{dg}{dt}(0)$ two ways. By definition:

$$\begin{aligned}\frac{dg}{dt}(0) &= \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} = \lim_{h \rightarrow 0} \frac{f(\mathbf{r}(h)) - f(\mathbf{r}(0))}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h} = D_{\mathbf{u}}f(\mathbf{x})\end{aligned}$$

By the Chain Rule:

$$\frac{dg}{dt}(0) = \nabla f(\mathbf{r}(0)) \cdot \mathbf{r}'(0) = \nabla f(\mathbf{x}) \cdot \mathbf{u}$$

Equate the two we obtain

The directional derivative of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ in the direction \mathbf{u} can be obtained as the dot product of the gradient vector ∇f and the direction vector \mathbf{u} :

$$D_{\mathbf{u}}f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{u}$$

Caution: there is variation in the definition of directional derivative. Some authors do not require \mathbf{u} to be a unit vector, and then there is variation in how the case of non-unit vectors is treated. However, this is not an issue when \mathbf{u} is a unit vector and we will always work with unit vectors when taking directional derivatives.

3.8 Caution on the extension to \mathbb{R}^n

Extending analysis from function $f : \mathbb{R} \rightarrow \mathbb{R}$ to functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is more involved than it might at first appear. One can see this from the example function

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

What is the limit of $f(x, y)$ as $(x, y) \rightarrow (0, 0)$? Since $f = 0$ when either $x = 0$ or $y = 0$ the limit $f(x, y)$ approaching the origin along either the x or y axis is 0. This is also the limit approaching the origin along any line $y = mx$. Hence it might seem that the $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0$. However, along the curve $x = y^2$, we have $f(x = y^2, y) = y^4/2y^4 = 1/2$. Hence, if the origin is approached along this curve the limit is $1/2$. Since one obtains different values depending on how $(0, 0)$ is approached, the limit does not, in fact, exist.

This illustrates that limits and continuity for functions on \mathbb{R}^n cannot be viewed from a one-dimensional perspective, but must be properly generalised using regions (called neighbourhoods) in \mathbb{R}^n . This will be covered in later Analysis modules and in Differentiation. While we will not define these things here, we will sometimes state properties that hold for continuous functions. You will just have to take this on faith for the present.

Fortunately, several of the most important aspects of multivariable calculus are “one dimensional” and follow easily from things you know. Nothing stops us from being able to define and do calculations using these quantities.

Quadric surfaces

A **quadric surface** is the set of points in \mathbb{R}^3 that satisfy a second-degree equation three variables x, y, z . The most general form of such an equation is:

$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J = 0$$

for constants A, \dots, J . In most cases (non-degenerate cases), by translation and rotation of coordinates it is possible to bring the equation into standard form of

$$Ax^2 + By^2 + Cz^2 + J = 0 \quad \text{or} \quad Ax^2 + By^2 + Iz = 0$$

Quadric surfaces are the generalisation to three dimensions of conic sections in two dimensions.

You should see that points satisfying an equation in three variables, e.g., $f(x, y, z) = 0$, is no different than the zero level set, or isosurface, of a function of three variables $f(x, y, z)$.

We will potentially be interested in the following surfaces and will use them as examples throughout the remainder of the module.

Ellipsoid:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Horizontal and vertical cuts are ellipses. For $a = b \neq c$ this is a **spheroid**. For $a = b = c$ this is a **sphere**.

Note that here, and below, we follow common practice and write the equation with the constant, or lower-order terms, on the right hand side of the equal sign. In this form it is evident that the ellipsoid is the $k = 1$ isosurface of the second-degree polynomial

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

Elliptic Paraboloid:

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

Horizontal cuts are ellipses and vertical cuts are parabolas.

Hyperbolic Paraboloid:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \frac{z}{c}.$$

Horizontal cuts are hyperbolas and vertical cuts are parabolas.

Hyperboloid of One Sheet:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

Horizontal cuts are ellipses. Vertical cuts are hyperbolas.

Hyperboloid of Two Sheets:

$$-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Horizontal cuts are ellipses (if they intersect the surface). Vertical cuts are hyperbolas.

We also will want to consider the degenerate case of circular cylinders.

Circular Cylinder:

$$x^2 + y^2 = a^2$$

In practice we will consider cylinders whose axis is parallel to the x or y coordinate axes.

Making contour plots

For simple functions you should be able to sketch contours and thus produce an approximate contour map. You should understand the relationship between graph of f and its contour map and you should be able to describe a function given a contour map.

In practice one often uses software to generate of contours. Algorithms for generating contours are non-trivial. Think a little about what you might do to numerically generate all *curves* at a given level for a function f . While commonly the contour levels correspond to an equal spacing in k , at times it might be more appropriate to choose a different spacing, e.g., powers of 10, $k = 1, 10, 100, \dots$. It is also common to plot contours using colour or grey scale values. Here each set \mathcal{L}_k is assigned a specific colour or grey value depending on k .

Another approach (far easier algorithmically) is to generate colour contour plots where one only need consider a grid of values (x_i, y_j) covering the region of interest. For each grid point one computes $f(x_i, y_j)$ and assigns a corresponding colour or grey level. It is not necessary to generate any curves in the plane – your eye will do that for you.

4 Week 8 Lecture 1: Multiple Integration 1

Introduction

Given a continuous function $f(x) > 0$, you know that the integral $\int_a^b f(x) dx$ can be interpreted as the area under the curve $f(x)$ from $x = a$ to $x = b$. This interpretation as area under the curve is intimately connected to with the definition of integration. Since the graph of a function of two variables $f(x, y) > 0$ is a surface above the xy -plane, it is natural to ask about the volume under the sheet. This volume will be expressed as a multiple integral.

