Phys 2920 (Math/Phys) Notes

David Murdock, TTU

December 19, 2011

Contents

Chapter 1

Preliminaries and Vector Algebra

1.1 Plane Polar Coordinates

Points in the (x, y) plane can be described by the coordinates ρ and ϕ. If we connect the point P to the origin with a straight segment, the segment has length ρ and is inclined at an angle ϕ from the $+x$ axis, as shown in Fig. 1.1.

$$x = \rho \cos\phi \qquad y = \rho \sin\phi$$

and we can go backwards via

$$\rho = \sqrt{x^2 + y^2} \qquad \tan\phi = \frac{y}{x}$$

where we have to choose ϕ appropriately to be in the correct quadrant!

This type of coordinate system will be extended to three dimensions in two different ways later on in the course.

1.2 Complex Numbers

Complex numbers use the "imaginary" number $i = \sqrt{-1}$. They have the form

$$z = a + bi$$

where a and b are real numbers.

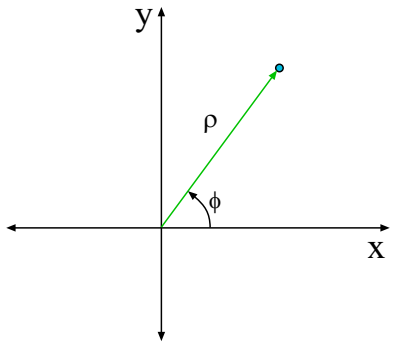

Figure 1.1: Plane polar coordinates.

They occur in many branches of mathematics and theoretical physics. While they make calculations easier in some instances, they are *essential* in the study of quantum mechanics because the central element in QM, the wave function is a *complex* function of space and time.

When we multiply two complex numbers, the real and imaginary parts mix. If

$$z_1 = x_1 + iy_1 \qquad \text{and} \qquad z_2 = x_2 + iy_2$$

then

$$z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1)$$

The magnitude or modulus of z is given by

$$|z| = \sqrt{a^2 + b^2}$$

One can represent the complex numbers as points in a plane, illustrated by an **Argand diagram**. If $z = a + bi$ then a and b give the horizontal and vertical coordinates of the point P (corresponding to z). The polar coordinates of P are given by

$$\rho = |z| = \sqrt{a^2 + b^2} \qquad \tan\phi = \frac{b}{a}$$

The **complex conjugate** of a complex number z, denoted by z^*, is the same as z but with the sign of the imaginary part changed. Thus,

$$z = a + bi \qquad \Longrightarrow \qquad z^* = a - bi$$

One can take the exponential function of a complex number. Here, the fundamental relation is

$$e^{i\theta} = \cos\theta + i\sin\theta$$

It follows that any complex number can be written in the form

$$z = \rho e^{i\phi}$$

where ρ and ϕ are as given above.

1.3 Scalars and Vectors

Quantities which can be given by a single number are called **scalars**. Quantities which must be specified by giving a magnitude and a direction are known as **vectors**.

Examples of scalars in physics are: Temperature, Energy the Electric Potential and Density. Examples of vectors in physics are velocity, force and the electric field.

Vectors can be represented on paper by an arrow. The length of the arrow gives an indication of the magnitude and the direction of the arrow gives... the direction of the vector!

1.4 Vectors: Basics

Vectors (the 2 and 3 dimensional ones!) can be represented on paper by an arrow pointing in the appropriate direction with a length indicative of its magnitude. It does not matter *where on the page* the vector is drawn; only the magnitude and direction matter.

When we give a vector a *name* it is good to distinguish it from a scalar as our problems will involve both kinds of quantities. In print (like here) I will put the name in boldface, thus $\mathbf{a}$ is the name of a vector. On paper and on a blackboard this is hard to do, so there one should put an arrow on top of the symbol, like: $\vec{a}$.

Practices vary for this; nowadays textbooks are *also* putting little arrows on top of the letters to emphasize to students that the things they're talking about have direction! So you might also see "$\vec{a}$" in printed material!

1.5 Basic Operations

Vectors are useful mathematical entities because we can add them together. On paper, the addition of vectors is done by joining them head–to–tail (keeping their directions and magnitudes the same).

Vector addition has the properties

- $\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$ (Commutative)
- $\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$ (Associative)
- $-\mathbf{b}$ is a vector equal in magnitude but opposite in direction to $\mathbf{b}$. Subtracting vectors means

$$\mathbf{a}-\mathbf{b}=\mathbf{a}+(-\mathbf{b})$$

Vectors can be multiplied by scalars. The vector $\lambda\mathbf{a}$ points in the same direction as $\mathbf{a}$ (opposite if λ is negative) but has $|\lambda|$ times the magnitude.

Scalar multiplication follows the rules

- $\lambda(\mu\mathbf{a})=(\lambda\mu)\mathbf{a}=\mu(\lambda\mathbf{a})$
- $\lambda(\mathbf{a}+\mathbf{b})=\lambda\mathbf{a}+\lambda\mathbf{b}$ (Distributive)

1.6 Components, Basis Vectors

In three dimensions, using the unit vectors along the x, y and z axes (called $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ here) we can express any other vector as a linear combination:

$$\mathbf{a}=a_x\,\hat{\mathbf{i}}+a_y\,\hat{\mathbf{j}}+a_z\,\hat{\mathbf{k}}$$

The notation for these unit vectors varies between books; also used are

$$\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}} \qquad \text{and} \qquad \hat{\mathbf{e}}_x,\hat{\mathbf{e}}_y,\hat{\mathbf{e}}_z$$

so be prepared for different notation!

In N dimensions (which we will be considering before long, we will generally use a set of N fundamental of **basis** vectors to construct all the others:

$$\mathbf{a}=a_1\hat{\mathbf{e}}_1+a_2\hat{\mathbf{e}}_2+\cdots+a_N\hat{\mathbf{e}}_N$$

We will want to use a set of basis vectors which **spans** the entire space (that is, can give any vector of interest by taking linear combinations) but which is not redundant, meaning specifically that the $\hat{\mathbf{e}}_i$'s are **linearly independent**. A set of vectors $\hat{\mathbf{e}}_i$ is linearly independent if

$$c_1\hat{\mathbf{e}}_1 + c_2\hat{\mathbf{e}}_2 + \cdots + c_N\hat{\mathbf{e}}_N = \mathbf{0}$$

implies that all the c_i's are zero.

The **magnitude** of the vector $\mathbf{a}$, in terms of its components is

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

1.7 Products of Vectors

The **dot product** (or **scalar product**) of two vectors is given by

$$\mathbf{a} \cdot \mathbf{b} = ab\cos\theta \tag{1.1}$$

The result of this operation (of combining two *vectors*) is a *scalar*. One can show that it is also given in terms of components of the two vectors by

$$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z \tag{1.2}$$

and we note that the magnitude of a vector is also given by

$$a = |\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}}$$

The **cross product** is given by

$$\mathbf{a} \times \mathbf{b} = (a_y b_z - a_z b_y)\,\hat{\mathbf{i}} + (a_z b_x - a_x b_z)\,\hat{\mathbf{j}} + (a_x b_y - a_y b_x)\,\hat{\mathbf{k}} \tag{1.3}$$

If you are already familiar with evaluating the determinant of a 3×3 matrix, this can be written:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} \tag{1.4}$$

And we note the for the cross product, two vectors combine to give another vector.

One can show that if $\mathbf{c} = \mathbf{a} \times \mathbf{b}$ then $\mathbf{c}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$; if the angle between $\mathbf{a}$ and $\mathbf{b}$ is θ, then $c = ab|\sin\theta|$. These two facts are not enough to determine $\mathbf{c}$; there remains an ambiguity in direction which is resolved by stating that the direction of $\mathbf{c}$ is given by the **right-hand rule**.

1.8 Vector Algebra Relations

The "scalar triple product" formed from the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$: $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ is of course, a scalar. One can show:

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

Note, the order is important in these expressions. This product can be written as $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ and equals the determinant given by

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = [\mathbf{a}, \mathbf{b}, \mathbf{c}] = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix}$$

Another useful vector relation is

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

Finally we consider a triple product of vectors, $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$. Since the cross product is not associative, we note

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$$

so lots of care is needed with the order and grouping when cross products are involved! Two useful formulae are:

$$\begin{aligned} \mathbf{a} \times (\mathbf{b} \times \mathbf{c}) &= (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c} \\ (\mathbf{a} \times \mathbf{b}) \times \mathbf{c} &= (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a} \end{aligned}$$

1.9 Some Physics

The dot product shows up in elementary physics in the definition of work. If a particle moves by some (small) displacement $d\mathbf{s}$ while a force $\mathbf{F}$ acts on it, the work done by the force is

$$dW = \mathbf{F} \cdot d\mathbf{s}$$

The cross product shows up in the definition of **torque** (also called **moment of the force** by some people. Torque is measured with respect to some origin; if a force $\mathbf{F}$ acts at point $\mathbf{r}$ then the torque exerted by that force is

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$$

It is also appears in the definition of angular momentum (also defined with respect to a given origin). If a particle of mass m has velocity $\mathbf{v}$ at point $\mathbf{r}$; its angular momentum is

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = \mathbf{r} \times (m\mathbf{v})$$

One sees a cross product in the expression for the magnetic force on a charged particle moving in a magnetic field. If a charge q has velocity $\mathbf{v}$ as it experiences a magnetic field $\mathbf{B}$, the magnetic force on the charge is

$$\mathbf{F}_{\text{mag}} = q\mathbf{v} \times \mathbf{B}$$

1.10 The Symbols δ_{ij} and ϵ_{ijk}

Before moving on to the next chapter I want to introduce some notational tricks which help in doing derivations with vectors. Recall the meaning of "Kronecker delta":

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

For example, if your basis vectors are orthonormal, then

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

and the usual dot product is

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i,j=1}^{3} a_i b_j \delta_{ij}$$

Another useful symbol with the name **Levi-Cevita tensor** is denoted by ϵ_{ijk} has the value 1, -1 or 0 depending on the values of (ijk) (where the indices have the values 1, 2, or 3. We have:

$$\epsilon_{123} = 1 \qquad \epsilon_{132} = -1 \qquad \epsilon_{312} = 1 \qquad \text{etc.}$$

where we get a change of sign if we switch two indices. If any two indices are the same, it is zero, hence

$$\epsilon_{133} = 0 \qquad \epsilon_{121} = 0 \qquad \epsilon_{222} = 0 \qquad \text{etc.}$$

With ϵ_{ijk}, we can write the cross product more compactly. If $\hat{\mathbf{e}}_i$ stands for the x, y and z unit vectors then

$$\mathbf{a} \times \mathbf{b} = \sum_{i,j=1}^{3} a_i b_j \epsilon_{ijk} \hat{\mathbf{e}}_k$$

or:

$$\mathbf{c} = \mathbf{a} \times \mathbf{b} \qquad \Longrightarrow \qquad c_k = \sum_{i,j=1}^{3} a_i b_j \epsilon_{ijk}$$

To save on writing in our derivations of vector relations, there is another convention which is common. If an index appears twice on one side of an equation (in a product) we can agree that we will *sum* on that index. Then with this convention, we would have written

$$\mathbf{c} = \mathbf{a} \times \mathbf{b} \qquad \Longrightarrow \qquad c_k = a_i b_j \epsilon_{ijk}$$

where in the second expression we know that we should sum on i and j.

With this notation, the dot product is simply

$$\mathbf{a} \cdot \mathbf{b} = a_i b_i$$

Several useful theorems help us to use the ϵ's in proving theorems. (In both we are using the new summation convention.)

$$\epsilon_{ijk}\epsilon_{lmk} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \tag{1.5}$$

$$\epsilon_{ijk}\epsilon_{ijl} = 2\delta_{kl} \tag{1.6}$$

$$\epsilon_{ijk}\epsilon_{ijk} = 6 \tag{1.7}$$

As example of the use this new notation we will prove the triple product rule

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

Imagine doing this proof longhand; you'd spend a lot of time writing out all the components, which occur in many combinations. Smart people use the δ's and ϵ's.

Start with the left side and use the ϵ's to write out the cross products:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = a_i (\mathbf{b} \times \mathbf{c})_j \, \epsilon_{ijk} \, \hat{\mathbf{e}}_k \; = \; a_i \, (b_l \, c_m \, \epsilon_{lmj}) \epsilon_{ijk} \, \hat{\mathbf{e}}_k$$

Switching indices on one of the ϵ's gives

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \; = \; a_i \, b_l \, c_m \, \epsilon_{lmj} \, \epsilon_{ijk} \, \hat{\mathbf{e}}_k \; = \; -a_i b_l c_m \epsilon_{lmj} \, \epsilon_{ikj} \, \hat{\mathbf{e}}_k$$

Using Eq. 1.5 gives

$$= -a_i b_l c_m (\delta_{il} \delta_{km} - \delta_{kl} \delta_{im}) \, \hat{\mathbf{e}}_k = [-a_i \, b_l \, \delta_{il} \, \delta_{km} \, c_m + a_i \, b_l \, c_m \, \delta_{kl} \, \delta_{im}] \, \hat{\mathbf{e}}_k$$

Use $a_i b_l \delta_{il} = \mathbf{a} \cdot \mathbf{b}$, etc. and get

$$\begin{aligned} &= \quad (-(\mathbf{a} \cdot \mathbf{b}) c_k + (\mathbf{a} \cdot \mathbf{c}) b_k) \hat{\mathbf{e}}_k \; = \; (\mathbf{a} \cdot \mathbf{c}) b_k \hat{\mathbf{e}}_k - (\mathbf{a} \cdot \mathbf{b}) c_k \hat{\mathbf{e}}_k \\ &= \quad (\mathbf{a} \cdot \mathbf{c}) \mathbf{b} - (\mathbf{a} \cdot \mathbf{b}) \mathbf{c} \end{aligned}$$

Chapter 2

Vector Spaces and Matrices

2.1 Vectors Spaces

A set of objects $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, ... forms a **vector space** if

- The set is closed under addition; addition is commutative and associative
- The set is closed under multiplication by a scalar, which is also distributive and associative:
  1. $\lambda(\mathbf{a}+\mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$
  2. $(\lambda+\mu)\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{a}$
  3. $\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a}$
- There is a null vector $\mathbf{0}$ such that $\mathbf{a}+\mathbf{0} = \mathbf{a}$ for all $\mathbf{a}$.
- All vectors have a negative $-\mathbf{a}$ such that $\mathbf{a}+(-\mathbf{a}) = \mathbf{0}$. And $-\mathbf{a} = (-1)\mathbf{a}$.

The scalars which multiply the vectors can be restricted to real numbers (giving a real vector space) but in general they are complex (giving a complex vector space).

2.2 Basis Vectors

2.3 Inner Product

It is useful to add the concept of an **inner product** to a vector space. The inner product of $\mathbf{a}$ and $\mathbf{b}$, denoted $\langle\mathbf{a}|\mathbf{b}\rangle$ gives a *scalar*, that is, a (complex) number.

The inner product has the properties

- $\langle\mathbf{a}|\mathbf{b}\rangle = \langle\mathbf{b}|\mathbf{a}\rangle^*$, it isn't quite commutative.
- It is distributive, that is,
$$\langle\mathbf{a}|\lambda\mathbf{b}+\mu\mathbf{c}\rangle = \lambda\langle\mathbf{a}|\mathbf{b}\rangle + \mu\langle\mathbf{a}|\mathbf{c}\rangle$$
From the first property it follows that
$$\langle\lambda\mathbf{a}+\mu\mathbf{b}|\mathbf{c}\rangle = \lambda^*\langle\mathbf{a}|\mathbf{c}\rangle + \mu^*\langle\mathbf{b}|\mathbf{c}\rangle$$

This is a generalization of the familiar dot product of three-dimensional vectors.

Two vectors **a** and **b** are **orthogonal** if $\langle \mathbf{a}|\mathbf{b}\rangle = 0$.

The **norm** of a vector is given by

$$\|\mathbf{a}\| = \langle \mathbf{a}|\mathbf{a}\rangle^{1/2}$$

so that we hereby restrict attention to vector spaces for which $\langle \mathbf{a}|\mathbf{a}\rangle \geq 0$. If $\langle \mathbf{a}|\mathbf{a}\rangle = 0$ the $\mathbf{a} = \mathbf{0}$.

Now suppose the $N-$dimensional vector space has a basis $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$,$\ldots \hat{\mathbf{e}}_N$ which has the (desirable) property that $\langle \hat{\mathbf{e}}_i|\hat{\mathbf{e}}_j\rangle = \delta_{ij}$. Then we say the basis is **orthonormal**. If the vectors **a** and **b** expressed in this basis are

$$\mathbf{a} = \sum_{i=1}^{N} a_i \hat{\mathbf{e}}_i \qquad \mathbf{b} = \sum_{i=1}^{N} b_i \hat{\mathbf{e}}_i$$

and *if* the basis is orthonormal, then we get

$$\langle \hat{\mathbf{e}}_i|\mathbf{a}\rangle = a_i \qquad \text{and} \qquad \langle \mathbf{a}|\mathbf{b}\rangle = \sum_{i=1}^{N} a_i^* b_i$$

Two inequalities follow from the definition of the inner product. The **Schwarz Inequality** is

$$|\langle \mathbf{a}|\mathbf{b}\rangle| \leq \|\mathbf{a}\|\ \|\mathbf{b}\| \tag{2.1}$$

The **triangle inequality** is

$$\|\mathbf{a}+\mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\| \tag{2.2}$$

That's all for math of the vectors within a vector space. Now we introduce the idea of linear operators, leading to the objects known as matrices

2.4 Linear Operators

A linear operator changes one vector into another, but in a specific way! For the operator $\mathcal{A}$, for every vector **x** it associates a vector **y** (in the same vector space), written:

$$\mathcal{A}\mathbf{x} = \mathbf{y}$$

A **linear operator** has the property

$$\mathcal{A}(\lambda\mathbf{a} + \mu\mathbf{b}) = \lambda\mathcal{A}\mathbf{a} + \mu\mathcal{A}\mathbf{b}$$

Now we introduce a *specific* basis $\hat{\mathbf{e}}_i$. If $\mathcal{A}$ acts on $\hat{\mathbf{e}}_j$ we have to get some linear combination of all the $\hat{\mathbf{e}}_i$'s. We write:

$$\mathcal{A}\hat{\mathbf{e}}_j = \sum_{i=1}^{N} \mathsf{A}_{ij}\hat{\mathbf{e}}_i \tag{2.3}$$

Eq. 2.3 defines the **matrix elements** of the matrix A for the linear transformation $\mathcal{A}$, for the basis $\hat{\mathbf{e}}_i$. We use this to show how the components of any vector transfrom under $\mathcal{A}$: If $\mathbf{y} = \mathcal{A}\mathbf{x}$, then

$$\mathbf{y} = \sum_{i=1}^{N} y_i\hat{\mathbf{e}}_i = \mathcal{A}\left(\sum_{j=1}^{N} x_j\hat{\mathbf{e}}_j\right) = \sum_{j=1}^{N} x_j \sum_{i=1}^{N} \mathsf{A}_{ij}\hat{\mathbf{e}}_i \tag{2.4}$$

Matching the coefficients of $\hat{\mathbf{e}}_i$ gives

$$y_i = \sum_{j=1}^{N} \mathsf{A}_{ij} x_i \tag{2.5}$$

Eq. 2.5 shows how the components of vectors transform for the linear transformation $\mathcal{A}$.

