Introduction

In the previous session we saw how differential and integral calculus allow us to characterise the way in which a quantity might vary: its instantaneous rate of change is captured by evaluating its derivative while the cumulative effects of its variation can be calculated by integration. Much of our time was devoted to the mechanics of carrying through these important formal operations. To keep things relatively simple we confined our attention to variations represented by functions of one independent variable, for example time or a single spatial dimension. In this session we will consider how we can extend these ideas to capture dependence on several variables. At any one moment, we live in, and hope to describe, a three-dimensional world; thus these independent variables might well be Cartesian spatial coordinates. In such a world the concept of direction is evidently of importance. Consequently we introduce the vector, a quantity with both magnitude and direction. We proceed fairly informally, discussing how the components of a vector change as the coordinate system to which they are referred is rotated. This helps us introduce the Cartesian tensor formalism. To exploit this fully we need to know something about matrices and determinants; we review this material and introduce one or two tricks that are omitted from introductory treatments. Once we have this under our belts we introduce the vector product; we also see how the Cartesian tensor notation reduces the algebraic overhad involved quite dramatically. This provides us with the tools to generalise the concepts of differentiation and integration to several dimensions; again the tensor notation allows us to circumvent much of the formal drudgery involved. (Some of these results turn out to have quite startling implications for the behaviour of functions of a complex variable.) Once we are familiar with these ideas we apply them to the solution of a real life problem, the description of rotations of a three-dimensional body. This provides us with a concrete example, in which many of the concepts and techniques of vector and matrix analysis can be seen in action. The results we derive are of fundamental importance in the analysis of navigational systems and their deployment in solutions of the SAR motion compensation problem. .

Cartesian coordinate systems, vectors and tensors

In two dimensions we are familiar with (*x,y*) coordinates; a point in the plane is labeled by its associated *x* and *y* values. It is customary to take the x direction 'horizontal' and the y direction 'vertical'. The distance between two points (x_1, y_1) and (x_2, y_2) is given by

$$s = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \qquad (1)$$

In three dimensions (x, y, z) are the coordinates; we adopt the convention that, on casting your left hand into an Ali G like gesture, your thumb is in the *z* direction, your index finger is in the *y* direction and your middle finger in the *x* direction. Distances are now calculated from

$$s = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \ . \qquad \text{`}(2)$$

A scalar quantity, such as temperature, has a magnitude but no associated direction. This may vary in space, and be represented by $T(x, y, z)$. Many quantities (e.g. force, displacement) have both a magnitude and direction. These are referred to as vectors, and canl be represented by lower-case bold type. Thus a displacement **d** from the origin may be achieved by travelling distances (d_x, d_y, d_z) in the x, y, and z directions respectively. These three numbers are referred to as the components of the vector. Unit basis vectors are identified as

$$\mathbf{i} \equiv (1,0,0), \mathbf{j} \equiv (0,1,0), \mathbf{k} \equiv (0,0,1); \qquad (3)$$

the vector **d** can then be represented as

$$\mathbf{d} = d_x\mathbf{i} + d_y\mathbf{j} + d_z\mathbf{k}\,. \tag{4}$$

The scalar product of two vectors is defined in terms of its components as

$$\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a} = a_x b_x + a_y b_y + a_z b_z\,; \tag{5}$$

this allows us to define the magnitude of a vector **a** through

$$|\mathbf{a}| = \sqrt{\mathbf{a}\cdot\mathbf{a}} \tag{6}$$

and the cosine of the angle Φ between two vectors **a**,**b** as

$$\cos\Phi = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|} \tag{7}$$

in particular we note that

$$\mathbf{i}\cdot\mathbf{i} = \mathbf{j}\cdot\mathbf{j} = \mathbf{k}\cdot\mathbf{k} = 1; \quad \mathbf{i}\cdot\mathbf{j} = \mathbf{i}\cdot\mathbf{k} = \mathbf{j}\cdot\mathbf{k} = 0\,. \tag{8}$$

If we now rotate the coordinate system, both the basis vectors and the vector components change; the vector itself remains unchanged. To see how this happens we consider a positive (anti-clockwise) rotation of θ of the basis vectors around the *z* axis (i.e. d_z,**k** remain unchanged) Simple trig tells us that

$$\begin{aligned}
&\mathbf{i}' = \mathbf{i}\cos\theta + \mathbf{j}\sin\theta, \quad \mathbf{j}' = -\mathbf{i}\sin\theta + \mathbf{j}\cos\theta \\
&d_x' = d_x\cos\theta + d_y\sin\theta, \quad d_y' = -d_x\sin\theta + d_y\cos\theta \\
&d_x'\mathbf{i}' + d_y'\mathbf{j}' = d_x\mathbf{i} + d_y\mathbf{j}
\end{aligned} \tag{9}$$