4.1 Multiple Integration

Consider a function f of two variables which is defined, and is positive, on the closed rectangle, i.e., $f : R \subset \mathbb{R}^2 \rightarrow \mathbb{R}^+$, where

$$R = [a, b] \times [c, d] = \{(x, y) \mid a \leq x \leq b, c \leq y \leq d\}$$

Our goal is to compute the volume V of the three dimensional solid Ω generated by the graph of f above this rectangle

$$\Omega = \{(x, y, z) \mid 0 \leq z \leq f(x, y), (x, y) \in R\}$$

This volume can be express as a double integral of f over the rectangle R , written

The **double integral** of f over the rectangle R is

$$V = \iint_R f dA$$

This is just notation. One can define the double integral as a limit of a double sum (in x and y directions) of rectangular columns, similar to how one defines a single integral for a function of one variable as the limit of a sum of rectangles. (See optional material at end of this week's notes, or better look at figures in Chapter 15 of Stewart *et al.*) While it is useful for you to look briefly at how double integrals are defined, the important thing is to be able to calculate such integrals using iterated integration.

4.2 Iterated Integration over a Rectangle

To compute double integrals, one relies on iterated or repeated one-variable integration as we now explain. Consider a fixed value of y between c and d . The area under the curve $f(x, y)$ from $x = a$ to $x = b$ with fixed y is given by

$$A(y) = \int_a^b f(x, y) dx$$

We use A for area. It depends on the fixed value of y , hence we write $A(y)$.

Such an integral is call **partial integration** – integration of a function of several variables with respect to one variable while treating any others as constants. You will recognise the analogy with partial differentiation.

Given that $A(y)$ is the area under the function at a given value of y , the volume will be the integral of these areas over y , that is

$$V = \int_c^d A(y) dy$$

Hence the volume, double integral, can be expressed as

$$V = \iint_R f \, dA = \int_c^d A(y) \, dy = \int_c^d \left[\int_a^b f(x, y) \, dx \right] dy$$

The inner integral gives the area under a curve at constant y . The outer integral integrates these areas over the range of y to give the volume.

We did not have to order the integrations such that the x -integral appeared on the inside. One could have equally put the y integration on the inside.

We now drop the specific interpretation of volume under a function and arrive at the following relationship between double integrals and iterated integrals.

The double integral of f over the rectangle $R = [a, b] \times [c, d]$ is given by iterated, also known as repeated, integration

$$\iint_R f \, dA = \int_c^d \int_a^b f(x, y) \, dx \, dy = \int_a^b \int_c^d f(x, y) \, dy \, dx$$

Discussion

- *Notation:* You will notice that we dropped the square brackets in boxed expression above. This is common. All of the following are used to denote iterated integration

$$\int_c^d \int_a^b f(x, y) \, dx \, dy, \quad \int_c^d \left[\int_a^b f(x, y) \, dx \right] dy,$$

$$\int_{y=c}^{y=d} \int_{x=a}^{x=b} f(x, y) \, dx \, dy, \quad \int_{y=c}^d dy \int_{x=a}^b dx f(x, y)$$

and other combinations of these forms.

- The boxed expression is known as *Fubini's Theorem*, or would be if we stated it as a theorem. Equality is guaranteed if f is a continuous function, and even under weaker conditions on f .
- Understanding the following is crucial to iterated integration: Integrals are nested. (Now just 2 levels, but in higher dimensions there will be more nesting.) Integrals are done from inside out. You work with one variable at a time treating any variables outside the current level as constants. As we generalise to other integration domains this will be key to your success. If you are ever confused by an iterated integral, explicitly include the square brackets showing the nesting.
- Sometimes the function $f(x, y)$ separates into the product of a function of x only and a function of y only. Letting $f(x, y) = g(x)h(y)$,

$$\int_c^d \int_a^b f(x, y) \, dx \, dy = \int_c^d \left[\int_a^b g(x)h(y) \, dx \right] dy$$

but $h(y)$ is a constant as far as the inner x integration is concerned, so it can be pulled out

$$\int_c^d h(y) \left[\int_a^b g(x) \, dx \right] dy$$

and now the whole x integral is a constant as far as the y integration is concerned so it can be pulled out

$$\left[\int_a^b g(x) \, dx \right] \int_c^d h(y) \, dy$$

Thus the double integral is said to **separate**

$$\int_c^d \int_a^b f(x, y) \, dx \, dy = \int_a^b g(x) \, dx \int_c^d h(y) \, dy$$

4.3 Multiple Integration in Three or More Variables

The extension of all the above to functions of more than two variables is straightforward. We will only explicitly consider functions of three variables, but functions of four, five, etc variables are essentially the same.

The following treatment is a natural extension of the above. The reader can fill in the details. Rather than consider volume under a surface, a useful physical example would be the total mass of a solid whose density (mass per unit volume) is given by $f(x, y, z)$. f is necessarily positive in this example. Assume the solid is in the shape of a rectangular box. We then have $f : B \subset \mathbb{R}^3 \rightarrow \mathbb{R}^+$, where B is the rectangular box

$$\begin{aligned} B &= [a, b] \times [c, d] \times [r, s] \\ &= \{(x, y, z) \mid a \leq x \leq b, c \leq y \leq d, r \leq z \leq s\}. \end{aligned}$$

The total mass of the solid given as a triple integral

$$\text{Total Mass} = \iiint_B f \, dV$$

Now forgetting the particular physical example, and the requirement that f be positive,

The **triple integral** of f over the rectangular box $B = [a, b] \times [c, d] \times [r, s]$ is given by iterated integration

$$\iiint_B f \, dV = \int_r^s \int_c^d \int_a^b f(x, y, z) \, dx \, dy \, dz$$

or other orderings of the x , y , and z integrals.

The same comments from Sec. 4.5 and 4.2 apply here as well.

4.4 Area and Volume

Suppose you want to compute the area of a region Ω in the plane. Using the formalism of double integrals, we can just leave out the function $f(x, y)$ to get the area of a region. Similarly one can obtain the volume of a region Ω in space as a triple integral – just leave out the function $f(x, y, z)$ in triple integration.