One can also talk about linear transformations between vector spaces with different dimensionality, but we won't; this is an interesting topic, but you will cover it in any *real* course in linear algebra.

Properties of linear operators:

- $(\mathcal{A}+\mathcal{B})\mathbf{x} = \mathcal{A}\mathbf{x} + \mathcal{B}\mathbf{x}$
- $(\lambda\mathcal{A})\mathbf{x} = \lambda(\mathcal{A}\mathbf{x})$
- $(\mathcal{A}\mathcal{B})\mathbf{x} = \mathcal{A}(\mathcal{B}\mathbf{x})$

Note that while the *addition* of linear operators is commutative, the *multiplication* (that is, successive application) of operators is *not*: $(\mathcal{A}\mathcal{B}\mathbf{x} \neq \mathcal{B}\mathcal{A}\mathbf{x}$. We also have the null and identity operators, which for all $\mathbf{x}$ give

$$\mathcal{O}\mathbf{x} = \mathbf{0} \qquad \mathcal{I}\mathbf{x} = \mathbf{x}$$

2.5 Matrices

In a *particular basis* $\hat{\mathbf{e}}_i$, both vectors and linear operators are described in terms of their components. As we've seen, the components transform as:

$$\mathbf{y} = \mathcal{A}\mathbf{x} \qquad \Longrightarrow \qquad y_i = \sum_{j=1}^{N} \mathsf{A}_{ij} x_j$$

and we write the set of numbers A_{ij} as

$$\mathsf{A} = \begin{pmatrix} \mathsf{A}_{11} & \mathsf{A}_{12} & \cdots & \mathsf{A}_{1N} \\ \mathsf{A}_{21} & \mathsf{A}_{22} & \cdots & \mathsf{A}_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \mathsf{A}_{N1} & \mathsf{A}_{N2} & \cdots & \mathsf{A}_{NN} \end{pmatrix}$$

We will assume that a matrix is "square" (same number of rows as columns) though if the linear transformation is between spaces of different dimensionality, it won't be.

Likewise we write out the components of the vector $\mathbf{x}$ as a column:

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

One can show:

- $(\mathsf{A}+\mathsf{B})_{ij} = \mathsf{A}_{ij} + \mathsf{B}_{ij}$
- $(\lambda\mathsf{A})_{ij} = \lambda\mathsf{A}_{ij}$

- $(\mathsf{AB})_{ij} = \sum_k \mathsf{A}_{ik}\mathsf{B}_{kj}$

The last of these gives the rule for **matrix multiplication**.

There is always a matrix such that $\mathsf{A} + \mathsf{0} = \mathsf{A}$. It's simply a matrix full of zeroes!

There is also always a matrix $\mathbf{1}$ for which $\mathbf{1}\mathsf{A} = \mathsf{A}\mathbf{1} = \mathsf{A}$. It is a matrix with 1's along the diagonal but zeroes everywhere else. Thus we have the **unit matrix**,

$$\mathbf{1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

2.6 Basic Operations on Matrices

In working with matrices, we will have reason to carry out the following operations on their elements:

Transpose: Interchange rows and columns. Denote the transpose of A by A^T. Then:

$$\mathsf{A}^\mathsf{T}_{ij} = \mathsf{A}_{ji}$$

The transpose of a product of matrices is *related to* a product of the transposes of the individual matrices; you have to reverse the order!

$$(\mathsf{AB})^\mathsf{T} = \mathsf{B}^\mathsf{T}\mathsf{A}^\mathsf{T}$$

Complex Conjugate: Take the complex conjugate of all elements; denote by A^*. Thus:

$$(\mathsf{A}^*)_{ij} = (\mathsf{A}_{ij})^*$$

The complex conjugate of a product doesn't change the order:

$$(\mathsf{AB})^* = \mathsf{A}^*\mathsf{B}^*$$

Hermitian Conjugate: Take the complex conjugate *and* the transpose! Denote by $\mathsf{A}^\dagger$. Thus:

$$\mathsf{A}^\dagger = (\mathsf{A}^*)^\mathsf{T} = (\mathsf{A}^\mathsf{T})^* \qquad (\mathsf{A}^\dagger)_{ij} = \mathsf{A}^*_{ji}$$

The Hermitian conjugate of a product also gives a reversed order:

$$(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$$

Note that when we take a vector inner product (using components of the vectors) we are *really* doing a sort of matrix multiplication; for real vectors it is:

$$\langle \mathbf{a}|\mathbf{b}\rangle = \begin{pmatrix} a_1 & a_2 & \cdots & a_N \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_N \end{pmatrix} = \mathsf{a}^\mathsf{T}\mathsf{b}$$

while for a complex vector space it is

$$\langle \mathbf{a}|\mathbf{b}\rangle = \begin{pmatrix} a_1^* & a_2^* & \cdots & a_N^* \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_N \end{pmatrix} = \mathsf{a}^\dagger\mathsf{b}$$

(Of course the second case includes the first one; for the real vector space, the Hermitian conjugate *is* the transpose.)

2.7 Inverse of a Matrix

A matrix A can have an **inverse**, that is, there may be some matrix A^{-1} for which

$$\mathsf{A}^{-1}\,\mathsf{A} = \mathsf{A}\,\mathsf{A}^{-1} = \mathbf{1}$$

(More on which matrices *have* an inverse shortly!) If we have the matrix equation $\mathsf{AB} = \mathsf{C}$ and the matrix B is unknown but matrix A has an inverse, we can find B by multiplying both sides of the equation *on the left* by A^{-1}:

$$\mathsf{AB} = \mathsf{C} \qquad \Longrightarrow \qquad \mathsf{B} = \mathsf{A}^{-1}\mathsf{C}$$

But we note that since (in general) *matrices do not commute* it would be *wrong* to write $\mathsf{B} = \mathsf{CA}^{-1}$ here. One must be careful about the order of matrix multiplication.

2.8 Trace of a Matrix

The trace of a (square) matrix is the sum of the diagonal elements:

$$\mathrm{Tr}(\mathsf{A}) = \mathsf{A}_{11} + \mathsf{A}_{22} + \cdots + \mathsf{A}_{NN} = \sum_{i=1}^{N} \mathsf{A}_{ii}$$

Thus the Trace operation takes a *matrix* and returns a *number*.

Some properties are:

- $\mathrm{Tr}(\mathsf{A} \pm \mathsf{B}) = \mathrm{Tr}(\mathsf{A}) \pm \mathrm{Tr}(\mathsf{B})$
- $\mathrm{Tr}(\mathsf{AB}) = \mathrm{Tr}(\mathsf{BA})$

More generally, the trace does not change if we do a *cyclic rotation* of the product of matrices. For three matrices, this gives:

$$\mathrm{Tr}(\mathsf{ABC}) = \mathrm{Tr}(\mathsf{CAB}) = \mathrm{Tr}(\mathsf{BCA})$$

And remember, the product of the matrices itself may change if we switch the order in this way, but its trace does not.

2.9 Determinant of a Matrix

The **determinant** is another operation which takes a matrix and returns a *number*. It is denoted $\det\mathsf{A}$ and is indicated with vertical bars on either side of the matrix entries:

$$\det \mathsf{A} = |\mathsf{A}| = \begin{vmatrix} \mathsf{A}_{11} & \mathsf{A}_{12} & \cdots & \mathsf{A}_{1N} \\ \mathsf{A}_{21} & \mathsf{A}_{22} & \cdots & \mathsf{A}_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \mathsf{A}_{N1} & \mathsf{A}_{N2} & \cdots & \mathsf{A}_{NN} \end{vmatrix}$$

Recall, for the matrix itself, we enclose the entries with *parentheses*.

It is a rather simple operation for a 2×2 or a 3×3 matrix but the simple pattern for these cases does *not* hold for bigger sizes!

Properties of the determinant:

1. The determinant of a transpose is the same that of the original matrix:
$$|\mathsf{A}^{\mathsf{T}}| = |\mathsf{A}|$$
2. The determinant of a complex conjugate and Hermitian conjugate satisfy
$$|\mathsf{A}^{\dagger}| = |(\mathsf{A}^{*})^{\mathsf{T}}| = |\mathsf{A}^{*}| = |\mathsf{A}|^{*}$$
3. If any two rows or columns of the matrix are interchanged, the determinant changes sign.
4. $|\lambda\mathsf{A}| = \lambda^{N}|\mathsf{A}|$, where N is the dimension of matrix A.
5. If any two rows or columns of A are the same, $|\mathsf{A}| = 0$.
6. If you add a constant multiple of one row or column to another row or column, the value of the determinant is unchanged.
7. For square matrices,
$$|\mathsf{AB}| = |\mathsf{A}|\,|\mathsf{B}| = |\mathsf{BA}|$$
so the value of the determinant is unchanged by *any* permutation of the order of the matrices in a product.

2.10 More On the Inverse of a Matrix

It follows from $\det(\mathbf{1}) = 1$ that the determinants of A and A^{-1} are inverses of each other in the normal numerical sense. This must mean that if the determinant of A is *zero*, then A cannot have an inverse. In that case we say that A is **singular**. But for a non-singular matrix A we can always find an inverse.

Computing the inverse of a matrix can be done by any computer and many calculators but it could be done by hand. It can be shown that
$$(\mathsf{A}^{-1})_{ik} = \frac{(\mathsf{C}^{\mathsf{T}}_{ik})}{|\mathsf{A}|} = \frac{\mathsf{C}_{ki}}{|\mathsf{A}|}$$
where C_{ik} is the cofactor of the ik element.

And there other ways to calculate it. (Explanation to be given in class on an algorithm which will produce the inverse from a given matrix.) It turns out that it requires a bit of care to program a computer to compute it *accurately*.

Example: Use the algorithm given in class to find the inverse of the matrix
$$\mathsf{A} = \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix}$$

Start by writing the given matrix and the unit matrix side by side:
$$\left(\begin{array}{ccc||ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 2 & -1 & 3 & 0 & 1 & 0 \\ 4 & 1 & 8 & 0 & 0 & 1 \end{array}\right)$$

Subtract 2 times the first row from the second row:

$$\left(\begin{array}{ccc||ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{array}\right)$$

Add the second row to the third row:

$$\left(\begin{array}{ccc||ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 0 & -1 & -6 & 1 & 1 \end{array}\right)$$

Take 2 times the third row and add to the first row; subract the third row from the second row. This gives:

$$\left(\begin{array}{ccc||ccc} 1 & 0 & 0 & -11 & 2 & 2 \\ 0 & -1 & 0 & 4 & 0 & -1 \\ 0 & 0 & -1 & -6 & 1 & 1 \end{array}\right)$$

Finally, multiply the second and third rows by -1:

$$\left(\begin{array}{ccc||ccc} 1 & 0 & 0 & -11 & 2 & 2 \\ 0 & 1 & 0 & -4 & 0 & 1 \\ 0 & 0 & 1 & -6 & -1 & -1 \end{array}\right)$$

And so we conclude that the inverse of the original matrix is

$$\mathsf{A}^{-1} = \begin{pmatrix} -11 & 2 & 2 \\ -4 & 0 & 1 \\ 6 & -1 & -1 \end{pmatrix}$$

We can check the result by multiplying the original matrix and this result and verifying that we do get the 3×3 unit matrix.

Properties of the inverse:

1. $(\mathsf{A}^{-1})^{-1} = \mathsf{A}$
2. $(\mathsf{A}^{\mathsf{T}})^{-1} = (\mathsf{A}^{-1})^{\mathsf{T}}$
3. $(\mathsf{A}^{\dagger})^{-1} = (\mathsf{A}^{-1})^{\dagger}$
4. $(\mathsf{AB})^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1}$
5. $(\mathsf{AB}\cdots\mathsf{G})^{-1} = \mathsf{G}^{-1}\cdots\mathsf{B}^{-1}\mathsf{A}^{-1}$

and as already mentioned,

$$|\mathsf{A}^{-1}| = \frac{1}{|\mathsf{A}|}$$

2.11 Special Types of Matrices

A **diagonal matrix** only has non-zero entries along the diagonal; that is, $\mathsf{A}_{ij} = 0$ if $i \neq j$.

A **symmetric matrix** is equal to its transpose: $\mathsf{A} = \mathsf{A}^{\mathsf{T}}$. So $\mathsf{A}_{ij} = \mathsf{A}_{ji}$ for all i, j.

An **antisymmetric matrix** equals the negative of its transpose: $\mathsf{A} = -\mathsf{A}^{\mathsf{T}}$. So $\mathsf{A}_{ij} = -\mathsf{A}_{ji}$ for all i, j. An antisymmetric matrix must have zeroes along the diagonal!

If $\mathsf{A}^{\mathsf{T}} = \mathsf{A}^{-1}$ then the matrix A is **orthogonal**. (And the inverse of an orthogonal matrix is also orthogonal!)

The determinant of on orthogonal matrix must be ± 1.

A special property of the transformation effected b an orthogonal matrix is that the norm of the vector is unchanged:

If $\mathsf{A}^\dagger = \mathsf{A}$ then the matrix A is **Hermitian**. (If $\mathsf{A}^\dagger = -\mathsf{A}$ then A is **anti-Hermitian**.)

If $\mathsf{A}^\dagger = \mathsf{A}^{-1}$, then the matrix A is **unitary**. (If A is real, then it is also orthogonal.)

If A is unitary, then $|\mathsf{A}|$ is in general a complex number with *magnitude* 1. A unitary matrix has the property that it leaves the magnitude of a complex vector unchanged.

If $\mathsf{A}\mathsf{A}^\dagger = \mathsf{A}^\dagger\mathsf{A}$ then A is **normal**.

2.12 Eigenvectors and Eigenvalues

Suppose a linear operator $\mathcal{A}$ transforms vectors $\mathbf{x}$ in an N–dimensinal space into other vectors in the same space (our normal situation). It *could* happen that $\mathbf{x}$ is transformed into a vector parallel to itself; that is it could hbe the case that

$$\mathcal{A}\mathbf{x} = \lambda\mathbf{x}$$

for some vectors $\mathbf{x}$ and the accompanying numbers λ. Then $\mathbf{x}$ is called an **eigenvector** of $\mathcal{A}$ and λ is an **eigenvalue** of $\mathcal{A}$. To the uninitiated, these might seems like odd things to look for, but they of fundamental importance in quantum mechanics. They also arise in a few areas of classical mechanics.

To compute these vectors we choose a particular basis for which the operator $\mathcal{A}$ has the representation A. But we can give a few facts about them before doing any computations:

• If A is nonsingular and A has eigenvectors $\mathbf{x}^i$ and corresponding eigenvalues λ_i, then A^{-1} has the same eigenvectors with eigenvalues $1/\lambda_i$.

• If A is Hermitian, its eigenvalues are real, and eigenvectors corresponding to *different* eigenvalues are orthogonal.

2.13 Determination of Eigenvectors and Eigenvalues

$$\mathsf{A}\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad \mathsf{A}\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \quad \Longrightarrow \quad (\mathsf{A} - \lambda\mathbf{1})\mathbf{x} = \mathbf{0}$$

Now if the last of these is written out longhand, we get

$$\begin{aligned} (\mathsf{A}_{11} - \lambda)x_1 &+ \mathsf{A}_{12}x_2 + \cdots + \mathsf{A}_{1N}x_N = 0 \\ \mathsf{A}_{12}x_1 &+ (\mathsf{A}_{22} - \lambda)x_2 + \cdots + \mathsf{A}_{1N}x_N = 0 \\ &\cdots \\ \mathsf{A}_{N1}x_{11} &+ \mathsf{A}_{N2}x_2 + \cdots + (\mathsf{A}_{NN} - \lambda)x_N = 0 \end{aligned} \tag{2.6}$$

What do we have here? This is a set of N **homogeneous** linear equations, that is, the constant term is zero in each of them. If for some λ there is a solution, then we will have an eigenvalue and an eigenvector.

Now we get a bit ahead of ourselves since we haven't yet discussed systems of linear equations, but turns out that such a system will have a solution if the determinant of the coefficients is zero:

$$\begin{vmatrix} \mathsf{A}_{11}-\lambda & \mathsf{A}_{12} & \cdots & \mathsf{A}_{1N} \\ \mathsf{A}_{21} & \mathsf{A}_{22}-\lambda & \cdots & \mathsf{A}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{A}_{N1} & \mathsf{A}_{N2} & \cdots & \mathsf{A}_{NN}-\lambda \end{vmatrix} = |\mathsf{A}-\lambda\mathbf{1}| = 0$$

For a 2×2 matrix this will give a quadratic equation for λ, which of course is easy to solve (and we get at most two possible values for λ), but a 3×3 matrix gives a cubic equation for λ and large matrices are even messier. Fortunately the computer packages are quite good at finding eigenvalues.

Having the eigenvalues λ_i, we can go on to solve the equations in 2.6 to get the eigenvector for each one.

One can show that the sum of the eigenvalues gives the trace:

$$\sum_{i=1}^{N} \lambda_i = \mathrm{Tr}(\mathsf{A})$$

where λ_i are the eigenvalues of A. This can give a check on on a computation of the eigenvalues.

One can also show that if matrices A and B commute (that is, $\mathsf{AB} = \mathsf{BA}$) then their eigenvalues are the same.

It may happen that a matrix has two or more eigenvalues which are the same, that is, the algebraic equation for λ has multiple roots. Then the corresponding eigenvectors are said to be **degenerate** and we can find linearly independent eigenvectors with the same eigenvalue.

2.14 Similarity Transformations

Recall that a vector space is a set of abstract objects, as are the linear operators which act upon them. We can express them as a set of numbers *only* when we make a particular choice of a basis.