These transformation properties are characteristic of both the basis vectors and vector components in a Cartesian coordinate system. To handle a more general rotation of the coordinate system we gradually introduce tensorial notation. Thus far we have labeled vector components with *x y z* and denoted the unit vectors by **i j k**. We could, however, denote the components of a vector **d** by $d_\alpha, \alpha = 1,2,3$. (This immediately clears the path to higher dimensional spaces,) The unit vectors might also be written as $\mathbf{e}_\alpha, \alpha = 1,2,3$. The vector **d** (c.f (4)) could then be represented as

$$\mathbf{d} = \sum_{\alpha=1}^{3} d_\alpha \mathbf{e}_\alpha \tag{10}$$

An awful lot of $\sum$ signs are dispensed with when if we adopt the Einstein summation convention, that repeated indices are summed over; we can then write

$$\mathbf{d} = d_\alpha \mathbf{e}_\alpha \tag{11}$$

Using this notation we write the scalar product (5) as

$$\mathbf{a}\cdot\mathbf{b} = a_\alpha b_\alpha\,. \tag{12}$$

The orthogonality of the basis vectors finds expression through

$$\mathbf{e}_\alpha \cdot \mathbf{e}_\beta = \delta_{\alpha b} \qquad (13)$$

where we have surreptitiously introduced the Kroenecker delta $\delta_{\alpha\beta} = 1, \alpha = \beta; \quad \delta_{\alpha\beta} = 0, \alpha \neq \beta$. The decomposition of a vector into its various components can be expressed as follows

$$\mathbf{a} = \mathbf{a} \cdot \mathbf{e}_\alpha \mathbf{e}_\alpha = a_\alpha \mathbf{e}_\alpha . \qquad (14)$$

Implicit in this a representation of the identity operator (which leaves all vectors unchanged)

$$\mathbf{1} = \mathbf{e}_\alpha \mathbf{e}_\alpha, \quad \mathbf{1} \cdot \mathbf{a} = \mathbf{a} \qquad (15)$$

an artifice that is sometimes referred to as the decomposition of unity.

Let us now consider the rotation of a Cartesian coordinate frame, initially spanned by the unit vectors $\mathbf{e}_\alpha$. The unit vectors in the rotated frame are represented by $\mathbf{e}_\alpha'$; these satisfy the analogous relations

$$\mathbf{e}_\alpha' \cdot \mathbf{e}_\beta' = \delta_{\alpha b} \qquad (16)$$

and provide an alternative decomposition of unity

$$\mathbf{1} = \mathbf{e}_\alpha \mathbf{e}_\alpha = \mathbf{e}_\alpha' \mathbf{e}_\alpha' . \qquad (17)$$

Using this we can now represent a vector in two different ways, referred to rotated and un-rotated bases

$$\begin{aligned} \mathbf{a} &= \mathbf{a} \cdot \mathbf{e}_\alpha \mathbf{e}_\alpha = a_\alpha \mathbf{e}_\alpha \\ &= \mathbf{a} \cdot \mathbf{e}_\alpha' \mathbf{e}_\alpha' = a_\alpha' \mathbf{e}_\alpha' \end{aligned} . \qquad (18)$$

Its components in these two bases are related as follows:

$$a_\alpha' = \mathbf{a} \cdot \mathbf{e}_\alpha' = \mathbf{a} \cdot \mathbf{e}_\beta \mathbf{e}_\beta \cdot \mathbf{e}_\alpha' = a_\beta l_{\beta\alpha} \qquad (19)$$

while the basis vectors satisfy

$$\mathbf{e}_\alpha' = \mathbf{e}_\beta \mathbf{e}_\beta \cdot \mathbf{e}_\alpha' = \mathbf{e}_\beta l_{\beta\alpha} \qquad (20)$$

much as we have in (9). Here we have identified $l_{\beta\alpha}$ as the (direction) cosine between the β axis in the un-rotated coordinate frame and the α axis in the rotated coordinate frame. In more formal treatments this transformation property is taken to define the components of a Cartesian vector a_β and is extended to define those of a Cartesian tensor $a_{\beta_1\beta_2\cdots\beta_n}$ through

$$a_{\alpha_1\cdots\alpha_n}' = a_{\beta_1\cdots\beta_n} l_{\beta_1\alpha_1} \cdots l_{\beta_n\alpha_n} . \qquad (21)$$