Denoting the area of a planar region Ω by $A(\Omega)$ and the volume of a space region by $V(\Omega)$, we have

Area of a region Ω in the plane

$$A(\Omega) = \iint_{\Omega} dA$$

Volume of a region Ω in space

$$V(\Omega) = \iiint_{\Omega} dV$$

4.5 "Derivation" of Multiple Integration (Optional)

This will be a quick, heuristic development of multiple integration. Consider a function f of two variables which is defined, and is positive, on the closed rectangle, i.e., $f : R \subset \mathbb{R}^2 \rightarrow \mathbb{R}^+$, where

$$R = [a, b] \times [c, d] = \{(x, y) \mid a \leq x \leq b, c \leq y \leq d\}$$

Our goal is to compute the volume V of the three dimensional solid Ω generated by the graph of f above this rectangle

$$\Omega = \{(x, y, z) \mid 0 \leq z \leq f(x, y), (x, y) \in R\}$$

To compute the volume we partition the rectangle R into subrectangles by partitioning separately each of the intervals $[a, b]$ and $[c, d]$ as follows. First partition $[a, b]$ into N equal subintervals by letting $x_0 = a, x_1 = a + \Delta x, \dots, x_i = a + i\Delta x, \dots, x_N = b$. This gives subintervals $[x_i, x_{i+1}]$ each with length $\Delta x = (b - a)/N$.

Similarly partition $[c, d]$ into M equal subintervals $[y_j, y_{j+1}]$ of length Δy by letting $y_0 = c, y_1 = c + \Delta y, \dots, y_j = c + j\Delta y, \dots, y_M = d$. Then form subrectangles R_{ij} of R by taking the Cartesian product of the x and y subintervals:

$$R_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$$

So each subrectangle has area $\Delta A = \Delta x \Delta y$.

We can now approximate the volume of the solid over each subrectangle as $f(x_i, y_j)\Delta A$. Summing these we get an approximation for the volume of Ω

$$V \approx \sum_{j=0}^{M-1} \sum_{i=0}^{N-1} f(x_i, y_j)\Delta A$$

For N and M large, the size of each subrectangle is small and f does not vary much over the subrectangle. Each term in the sum then corresponds to a subvolume that is very tall compared with its base dimensions. Now take the limit as the number of elements N and M goes to infinity, with the sizes Δx and Δy going to zero. We thus get the exact volume

$$V = \lim_{M, N \rightarrow \infty} \sum_{j=0}^{M-1} \sum_{i=0}^{N-1} f(x_i, y_j)\Delta A$$

The volume under a function of two variables is just one of numerous quantities one wants to compute in this way. In general there is no reason to restrict functions taking on positive values, so we now drop this restriction and define

The **double integral** of f over the rectangle R is

$$\iint_R f \, dA = \lim_{M, N \rightarrow \infty} \sum_{j=0}^{M-1} \sum_{i=0}^{N-1} f(x_i, y_j)\Delta A$$

Comments

- We have assumed here that the limit exists. If the limit exists, f is said to be **integrable** and in later courses you will learn that f is continuous (in the multivariable sense which we have not defined), then it is integrable.
- It is unnecessary that R be partitioned into equal rectangles, all with the same area. In fact it is unnecessary that the partition of R use rectangles at all, we did this here for convenience.
- We chose to approximate the subvolumes using f evaluated at (x_i, y_j) , but we could have used any sample point within R_{ij} . It does not matter where we evaluate f within each element since in the limit their size goes to zero.

5 Week 8 Lecture 2: Multiple Integration 2

5.1 Iterated Integration over General Domains

Now the fun starts. We are going to consider integration of functions over more general domains, first in 2D then in 3D. Doing integrals is not going to be the issue – getting the geometry right is.

The notes are short for this lecture, because there is not a lot to explain. One must do lots of examples.

Two variables

Let Ω denote the region or domain in \mathbb{R}^2 on which f is defined and over which we wish to integrate.

A **Type I** region can be expressed

$$\Omega = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, g_1(x) \leq y \leq g_2(x)\}$$

where g_1 and g_2 are continuous functions on $[a, b]$. We evaluate the double integral of f over Ω as an iterated integral in the following way

$$\iint_{\Omega} f dA = \int_a^b \left[\int_{g_1(x)}^{g_2(x)} f(x, y) dy \right] dx$$

Let us elaborate by examining the nesting from inside to outside. Within the brackets, x is treated as a constant. Its actual value is not known, but everywhere it appears it acts like a constant. Hence the limits of the y -integral are the “constants” $g_1(x)$ and $g_2(x)$. Assuming one can partially integrate $f(x, y)$ with respect to y , then one can evaluate the inner brackets resulting in a function of x only. Then the outer integral is a standard integral over the interval $[a, b]$.

A **Type II** region or domain is one that can be expressed

$$\Omega = \{(x, y) \in \mathbb{R}^2 \mid h_1(y) \leq x \leq h_2(y), c \leq y \leq d\}$$

where h_1 and h_2 are continuous functions on $[c, d]$. In this case we evaluate the double integral of f over Ω as iterated integral in the following way

$$\iint_{\Omega} f dA = \int_c^d \left[\int_{h_1(y)}^{h_2(y)} f(x, y) dx \right] dy$$

Comments

- There are two aspects to iterated integrals over these types of domains. First setting up the integration correctly and second carrying out the integration correctly. One simply needs to practice a lot.
- My advice is, given some specification of the domain Ω , first write it explicitly as a set either of Type I or Type II. It will then be immediately obvious which integral is on the inside and which is on the outside and what the limits of integration are. If you cannot express Ω in this form, you won't be able to write the integral.
- Many domains can equally be written in Type I or Type II format. One way may be easier to calculate in the end, but it is not always obvious when you start. Again, practice helps.

Three variables

Now let Ω be a region or domain in \mathbb{R}^3 on which f is continuous and over which we wish to integrate. In analogy with the 2D case we could define 6 types of regions. We won't do this, but instead look at a one case and let you deduce the others.

Suppose the set Ω can be written

$$\Omega = \{(x, y, z) \in \mathbb{R}^3 \mid a \leq x \leq b, g_1(x) \leq y \leq g_2(x), \\ u_1(x, y) \leq z \leq u_2(x, y)\}$$

Then

$$\iiint_{\Omega} f dV = \int_a^b \left[\int_{g_1(x)}^{g_2(x)} \left[\int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z) dz \right] dy \right] dx$$

The inner most integral is a partial integral over z . Here x and y , and hence integration limits $u_1(x, y)$ and $u_2(x, y)$, are viewed as constants. The result will be a function of (x, y) . The middle integral is a partial integral over y with x treated as constant, followed by the outer definite integral over x with limits a and b .