If we have some new basis vectors $\mathbf{e}'_j$ related to some old basis vectors $\mathbf{e}_i$ by

$$\mathbf{e}'_j = \sum_{i=1}^{N} \mathsf{S}_{ij}\mathbf{e}_i$$

then one can show that the new components of the vector $\mathbf{x}$ for the two choices of basis are related by

$$x_i = \sum_{j=1}^{N} \mathsf{S}_{ij}x'_j \qquad \text{or} \qquad \mathbf{x} = \mathsf{S}\mathbf{x}' \qquad \text{or} \qquad \mathbf{x}' = \mathsf{S}^{-1}\mathbf{x}$$

S is the transformation matrix associated with the change in basis. Since the $\mathbf{e}'_j$ are linearly independent, S is non-singular and so it *does* have an inverse S^{-1}.

We also need to transform the the matrix expression of linear operators. One can show

$$\mathsf{A}' = \mathsf{S}^{-1}\mathsf{A}\mathsf{S}$$

The change of basis is an important operation in quantum mechanics, where can use coordinates or momenta to label the components of the quantum mechanical "vectors", i.e. the quantum states.

S is called a **similarity transform**. Note that after a transformation with S the vectors and operators still have the same *relations* to one another.

Some facts about a transformation with S:

- If $\mathsf{A} = \mathbf{1}$ then $\mathsf{A}' = \mathbf{1}$.
- $|\mathsf{A}'| = |\mathsf{A}|$; the value of the determinant is unchanged.
- The eigenvalues of A' are the same as those of A.
- $\mathrm{Tr}\,(\mathsf{A}') = \mathrm{Tr}\,(\mathsf{A})$

2.14.1 Unitary Transforms

An important class of similarity transforms is one for which S is unitary: $\mathsf{S}^\dagger = \mathsf{S}^{-1}$. Then

$$\mathsf{A}' = \mathsf{S}^{-1}\mathsf{A}\mathsf{S} = \mathsf{S}^\dagger\mathsf{A}\mathsf{S}$$

These are important because if the original basis $\hat{\mathbf{e}}_i$ is orthonormal, then the transformed basis $\hat{\mathbf{e}}'_j$ is also orthonormal:

$$\langle\hat{\mathbf{e}}_i|\hat{\mathbf{e}}_j\rangle = \delta_{ij} \qquad \Longrightarrow \qquad \langle\hat{\mathbf{e}}'_i|\hat{\mathbf{e}}'_j\rangle = \delta_{ij}$$

For unitary transforms, we have:

- If A is Hermitian, then A' is Hermitian
- If A is unitary, then A' is unitary.

2.15 Diagonalization of Matrices

A diagonal matrix is clearly a very simple kind of matrix! If we begin with a more complicated matrix A, it is useful procedure to find a similarity transformation which can transform it into a *diagonal* matrix A'. And the eigenvectors of the matrix A are used for this!

Suppose we begin with a matrix A and choose the *eigenvectors* of A as the new basis. Thus for

$$\mathsf{A}\mathbf{x}^j = \lambda_j\mathbf{x}^j$$

we want to find an S such that

$$\mathbf{x}^j = \sum_{i=1}^{N} \mathsf{S}_{ij}\mathbf{e}_i$$

But this equation tells us that the columns of S are the *components of the eigenvectors themselves*. With S constructed this way, one can show

$$\mathsf{A}'_{ij} = (\mathsf{S}^{-1}\mathsf{A}\mathsf{S})_{ij} = \lambda_j\delta_{ij}$$

Then A' is diagonal and its elements are the eigenvalues of the original matrix (and of course the matrix A' itself). Thus with this S,

$$\mathsf{A}' = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_N \end{pmatrix}$$

Can we always do this? If A has N distinct eigenvalues then it *can* be diagonalized by this procedure. But in general, if a square matrix has degenerate eigenvalues, it may not have N independent eigenvectors; if it doesn't then it can't be diagonalized.

For matrices which are Hermitian or unitary, the N eigenvectors *are* linearly independent and so one can choose an orthonormal set. If the columns of S are made of orthonormal vectors, one can show:

$$(\mathsf{S}^\dagger\mathsf{S})_{ij} = \delta_{ij} \qquad \Longrightarrow \qquad \mathsf{S}^\dagger\mathsf{S} = \mathbf{1}$$

and then

$$\mathsf{A}' = \mathsf{S}^{-1}\mathsf{A}\mathsf{S} = \mathsf{S}^\dagger\mathsf{A}\mathsf{S}$$

2.16 Simultaneous Linear Equations

This is probably the topic where you first heard about matrices and determinants. We've left it for last!

In physical applications we often have a set of M linear equations for N unknowns $(x_1, \ldots x_N)$. The set would look like:

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1N}x_N &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2N}x_N &= b_2 \\ \cdots\cdots + \cdots\cdots + \ddots + \cdots\cdots &= \cdots \\ A_{M1}x_1 + A_{M2}x_2 + \cdots + A_{MN}x_N &= b_M \end{aligned}$$

There can be a unique solution, no solution or an infinite number of solutions! In any case, the set of equations is conveniently expressed in matrix/vector form as

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{12} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & \cdots & \cdots & A_{MN} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix} \qquad \text{or} \qquad \mathsf{A}\mathbf{x} = \mathbf{b} \tag{2.7}$$

Generally speaking, if $M < N$ there is an infinity of solutions. If $M > N$, then the system is over-determined and there is *no* solution. We will only consider the simple but common case $M = N$.

With $M = N$ the matrix A in 2.7 is square and we can find its determinant. If $|\mathsf{A}| \neq 0$ then A had an inverse and we can get $\mathbf{x}$ from

$$\mathbf{x} = \mathsf{A}^{-1}\mathbf{b}$$

but if $|\mathsf{A}| = 0$ then there are infinitely many solutions for $\mathbf{x}$.

This is probably the preferred method of solution if you have a computer handy. If not, then due to the computation demands of finding A^{-1} we would need an easier method.

The method that may be familiar to you from high school or someplace is called **Cramer's Rule**; it involves taking the ratios of determinants of matrices made from the coefficients in the problem, and thse ratios give the solution for the unknowns. For clarity, we'll show why it works with a set of 3 equations for 3 unknowns; the pattern will then be obvious. Start with

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + A_{13}x_3 &= b_1 \\ A_{21}x_1 + A_{22}x_2 + A_{23}x_3 &= b_2 \\ A_{31}x_1 + A_{32}x_2 + A_{33}x_3 &= b_3 \end{aligned}$$

If we form the matrix A from the coefficients we find:

$$|\mathsf{A}| = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = \frac{1}{x_1} \begin{vmatrix} b_1 & A_{12} & A_{13} \\ b_2 & A_{22} & A_{23} \\ b_3 & A_{32} & A_{33} \end{vmatrix}$$

How did we get the last equality? First, we multiplied the first colmun by x_1 (whatever it turns out to be!) and then divided by the same, with no net change. Then we took x_2 times the second column and add it to the first column. Then we took x_3 times the last column and added it to the first column. Both of these have no net effect on the determinant, from the properties of the determinant. But now by the original set of equations, the first column is equal to b_1, b_2, b_3. And thus we get the last expression. If we make the definitions

$$\Delta_1 = \begin{vmatrix} b_1 & A_{12} & A_{13} \\ b_2 & A_{22} & A_{23} \\ b_3 & A_{32} & A_{33} \end{vmatrix} \qquad \Delta_2 = \begin{vmatrix} A_{11} & b_1 & A_{13} \\ A_{21} & b_2 & A_{23} \\ A_{31} & b_3 & A_{33} \end{vmatrix} \qquad \Delta_3 = \begin{vmatrix} A_{11} & A_{12} & b_1 \\ A_{21} & A_{22} & b_2 \\ A_{31} & A_{32} & b_3 \end{vmatrix},$$

then we have

$$|\mathsf{A}| = \frac{1}{x_1}\Delta_1 \qquad |\mathsf{A}| = \frac{1}{x_2}\Delta_2 \qquad |\mathsf{A}| = \frac{1}{x_3}\Delta_3$$

which as if by magic can be rearranged to give

$$x_1 = \frac{\Delta_1}{|\mathsf{A}|} \qquad x_2 = \frac{\Delta_2}{|\mathsf{A}|} \qquad x_3 = \frac{\Delta_3}{|\mathsf{A}|} \tag{2.8}$$

which gives a straightforward way to get the solution as long as we can take all the determinants.

The extension to more equations and more variables (as long as there are the same number of each) is clear. Eqs. 2.8 can not be evaluated if $|\mathsf{A}| = 0$ and indeed in that case there is no unique solution.

Chapter 3

Vector Analysis

We now consider only the mundane vectors we used in basic physics: the three-dimensional ones made out of $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ and which are represented by arrows! But we will now use them in function relations and analyze those functions with the ideas of calculus.

3.1 Vector Functions

Vectors may themselves be *functions* of numbers or indeed functions of other *vectors*, such as the coordinate $\mathbf{r}$ in a given coordinate system.

In the first case, we evaluate (say) the vector $\mathbf{a}$ at each value of t, written $\mathbf{a}(t)$. The most familiar example of this is the evaluation of the location $\mathbf{r}$ at each time t; the function $\mathbf{r}(t)$ is the **trajectory** of the moving particle.

Or it could be the case that at each point in space (specified by $\mathbf{r}$ or (x, y, z)) we can give the value a scalar ϕ or a vector $\mathbf{a}$. These are respectively scalar or vector **fields**, written as

$$\phi(\mathbf{r}) \qquad \text{and} \qquad \mathbf{a}(\mathbf{r})$$

With this concept, we have a dependence of ϕ on three independent variables: $\phi = \phi(x, y, z)$; for the vector field, we have *three* components depending on *three* coordinates:

$$\mathbf{a}(\mathbf{r}) = a_x(x, y, z)\,\hat{\mathbf{i}} + a_y(x, y, z)\,\hat{\mathbf{j}} + a_z(x, y, z)\,\hat{\mathbf{k}}$$

Whoa! Lots of opportunities for fun here!

In this framework, we are still interested in the calculus ideas of "rates of change" and of integrals (infinite summations), but they need to be made more elaborate.

To start with, we will consider the case that a vector is a function of a single scalar parameter which we might think of as the time t (but which we'll call u). For this case we have

$$\mathbf{a}(u) = a_x(u)\,\hat{\mathbf{i}} + a_y(u)\,\hat{\mathbf{j}} + a_z(u)\,\hat{\mathbf{k}}$$

3.2 Differentiation of Vectors

The derivative of the vector $\mathbf{a}(u)$ with respect to the parameter u is

$$\frac{d\mathbf{a}}{du} = \lim_{\Delta u \to 0} \frac{\mathbf{a}(u + \Delta u) - \mathbf{a}(u)}{\Delta u} = \frac{da_x}{du}\,\hat{\mathbf{i}} + \frac{da_y}{du}\,\hat{\mathbf{j}} + \frac{da_z}{du}\,\hat{\mathbf{k}} \tag{3.1}$$

Of course, this operation is familiar from kinematics, where we all learned:

$$\mathbf{v} \;=\; \frac{d\mathbf{r}}{dt} \;=\; \frac{dx}{dt}\,\hat{\mathbf{i}} + \frac{dy}{dt}\,\hat{\mathbf{j}} + \frac{dz}{dt}\,\hat{\mathbf{k}} \;=\; v_x\,\hat{\mathbf{i}} + v_y\,\hat{\mathbf{j}} + v_z\,\hat{\mathbf{k}}$$

and

$$\mathbf{a} \;=\; \frac{d\mathbf{v}}{dt} \;=\; \frac{dv_x}{dt}\,\hat{\mathbf{i}} + \frac{dv_y}{dt}\,\hat{\mathbf{j}} + \frac{dv_z}{dt}\,\hat{\mathbf{k}} = a_x\,\hat{\mathbf{i}} + a_y\,\hat{\mathbf{j}} + a_z\,\hat{\mathbf{k}}$$

Old stuff. But we should appreciate here that when we took the derivatives of $\mathbf{v}$ and $\mathbf{a}$, we used the fact that the unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ *did not change with time*.

Eq. 3.1 above assumes that the unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ $\hat{\mathbf{k}}$ do *not* depend on u, or if the time t is the parameter, on the time. This seems obvious, but if we had another scheme for a set of "standard" unit vectors, this would not be the case. In plane polar coordinates we can consider the unit vectors $\hat{\mathbf{e}}_\rho$ and $\hat{\mathbf{e}}_\phi$ which point outward from the origin and perpendicular to this, in the counterclockwise direction. These

The particle's location in the plane is given by $\mathbf{r} = \rho\hat{\mathbf{e}}_\rho$, where the particle's polar coordinates are (ρ, ϕ). The new unit vectors are given by

$$\hat{\mathbf{e}}_\rho = \cos\phi\,\hat{\mathbf{i}} + \sin\phi\,\hat{\mathbf{j}} \qquad\qquad \hat{\mathbf{e}}_\phi = -\sin\phi\,\hat{\mathbf{i}} + \cos\phi\,\hat{\mathbf{j}}$$

and now if we take the time derivative of these vectors we get

$$\frac{d\hat{\mathbf{e}}_\rho}{dt} = \dot{\phi}\hat{\mathbf{e}}_\phi \qquad\qquad \frac{d\hat{\mathbf{e}}_\phi}{dt} = -\dot{\phi}\hat{\mathbf{e}}_\rho$$

where $\dot{\phi} = \frac{d\phi}{dt}$. This gives

$$\begin{aligned}\mathbf{v} \;&=\; \frac{d\mathbf{r}}{dt} \;=\; \frac{d\rho}{dt}\hat{\mathbf{e}}_\rho + \rho\frac{d\hat{\mathbf{e}}_\rho}{dt} \;=\; \dot{\rho}\hat{\mathbf{e}}_\rho + \rho(\dot{\phi}\,\hat{\mathbf{e}}_\phi) \\ &=\; \dot{\rho}\hat{\mathbf{e}}_\rho + \rho\dot{\phi}\,\hat{\mathbf{e}}_\phi\end{aligned}$$

This is the expression for velocity using polar coordinates for both the coordinates and the unit vectors.

3.2.1 Product Rules

When a scalar ϕ and vector $\mathbf{a}$ are functions of the parameter u, we have

$$\begin{aligned}\frac{d}{du}(\phi\mathbf{a}) \;&=\; \phi\frac{d\mathbf{a}}{du} + \frac{d\phi}{du}\mathbf{a} \\ \frac{d}{du}(\mathbf{a}\cdot\mathbf{b}) \;&=\; \mathbf{a}\cdot\frac{d\mathbf{b}}{du} + \frac{d\mathbf{a}}{du}\cdot\mathbf{b} \\ \frac{d}{du}(\mathbf{a}\times\mathbf{b}) \;&=\; \mathbf{a}\times\frac{d\mathbf{b}}{du} + \frac{d\mathbf{a}}{du}\times\mathbf{b}\end{aligned}$$

In the last of these note how the order is preserved; this is important since the cross product is not commutative.

3.3 Space Curves

A general curve in space can be given **parametrically** by

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

This form of a curve is useful for analyzing a trajectory in mechanics; it is also useful when we want to evaluate in integral along a path; more on that later.

While the parameter u could be anything, it is often useful to use the arclength of the curve as a parameter. The arclength s (starting from the point specified by u_1) is

$$s = \int_{u_1}^{u_2} \sqrt{\frac{d\mathbf{r}}{du} \cdot \frac{d\mathbf{r}}{du}}\, du$$

There are lots of juicy mathematical issues in the analysis of curves in space. This material falls under the heading of (classical) **differential geometry** which does have some use in physics but not enough that we need to go there right now!

Similarly, a surface can be specified by

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

3.4 Scalar and Vector fields

Now we return to the concept that a scalar (say, ϕ) or a vector (say, $\mathbf{a}$) can be defined at each point in space so that we deal with the functions

$$\phi(\mathbf{r}) \qquad \text{and} \qquad \mathbf{a}(\mathbf{r})$$

which are scalar and vector fields. (Actually, a field might depend on the space *and* time coordinate: $\phi(\mathbf{r}, t)$. That does arise in physics but here we will ignore any possible t dependence.)

We are immediately interested in space derivatives (of some kind) for these functions, and as already mentioned there are lots of choices!

3.5 Vector Operators

Now we consider the idea of a scalar or vector field, where the scalar ϕ or the vector $\mathbf{a}$ depends on the vector $\mathbf{r}$, that is, on the 3 coordinates (x, y, z).

The common vector (derivative) operators all use the ∇ operator (called "del" or "nabla"), defined by

$$\nabla \equiv \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z}$$

3.5.1 Gradient

The gradient of the scalar field $\phi(x, y, z)$ is defined by

$$\text{grad}\ \phi = \nabla\phi = \hat{\mathbf{i}}\frac{\partial\phi}{\partial x} + \hat{\mathbf{j}}\frac{\partial\phi}{\partial y} + \hat{\mathbf{k}}\frac{\partial\phi}{\partial z} \tag{3.2}$$

What does one *do* with the gradient? This operation produces the change in ϕ in a certain direction. In calculus, the rate of change is unambiguously written as $\frac{d\phi}{dx}$ (loosely speaking, step by dx and get a change $d\phi$) but here we need to specify the *direction* in which we step by some distance ds.

One can show that if we step by a small distance ds in the direction of the unit vector $\hat{\mathbf{a}}$, then

$$\frac{d\phi}{ds} = \nabla\phi \cdot \hat{\mathbf{a}}$$

To see this, write

$$d\phi = \frac{\partial\phi}{\partial x}dx + \frac{\partial\phi}{\partial y}dy + \frac{\partial\phi}{\partial z}dz = \nabla\phi \cdot d\mathbf{r} \ ,$$

so with $d\mathbf{r} = ds \cdot \hat{\mathbf{a}}$ we get the advertised result.

From this it follows that at a given point, the direction in ψ gives the greatest change per unit displacement ($\frac{d\phi}{ds}$) *is* the direction of the gradient, and the magnitude of that rate of change is $\left|\frac{d\phi}{ds}\right| = |\nabla\phi|$.

3.5.2 Divergence

The divergence of a vector field $\mathbf{a}(x, y, z)$ is given by

$$\text{div } \mathbf{a} = \nabla \cdot \mathbf{a} = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \tag{3.3}$$

In words, the divergence tells how much a vector function spreads out, or, well, *diverges* from a given point.

3.5.3 Curl

The curl of a vector field $\mathbf{a}(x, y, z)$ (also called the rotation or circulation of the field) is given by

$$\text{curl } \mathbf{a} = \nabla \times \mathbf{a} = \left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right)\hat{\mathbf{i}} + \left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right)\hat{\mathbf{j}} + \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right)\hat{\mathbf{k}} \tag{3.4}$$

a formula which is more easily remembered with the schematic expression:

$$\text{curl } \mathbf{a} = \nabla \times \mathbf{a} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ a_x & a_y & a_z \end{vmatrix} \tag{3.5}$$

3.6 Vector Operator Formulae

If we perform a vector operation on a *product* of two scalar or vector functions we can show that the result can be written in a form similar to the "product rule" of elementary calculus. Here though there are more possibilities for the product of tow functions (either scalar or vector) and the derivative operator.