5.2 Area and volume elements

We began the topic integration by first recalling the finding the area under a curve for a function of a single variable,

$$\text{Area under curve} = \int_a^b f(x) dx$$

The notation for integration is chosen to evoke that the area under a curve is the sum of (infinitely many) rectangles of height $f(x)$ and infinitesimal width dx .

We then introduced double integration by considering the volume under a surface for a function of two variables, and wrote

$$\text{Volume under surface} = \iint_R f dA$$

While we did not formally define the double integral, the notation is similarly meant to evoke that the volume is given by a sum of (infinitely many) boxes of height f and infinitesimal base area dA .

We then stated that the double integral can be evaluated by iterated integrals

$$\iint_R f dA = \int_c^d \int_a^b f(x, y) dx dy$$

Comparing the two sides, it is evident that the infinitesimal base area dA on the left hand side becomes $dx dy$ on the right hand side. Hence,

$$dA = dx dy$$

The correct way to think about this is that an infinitesimal element of area dA in the plane is given by the product of infinitesimals dx and dy , where x and y are Cartesian coordinates.

Similarly for integration in 3D,

$$dV = dx dy dz$$

It is possible to make these statements rigorous, and derive the results through a systematic approach, but this is beyond the scope of MA133. We will have to rely on heuristic understanding.

We are going to consider important non-Cartesian coordinate systems and address the question: what are the infinitesimal elements of area dA and volume dV in terms of infinitesimal coordinate increments in these coordinate systems?

6 Week 8 Lecture 3: Integration in Special Coordinates

Introduction

Many problems naturally involve symmetry. One should exploit it where possible and this often means using coordinate systems other than Cartesian coordinates.

6.1 Polar Coordinates

You know the relationship between polar coordinates (r, θ) and Cartesian coordinates (x, y) :

$$\begin{aligned}x &= r \cos \theta & r^2 &= x^2 + y^2 \\y &= r \sin \theta & \tan \theta &= \frac{y}{x}\end{aligned}$$

It is useful to view the polar coordinate system in terms of a polar grid consisting of curves of constant r -coordinate – circles centred on the origin, and curves of constant θ -coordinate – radial lines.

Now consider the equivalent of a rectangular region R in polar coordinates.

$$R = \{(r, \theta) \mid a \leq r \leq b, \alpha \leq \theta \leq \beta\}$$

This is referred to as a **polar rectangle**. Of course this is not a rectangle in the plane, but typically a wedge (although these include disks and annuli).

In polar coordinates the relationship between the double integral and iterated integrals

In polar coordinates:

$$\iint_R f \, dA = \int_{\alpha}^{\beta} \int_a^b f(r \cos \theta, r \sin \theta) r \, dr \, d\theta$$

where

$$R = \{(r, \theta) \mid a \leq r \leq b, \alpha \leq \theta \leq \beta\}$$

The key feature that distinguishes integration in polar coordinates from integration in Cartesian coordinates is that now the infinitesimal area element is no longer independent of position. It depends on coordinate r .

Key fact, in polar coordinates:

$$dA = r \, dr \, d\theta$$

This is the relationship between the infinitesimal area in polar coordinates and the infinitesimal changes in the coordinates dr and $d\theta$. You should understand this as follows. If one makes a small (infinitesimal) change dr in the r coordinate, then the area element will be proportional to that change, while if one makes an small (infinitesimal) change $d\theta$ in the θ coordinate, then the area element will be proportional $r d\theta$.

6.2 Iterated Integration in Cylindrical Coordinates

Cylindrical coordinates (r, θ, z) are a three dimensional coordinate systems composed of polar coordinates (r, θ) in the plane and a Cartesian coordinate z in the third direction, generally thought of as the vertical direction. The relationship between cylindrical and Cartesian coordinates is

$$\begin{aligned}x &= r \cos \theta & r &= (x^2 + y^2)^{1/2} \\y &= r \sin \theta & \tan \theta &= \frac{y}{x} \\z &= z & z &= z\end{aligned}$$

One could give a different symbol ξ to the vertical coordinate in cylindrical coordinates and write relationship between this cylindrical coordinate and the Cartesian coordinate is $\xi = z$, but this is silly.

It is useful to view the cylindrical coordinate system in terms of a grid consisting of surfaces of constant r -coordinate – cylinders centred on the z -axis, surfaces of constant θ -coordinate – radial half-planes, and surfaces of constant z -coordinate – horizontal planes.

We suppose have a function f defined on a cylindrical wedge given by

$$\Omega = \{(r, \theta, z) \mid a \leq r \leq b, \alpha \leq \theta \leq \beta, c \leq z \leq d\}.$$

and we want to compute the triple of f over Ω .

In cylindrical coordinates:

$$\iiint_{\Omega} f \, dV = \int_c^d \int_{\alpha}^{\beta} \int_a^b f(r \cos \theta, r \sin \theta, z) r \, dr \, d\theta \, dz$$

where

$$\Omega = \{(r, \theta, z) \mid a \leq r \leq b, \alpha \leq \theta \leq \beta, c \leq z \leq d\}.$$

Key point is the expression of the infinitesimal volume element dV in terms of infinitesimal changes in the coordinates.

Key fact, in cylindrical coordinates:

$$dV = r \, dr \, d\theta \, dz$$

Discussion

- In problems with symmetry, the integral separates in the corresponding coordinate, or coordinates, and the integrals simplify.
- Sometimes we use special coordinates because of geometry, that is the region of integration suggest special coordinates, but equally often we use special coordinates because the integrand simplifies in these coordinates. Another way to say this is that in many cases the function we wish to integrate arises because a problem depends only on distance from the origin (polar coordinates in 2D and spherical coordinates in 3D) or because a problem depends only on distance from an axis (cylindrical coordinates).
- In our statements of formulas for iterated integration we took the regions to be simple and hence the limits of integration were constants. However, just as with integration in Cartesian coordinates, the regions need not be this simple. In general, in iterated integration the limits of any inner integrals may depend on coordinates of any of the outer integrals.

6.3 Iterated Integration in Spherical Coordinates (Mostly Optional)

Spherical coordinates (r, θ, ϕ) are a three dimensional coordinate systems where r is the distance from the origin (three-dimensional radial coordinate), θ is same angle as in cylindrical coordinates (corresponds to longitude where the x -axis is zero longitude), and ϕ is the angle from the vertical (angle from the north pole or co-latitude). The relationship between spherical and Cartesian coordinates is

$$\begin{aligned} x &= r \sin \phi \cos \theta & r &= (x^2 + y^2 + z^2)^{1/2} \\ y &= r \sin \phi \sin \theta & \tan \phi &= \frac{(x^2 + y^2)^{1/2}}{z} \\ z &= r \cos \phi & \tan \theta &= \frac{y}{x} \end{aligned}$$

Some authors give a different symbol ρ to the radial coordinate in spherical coordinates but we will not do this. (*Be warned, many authors use θ and ϕ in exactly the reverse of the roles here.*) The ranges of spherical coordinates are

$$r \geq 0 \quad 0 \leq \theta \leq 2\pi \quad 0 \leq \phi \leq \pi$$

We suppose have a function f of spherical coordinates $f(r, \theta, \phi)$ defined on a spherical wedge

$$\Omega = \{(r, \theta, \phi) \mid a \leq r \leq b, \alpha \leq \theta \leq \beta, \gamma \leq \phi \leq \delta\}.$$

and we want to compute the triple of f over Ω .

In spherical coordinates:

$$\iiint_{\Omega} f \, dV = \int_{\gamma}^{\delta} \int_{\alpha}^{\beta} \int_a^b f(r, \theta, \phi) r^2 \sin \phi \, dr \, d\theta \, d\phi$$

where

$$\Omega = \{(r, \theta, \phi) \mid a \leq r \leq b, \alpha \leq \theta \leq \beta, \gamma \leq \phi \leq \delta\}.$$

Note that in the above we have written $f(r, \theta, \phi)$, meaning that the function is already expressed in spherical coordinates. If the function is given in Cartesian coordinates, one needs to first substitute for x , y and z in terms of r , θ and ϕ .

Key point is the expression of the infinitesimal volume element dV in terms of infinitesimal changes in the coordinates.

Key fact, in spherical coordinates:

$$dV = r^2 \sin \phi \, dr \, d\theta \, d\phi$$

Integrating function of r only (Not Optional)

Suppose the function we wish to integrate is a function of r only in spherical coordinates. Suppose further that we want to integrate this function over a sphere of radius R , then the triple integral becomes

$$\begin{aligned} \iiint_{\Omega} f \, dV &= \int_0^{\pi} \int_0^{2\pi} \int_0^R f(r) r^2 \sin \phi \, dr \, d\theta \, d\phi = \int_0^{\pi} \sin \phi \, d\phi \int_0^{2\pi} d\theta \int_0^R f(r) r^2 \, dr \\ &= 2 \cdot 2\pi \cdot \int_0^R f(r) r^2 \, dr = 4\pi \int_0^R f(r) r^2 \, dr \end{aligned}$$

If Ω is a sphere of radius R and f a function of r only, then

$$\iiint_{\Omega} f \, dV = 4\pi \int_0^R f(r) r^2 \, dr$$

Some pictures

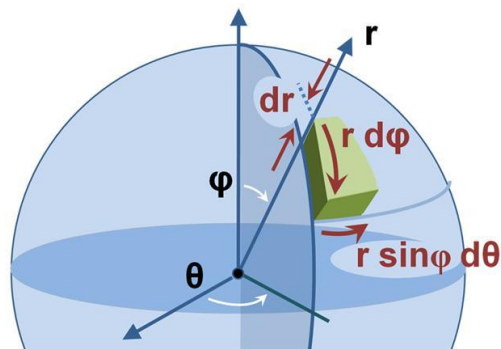
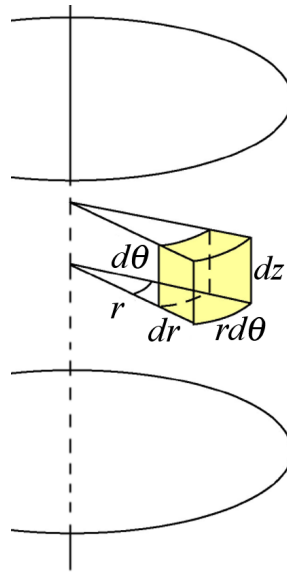


Figure 4: Volume elements in cylindrical and spherical coordinate systems. (Cylindrical case taken from MIT Physics 8.01 course notes. Spherical case from Wikipedia.)

7 Week 9 Lecture 1: Vector Fields

7.1 Introduction

We have considered vector functions

$$\mathbf{r} : U \subseteq \mathbb{R} \rightarrow \mathbb{R}^n$$

and their relationship to curves in \mathbb{R}^n .

We have considered functions of several variables

$$f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$$

and how to differentiate such functions (partial derivatives) and integrate such functions (multiple integrals).

We will now consider vector fields

$$\mathbf{F} : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$$

The function \mathbf{F} is viewed as assigning an n -vector to each point in \mathbb{R}^n . This is called a **vector field on \mathbb{R}^n** .

There are many situations in which a vector is associated to each point in some region of space. Familiar examples come from fluid motion such as the wind or the motion of water in a river. Wind has both a magnitude and a direction (blowing North-East at 18 miles/hour) and hence is a vector quantity whose value generally varies with location. Other applications include gravitational fields, electric fields, and magnetic fields. These are all vector fields.

We typically use boldface to denote vector fields, (usually \mathbf{F} , \mathbf{V} , \mathbf{u} or \mathbf{v}), to emphasise that to each point in U the function \mathbf{F} assigns a vector. We will primarily be interested in vector fields on the plane ($n = 2$) or in space ($n = 3$).

A few examples and notation

- The planar vector field,

$$\mathbf{F}(x, y) = -y\mathbf{i} + x\mathbf{j}$$

corresponds to vectors pointing counterclockwise around the origin.

The vector field,

$$\mathbf{F}(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

corresponds to vectors in space pointing away from the origin.

- More generally, we can write our vector fields on \mathbb{R}^2 as

$$\mathbf{F}(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$$

where $P(x, y)$ and $Q(x, y)$ are component functions, each a real valued function of two variables (coordinates).