One can show:

$$\begin{aligned}
\nabla(\phi\psi) &= \phi\nabla\psi + \psi\nabla\phi \\
\nabla(\mathbf{a}\cdot\mathbf{b}) &= \mathbf{a}\times(\nabla\times\mathbf{b}) + \mathbf{b}\times(\nabla\times\mathbf{a}) + (\mathbf{a}\cdot\nabla)\mathbf{b} + (\mathbf{b}\cdot\nabla)\mathbf{a} \\
\nabla\cdot(\phi\mathbf{a}) &= \phi\nabla\cdot\mathbf{a} + \mathbf{a}\cdot\nabla\phi \\
\nabla\cdot(\mathbf{a}\times\mathbf{b}) &= \mathbf{b}\cdot(\nabla\times\mathbf{a}) - \mathbf{a}\cdot(\nabla\times\mathbf{b}) \\
\nabla\times(\phi\mathbf{a}) &= \nabla\phi\times\mathbf{a} + \phi\nabla\times\mathbf{a} \\
\nabla\times(\mathbf{a}\times\mathbf{b}) &= \mathbf{a}(\nabla\cdot\mathbf{b}) - \mathbf{b}(\nabla\cdot\mathbf{a}) + (\mathbf{b}\cdot\nabla)\mathbf{a} - (\mathbf{a}\cdot\nabla)\mathbf{b}
\end{aligned}$$

3.7 Double Operations

There are five ways to combine the grad, div and curl operations (operating on a scalar or vector field) and the results are interesting. Two of them in fact give zero. For any sensible fields ϕ and $\mathbf{a}$, we have

$$\begin{aligned}
\nabla\times\nabla\phi &= \mathbf{0} \\
\nabla\cdot(\nabla\times\mathbf{a}) &= 0
\end{aligned}$$

In words, the curl of a gradient is zero, and the divergence of a curl is zero.

The other three possible relations need to done individually. First, after taking the gradient of a scalar field one can then take the divergence. We get:

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

The operator ∇^2, which here operates on the scalar ϕ, and often pronounced as "del squared" is of great importance in physics. It appears in classical wave equations and in the quantum (Schrödinger) wave equation.

You sometimes see ∇^2 operating on a *vector*. When the vector is expressed in Cartesian coordinates, the meaning of this is

$$\nabla^2\mathbf{a} = (\nabla^2 a_x)\,\hat{\mathbf{i}} + (\nabla^2 a_y)\,\hat{\mathbf{j}} + (\nabla^2 a_z)\,\hat{\mathbf{k}}$$

where we note that for example $\nabla^2 a_x$ really contains three terms, so there are actually nine second derivatives contained here!

The gradient of a divergence, $\nabla(\nabla\cdot\mathbf{a})$ is *not* necessarily zero but it doesn't come up very much in physics. For the record, in Cartesian coordinates it is given by

$$\begin{aligned}
\nabla(\nabla\cdot\mathbf{a}) &= \left(\frac{\partial^2 a_x}{\partial x^2} + \frac{\partial^2 a_y}{\partial x\partial y} + \frac{\partial^2 a_z}{\partial x\partial z}\right)\hat{\mathbf{i}} + \left(\frac{\partial^2 a_x}{\partial y\partial z} + \frac{\partial^2 a_y}{\partial y^2} + \frac{\partial^2 a_z}{\partial y\partial z}\right)\hat{\mathbf{j}} \\
&\quad + \left(\frac{\partial^2 a_x}{\partial z\partial x} + \frac{\partial^2 a_y}{\partial z\partial y} + \frac{\partial^2 a_z}{\partial z^2}\right)\hat{\mathbf{k}}
\end{aligned}$$

Finally, one can take the curl of the curl of a vector field. One gets

$$\nabla\times(\nabla\times\mathbf{a}) = \nabla(\nabla\cdot\mathbf{a}) - \nabla^2\mathbf{a} \tag{3.6}$$

This isn't so bad, but some caution is needed; we will soon look at these vector operators as they are used in other coordinate systems where the unit vectors depend on the coordinates and then the expression for $\nabla^2\mathbf{a}$ given above is not useful; in that case, Eq. 3.6 is still true but we need another way of finding $\nabla^2\mathbf{a}$.

3.8 Examples

Example: Evaluate $\nabla \cdot (r^3 \mathbf{r})$.

Here the $\mathbf{r}$ stands for

$$\mathbf{r} = r\hat{r} = x\,\hat{\mathbf{i}} + y\,\hat{\mathbf{j}} + z\,\hat{\mathbf{k}} \qquad \text{and} \qquad r = \sqrt{x^2 + y^2 + z^2}$$

so that

$$r^3\mathbf{r} = (x^2 + y^2 + z^2)^{3/2}(x\,\hat{\mathbf{i}} + y\,\hat{\mathbf{j}} + z\,\hat{\mathbf{k}})$$

To find $\nabla \cdot (r^3\mathbf{r})$ we need to take $\frac{\partial}{\partial x}$ of the x component, $\frac{\partial}{\partial y}$ of the y component, etc. The first of these is:

$$\frac{\partial}{\partial x}\left[x(x^2 + y^2 + z^2)^{3/2}\right] = (x^2 + y^2 + z^2)^{3/2} + x\tfrac{3}{2}(2x)(x^2 + y^2 + z^2)^{1/2}$$

Using $r = \sqrt{x^2 + y^2 + z^2}$, this simplifies to

$$\frac{\partial}{\partial x}\left[x(x^2 + y^2 + z^2)^{3/2}\right] = r^3 + 3x^2 r$$

It is fairly clear that we will also get

$$\frac{\partial}{\partial y}\left[y(x^2 + y^2 + z^2)^{3/2}\right] = r^3 + 3y^2 r \qquad \text{and} \qquad \frac{\partial}{\partial z}\left[z(x^2 + y^2 + z^2)^{3/2}\right] = r^3 + 3z^2 r$$

Adding these three terms together gives the divergence,

$$\nabla \cdot (r^3\mathbf{r}) = 3r^3 + 3(x^2 + y^2 + z^2)r = 3r^3 + 3r^3 = \boxed{6r^3}$$

3.9 Cylindrical Polar Coordinates

An alternative to the usual system of Cartesian coordinates (x, y, z) is shown in Fig. 3.1. Instead, we uniquely specify a point by joining it to the z axis with a line segment; the length of that segment is ρ, the angle it makes with the $+x$ axis is ϕ and its z coordinate is used as before; the point is specified with the set (ρ, ϕ, z), the **cylindrical polar coordinates** of the point.

The Cartesian and cylindrical polar coordinates are related by

$$x = \rho\cos\phi \qquad y = \rho\sin\phi \qquad z = z \tag{3.7}$$

with $0 \le \rho \le \infty$, $0 \le \phi < 2\pi$, and $-\infty < z < \infty$.

We could then write a coordinate vector as

$$\mathbf{r} = \rho\cos\phi\,\hat{\mathbf{i}} + \rho\sin\phi\,\hat{\mathbf{j}} + z\,\hat{\mathbf{k}}$$

but it is useful to new unit vectors to go along with the new coordinates. The unit vector associated with a coordinate must point in the direction we go when when change *only* that particular coordinate. We get these by taking

$$\frac{\partial \mathbf{r}}{\partial r}, \quad \frac{\partial \mathbf{r}}{\partial \phi} \quad \text{and} \quad \frac{\partial \mathbf{r}}{\partial z}$$

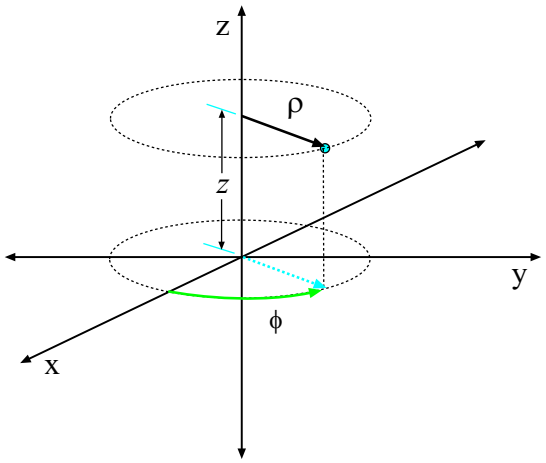

Figure 3.1: Cylindrical polar coordinates ρ, ϕ and z.

and making a unit vector of the result. One finds:

$$\hat{\mathbf{e}}_\rho = \cos\phi\,\hat{\mathbf{i}} + \sin\phi\,\hat{\mathbf{j}} \qquad \hat{\mathbf{e}}_\phi = -\sin\phi\,\hat{\mathbf{i}} + \cos\phi\,\hat{\mathbf{j}} \qquad \hat{\mathbf{z}} = \hat{\mathbf{k}} \tag{3.8}$$

These three (unit) vectors are mutually orthogonal, which was assure by the choice of the coordinate system itself. But the vectors *depend on the coordinates themselves*, a situation we did not have with our Cartesian unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$. We must keep this in imnd when we work with the new unit vectors.

Eqs. 3.8 can be inverted to give the Cartesian unit vectors in terms of the cylindrical ones:

$$\hat{\mathbf{i}} = \cos\phi\,\hat{\mathbf{e}}_\rho + \sin\phi\,\hat{\mathbf{e}}_\phi \qquad \hat{\mathbf{j}} = \sin\phi\,\hat{\mathbf{e}}_\rho + \cos\phi\,\hat{\mathbf{e}}_\phi \qquad \hat{\mathbf{k}} = \hat{\mathbf{e}}_z \tag{3.9}$$

We will want to evaluate integrals on lines, surfaces and volumes using these coordinates, and we will need expressions for the line, surface and volume elements in terms of them. In Cartesian coordinates we had the relatively simple ideas that a small displacement (path element) is

$$d\mathbf{r} = dx\,\hat{\mathbf{i}} + dy\,\hat{\mathbf{j}} + dz\,\hat{\mathbf{k}}$$

An area element that is perpendicular to the z direction (for example) is given by

$$d\mathbf{a} = dx\,dy\,\hat{\mathbf{z}}$$

and a small element of volume is given by $dV = dx\,dy\,dz$.

$$d\mathbf{r} = d\rho\,\hat{\mathbf{e}}_\rho + \rho\,d\phi\,\hat{\mathbf{e}}_\phi + dz\,\hat{\mathbf{e}}_z \tag{3.10}$$

$$dV = \rho\,d\rho\,d\phi\,dz \tag{3.11}$$

A surface of constant ρ (a cylinder) everywhere has its normal parallel to $\hat{\mathbf{e}}_\rho$. A small area element on this surface is given by $da_\rho = \rho\,d\phi\,dz$. Similarly, if we consider surfaces of constant ϕ (a half-plane) and z (a plane) we find that the (vector) area elements have components

$$da_\rho = \rho\,d\phi\,dz \qquad da_\phi = d\rho\,dz \qquad da_z = \rho\,d\rho\,d\phi \tag{3.12}$$

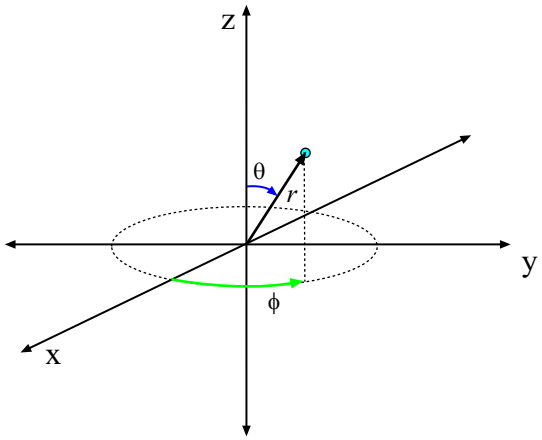

Figure 3.2: Spherical (polar) coordinates r, θ and ϕ.

3.10 Vector Operators in Cylindrical Coordinates

A vector field expressed *completely* in terms of cylindrical coordinates and the cylindrical units vectors has the form

$$\mathbf{a} = a_\rho(\rho,\phi,z)\,\hat{\mathbf{e}}_\rho + a_\phi(\rho,\phi,z)\,\hat{\mathbf{e}}_\phi + a_z(\rho,\phi,z)\,\hat{\mathbf{e}}_z$$

where there's nary an x, y or z to be seen!

We would like to have expressions for the vector derivative operators which can be used with a vector field given in this form.

$$\begin{aligned}
\nabla\Phi &= \frac{\partial\Phi}{\partial\rho}\hat{\mathbf{e}}_\rho + \frac{1}{\rho}\frac{\partial\Phi}{\partial\phi}\hat{\mathbf{e}}_\phi + \frac{\partial\Phi}{\partial z}\hat{\mathbf{e}}_z \\
\nabla\cdot\mathbf{a} &= \frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho a_\rho) + \frac{1}{\rho}\frac{\partial a_\phi}{\partial\phi} + \frac{\partial a_z}{\partial z} \\
\nabla\times\mathbf{a} &= \left(\frac{1}{\rho}\frac{\partial a_z}{\partial\phi} - \frac{\partial a_\phi}{\partial z}\right)\hat{\mathbf{e}}_\rho + \left(\frac{\partial a_\rho}{\partial z} - \frac{\partial a_z}{\partial\rho}\right)\hat{\mathbf{e}}_\phi + \frac{1}{\rho}\left[\frac{\partial}{\partial\rho}(\rho a_\phi) - \frac{\partial a_\rho}{\partial\phi}\right]\hat{\mathbf{e}}_z \\
\nabla^2\Phi &= \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\Phi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\Phi}{\partial\phi^2} + \frac{\partial^2\Phi}{\partial z^2}
\end{aligned}$$

3.11 Spherical Coordinates

Another common alternative to the Cartesian coordinates (x,y,z) is shown in Fig. 3.2. Here, we uniquely specify a point by joining it to the origin with a line segment; the length of that segment is r, the angle it makes with the $+z$ axis is θ and if we join the point to the z axis with a shortest line segment, the angle which *that* segment makes with the $+x$ axis is ϕ; then the point is specified with the set (r,θ,ϕ), the **spherical polar coordinates** of the point.

The Cartesian and spherical coordinates are related by

$$x = r\sin\theta\cos\phi \qquad y = r\sin\theta\sin\phi \qquad z = r\cos\theta \tag{3.13}$$

As with cylindrical coordinates, we can form unit vectors for spherical coordinates (unit vectors pointing in the direction of displacement if only one of the coordinates changes). The relations are:

$$\begin{aligned}\hat{\mathbf{e}}_r &= \sin\theta\cos\phi\,\hat{\mathbf{i}} + \sin\theta\sin\phi\,\hat{\mathbf{j}} + \cos\theta\,\hat{\mathbf{k}}\\ \hat{\mathbf{e}}_\theta &= \cos\theta\cos\phi\,\hat{\mathbf{i}} + \cos\theta\sin\phi\,\hat{\mathbf{j}} - \sin\theta\,\hat{\mathbf{k}}\\ \hat{\mathbf{e}}_\phi &= -\sin\phi\,\hat{\mathbf{i}} + \cos\phi\,\hat{\mathbf{j}}\end{aligned}$$

with inverse relations

$$\begin{aligned}\hat{\mathbf{i}} &= \sin\theta\cos\phi\,\hat{\mathbf{e}}_r + \cos\theta\cos\phi\,\hat{\mathbf{e}}_\theta - \sin\phi\,\hat{\mathbf{e}}_\phi\\ \hat{\mathbf{j}} &= \sin\theta\sin\phi\,\hat{\mathbf{e}}_r + \cos\theta\sin\phi\,\hat{\mathbf{e}}_\theta + \cos\phi\,\hat{\mathbf{e}}_\phi\\ \hat{\mathbf{k}} &= \cos\theta\,\hat{\mathbf{e}}_r - \sin\theta\,\hat{\mathbf{e}}_\theta\end{aligned}$$

The line elements and the volume element in spherical coordinates are given by

$$d\mathbf{r} = dr\,\hat{\mathbf{e}}_r + r\,d\theta\,\hat{\mathbf{e}}_\theta + r\sin\theta\,d\phi\,\hat{\mathbf{e}}_\phi \tag{3.14}$$

$$dV = r^2\sin\theta\,dr\,d\theta\,d\phi \tag{3.15}$$

A surface of constant r (a sphere) everywhere has its normal parallel to $\hat{\mathbf{e}}_r$. A small area element on this surface is given by $da_r = r^2\sin\theta\,d\theta\,d\phi$. Similarly, if we consider surfaces of constant θ (a cone) and ϕ (a half–plane) we find that the (vector) area elements have components

$$da_r = r^2\sin\theta\,d\theta\,d\phi \qquad da_\theta = r\sin\theta\,dr\,d\phi \qquad da_\phi = r\,dr\,d\theta \tag{3.16}$$

3.12 Vector Operators in Spherical Coordinates

And now we consider scalar and vector field expressed *completely* in terms of spherical coordinates and the spherical unit vectors, in the form

$$\Phi = \Phi(r,\theta,\phi) \qquad \mathbf{a} = a_r(r,\theta,\phi)\,\hat{\mathbf{r}} + a_\theta(r,\theta,\phi)\,\hat{\mathbf{e}}_\theta + a_\phi(r,\theta,\phi)\,\hat{\mathbf{e}}_\phi\ .$$

We would like to have expressions for the vector derivative operators which can be used with a vector field given in this form. One can show that they are:

$$\begin{aligned}\nabla\Phi &= \frac{\partial\Phi}{\partial r}\hat{\mathbf{e}}_r + \frac{1}{r}\frac{\partial\Phi}{\partial\theta}\hat{\mathbf{e}}_\theta + \frac{1}{r\sin\theta}\frac{\partial\Phi}{\partial\phi}\hat{\mathbf{e}}_\phi\\ \nabla\cdot\mathbf{a} &= \frac{1}{r^2}\frac{\partial}{\partial r}(r^2a_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta a_\theta) + \frac{1}{r\sin\theta}\frac{\partial a_\phi}{\partial\phi}\\ \nabla\times\mathbf{a} &= \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta\,a_\phi) - \frac{\partial a_\theta}{\partial\phi}\right]\hat{\mathbf{e}}_r + \frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial a_r}{\partial\phi} - \frac{\partial}{\partial r}(ra_\phi)\right]\hat{\mathbf{e}}_\theta + \frac{1}{r}\left[\frac{\partial}{\partial r}(ra_\theta) - \frac{\partial a_r}{\partial\theta}\right]\hat{\mathbf{e}}_\phi\\ \nabla^2\Phi &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Phi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Phi}{\partial\phi^2}\end{aligned}$$

Example: Evaluate $\nabla\cdot(r^3\mathbf{r})$.