Similarly for a vector field on \mathbb{R}^3

$$\mathbf{F}(x, y, z) = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}$$

where P , Q , and R are real valued functions of three variables (coordinates).

- Since a point in the plane or in space can be written as a vector $\mathbf{x} = (x, y)$ or $\mathbf{x} = (x, y, z)$, we will sometimes write our vector fields compactly as

$$\mathbf{F}(\mathbf{x})$$

You should always think of this as \mathbf{x} specifies the coordinates of a point, and \mathbf{F} is the vector associated with that point.

- An example we have already seen is the gradient of a function f of several variables. The gradient is a vector field since $\nabla f(\mathbf{x})$ is a vector whose value depends on the point \mathbf{x} . On the plane for example

$$\mathbf{F}(\mathbf{x}) = \nabla f(\mathbf{x}) = \nabla f(x, y) = \frac{\partial f}{\partial x}(x, y)\mathbf{i} + \frac{\partial f}{\partial y}(x, y)\mathbf{j}$$

- Systems of ordinary differential equations

$$\dot{x}_1 = f_1(x_1, \dots, x_n)$$

\vdots

$$\dot{x}_n = f_n(x_1, \dots, x_n)$$

can be viewed as vector fields on the corresponding phase space. To each point in phase space $\mathbf{x} = (x_1, \dots, x_n)$ there is an associated derivative vector,

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) = (f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n))$$

We return to this important example next week.

Comment on scalar fields

In the context of what we will now be discussing, one can view a real valued function f of several variables as a **scalar field**. That is, to each point (x_1, x_2, \dots, x_n) in \mathbb{R}^n , the function f assigns a scalar value.

7.2 Differential calculus of vector fields

There are two important differential operations that commonly arise for vector fields. These are known as *divergence* and *curl*. One initially just learns how to compute these, and then with time their role in applications will become evident.

7.2.1 Divergence

Let \mathbf{F} be a vector field defined on \mathbb{R}^3

$$\mathbf{F}(x, y, z) = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}$$

Then the divergence of \mathbf{F} is defined to be

$$\operatorname{div} \mathbf{F} = \nabla \cdot \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$$

Note that both $\operatorname{div} \mathbf{F}$ and $\nabla \cdot \mathbf{F}$ are used to denote divergence.

For a vector field \mathbf{F} defined on \mathbb{R}^2

$$\mathbf{F}(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$$

The divergence is

$$\operatorname{div} \mathbf{F} = \nabla \cdot \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}$$

As you can surmise, the divergence in any dimension is just the sum of the partial derivatives of the component functions. We will only consider the cases $n = 2$ and $n = 3$ as defined above.

Note that divergence is a real value, i.e., a scalar. Hence, divergence takes a vector field and gives a scalar field.

7.2.2 Curl

Curl is defined for vector fields \mathbf{F} on \mathbb{R}^3 . The definition is best expressed using the determinant notation used for the cross product. If

$$\mathbf{F}(x, y, z) = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}$$

Then the curl of \mathbf{F} is defined to be

$$\operatorname{curl} \mathbf{F} = \nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$$

From completeness we write out the expression for curl,

$$\text{curl } \mathbf{F} = \nabla \times \mathbf{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) \mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) \mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathbf{k}$$

but is it not recommended that you memorise the above line, but rather use the determinant notation.

Similar to the case of divergence, both $\text{curl } \mathbf{F}$ and $\nabla \times \mathbf{F}$ are used to denote the curl of \mathbf{F} .

Curl is a more involved calculation than divergence, but with a little practice you will be easily able to master it.

7.2.3 "Curl" of 2D vector field

Although curl is defined for 3D vector fields, the following is common and useful. Suppose we have a 2D vector field

$$\mathbf{F}(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$$

We can think of this as a 3D vector field with a zero third component and no dependence on z

$$\mathbf{F}(x, y, z) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j} + 0\mathbf{k}$$

Plugging this into the definition of curl we obtain

$$\text{curl } \mathbf{F} = \nabla \times \mathbf{F} = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \mathbf{k}$$

Hence, $\text{curl } \mathbf{F}$ has component only in the \mathbf{k} direction and it is only a function of (x, y) . It can be useful to consider just this component of the curl, specifically

$$h(x, y) = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) = \mathbf{k} \cdot (\text{curl } \mathbf{F})$$

7.3 Vector calculus identities

There are a number of identities that arise and are important in vector calculus. We will only consider a few of the most important ones.

div curl \mathbf{F}

The divergence of the curl of a vector field is necessarily zero.

$$\text{div } (\text{curl } \mathbf{F}) = \nabla \cdot (\nabla \times \mathbf{F}) = 0$$

curl grad f

The curl of the gradient of a function of three variables is necessarily zero.

$$\text{curl } (\text{grad } f) = \nabla \times (\nabla f) = 0$$

Proofs of the above two statements are left to the reader. Proofs follow by direct application of the definitions.

Laplacian

The divergence of the gradient of a function f of several variables is called the Laplacian and is denoted $\nabla^2 f$

$$\nabla^2 f = \nabla \cdot (\nabla f)$$

Note that the Laplacian is a scalar, just as is f . We have already seen the Laplacian, although we did not give it a name.

The definition applies in any dimension, although as usual here we will be interested in the cases $n = 2$ and $n = 3$. Using the definitions of div and grad one finds that the Laplacian is just the sum of second derivatives, e.g., for $n = 3$

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

7.4 Final Remarks

At this point you should just learn how to compute div and curl from vector fields. Together with grad for scalar fields, these form the basis of differential vector calculus. The only two vector identities that you need to know now are "div curl = 0" and "curl grad = 0". These are statements of fact you should just know. You should know that the Laplacian is the div of grad.

Note that we are here working only in Cartesian coordinates. Div, grad, curl take different forms in different coordinate systems. However, the vector identities hold independently of coordinate system.

It is highly recommended that you look at the Wikipedia page on Vector Calculus Identities.

8 Week 9 Lecture 2: Line Integrals I

Introduction

We return to parametrised curves and consider integration along such curves. We already saw this when we integrated along a curve to find its length. Here we generalise this with particular emphasis on integrating over vector fields.