Recall (?) that we did this example using Cartesian coordinates; as one might expect, it is much easier when we use spherical coordinates throughout. Write the vector field as:

$$\mathbf{a} = r^3\mathbf{r} = (r^3)r\,\hat{\mathbf{r}} = r^4\,\hat{\mathbf{e}}_r$$

so that there is only *one* coefficient function here,

$$a_r = r^4$$

Then the formula for the divergence in spherical coordinates gives

$$\begin{aligned} \nabla \cdot \mathbf{a} &= \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 a_r) \\ &= \frac{1}{r^2}\frac{\partial}{\partial r}(r^6) = \frac{1}{r^2}(6r^5) = \boxed{6r^3} \end{aligned}$$

which is the same answer we got before.

3.13 Line, Surface and Volume Integrals

Opening up our space to three dimensions and our objects to vectors gives more options for performing integrations (infinite summations of vanishingly small entities).

3.13.1 Line Integrals

Possibilities are

$$\int_C \phi \, d\mathbf{r}\,, \qquad \int_C \mathbf{a} \cdot \, d\mathbf{r}\,, \qquad \int_C \mathbf{a} \times \, d\mathbf{r}$$

For example, the second of these means to consider the limiting process as we split up the curve C into N displacements $d\mathbf{r}_i$ located at (around) the coordinate (x^*, y^*, z^*) and then take

$$\sum_i^N \mathbf{a}(x^*, y^*, z^*) \cdot d\mathbf{r}_i$$

in the limit of large N and small intervals. But how we evaluate such a thing?

In basic calculus, we have the Fundamental Theorem,

$$\int_a^b f(x)\, dx = F(x)\Big|_a^b = F(b) - F(a) \qquad \text{where} \qquad \frac{dF}{dx} = f(x)$$

Here, the step $d\mathbf{r}$ can be written $d\mathbf{r} = dx\,\hat{\mathbf{i}} + dy\hat{\mathbf{j}} + dz\,\hat{\mathbf{k}}$, but more care is needed because the dx, dy and dz are *related* as we step along the curve. Likewise, the argument is a function of three variables given by the particular path. So it's not so simple.

Generally, we have to the reduce the integral down to *one* variable which will be some parameter giving the point along the curve C.

Before giving an example, we state some (obvious?) facts about line integrals:

$$\int_A^B \mathbf{a} \cdot d\mathbf{r} = -\int_B^A \mathbf{a} \cdot d\mathbf{r}$$

$$\int_A^B \mathbf{a} \cdot d\mathbf{r} = \int_A^P \mathbf{a} \cdot d\mathbf{r} + \int_P^B \mathbf{a} \cdot d\mathbf{r}$$

where the integral are all done along the same curve C and the point P is on the curve between A and B.

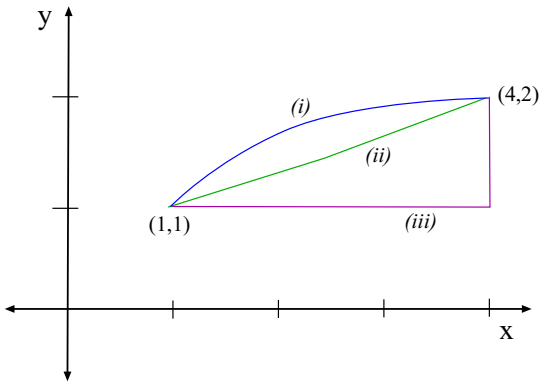

Figure 3.3: Line integral evaluated on three different paths.

3.13.2 Example of Line Integrals

RHB give a useful example of evaluating a line integral by different methods.

Example: Here we want to evaluate

$$I = \int_C \mathbf{a} \cdot d\mathbf{r} \qquad \text{where} \qquad \mathbf{a} = (x+y)\,\hat{\mathbf{i}} + (y-x)\,\hat{\mathbf{j}}$$

along each of the paths in Fig. 3.3, which are

(i) the parabola $y^2 = x$ from $(1,1)$ to (4.2).
(ii) the curve $x = 2u^2 + u + 1$, $y = 1 + u^2$ from $(1,1)$ to $(4,2)$
(iii) the line $y = 1$ from $(1,1)$ to $(4,1)$ followed by the line $x = 4$ from $(4,1)$ to $4,2)$.

In all the cases, we have $d\mathbf{r} = dx\,\hat{\mathbf{i}} + dy\,\hat{\mathbf{j}}$, so that in general

$$I = \int_C \mathbf{a} \cdot d\mathbf{r} = \int_C [(x+y)\,dx + (y-x)\,dy] \tag{3.17}$$

but the different paths necessitate different steps to go further!

(i) Along the parabola $y^2 = x$, the differentials are related by $2y\,dy = dx$. Substitute for x and dx in 3.17; for the both integrals –since everything is now in terms of y— the limits are now $y = 1$ to $y = 2$. This gives:

$$I = \int_1^2 (y^2 + y)\,dy + \int_1^2 (y - y^2)\,dy = \boxed{\tfrac{34}{3}} \tag{3.18}$$

where I assumed that in the final step you can work out such a simple standard integral!

(ii) With the parametric representation of the path, ($x = 2u^2 + u + 1$ and $y = 1 + u^2$) we have

$$dx = (4u+1)\,du \qquad\qquad dy = 2u\,du$$

and on the curve the parameter u goes from 0 to 1. Substituting all of this, 3.17 becomes

$$I = \int_0^1 \left[(2u^2 + u + 1 + 1 + u^2)(4u+1) + (1 + u^2 - 2u^2 - u - 1)\,(2u)\right] du = \boxed{\tfrac{32}{3}} \tag{3.19}$$

where again I omit the steps by which we get the final answer from a simple standard integral.

(iii) For the third path, each section is simple (since either x or y is held constant) but we have to do them separately.

One the first segment, y is constant at $y = 1$ and $dy = 0$ so that the second term in 3.17 is zero. The integration limits are $x = 1$ to $x = 4$ so this part gives

$$I_1 = \int_1^4 (x + 1)\, dx = \tfrac{21}{2}$$

On the second section, x is constant at $x = 4$ and $dx = 0$ so that the first term in 3.17 is zero. The integration limits are $y = 1$ to $y = 2$ so for this part we have

$$I_2 = \int_1^2 (y - 4)\, dy = -\tfrac{5}{2}$$

and the sum of the two parts gives

$$I = I_1 + I_2 = \tfrac{21}{2} - \tfrac{5}{2} = \boxed{8}$$

3.13.3 More On Line Integrals

If the curve is very kinky it may not be representable as a simple function $y(x)$ or $x(y)$. In that case it can be subdivided, or else one should use a parametric representation.

We note that in general (as we saw in the example above) the value of the line integral depends on the particular path taken between the endpoints. Physicists often deal with the case where it does *not* depend on it (more on that later) but one can't assume that!

Some notation: One can consider a line integral which ends up at the starting point! Then there really is not beginning and end point of the curve, it is just one path but the *sign* of the result does depend on the *direction* of the path; the convention is the counter-clockwise sense, whatever that may mean for the particular problem!

For a closed path we write, e.g.

$$I = \oint \mathbf{a} \cdot d\mathbf{r}$$

There are situations where an integral on a closed path necessarily gives zero.

3.13.4 Line Integrals in Physics

So far you have seen the definition of the work done by a force $\mathbf{F}$ of a particle which moves on a path on the curve C:

$$W = \int_C \mathbf{F} \cdot d\mathbf{r}$$

For *some* forces (i.e. the conservative ones) this is equal to the (negative of) the change in potential energy.

Ampere's law in electromagnetism involves a line integral of the magnetic field over a *closed* path:

$$\oint_C \mathbf{B} \cdot d\mathbf{r} = \mu_0 I$$

where I is the total electric current passing through the loop. (And μ_0 is a constant.) If there is no net current though the loop, then $\oint B \cdot d\mathbf{r} = 0$.

The force on a finite length of a current–carrying wire is

$$\mathbf{F} = I \int_C d\mathbf{r} \times \mathbf{B}$$

and the force on a closed current loop is

$$\mathbf{F} = I \oint_C d\mathbf{r} \times \mathbf{B}$$

3.13.5 Green's Theorem in a Plane

A possibly useful theorem is the case the a closed path C in a plane (say the xy plane) encloses a region R. Then we have

$$\oint_C (P\,dx + Q\,dy) = \int\int_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy$$

3.13.6 Conservative Fields

For a certain class of vector fields, the line integral $\int_A^B \mathbf{a} \cdot d\mathbf{r}$ has an important property: It is independent of the path C taken to get from A to B, and as a consequence, the integral $\oint_C \mathbf{a} \cdot d\mathbf{r}$ around closed loop C is zero. A very important theorem tells us that the following conditions on a vector field $\mathbf{a}$ are *equivalent*:

- The integral $\int_A^B \mathbf{a} \cdot d\mathbf{r}$ is independent of the path taken from A to B. Equivalently, the integral $\oint_C \mathbf{a} \cdot d\mathbf{r}$ around any closed loop C is zero.
- There exists a function ϕ of position such that $\mathbf{a} = \nabla\phi$.
- $\nabla \times \mathbf{a} = \mathbf{0}$.

3.14 Surface Integrals

Possibilities are

$$\int_S \phi\, d\mathbf{S}\,, \qquad \int_S \mathbf{a} \cdot\, d\mathbf{S}\,, \qquad \int_S \mathbf{a} \times\, d\mathbf{S}$$

While the meaning of these is pretty clear, we realize that there's a remaining ambiguity about the direction of the surface element $d\mathbf{S}$. Making one choice or the other amounts to a change in overall sign for the surface integral (which is not really profound) but it's useful to have some sort of convention!

For a closed surface, there is an *outward* and an *inward* direction for the surface vector and the choice will be clear. For an open surface, we are often also thinking of evaluating a line integral on the boundary of the surface, which is a closed path. In that case, when we choose the direction in which we want to go around the closed path and the direction of the surface vector $d\mathbf{S}$ is given by the right-hand rule, as shown in Fig. 3.4.

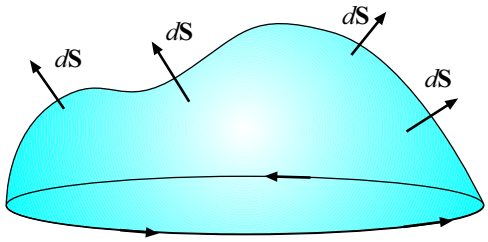

Figure 3.4: Direction of surface vector goes with the sense of the line integral on the boundary, by the right-hand rule.

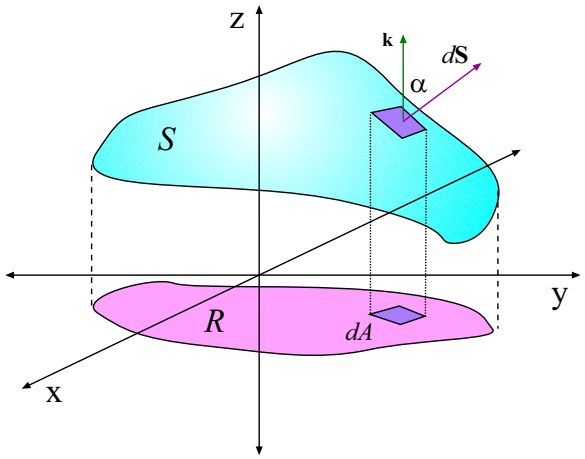

Figure 3.5: Setting up surface integral for an arbitrary surface.

3.14.1 A General Method

While a surface integral is fairly easy to do if the surface in question is one where a coordinate is constant for a curvilinear system (for example, the conical surface where θ is constant in spherical coordinates) we would like to have a method good for an arbitrary surface.

Consider the situation shown in Fig. 3.5. The strategy is to project the surface S down to the xy plane where a region R is swept out; eventually we want to do the integral over R.

A path of surface S has surface element $d\mathbf{S}$ which makes an angle α with the z direction. The projection of this area element in the xy plane is dA. These infinitesimal areas are related by $dA = \cos\alpha\, dS$. If $\hat{\mathbf{n}}$ is the unit normal to the surface for this patch (that is, $d\mathbf{S} = dS\,\hat{\mathbf{n}}$) then $\cos\alpha = \hat{\mathbf{n}} \cdot \hat{\mathbf{k}}$, or

$$dS = \frac{dA}{(\hat{\mathbf{n}} \cdot \hat{\mathbf{k}})} \qquad \text{and} \qquad d\mathbf{S} = \frac{\hat{\mathbf{n}}\, dA}{(\hat{\mathbf{n}} \cdot \hat{\mathbf{k}})}$$

From our study of the gradient, we know that if the surface is given by the equation $f(x, y, z,) = 0$ then the unit normal at any point in given by

$$\hat{\mathbf{n}} = \frac{\nabla f}{|\nabla f|}$$

With this, the first of the expression for dS can be written

$$dS \;=\; \frac{dA}{(\hat{\mathbf{n}} \cdot \hat{\mathbf{k}})} \;=\; \frac{|\nabla f| dA}{\nabla f \cdot \hat{\mathbf{k}}} \;=\; \frac{|\nabla f|\, dA}{\left(\frac{\partial f}{\partial z}\right)}$$

The last expression can, using the defining relation for the surface, be written as a function of x and y and in this way the original integral involving a vector field and the surface S can be made into an integral over the region R in the xy plane.

As a simple (maybe too simple) example of the general method, we find the area of a hemisphere of radius a. The hemisphere is centered at the origin and we consider the part with $z > 0$. On this surface S we want to evaluate $\int_S dS$. Of course we have to get half the area of a sphere, $2\pi a^2$!

The simplest way to do the integral is to use spherical coordinates, as S is a surface on constant r. The surface element for a constant-r surface is

$$da_r \hat{\mathbf{e}}_r = r^2 \sin\theta \, d\theta \, d\phi \, \hat{\mathbf{e}}_r$$

which (on the surface where $r = a$) has magnitude $a^2 \sin\theta \, d\theta \, d\phi$. For the upper hemisphere, θ goes from 0 to $\pi/2$ and ϕ goes from 0 to 2π. The area is then

$$S = \int_S da_r = \int_0^{2\pi} \int_0^{\pi/2} a^2 \sin\theta \, d\theta \, d\phi = (2\pi)a^2 \int_0^{\pi/2} \sin\theta \, d\theta$$

The sin integral is easy and we get

$$S = (2\pi)a^2(2) = 4\pi a^2$$

Now we work the example by projecting onto the xy plane. The function defining the surface is

$$f(x, y, z) = x^2 + y^2 + z^2 - a^2 = 0$$

which gives

$$\nabla f = 2x \, \hat{\mathbf{i}} + 2y \, \hat{\mathbf{j}} + 2z \, \hat{\mathbf{k}} = 2\mathbf{r} \qquad \Longrightarrow \qquad |\nabla f| = 2|\mathbf{r}| = 2a$$

Also,

$$\frac{\partial f}{\partial z} = 2z = 2\sqrt{a^2 - x^2 - y^2} \qquad \Longrightarrow \qquad dS = \frac{|\nabla f| \, dA}{\left(\frac{\partial f}{\partial z}\right)} = \frac{2a \, dA}{2\sqrt{a^2 - x^2 - y^2}}$$

And using this, the integral over S becomes an integral over the region R (a circle of radius a) in the xy plane:

$$\text{Area} = \int_S dS = \int\int_R \frac{|\nabla f| \, dA}{\partial f / \partial z} = \int\int_R \frac{a \, dA}{\sqrt{a^2 - x^2 - y^2}}$$

At this point, we realize that this integral will be much easier to work using plane polar coordinates (ρ, ϕ) for which the area element is $\rho \, d\rho \, d\phi$. This gives

$$\text{Area} = a \int_0^{2\pi} \frac{1}{\sqrt{a^2 - \rho^2}} \rho \, d\rho \, d\phi = 2\pi a(-1)\sqrt{a^2 - \rho^2}\Big|_0^a = 2\pi a(-1)(0 - a) = 2\pi a^2$$

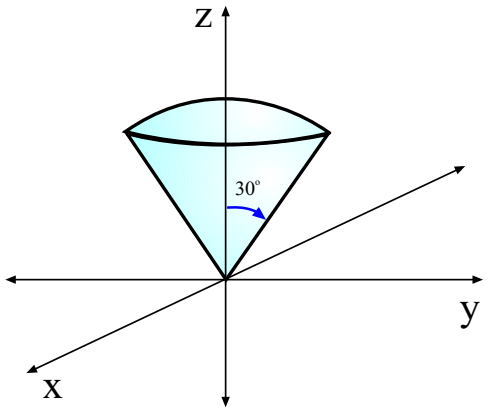

Figure 3.6: Ice cream cone surface for the example.

3.14.2 Integrals on Coordinate Surfaces

If the surface of integration is one where a particular coordinate of a curvilinear system is constant, the integral is usually much easier, and we make use the area elements da_i for the surface in question. An example that will give us lots of practice is (part of) a problem from our EM text:

Example: For the vector field

$$\mathbf{v} = r^2 \sin\theta\, \hat{\mathbf{e}}_r + 4r^2 \cos\theta\, \hat{\mathbf{e}}_\theta + r^2 \tan\theta\, \hat{\mathbf{e}}_\phi \ ,$$

evaluate the integral $\oint_S \mathbf{v} \cdot d\mathbf{S}$ for the closed "ice cream cone" shown in Fig. 3.6. (The top surface is spherical with radius R and centered at the origin.)