8.1 Line Integrals

The basic line integral can be motivated as follows. Given an interval $[a, b]$ and a function $f(x)$ which is positive over the interval, $\int_a^b f(x) dx$ is the area under the graph $y = f(x)$. Intuitively one understands that $f(x) dx$ is the area of a tall skinny rectangle of height $f(x)$ and width dx and \int_a^b means "add these up" for x 's in the interval $[a, b]$.

Why restrict ourselves to just integrating along straight lines? We know how to work with curves so let us generalise and consider a curve in the plane and a function $f(x, y)$ that is positive in some region containing the curve. A surface is formed by f over the curve. Think of a curtain hanging down from f to the curve. We want to compute the area of this curtain by integration.

Recall the length of a curve \mathcal{C} is given by

$$\ell(\mathcal{C}) = \int_{\mathcal{C}} ds$$

where ds is the infinitesimal arc length, or distance, along the curve. Thus to find the area of the curtain formed from f over \mathcal{C} , we simply multiply the height f times the infinitesimal arc length ds and integrate over the curve

$$\int_{\mathcal{C}} f ds$$

In practice, such an integral is evaluate by parametrising the curve. Given a parametrisation of the curve $\mathbf{r}(t)$, $t \in [a, b]$, the infinitesimal arc length ds can be expressed in terms of the infinitesimal change dt via

$$ds = \|\mathbf{r}'(t)\| dt$$

so that arc length is in practice computed using

$$\int_{\mathcal{C}} ds = \int_a^b \|\mathbf{r}'(t)\| dt$$

Thus, in practice to find the area of the curtain formed from f over \mathcal{C} , we use

$$\int_{\mathcal{C}} f ds = \int_a^b f(\mathbf{r}(t)) \|\mathbf{r}'(t)\| dt$$

It is not necessary to restrict to positive functions, nor does the method depend on dimension. The relationship $ds = \|\mathbf{r}'(t)\| dt$ holds in any dimension. Thus we can go directly to the general formula.

Given $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ and $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$ a parametrisation of a curve \mathcal{C} lying in U , the **line integral** of f along a curve \mathcal{C} is given by

$$\int_{\mathcal{C}} f ds = \int_a^b f(\mathbf{r}(t)) \|\mathbf{r}'(t)\| dt$$

one must parametrise the curve in a sensible way.

8.2 Line Integrals for Vector Fields

Given a vector field \mathbf{F} , it frequently occurs that one wants to compute a line integral where the function f is

$$f = \mathbf{F} \cdot \mathbf{T}$$

where \mathbf{T} is the unit tangent vector to the curve \mathcal{C} . Examples of this type of integration are work and circulation discussed below. Hence we need to evaluate

$$\int_{\mathcal{C}} \mathbf{F} \cdot \mathbf{T} ds$$

To derive a useful formula for such an integral we recall that

$$\mathbf{T} = \frac{\mathbf{r}'}{\|\mathbf{r}'\|}$$

Thus we can write

$$\mathbf{F} \cdot \mathbf{T} ds = \mathbf{F} \cdot \frac{\mathbf{r}'}{\|\mathbf{r}'\|} \|\mathbf{r}'(t)\| dt = \mathbf{F} \cdot \mathbf{r}' dt$$

The right-most expression is what we will use in practice to evaluate this type of line integral. However, it is common to write $\mathbf{r}' dt$ as $d\mathbf{r}$.

Let \mathbf{F} be a vector field define in some region of \mathbb{R}^n , and let $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$ be a parametrisation of a curve \mathcal{C} in this region,

The **line integral of \mathbf{F} along \mathcal{C}** is

$$\int_{\mathcal{C}} \mathbf{F} \cdot \mathbf{T} ds = \int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt$$

One important feature of line integrals of vector fields is that they are *not independent of the orientation of the curve*. The reason is that if one reverses the orientation of a curve, then the tangent vector changes sign. Denoting $-\mathcal{C}$ as the curve \mathcal{C} with the opposite orientation, then

$$\int_{-\mathcal{C}} \mathbf{F} \cdot \mathbf{T} ds = - \int_{\mathcal{C}} \mathbf{F} \cdot \mathbf{T} ds$$

8.3 Fundamental Theorem of Line Integrals

When we introduced vector fields it was noted that an important class of vector fields was that obtained as the gradient of a function of several variables: $\mathbf{F} = \nabla f$. Such vector fields are called **conservative vector fields**. They are important because they arise in practice and because the following holds

Fundamental Theorem of Line Integrals (FTLI). Let \mathcal{C} be a regular curve parametrised by $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$ and let f be a differentiable function whose gradient vector is continuous on \mathcal{C} , then

$$\int_{\mathcal{C}} \nabla f \cdot d\mathbf{r} = f(\mathbf{r}(b)) - f(\mathbf{r}(a))$$

Note the analogy with the Fundamental Theorem of Calculus (FTC)

$$\int_a^b F'(t) dt = F(b) - F(a)$$

Proving the FTLI is not difficult as it primarily relies on the Chain Rule and the FTC. The manipulations are

$$\begin{aligned} \int_{\mathcal{C}} \nabla f \cdot d\mathbf{r} &= \int_a^b \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt \\ &= \int_a^b \frac{d}{dt} f(\mathbf{r}(t)) dt = f(\mathbf{r}(b)) - f(\mathbf{r}(a)) \end{aligned}$$

The FTLI tells us that if we know our vector field \mathbf{F} is a conservative vector field, and hence given by gradient of some function f , then we can evaluate any line integral of \mathbf{F} over \mathcal{C} simply by evaluating f at the end points of \mathcal{C} . Call these points \mathbf{r}_a and \mathbf{r}_b . The importance is not just that it simplifies our calculations, but the fact that since the integral depends only on the end points, it in fact must be the same for any curve that starts \mathbf{r}_a and ends at \mathbf{r}_b . That is, if $\mathbf{F} = \nabla f$, then

$$\int_{\mathcal{C}_1} \mathbf{F} \cdot d\mathbf{r} = \int_{\mathcal{C}_2} \mathbf{F} \cdot d\mathbf{r},$$

for any two \mathcal{C}_1 and \mathcal{C}_2 that start at \mathbf{r}_a and end at \mathbf{r}_b . The line integral is said to be **path independent**.

Note in particular that if $\mathbf{F} = \nabla f$, then the integral around any closed curve will be zero because $\mathbf{r}_a = \mathbf{r}_b$ for a closed curve. We will write this as

$$\oint_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = 0$$

where the circle on the integral symbol indicates that the curve \mathcal{C} is closed.

With some mild conditions, it can be shown that if all line integrals of a vector field \mathbf{F} are path independent, or equivalently if the line integral around all closed curves is zero, then \mathbf{F} is a conservative vector field and there is a function f such that $\mathbf{F} = \nabla f$.

The converse is generally easier, although perhaps less important. If the line integral of \mathbf{F} around a close path is not zero, then \mathbf{F} is definitely not a conservative vector field and it cannot be expressed as the gradient of a function.

9 Week 9 Lecture 3: Line Integrals II

9.1 Work and potential energy

Work is an important physical concept that you can learn more about in a mechanics module. It is a classic example of a case where one needs to do line integrals of a vector field. If a force $\mathbf{F}(\mathbf{r})$ acts on a point particle and the particle moves from position \mathbf{r}_a to \mathbf{r}_b along a curve \mathcal{C} , then the work W_{ab} done by the force on the particle is

$$W_{ab} = \int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$$

This definition is independent of whether or not the force $\mathbf{F}(\mathbf{r})$ is a conservative vector field.

In many situations (gravitational fields for example), but not all, the force $\mathbf{F}(\mathbf{r})$ is a conservative vector field. It can thus be written as a gradient of a function. One typically defines the function so that $\mathbf{F} = -\nabla V$ where V is a **potential**, and more specifically in this case, V is **potential energy**.

The work done in moving from \mathbf{r}_a to \mathbf{r}_b is given in terms of the potential at the end points

$$W_{ab} = (-V(\mathbf{r}_b)) - (-V(\mathbf{r}_a)) = V(\mathbf{r}_a) - V(\mathbf{r}_b)$$

independently of how the particle moved from \mathbf{r}_a to \mathbf{r}_b .

You should visualise the potential V as the height of a hill, or more general landscape. Assume that $V(\mathbf{r}_a) > V(\mathbf{r}_b)$. This means the particle starts out at some high point and moves to some lower point. The work done on the particle is $W_{ab} = V(\mathbf{r}_a) - V(\mathbf{r}_b) > 0$, independently of the path followed from \mathbf{r}_a to \mathbf{r}_b . W_{ab} is the energy that can be extracted from the particle as it moves downhill from \mathbf{r}_a to \mathbf{r}_b .

Contrarily, if a particle starts at \mathbf{r}_b then one must expend energy to push it uphill to \mathbf{r}_a . We must input energy equal to W_{ab} . All work (or energy) differences are encoded in the potential V and are independent of the path taken by the particle. Informally, the force conserves mechanical energy by converting work to potential energy, and back. Gravitational and Coulomb forces are two examples of conservative forces that are frequently described in terms of potentials.

9.2 Circulation

For many vector fields, line integrals around closed curves have physical significance. In fluid dynamics, for example, such integrals give what is known as the **circulation** of the fluid around the curve. In electricity and magnetism, such integrals appear in the integral statement of Maxwell's equations and correspond to a circulation of electric or magnetic fields.

We will focus on the fluids case. Let \mathbf{v} be a vector field corresponding to the velocity of fluid in some region of space (or could be confined to a plane). Then the circulation Γ of \mathbf{v} over a closed curve \mathcal{C} is

$$\Gamma = \oint_{\mathcal{C}} \mathbf{v} \cdot d\mathbf{r}$$

Intuitively this integral corresponds to the net amount the fluid is circulating around the curve.

Knowing the circulation around a body such as a wing or a spinning ball, one can calculate the **lift force** on the body. In the case of a wing, the lift force is what holds the aeroplane up. In the case of a spinning ball, the lift force gives rise to a deflection, or bending, of its path through the air.

9.3 Relationship between line integrals and double integrals

We end our brief tour of vector analysis with a look into the relationship between line a integral along a closed curve and a double integral over the region enclosed by the curve. This is the simplest case of a set of deep relationships between certain line, surface and volume integrals.

Let $\mathbf{F}(x, y)$ be a vector field in the plane. As previously discussed, we can view this as a three dimensional vector field with zero component in the \mathbf{k} direction. In this way we can then take the curl of this vector field. It will have component only in the \mathbf{k} direction. (You should view this geometrically as a vector field in the plane whose the curl is a vector pointing out of the plane.)

It is a theorem that

$$\oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_{\Omega} (\text{curl } \mathbf{F}) \cdot \mathbf{k} dA$$

where Ω is the region bounded by the close curve \mathcal{C} . The orientation of the curve is what is call *positive*, but simply think of this as "counterclockwise". There are mild conditions, such as the curve \mathcal{C} must be piecewise regular and curl \mathbf{F} must be defined, but we will not worry about these. This relationship between the line integral of a vector field and a double integral over a region is the vector form of Green's Theorem, or a baby version of Stokes' Theorem.

(Optional)

For fun, we state integral formulas for Stokes' Theorem and the Divergence Theorem.

Stokes' Theorem states

$$\oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_S (\text{curl } \mathbf{F}) \cdot \mathbf{n} dS$$

The left-hand side is just a line integral over closed curve \mathcal{C} that you know how to do. The right-hand side is a surface integral over any surface S that is bounded by \mathcal{C} . \mathbf{n} is a unit normal vector to the surface. There are some mild conditions of the curve and surface for this to make sense.

In the special case where \mathbf{F} is a two-dimensional vector field and \mathcal{C} is in the plane, then the surface S can be taken to be the flat surface in the plane, i.e., a region in the plane and \mathbf{n} become \mathbf{k} and we get the special case above.

The Divergence Theorem, also known as Gauss's Theorem states

$$\oiint_S (\mathbf{F} \cdot \mathbf{n}) dS = \iiint_{\Omega} (\text{div } \mathbf{F}) dV$$

The right-hand side is a triple integral over a region Ω . Given \mathbf{F} , this is in principle something you now know how to calculate. The left-hand side is the integral over a closed surface S that bounds the region Ω . \mathbf{n} is the outward normal vector to this surface.