The surface here has two parts; the "cone" part is a surface of constant θ, namely $\theta = \pi/6$, on which r runs from 0 to R and ϕ runs from 0 to 2π. The area element for this part is

$$d\mathbf{a}_\theta \ = \ r \sin\theta\, dr\, d\phi\, \hat{\mathbf{e}}_\theta \big|_{\theta=\pi/6} \ = \ r \sin(\pi/6)\, dr\, d\phi\, \hat{\mathbf{e}}_\theta$$

(The unit vector $\hat{\mathbf{e}}_\theta$ does point outward for this surface, which is what we want.) The surface integral picks out only the $\hat{\mathbf{e}}_\theta$ part of the field $\mathbf{v}$; we get:

$$\begin{aligned} \int_{S,\ \text{cone}} \mathbf{v} \cdot d\mathbf{S} \ &= \ \int_0^{2\pi} d\phi \int_0^R dr\, (r \sin(\pi/6)(4r^2 \cos(\pi/6)) \\ &= \ 4(2\pi) \sin(\pi/6) \cos(\pi/6) \int_0^R r^3\, dr = 8\pi \cdot \tfrac{1}{2} \cdot \tfrac{\sqrt{3}}{2} \cdot \frac{R^4}{4} = \frac{\pi\sqrt{3}R^4}{2} \end{aligned}$$

Now do the "ice cream" part of the surface, which is a surface of constant r, i.e. $r = R$, on which θ runs from 0 to $\pi/6$ and ϕ runs from 0 to 2π. Here the area element is

$$d\mathbf{a}_r = r^2 \sin\theta\, d\theta\, d\phi\, \hat{\mathbf{e}}_r \big|_{r=R} = R^2 \sin\theta\, d\theta\, d\phi\, \hat{\mathbf{e}}_r$$

Then the surface integral for the cap is:

$$\begin{aligned}\int_{S,\text{ cap}} \mathbf{v}\cdot d\mathbf{S} &= \int_0^{2\pi} d\phi \int_0^{\pi/6} d\theta (r^2\sin\theta)(R^2\sin\theta) \\ &= R^4(2\pi)\int_0^{\pi/6}\sin^2\theta\, d\theta = 2\pi R^4\left[-\tfrac{1}{2}\cos\theta\,\sin\theta+\tfrac{1}{2}\theta\right]\Big|_0^{\pi/6} \\ &= 2\pi R^4\tfrac{1}{2}\left[-\frac{\sqrt{3}}{4}+\frac{\pi}{6}\right]\end{aligned}$$

The integral over the whole surface is the sum of the integrals from the two parts, thus:

$$\oint_S \mathbf{v}\cdot d\mathbf{S} = 2\pi R^4\left[\frac{\sqrt{3}}{4}-\frac{\sqrt{3}}{8}+\frac{\pi}{12}\right] = \boxed{\frac{\pi R^4}{12}[2\pi+3\sqrt{3}]}$$

3.14.3 Applications in Physics

Possibly the most famous appearance of a surface integral is in Gauss' law, a relation between the electric field integrated over a closed surface and the total charge contained within that surface:

$$\oint_S \mathbf{E}\cdot d\mathbf{S} = \frac{q_{\text{tot}}}{\epsilon_0}$$

In general, a vector field dotted with the surface area element and integrated over a surface is known as a **flux**; for a general surface $\int_S \mathbf{E}\cdot d\mathbf{S}$ is called **electric flux**.

3.14.4 Surface Integrals on Spherical Surfaces

In general, for problems in spherical coordinates we are evaluating integrals of the form

$$\int_0^{2\pi}\int_0^{\pi} f(r,\theta,\phi)\, r^2\, dr\, \sin\theta\, d\theta\, d\phi$$

3.15 Volume Integrals

Possibilities are

$$\int_V \phi\, dV\,, \qquad \int_V \mathbf{a}\, dV$$

The last case can be split up as

$$\int_V \mathbf{a}\, dV = \hat{\mathbf{i}}\int_V a_x\, dV + \hat{\mathbf{j}}\int_V a_y\, dV + \hat{\mathbf{k}}\int_V a_z\, dV$$

Again, note that we can factor out the unit vectors for Cartesian coordinates because they are *constant*.

Obvious examples of volume integrals in physics are the total mass and total charge of an object:

$$M = \int_V \rho_{\text{mass}}(\mathbf{r})\, dV \qquad\qquad Q = \int_V \rho_{\text{ch}}(\mathbf{r})\, dV$$

Another is the moment of inertia of a mass about a particular axis: if $s(\mathbf{r})$ is the distance of the mass point at $\mathbf{r}$ *from the axis*, then the moment of inertia is

$$I = \int_V \rho_{\text{mass}}(\mathbf{r})\, s(\mathbf{r})^2\, dV \tag{3.20}$$

Example: Find the moment of inertia of a uniform sphere of radius R about an axis through its center.

For a mass element at (r, θ, ϕ), the distance from the axis is $s(\mathbf{r}) = r\sin\theta$. The volume element in spherical coordinates is $dV = r^2 \sin\theta\, d\theta\, d\phi$ and for the integral over the volume of the sphere, r goes from 0 to R, θ goes from 0 to π and ϕ goes from 0 to 2π. Then if the (uniform) mass density of the material is ρ then using Eq. 3.20 we have

$$I = \int_0^{2\pi}\int_0^{\pi}\int_0^{R} (r\sin\theta)^2 r^2 \sin\theta\, d\theta\, d\phi = 2\pi\rho \int_0^{\pi}\int_0^{R} r^4 \sin^3\theta\, dr\, d\theta$$

For the second step we have brought the constant ρ outside the integral and done the ϕ integral, which gives a factor of 2π. For the θ integral we have

$$\int_0^{\pi} \sin^2\theta \sin\theta\, d\theta = \int_{-1}^{1} (1 - x^2) dx$$

where we have changed variables to $x = \cos\theta$ and then used $\sin^2\theta = (1 - x^2)$. This is then

$$\Longrightarrow \qquad = x - \frac{x^3}{3}\bigg|_{-1}^{1} = 2 - \tfrac{2}{3} = \tfrac{4}{3}$$

The r integral is just

$$\int_0^R x^4\, dr = \frac{x^5}{5}\bigg|_0^R = \frac{R^5}{5}$$

Putting all of these factors together, we have

$$I = \frac{8\pi\rho R^5}{15}$$

but to express it in terms of the mass M of the sphere, use

$$\rho = \frac{M}{\frac{4}{3}\pi R^3}$$

which gives

$$I = \frac{8\pi R^5}{15} \cdot \frac{M}{\frac{4}{3}\pi R^3} = \tfrac{2}{5} M R^2$$

3.16 Integration Theorems

3.16.1 The Divergence Theorem

Given some vector field $\mathbf{a}(\mathbf{r})$, and a volume V bounded by a closed surface S, we can evaluate the volume integral $\int_V \nabla \cdot \mathbf{a}\, dV$ and the surface integral $\oint_S \mathbf{a} \cdot d\mathbf{S}$. *These give the same result:*

$$\int_V \nabla \cdot \mathbf{a}\, dV = \oint_S \mathbf{a} \cdot\, d\mathbf{S} \tag{3.21}$$

If ϕ and ψ are two scalar fields, one can show:

$$\oint_S (\phi \nabla\psi - \psi\nabla\phi) \cdot d\mathbf{S} = \int_V (\phi\nabla^2\psi - \psi\nabla^2\phi)\, dV \tag{3.22}$$

One can also show (for a vector field $\mathbf{b}$):

$$\int_V \nabla\phi\, dV = \oint_S \phi\, d\mathbf{S} \qquad\qquad \int_V \nabla \times \mathbf{b}\, dV = \oint_S d\mathbf{S} \times \mathbf{b} \tag{3.23}$$

Example: Using the divergence theorem, prove the first of the theorems in Eq. 3.23

Note that the integrals on both sides of this equation give *vectors*, while the divergence theorem relates two *scalars*.

Let the vector field $\mathbf{a}$ in the divergence theorem be given by

$$\mathbf{a} = \phi\mathbf{c}$$

where ϕ is any well–behaved scalar field and $\mathbf{c}$ is a *constant* vector, that is, it has no dependence on the coordinates. Then by one of the product rules, $\nabla \cdot \mathbf{a} = \mathbf{c} \cdot \nabla\phi$, and using this in the divergence theorem gives (for a volume V and its bounding surface S),

$$\int_V \mathbf{c} \cdot \nabla\phi\, dV = \oint_S \phi\mathbf{c} \cdot d\mathbf{S}$$

As it is a constant, $\mathbf{c}$ can go outside the integrals as:

$$\mathbf{c} \cdot \int_V \nabla\phi\, dV = \mathbf{c} \cdot \oint_S \phi\, d\mathbf{S}$$

Now if we have an equation of this form ($\mathbf{c}$ dotted with vectors on both sides of the equation, where $\mathbf{c}$ is arbitrary) then we can conclude that the vectors *themselves* are equal. To be more definite, if we were to choose $\mathbf{c} = \hat{\mathbf{i}}$, it would show that the x components of those vectors were equal and similarly for the y and z components. Thus we can conclude that the two vectors are equal, and then

$$\int_V \nabla\phi\, dV = \oint_S \phi\, d\mathbf{S}$$

3.16.2 The Divergence Theorem and Physics

Suppose we know that for some fields ϕ and $\mathbf{a}$, and for an *arbitrary* volume V,

$$\int \phi \, dV = \oint_S \mathbf{a} \cdot d\mathbf{S} = \int_V \nabla \cdot \mathbf{a} \, dV \tag{3.24}$$

where the second equality is *necessarily* true from the divergence theorem.

Now if the first and last expressions in 3.24 are equal for *any* volume V then we are free to choose a very small volume V, centered at the point $\mathbf{r}$. Over such a small volume the values of the integrands are essentially constant, giving us $\phi \, dV = \nabla \cdot \mathbf{a}$, or:

$$\phi(\mathbf{r}) = \nabla \cdot \mathbf{a}$$

In this way, an *integral* relation between scalar and vector fields has become a *differential* relation.

In this way the Maxwell equations of electrodynamics have both integral and differential expressions. One of those equations is

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = \frac{q_{\text{tot}}}{\epsilon_0} = \frac{1}{\epsilon_0} \int_V \rho(\mathbf{r}) \, dV$$

where q_{tot} is the total charge enclosed by the surface. But the divergence theorem allows us the write the surface as a volume integral and so

$$\int_V \nabla \cdot \mathbf{E} \, dV = \frac{q_{\text{tot}}}{\epsilon_0} = \frac{1}{\epsilon_0} \int_V \rho(\mathbf{r}) \, dV$$

Since this is true for *any* volume, we can equate the arguments of the volume integrals *at any value of* $\mathbf{r}$:

$$\nabla \cdot \mathbf{E} = \frac{\rho(\mathbf{r})}{\epsilon_0}$$

Also from electromagnetism is the principle of the conservation of charge. If $\mathbf{J}$ is the current density then the integral $\oint_S \mathbf{J} \cdot d\mathbf{S}$ gives the (outward) rate of passage of electric charge through a closed surface S. Because of charge conservation, this rate of the charge flow is equal to the rate at which charge is lost from the interior of S. Again, if q_{tot} is the total amount of charge inside S, then this is stated as:

$$\oint_S \mathbf{J} \cdot d\mathbf{S} \;=\; -\frac{dq_{\text{tot}}}{dt} \;=\; -\frac{d}{dt} \int_V \rho(\mathbf{r}, t) \, dV \;=\; -\int_V \frac{\partial \rho}{\partial t} \, dV$$

Using the divergence theorem, the surface integral is made into a volume integral, giving:

$$\int_V \nabla \cdot \mathbf{J} \, dV = -\int_V \frac{\partial \rho}{\partial t} \, dV$$

and since is true for any volume V, we can give the differential relation,

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t}$$

3.16.3 Stokes' Theorem

For this theorem we consider an *open* surface S bounded by a closed curve C and a vector field $\mathbf{a}$. As discussed, before when doing integrals on *both* the surface and its bounding curve, we want to be consistent with our choice of direction for the area element $d\mathbf{S}$ and the sense in which we go around the curve C. We agree that we'll follow the right-hand rule.

Having agreed to this, we have an equality between a surface integral and and a line integral:

$$\int (\nabla \times \mathbf{a}) \cdot \, d\mathbf{S} = \oint_C \mathbf{a} \cdot \, d\mathbf{r} \tag{3.25}$$

One can also show the related theorems (with scalar field ϕ):

$$\int_S d\mathbf{S} \times \nabla\phi = \oint_C \phi \, d\mathbf{r} \qquad\qquad \int_S (d\mathbf{S} \times \nabla) \times \mathbf{a} = \oint_C d\mathbf{r} \times \mathbf{a}$$

A special case of Stokes' Theorem is for a plane surface S and its (planar) bounding curve C. Here the surface elements $d\mathbf{S}$ only have a z component, so that the surface integral in Stokes' Theorem is $\int_S (\nabla \times a)_z dS$. Also, the curve C is limited to line elements $d\mathbf{r} = dx\,\hat{\mathbf{i}} + dy\,\hat{\mathbf{j}}$, so that Stokes' Theorem gives

$$\int_S \left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y} \right) dS = \oint_C (a_x \, dx + a_y \, dy)$$

But this is exactly the form of "Green's theorem in a plane", given earlier, with P replaced by a_x and Q replaced by a_y.

3.16.4 Stokes' Theorem and Physics

The most famous example of the use of Stokes' Theorem in physics is probably the relation between magnetic fields and currents known as **Ampere's Law**. To use this law, consider a set of currents (such as might be flowing though a set of wires and consider a closed path C which might loop around some of those wires and a surface bounded by that path. The currents give rise to a magnetic field $\mathbf{B}$ and on the curve C we can calculate the line integral $\oint \mathbf{B} \cdot d\mathbf{r}$. If I_{enc} is the total current enclosed by the loop, then we have

$$\oint_C \mathbf{B} \cdot d\mathbf{r} = \mu_0 I_{\text{enc}} \tag{3.26}$$

But if we draw a *simple* surface S which is bounded by the curve C, then the current passing through C is the same as the integral of the *current density* $\mathbf{J}$ over this surface:

$$I_{\text{enc}} = \int_S \mathbf{J} \cdot d\mathbf{S}$$

Also, by Stokes' Theorem, the line integral in 3.26 can be made into a surface integral on S:

$$\oint_C \mathbf{B} \cdot d\mathbf{r} = \int_S (\nabla \times \mathbf{B}) \cdot d\mathbf{S}$$

Combining these gives

$$\int_S (\nabla \times \mathbf{B}) \cdot d\mathbf{S} \;=\; \mu_0 \int_S \mathbf{J} \cdot d\mathbf{S}$$

The fact that this holds for *any* surface S lets us use an argument similar to when we used for Gauss' law to conclude that the integrands here are themselves equal at all points in space; thus

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$$

which we might call Ampere's law in differential form.

A word of caution is needed here; the preceding discussion is only valid for *static* (i.e. constant) currents and charge densities, i.e. what we call **electrostatics**. If either of these changes and then original expression of Ampere's law needs changing and then so does the differential relation. The more general version is one of the (true) Maxwell equations.

3.17 The Dirac Delta Function

I had this section here because the delta function is now becoming a standard part of the undergraduate curriculum and because one of its uses is to resolve an "inconsistency" in the use of the divergence theorem. I will follow the approach of Griffiths in him EM book. So you'll see it again, but it will be easier then!

3.17.1 A Paradox?

Consider the vector field (in spherical coordinates)

$$\mathbf{F}(\mathbf{r}) = \frac{\hat{\mathbf{r}}}{r^2}$$

if we integrate this field over the surface of sphere of radius R (centered at the origin) we get

$$\oint_S \mathbf{F} \cdot d\mathbf{S} = \int_0^{2\pi} \int_0^{\pi} \frac{1}{r^2} r^2 \, dr \, \sin\theta \, d\theta \, d\phi \Big|_{r=R} = 4\pi \tag{3.27}$$

But if we take the divergence (and throw away all caution) we get

$$\nabla \cdot \mathbf{F} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{1}{r^2} \right) = \frac{1}{r^2} \cdot 0 = 0 \qquad (!!!)$$

and the integral of the divergence over the interior gives

$$\int_V (\nabla \cdot \mathbf{F}) \, dV = 0 \qquad (!!!) \tag{3.28}$$

But by the divergence theorem, the two integrals in 3.27 and 3.28 should be equal, shouldn't they? So is 4π really equal to zero?

The problem is in the volume integral in 3.28. The divergence of $\mathbf{F}$ is indeed zero *everywhere except the origin* where the original function and its derivatives are really undefined. That's only one point but it's enough to render the divergence theorem unusable; the functions have to be nicely behaved *everywhere* in the interior, so this is not really a paradox, merely a bad experience from not listening to the mathematicians. We will find that we can get out divergence theorem back with a bit of unconventional mathematics.

3.17.2 A Sequence of Functions

We consider the set of functions defined by

$$f_a(x) = \begin{cases} 0 & |x| > \frac{1}{a} \\ \frac{a}{2} & |x| < \frac{1}{a} \end{cases}$$

This function is a step of width $\frac{2}{a}$ and height $\frac{a}{2}$ so that the area under the curve is 1 for any value of a, but as we consider increasing values of a the function becomes taller and thinner. What happens in the limit as $a \to \infty$?

Loosely speaking, the function turns into a spike which has infinite height, zero width but maintains its unit normalization; actually, the function has *no* limit. But when the function is used inside an integral, it does. Consider the integral of our $f_a(x)$'s times some *well–behaved* function $g(x)$:

$$\lim_{a\to\infty} \int_{-\infty}^{\infty} g(x) f_a(x)\, dx$$

As a gets very large and $f_a(x)$ becomes very thin, the only contributions to the integral come from values of x close to 0 (that is, from $g(0)$). In the limit, the factor $g(x)$ can be taken *outside* the integral as the constant $g(0)$ leaving:

$$\lim_{a\to\infty} \int_{-\infty}^{\infty} g(x) f_a(x)\, dx = g(0) \lim_{a\to\infty} \int_{-\infty}^{\infty} f_a(x)\, dx = g(0)$$

If we give the limit of the set of functions a name:

$$\delta(x) \equiv \lim_{a\to\infty} f_a(x)$$

then we can write

$$\int_{-\infty}^{\infty} g(x)\delta(x)\, dx = g(0)$$

We can put the long thin spike anywhere else,

$$\delta(x-c) = \lim_{a\to\infty} f_a(x-c)$$

and we then find

$$\int_{-\infty}^{\infty} g(x)\delta(x-c)\, dx = g(c)$$

so that the "function" $\delta(x-c)$ picks out the value of $g(x)$ at $x = c$. More generally,

$$\int_a^b g(x)\delta(x-c)\, dx = \begin{cases} g(c) & \text{if } b < c < a \\ 0 & \text{otherwise} \end{cases}$$

3.17.3 Three–Dimensional Delta Function

Generally we will be integrating in more than one dimension, with functions of $\mathbf{r}$ (3 coordinates). In Cartesian coordinates, the "function" which picks out the value a (well–behaved) function at the position $\mathbf{c}$ is

$$\delta^3(\mathbf{r}-\mathbf{c}) = \delta(x-c_x)\delta(y-c_y)\delta(z-c_z)$$

which gives

$$\int_V g(\mathbf{r})\delta^3(\mathbf{r}-\mathbf{c})\, dV = \begin{cases} g(\mathbf{c}) & \text{if } V \text{ includes } \mathbf{c} \\ 0 & \text{otherwise} \end{cases}$$

3.17.4 Resolving the Paradox

Now we return to the suspect volume integral of Eq. 3.28. We want to patch it up so that it gives 4π in accordance with the divergence theorem. We really need to patch up the result for $\nabla \cdot \frac{\hat{\mathbf{r}}}{r^2}$. It is zero everywhere but at the origin, where it is not defined. But we can (using the delta function) we make a definition for all points of the interior. If we define:

$$\nabla \cdot \frac{\hat{\mathbf{r}}}{r^2} = 4\pi\delta^3(\mathbf{r}) \tag{3.29}$$

then we get (for the sphere of radius R or in fact any volume V which includes the origin):

$$\int_V (\nabla \cdot \frac{\hat{\mathbf{r}}}{r^2})\, dV = \int 4\pi\delta^3(\mathbf{r})\, dV = 4\pi$$

Chapter 4

Complex Variables

We have covered the most basic facts about complex numbers in the *Preliminaries* section. In this part of the class we explore functions of complex variables and then doing calculus (derivatives and integrals of various sorts) with these functions.

One might think that we just have a simple extension of the basic calculus of real functions, but that turns out not to be the case. Again we need to re-do first–year calculus!

4.1 Complex Numbers: Basic Operations

A complex number is formed using the "imaginary" number i, which has the property $i^2 = -1$. (It is also true that $(-i)^2 = -1$.) For real numbers x and y, a complex number is

$$z = x + iy$$

and we write $x = \mathrm{Re}(z)$ and $y = \mathrm{Im}(z)$. The **complex conjugate** of a complex number z, denoted z^* (and in some books by $\bar{z}$), is the same as z but with the sign of the imaginary part changed:

$$z = x + iy \qquad \Longrightarrow \qquad z^* = x - iy$$

The modulus of z, denoted by $|z|$, is

$$z = x + iy \qquad \Longrightarrow \qquad |z| = \sqrt{x^2 + y^2}$$

The sum and product of the complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ are

$$z_1 + z_2 = (x_1 + x_2) + i(y_2 + y_2) \qquad\qquad z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_1 + x_2 y_2)$$

which gives

$$z z^* = (x + iy)(x - iy) = x^2 + y^2 = |z|^2$$

The complex conjugate operation "distributes" over sums and products:

$$(z_1 \pm z_2)^* = z_1^* \pm z_2^* \qquad (z_1 z_2)^* = z_1^* z_2^* \qquad \left(\frac{z_1}{z_2}\right)^* = \frac{z_1^*}{z_2^*}$$

All complex numbers (except for 0) have a multiplicative inverse so that (as usual) division just means multiplying by the inverse. When dividing, it is often useful to get the result in the form $x + iy$. This can be done by multiplying top and bottom by the complex conjugate of the denominator:

$$\frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2} \cdot \frac{x_2 - iy_2}{x_2 - iy_2} = \frac{(x_1 x_2 + y_1 y_2) + i(x_2 y_1 - x_1 y_2)}{x_2^2 + y_2^2}$$

4.1.1 Polar Representation; Relation to Trig Functions

The exponential function has a important role in the theory of complex variables. It is defined by:

$$e^z = \exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \tag{4.1}$$

where $0! = 1$. One can show that the sum in 4.1 converges for any z. Two more properties of *real* exponentials can also be shown to hold here,

$$e^{z_1} e^{z_2} = e^{z_1+z_2} \qquad \text{and} \qquad (e^z)^n = e^{nz}$$

for positive integers n.

If z is purely imaginary, say $z = i\theta$ with θ real, one can show that the sum in 4.1 gives

$$e^{i\theta} = \left(1 - \tfrac{\theta^2}{2!} + \tfrac{\theta^4}{4!} + \cdots\right) + i\left(\tfrac{\theta}{1} - \tfrac{\theta^3}{3!} + \tfrac{\theta^5}{5!}\cdots\right)$$

which we *define* as

$$e^{i\theta} = \cos\theta + i\sin\theta$$

so that $\cos\theta$ and $\sin\theta$ now have firm mathematical definitions as

$$\cos\theta \equiv 1 - \tfrac{\theta^2}{2!} + \tfrac{\theta^4}{4!} + \cdots \qquad\qquad \sin\theta \equiv \tfrac{\theta}{1} - \tfrac{\theta^3}{3!} + \tfrac{\theta^5}{5!}\cdots$$

From this, multiplying by the real number r gives

$$re^{i\theta} = r(\cos\theta + i\sin\theta) = r\cos\theta + ir\sin\theta = x + iy$$

so that any complex number can be represented as

$$z = re^{i\theta} \qquad \text{with} \quad r = |z| \quad \text{and} \quad \tan\theta = \frac{y}{x}$$

and we choose the proper quadrant for θ; for now, choose $-\pi < \theta \leq \pi$. We note that without some kind of restriction, there is no single value for the exponent, since

$$z = re^{i\theta} = re^{i(\theta+2n\pi)} \; ;$$

this ambiguity will give us major headaches shortly.

In polar form, multiplication and division are easy; for $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$,

$$z_1 z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)} \qquad \text{and} \qquad \frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)}$$

With the above properties of the complex exponential, we very easily get, for real θ and integer n,

$$e^{in\theta} = (\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$$

a result which sometimes is called "De Moivre's theorem". For $n = 2$ it gives us

$$(\cos^2\theta - \sin^2\theta) + i(2\sin\theta\cos\theta) = \cos 2\theta + i\sin 2\theta$$

which gives the double–angle identities

$$\cos 2\theta = \cos^2\theta - \sin^2\theta \qquad\qquad \sin 2\theta = 2\sin\theta\cos\theta$$

In fact, many of the identities of trigonometry can be shown much more easily using complex exponentials.

It is true that for *real* x, and *complex* c one can show:

$$\frac{d}{dx}e^{cx} = ce^{cx}$$

(a fact which we'll use below in working out an integral) but as for taking a derivative with respect number z, as in $\frac{d}{dz}e^{cz}$ we will have to be much more careful about what this operation means; more on that later!

One can show that the equation $z^n = 1$ (for positive integer n) has n roots, given by

$$z = e^{2\pi ik/n} \qquad \text{where} \quad k = 0, 1, 2, \ldots, n-1$$

4.1.2 Applications to Basic Integrals

Example: Evaluate the integral $I = \int e^{ax}\cos bx\, dx$.

While the standard techniques can be of use here, we will use complex variable to get it into a form so that we just need to know one of the "elementary forms".

We note that we can write the cosine in the integrand as the Real part of a complex exponential; this gives

$$e^{ax}\cos bx = \mathrm{Re}[e^{ax}(\cos bx + i\sin bx)] = \mathrm{Re}[e^{ax}e^{ibx}] = \mathrm{Re}\left[e^{(a+ib)x}\right]$$

With this substitution, the integral is:

$$I = \int e^{ax}\cos bx\, dx = \mathrm{Re}\int e^{(a+ib)}x\, dx = \mathrm{Re}\left(\frac{e^{a+ib}}{(a+ib)}\right) + C$$

and we're done! But we want to do some algebra to get a more convenient form:

$$I = \mathrm{Re}\left(\frac{e^{a+ib}}{(a+ib)}\cdot\frac{(a-ib)}{(a-ib)}\right) = \frac{1}{a^2+b^2}\mathrm{Re}\left((a-ib)e^{ax}(\cos bx + i\sin bx)\right)$$

from which we can read off the real part of the stuff inside the parentheses to get:

$$I = \frac{e^{ax}}{a^2+b^2}\left(a\cos bx + b\sin bx\right) + C$$

Had we taken the Imaginary parts we would have arrived at:

$$\int e^{ax}\sin bx\, dx = \frac{e^{ax}}{a^2+b^2}(-b\cos bx + a\sin bx)$$

4.1.3 Hyperbolic Functions — A Review, Maybe

Since we will see a close relation between the trig and hyperbolic functions when we consider functions of complex variables, we will do a review of the hyperbolic functions (of real variables). If scant attention was paid to them in your calculus class, this review will be useful!

The hyperbolic functions $\sinh x$, $\cosh x$, etc. are defined as:

$$\cosh x = \frac{e^x + e^{-x}}{2} \qquad \sinh x = \frac{e^x - e^{-x}}{2} \qquad \tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4.2}$$

with the others given by

$$\operatorname{sech} x = \frac{1}{\cosh x} \qquad \operatorname{csch} x = \frac{1}{\sinh x} \qquad \coth x = \frac{1}{\tanh x} = \frac{\cosh x}{\sinh x}$$

The hyperbolic functions obey some identities which look very much like corresponding ones for the trig functions; watch out though, because there is usually a difference of sign someplace! Some of these include:

$$\cosh^2 x - \sinh^2 x = 1 \qquad 1 - \tanh^2 x = \operatorname{sech}^2 x \qquad \coth^2 x - 1 = \operatorname{csch}^2 x$$

$$\sinh(x_1 \pm x_2) = \sinh x_1 \cosh x_2 \pm \cosh x_1 \sinh x_2 \qquad \cosh(x_1 \pm x_2) = \cosh x_1 \cosh x_2 \pm \sinh x_1 \sinh x_2$$

$$\tanh(x_1 \pm x_2) = \frac{\tanh x_1 \pm \tanh x_2}{1 \pm \tanh x_1 \, \tanh x_2}$$

The inverses of the hyperbolic functions are denoted by

$$\sinh^{-1} x, \qquad \cosh^{-1} x, \qquad \tanh^{-1} x\,, \quad \text{etc.}$$

Here, $y = \sinh^{-1} x$ means $x = \sinh y$. Note, for the inverse cosh function one must make a choice of sign; we choose $\cosh^{-1} x \geq 0$.

It is possible to get a "closed form" for the inverse hyperbolic functions. One can show:

$$\cosh^{-1} x = \ln(x + \sqrt{x^2 - 1}) \qquad \sinh^{-1} x = \ln(x + \sqrt{x^2 + 1}) \qquad \tanh^{-1} x = \tfrac{1}{2} \ln\left(\frac{1+x}{1-x}\right)$$

The derivatives of the hyperbolic functions are similar to those of the trig functions, but as always, watch the signs!

$$\frac{d}{dx}(\cosh x) = \sinh x \qquad \frac{d}{dx}(\sinh x) = \cosh x \qquad \frac{d}{dx}(\tanh x) = \operatorname{sech}^2 x$$

$$\frac{d}{dx}\left(\cosh^{-1}\frac{x}{a}\right) = \frac{1}{\sqrt{x^2 - a^2}} \qquad \frac{d}{dx}\left(\sinh^{-1}\frac{x}{a}\right) = \frac{1}{\sqrt{x^2 + a^2}} \qquad \frac{d}{dx}\left(\tanh^{-1}\frac{x}{a}\right) = \frac{a}{a^2 - x^2}$$

4.1.4 Complex Powers and Logarithms

We note that for real variables, the equation $y = e^x$ can be inverted so that for a given y it has *one* solution $x = \ln y$. But we found out that different (complex) exponents of e can give the same number, namely those which differ by multiples of $2\pi i$:

$$z = re^{i\theta} = re^{i(\theta + 2\pi)} = re^{i(\theta + 4\pi)} = \ldots$$

Because of this, the complex exponential does *not* invert so easily!

4.2 Functions of Complex Variables

4.2.1 Notation

Just as with functions of real variables we let the independent variable be called "x", the dependent variable called "y" with $y = f(x)$, the most common choice of letters for functions of complex variables is z for the independent variable and the dependent variable called w with $w = f(z)$.

As usual, $z = x+iy$ and w will also have real and imaginary parts, for which we write: $w = u+iv$.

4.2.2 The Elementary Functions

$$\sin(iz) = i\sinh z \qquad \cos(iz) = \cosh z \qquad \tan(iz) = i\tanh z$$
$$\sinh(iz) = i\sin z \qquad \cosh(iz) = \cos z \qquad \tanh(iz) = i\tan z$$

$$\sin^{-1}(z) = \frac{1}{i}\ln(iz + \sqrt{1-z^2}) \qquad \cos^{-1}(z) = \frac{1}{i}\ln(z + \sqrt{z^2-1}) \qquad \tan^{-1}(z) = \frac{1}{2i}\ln\left(\frac{1+iz}{1-iz}\right)$$

Examples: Evaluating functions of complex variables

Some examples I came up with to illustrate how your calculator *knows* what to do with complex variables.

(i) Find $\cos(2\pi + i)$.

Use the definition of the cosine function,

$$\cos z \equiv \frac{e^{iz} + e^{-iz}}{2} \qquad \Longrightarrow \qquad \cos(2\pi + i) = \frac{e^{i(2\pi+i)} + e^{-i(2\pi+i)}}{2}$$

In the numerator of the last expression we can pull out the factors

$$e^{2\pi i} = 1 \qquad \text{and} \qquad e^{-2\pi i} = 1$$

and combining the factors of i in the exponents gives

$$\Rightarrow = \frac{e^{-1} + e^{1}}{2} = \tfrac{1}{2}\left(e + \frac{1}{e}\right) = \frac{e^2+1}{2e} \approx 1.543$$

(ii) Find $\cosh(2\pi + i)$.

Here it might be easiest to use the "angle addition" rule for cosh,

$$\cosh(z_1 + z_2) = \cosh z_1 \cosh z_2 + \sinh z_1 \sinh z_2$$

to get

$$\cosh(2\pi + i) = \cosh(2\pi)\cosh(i) + \sinh(2\pi)\sinh(i) = \cos(1)\cosh(2\pi) + i\sin(1)\sinh(2\pi)$$

where in the last step we've used

$$\cos(iz) - \cos(z) \qquad \text{and} \qquad \sinh(iz) = i\sin z$$

In the last form we can use a $10 calculator to get

$$\cosh(2\pi + i) \approx 1.45 + i\,2.25$$

(iii) Find $\ln(i+1)$.

Write $1+i$ in polar form:

$$1 + i = \sqrt{2}\, e^{i(\frac{\pi}{4}+2\pi k)}$$

so the multi-valued ln function is

$$\ln\left(\sqrt{2}\, e^{i\frac{\pi}{4}+2\pi k}\right) = \mathrm{Ln}\,(\sqrt{2}) + i(\frac{\pi}{4} + 2\pi k) \qquad k = 0, 1, 2, \ldots$$

As our calculator will choose the branch with $k = 0$, this gives

$$\ln(1+i) = \mathrm{Ln}\,(\sqrt{2}) + i\frac{\pi}{r} \approx 0.3466 + i\,0.7854$$

(iv) Find $\sin^{-1}(17)$.

Here, use the "closed–form" form formula for the $\sin^{-1}$ function. Get:

$$\sin^{-1}(17) = \frac{1}{i}\ln(i17 + \sqrt{1-(17)^2}) = \frac{1}{i}\ln(i17 + i16.97)$$

You'll note here that chose the positive sign to go with i for the square root of the negative number. The other choice would give another "branch"; this choice is the one our calculator takes! Reducing it further:

$$\Longrightarrow \qquad = \frac{1}{i}\ln(33.97\,i) = \frac{1}{i}\ln(33.97e^{i\pi/2} = \frac{1}{i}\left[\mathrm{Ln}\,(33.97) + i\frac{\pi}{2}\right]$$

where for the log we choose the principal value, again as our calculator will do. Finally,

$$\Longrightarrow \qquad = \frac{\pi}{2}\mathrm{Ln}\,(33.97) \approx 1.57 - i\,3.52$$

(v) Find 2^π. Comment on the difference with $2^3 = 8$.

From the definition of powers, we have

$$2^\pi = e^{\pi \ln 2}$$

But $\ln 2$ is multivalued,

$$\ln 2 = \ln(2e^{i2\pi k}) = \mathrm{Ln}\,2 + i2\pi k \quad k = 0, 1, 2, \ldots$$

This gives

$$2^\pi = e^{\pi[\mathrm{Ln}\,2 + i2\pi k]} = e^{\pi \mathrm{Ln}\,2} e^{i2\pi^2 k} \approx (8.82)[\cos(2\pi^2 k) + i\sin(2\pi^2 k)]$$

which has many different messy values.

So why doesn't 2^3 give the same problems? Solving that problem in the same way gives

$$2^3 = e^{3\mathrm{Ln}\,2} e^{i2(3)\pi k} = e^{3\mathrm{Ln}\,2} e^{i6\pi k} = e^{3\mathrm{Ln}\,2} \cdot 1 = 8$$

The *integer* power gives only one answer.

4.2.3 Basic Types of Functions

4.2.4 Branch Points and Branch Cuts

The simplest example of the complications which arise in multi-valued functions (and what to do about it) is the function $f(z) = z^{1/2}$, which among friends can be written $f(z) = \sqrt{z}$. We already know that here there is a rather arbitrary choice to make for the sign of the function. For real variables we specify that we take the *positive* root of the equation $y^2 = x$ when we write $y = \sqrt{x}$ but for complex variables we always want to consider the entire complex plane so the choice is not so clear.

We want a function of z to change in a continuous way when z is changed continuously. We consider a trip around the origin in the z plane and the corresponding path in the w plane.

4.3 Derivatives of Complex Functions

What does it mean to take the derivative of the function $f(z)$, that is, what do we mean by $f'(z)$? As with real variables, it *does* mean to take the limit of the ratio

$$\frac{f(z + \Delta z) - f(z)}{\Delta z} \tag{4.3}$$

as Δz "gets small", but more precision in the definition is required for complex variables, because Δz can get small in many ways (whereas the real Δx has only one way to get small). It might have only a real part and shrink *that* or just an imaginary part or shrink both parts at once.

If the limit in 4.3 exists *irregardless*, I say *irregardless* of the way that Δz goes to zero then we say that $f(z)$ is **differentiable** at z. If $f(z)$ is differentiable at all points in some region $\mathcal{R}$ of the complex plane, we say that $f(z)$ is **analytic** in $\mathcal{R}$.

4.3.1 Cauchy–Riemann Equations

One can show that if $w = f(z) = u(x, y) + iv(x, y)$ is analytic in some region $\mathcal{R}$ then in $\mathcal{R}$ u and v will satisfy the **Cauchy–Riemann equations**,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad\qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{4.4}$$

4.3.2 Rules for Differentiation

Fortunately, if we have a set of functions which *are* differentiable we can use the familiar rules for derivatives that we know from elementary calculus. These now look like:

$$\frac{d}{dz}\{f(z) + g(z)\} \;=\; \frac{d}{dz}f(z) + \frac{d}{dz}g(z) \;=\; f'(z) + g'(z)$$

$$\frac{d}{dz}\{cf(z)\} \;=\; c\frac{d}{dz}f(z) = cf(z)$$

$$\frac{d}{dz}\{f(z)g(z)\} \;=\; f(z)\frac{d}{dz}g(z) + g(z)\frac{d}{dz}f(z) \;=\; f(z)g'z) + g(z)f'(z)$$

$$\text{If} \quad w = f(\zeta) \quad \text{and} \quad \zeta = g(z) \quad \text{then} \quad \frac{dw}{dz} = \frac{dw}{d\zeta} \cdot \frac{d\zeta}{dz} = f'(g(z))\, g'(z)$$

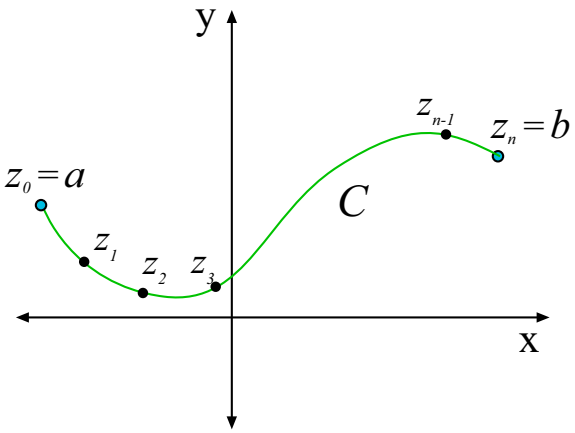

Figure 4.1: A complex line integral.

It is also true that the derivatives of the elementary functions of complex variables have the same form as the ones we know from elementary (real–variable) calculus. For example,

$$\frac{d}{dz}z^n = nz^{n-1} \qquad \frac{d}{dz}e^z = e^z \qquad \frac{d}{dz}a^z = a^z \ln a$$

$$\frac{d}{dz}\sin z = \cos z \qquad \frac{d}{dz}\cos z = -\sin z \qquad \frac{d}{dz}\tan z = \sec^2 z$$

4.3.3 Singular Points

If a function $f(z)$ is not analytic at a point, the point is called a **singular point** or **singularity** of the function. We can distinguish various types of singularities (that is, give the *reasons* why the function fails to be analytic). For isolated singularities, the types are:

- **Poles**: These occur basically because there is a polynomial in a denominator which is zero at some point. We can distinguish the **order** of the pole; if we find an integer n such that

$$\lim_{z\to z_0}(z - z_0)^n f(z) = A \neq 0$$

then the point z_0 is a **pole of order** n. If $n = 1$, the point is called a **simple pole**.

4.4 Complex Integration; Cauchy's Theorem

We consider a curve C in the complex (z) plane with endpoints a and b. With $z_0 \equiv a$ and $z_n \equiv b$, This process is illustrated in Fig. 4.1

From this curve we select a sequence of n points

$$z_0 = a,\ z_1,\ z_2, \ldots z_{n-1},\ z_n = b$$

and evaluate the sum

$$\sum_{k=1}^{n} f(\xi_k)(z_k - z_{k-1}) = \sum_{k=1}^{n} f(\xi_k)\Delta z_k$$

We then consider the limit where $n \to \infty$ and the size of the largest increment Δz_k goes to zero. If this limit exists, it defines the line integral of $f(z)$ along the path C from a to b:

$$\int_C f(z)\,dz = \lim_{\substack{n\to\infty \\ \Delta z\to 0}} \sum_{k=1}^{n} f(\xi_k)\Delta z_k$$

Cauchy's theorem says that if $f(z)$ is analytic in a region$\mathcal{R}$ and on its boundary C, then

$$\oint_C f(z)\,dz = 0$$

Yet one more surprising property of analyticity!

It follows from Cauchy's theorem that if a and b are any two points in $\mathcal{R}$ then

$$\int_a^b f(z)\,dz$$

is *independent* of the path in $\mathcal{R}$ joining z and b.

It also follows that if a and z are any two points in $\mathcal{R}$ and

$$G(z) = \int_a^z f(\zeta)\,d\zeta$$

then $G(z)$ is analytic in $\mathcal{R}$ and $G'(z) = f(z)$.

A more useful statement of this property is: If a and b are any two points in $\mathcal{R}$, and $F'(z) = f(z)$, then

$$\int_a^b f(z)\,dz = F(b) - F(a)$$

so that when the integrand is nice and analytic in the region of interest, integration is just as easy as it is for real variables.

4.5 Cauchy's Integral Formula

$$f(a) \;=\; \frac{1}{2\pi i}\oint_C \frac{f(z)}{z-a}\,dz \tag{4.5}$$

$$f^{(n)}(a) \;=\; \frac{n!}{2\pi i}\oint_C \frac{f(z)}{(z-a)^{n+1}}\,dz \qquad n = 1, 2, 3, \ldots \tag{4.6}$$

4.6 Infinite Series; Taylor and Laurent Series

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots \qquad \cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots$$

$$\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots \qquad \tan^{-1} z = z - \frac{z^3}{3} + \frac{z^5}{5} - \cdots$$

$$(1+z)^p = 1 + pz + \frac{p(p-1)}{2^!}z^2 + \cdots + \frac{p(p-1)\cdots(p-n-1)}{n!}z^n + \cdots$$

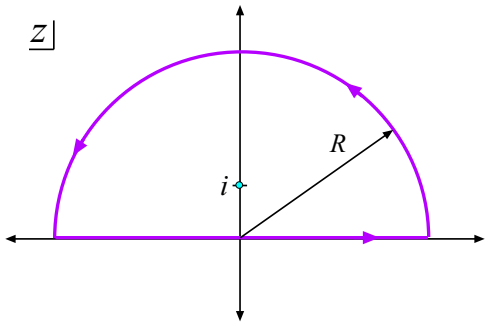

Figure 4.2: Contour for the integral $\int_C \frac{1}{x^2+1}\,dz$. Contour encloses a pole at $z = i$.

4.7 The Residue Theorem

The residue associated with the singularity a (which in our examples is nearly always a pole of some order) is the coefficient of $\frac{1}{(z-a)}$ in the Laurent series about a. In general we can find the residue using the general formula:

$$a_{-1} \;=\; \lim_{z\to a} \frac{1}{(k-1)!}\frac{d^{k-1}}{dz^{k-1}}\{(z-a)^k f(z)\} \tag{4.7}$$

For a pole of order 1 we can use the simpler formula

$$a_{-1} \;=\; \lim_{z\to a}(z-a)f(z) \tag{4.8}$$

The **residue theorem** says that if $f(z)$ is single-valued and analytic inside a simple closed curve C except at the singularities a, b, c, ... inside C which have residues a_{-1}, b_{-1}, c_{-1}, ... then

$$\oint_C f(z)\,dz \;=\; 2\pi i(a_{-1} + b_{-1} + c_{-1} + \cdots) \tag{4.9}$$

Example: Do the integral

$$\int_0^\infty \frac{1}{x^2+1}\,dx$$

using the residue theorem.

This one of course is elementary and can done without complex variables, but it will provide a check on the method. First, we note that since the integrand is an even function of x, we have

$$\int_0^\infty \frac{1}{x^2+1} = \frac{1}{2}\int_{-\infty}^\infty \frac{1}{x^2+1}$$

and the latter integral can be expanded into a contour integral $\oint_C \frac{1}{z^2+1}\,dz$ which runs along the real axis (from $-R$ to R) and then returns along a huge semi-circle of radius R, as shown in Fig. 4.2. The big semi-circle is not part of the original integral and will have to be dealt with later.

The integrand, $\frac{1}{z^2+1}$, has poles at $\pm i$, and clearly they are *simple* poles. Only the pole at $z = i$ lies inside our contour so we just need to find the residue there. The simple-pole-residue formula gives

$$a_{-1} = \lim_{z\to i}\,(z-i)\frac{1}{z^2+1} = \lim_{z\to i}\,\frac{1}{2z} = \frac{1}{2i}$$

then the residue theorem gives

$$\oint_C \frac{1}{z^2+1}\,dz = 2\pi i(a_{-1}) = 2\pi i\frac{1}{2i} = \pi$$

But we now have to deal with the curved path; we will do it in detail here and not go into as much detail in the other examples. On the curvy part we have $z = Re^{i\theta}$ with $\theta : 0 \to \pi$. On this part of the path,

$$f(z) = \frac{1}{R^2e^{2i\theta}+1} \qquad |f(z)| = \left|\frac{1}{R^2e^{2i\theta}+1}\right| = \frac{1}{|R^2e^{2i\theta}+1|}$$

In the denominator of the last expression, use an inequality for complex numbers

$$|z_1 + z_2| \geq |z_1| - |z_2|$$

to make the denominator *smaller* and hence make to whole expression *bigger*:

$$|f(z)| \;\leq\; \frac{1}{|R^2e^{2i\theta}+1|} \;\leq\; \frac{1}{|R^2e^{2i\theta}|-1} \;=\; \frac{1}{R^2-1}$$

At this point we realize that in the limit $R \to \infty$ (big numbers) $R^2/2$ is certainly less than $R^2 - 1$, so that our inequality is now

$$|f(z)| < \frac{1}{R^2/2} = \frac{2}{R^2}$$

and, calling the curvy path Γ, the length of Γ is πR and the integral on Γ is bounded by

$$\int_\Gamma f(z)\,dz \;\leq\; |f(z)|_{\max}L_\Gamma \;\leq\; \frac{2}{R^2}\pi R \;=\; \frac{2\pi}{R}$$

which vanishes as $R \to \infty$. Thus the integral on the curved part contributes nothing and we are done with the problem. We have shown

$$\int_0^\infty \frac{1}{x^2+1}\,dx = \frac{1}{2}\int_{-\infty}^\infty \frac{1}{x^2+1}\,dx = \frac{1}{2}\oint \frac{dz}{z^2} = \tfrac{1}{2}\pi = \frac{\pi}{2}$$

Is that right? You bet it is:

$$\int_0^\infty \frac{1}{x^2+1}\,dx = \tan^{-1}x\Big|_0^\infty = \frac{\pi}{2} - 0 = \frac{\pi}{2}$$

But of course the new method will give results where we *don't* know how to do the integral by elementary means.

Example: Do the integral

$$\int_0^\infty \frac{1}{x^4+1}\,dx$$

Here the the integrand is *not* an elementary form and though one might be able to do a complex factorization of the denominator and use "partial fractions", this may not be worth the effort, and in fact this one *is* probably easier to do by contour integration.

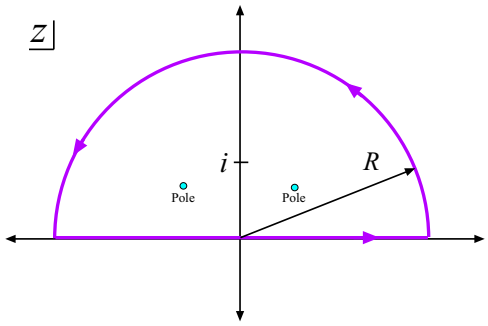

Figure 4.3: Contour for the integral $\int_C \frac{1}{x^4+1}\,dz$. Contour encloses poles at $z = e^{\pi i/4}$ and$z = e^{3\pi i/4}$.

Again, since the integrand is an even function of x we will use

$$\int_0^\infty \frac{1}{x^4+1}\,dx = \frac{1}{2}\int_{-\infty}^\infty \frac{1}{x^4+1}\,dx$$

with the intention of extending the second integral to a contour in the complex plane.

The integrand has poles at the values of z where the denominator is zero. This occurs where $z^4 = -1$, or

$$z \;=\; e^{\pi i/4}, \quad e^{3\pi i/4}, \quad e^{5\pi i/4}, \quad e^{7\pi i/4}$$

of which the first two are in the upper half–plane. If we enlarge the contour to include a large semi-circle in the upper half–plane to close the path, we note that it encloses *two* poles. This is shown in Fig. 4.3. Again, if we can deal with (or better, ignore) the integral along the curvy part then the residue theorem will give us the answer.

Calculate the residues at the two poles:

Pole at $e^{\pi i/4}$: It is clearly a simple pole, so use

$$a_{-1} \;=\; \lim_{z\to e^{\pi i/4}} (z - e^{\pi i/4})\frac{1}{z^4+1} \;=\; \lim_{z\to e^{\pi i/4}} \frac{1}{4z^3} \;=\; \frac{1}{4e^{3\pi i/4}}$$

Pole at $e^{3\pi i/4}$: Again a simple pole, so use

$$a_{-1} \;=\; \lim_{z\to e^{3\pi i/4}} (z - e^{3\pi i/4})\frac{1}{z^4+1} \;=\; \lim_{z\to e^{3\pi i/4}} \frac{1}{4z^3} \;=\; \frac{1}{4e^{9\pi i/4}} \;=\; \frac{1}{4e^{\pi i/4}}$$

Note, in both cases, l'Hospital's rule was used for the limit. Using our results for the residues at the two poles, the residue theorem gives the value of the contour integral on C:

$$\oint_C \frac{1}{z^4+1}\,dz = 2\pi i \sum a_{-1} = 2\pi i \left(e^{-3\pi i/4} + e^{-\pi i/4}\right)$$

Then a little trig gives

$$\oint_C \frac{1}{z^4+1}\,dz = 2\pi i \tfrac{1}{4}\left(-\tfrac{1}{\sqrt{2}} - i\tfrac{1}{\sqrt{2}} + \tfrac{1}{\sqrt{2}} - i\tfrac{1}{\sqrt{2}}\right) = \frac{i\pi}{2}(-i\sqrt{2}) = \frac{\pi}{\sqrt{2}}$$

We have to worry about what happens on the curvy part, but in fact because the denominator contains an even higher power of z than the last example, the integral will vanish in the limit

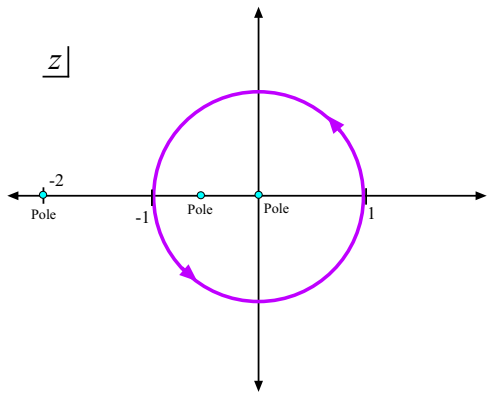

Figure 4.4: Contour for integration example and the poles of the integrand. The poles are at $z = 0$, $z = -\frac{1}{2}$ and $z = -2$.

$R \to \infty$ for the same reason. So the integral on the *whole* contour equals the integral just along the real axis.

Finally the value of the original integral is

$$\int_0^\infty \frac{1}{x^4+1}\,dx \;=\; \tfrac{1}{2}\int_{-\infty}^\infty \frac{1}{x^4+1}\,dx \;=\; \tfrac{1}{2}\oint_C \frac{1}{z^4+1}\,dz \;=\; \tfrac{1}{2}\frac{\pi}{\sqrt{2}} \;=\; \frac{\pi}{2\sqrt{2}}$$

Example: Do the integral

$$\int_0^{2\pi} \frac{\cos 3\theta}{5+4\cos\theta}\,d\theta$$

The trick here is to make the real integral into a contour integral by changing the integration over $\theta : 0 \to 2\pi$ into a contour around the unit circle via the substitution

$$z = e^{i\theta} \qquad \Longrightarrow \qquad dz = ie^{i\theta}\,d\theta = iz\,d\theta$$

We will also need

$$\cos\theta = \frac{z+z^{-1}}{2} \qquad \text{and} \qquad \cos 3\theta = \frac{e^{3i\theta}+e^{-3i\theta}}{2} = \frac{z^3+z^{-3}}{2}$$

and the path of integration C in the complex plane is the unit circle, as shown in Fig. 4.4. Making these substitutions, we get

$$\int_0^{2\pi} \frac{\cos 3\theta}{5+4\cos\theta}\,d\theta \;=\; \oint_C \frac{(z^3+z^{-3})/2}{5+2(z+z^{-1})}\frac{dz}{iz} \;=\; \frac{1}{2i}\oint_C \frac{z^3+z^{-3}\,dz}{(5z+2z^2+2)}$$

Then, factoring the denominator and multiplying top and bottom by z^3 gives

$$\int_0^{2\pi} \frac{\cos 3\theta}{5+4\cos\theta}\,d\theta \;=\; \frac{1}{2i}\oint_C \frac{(z^6+1)\,dz}{z^3(2z+1)(z+2)}$$

which can be evaluated by the residue theorem.

The denominator in the integrand has zeroes at

$$z = 0, \qquad z = -\tfrac{1}{2} \qquad z = -2$$

so there are poles of the integrand at these points (as shown in Fig. 4.4) but only the first two lie within the contour C, so it is only at these points that we need to find the residues.

Pole at $z = -\frac{1}{2}$: This is a pole of order 1 and we find the residue from:

$$a_{-1} = \lim_{z\to -\frac{1}{2}} (z+\tfrac{1}{2}) \frac{(z^6+1)}{z^3(2z+1)(z+2)} = \lim_{z\to -\frac{1}{2}} \frac{(z^6+1)}{z^3(2)(z+2)}$$

At this point we can just plug in the value, to get

$$a_{-1} = \frac{\frac{1}{64}+1}{\left(-\frac{1}{8}\right)(2)(3/2)} = -\frac{65}{24}$$

Pole at $z = 0$: This is a pole of order 3, so we have to do a little more work. The residue formula gives

$$a_{-1} = \lim_{z\to 0} \frac{1}{2}\frac{d^2}{dz^2}\left[\frac{z^3(z^6+1)}{z^3(2z+1)(z+2)}\right]$$

We just have to do the grueling work of taking the derivatives. Omitting some tedious steps, we get:

$$a_{-1} = \lim_{z\to 0} \frac{1}{2}\frac{d}{dz}\left[\frac{6z^2}{2z^2+5z+2} - \frac{(4z+5)(z^6+1)}{(2z+5z+1)^2}\right] = \cdots = \frac{21}{8}$$

But that completes all the work we need to do. Using the residue theorem, the original integral is

$$\int_0^{2\pi} \frac{\cos 3\theta}{5+4\cos\theta}\, d\theta = \frac{1}{2i}\oint_C \frac{(z^6+1)\, dz}{z^3(2z+1)(z+2)} = \frac{1}{2i}(2\pi i)\left(-\frac{65}{24}+\frac{21}{8}\right) = \pi\frac{(63-65)}{24} = -\frac{\pi}{12}$$

If I'm lyin', I'm dyin': Check this with Maple, which gives the same exact result.