The basis vectors in the rotated and un-rotated frames are orthogonal (as you would expect): this implies the following relation for the direction cosines

$$\delta_{a\gamma} = \mathbf{e}_\alpha \cdot \mathbf{e}_\gamma = \mathbf{e}_\alpha \cdot \mathbf{e}_\beta '\mathbf{e}_\beta ' \cdot \mathbf{e}_\gamma = l_{\beta\alpha} l_{\gamma\beta} \tag{22}$$

From this we see that the Kroenecker delta is invariant under rotations

$$\delta_{\alpha\beta}' = l_{\zeta\alpha} l_{\xi\beta} \delta_{\zeta\xi} = l_{\zeta\alpha} l_{\zeta\beta} = \delta_{\alpha\beta} \tag{23}$$

and so can be used to project out the rotationally invariant (or scalar) part of a tensor

$$A^{(0)} = a_{\alpha\,\beta} \delta_{\alpha\beta} = a_{\alpha\alpha} . \tag{24}$$

(You might like to check that this is indeed invariant under rotation of the coordinate frame) In much the same vein we find that the scalar product (12) is invariant; this is, after all, why it is called a scalar product.

Matrices and determinants

The nine direction cosines $l_{\alpha\beta}$ can be presented as a 3X3 array or matrix:

$$\mathbf{L} = \begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \tag{25}$$

(We will denote matrices by bold upper case letters). Here α and β label the row and column in which the element $l_{\alpha\beta}$ occurs. The properties of such arrays have been studied in considerable detail, as they appear in several guises. In our statistical work, for example, we will frequently encounter covariance matrices – arrays of expectation values of products of random variables. The manipulation of vectors and tensors is also facilitated by a knowledge of the properties of matrices. We will now review the more pertinent of these. Matrices need not be 3X3 arrays; they can have any number of rows and columns. Thus they need not be square; we will, however, confine our attention to those that are.

Two *NXN* matrices **P** and **Q** with elements P_{ij}, Q_{ij} respectively can be combined to form a third, known as their product. This is defined by

$$(\mathbf{P} \cdot \mathbf{Q})_{ij} = \sum_{k=1}^{N} P_{ik} Q_{kj} = P_{ik} Q_{kj} \tag{26}$$

where we have again invoked the summation convention. The product of any number of matrices can be built up in this way; squares , cubes and higher powers of a single matrix can also be constructed:

$$\begin{aligned} (\mathbf{P} \cdot \mathbf{Q} \cdot \mathbf{R})_{ij} &= P_{ik} Q_{kl} R_{lj} \\ \mathbf{P}^n &= \mathbf{P} \cdot \mathbf{P}^{n-1} \end{aligned} \tag{27}$$

Power series representations of functions of matrices can be constructed in this way, and will crop up later; the exponential function is particularly useful.

$$\exp(\mathbf{P}) = \sum_{n=0}^{\infty} \frac{\mathbf{P}^n}{n!} \tag{28}$$

For a general **P** the output of this exponential function is a bit of a mess; for some special cases however, it is possible to write exp(**P**) in a nice closed form e.g.

$$\begin{aligned}
&\mathbf{P} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad \mathbf{P}\cdot\mathbf{P} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \\
&\mathbf{P}^{2n} = (-1)^n \mathbf{1}; \quad \mathbf{P}^{2n+1} = (-1)^n \mathbf{P} \\
&\exp(\theta\mathbf{P}) = \cos(\theta)\mathbf{1} + \sin(\theta)\mathbf{P} \equiv \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}
\end{aligned} \tag{29}$$

We also note that the matrix products $\mathbf{P}\cdot\mathbf{Q}$ and $\mathbf{Q}\cdot\mathbf{P}$ need not be equal. Matrices for which $\mathbf{P}\cdot\mathbf{Q}$ and $\mathbf{Q}\cdot\mathbf{P}$ are equal are said to commute; should they not be equal we define the commutator of **P** and **Q** as

$$[\mathbf{P},\mathbf{Q}] = \mathbf{P}\cdot\mathbf{Q} - \mathbf{Q}\cdot\mathbf{P}\,. \tag{30}$$

Pre- and post- multiplication by the *NXN* identity matrix **1,** with diagonal elements equal to unity and off-diagonal elements equal to zero, leaves the value of an *NXN* matrix unchanged; **1** therefore commutes with any such matrix. You may like to verify that while matrices whose off-diagonal elements vanish but whose diagonal elements differ in value commute with each other, but need not commute with a general matrix.

In gaining familiarity with matrix manipulations it helps to try things out with small matrices e.g. 2X2 or 3X3. So, for example, we could try out the rule (26) by forming the product of the two matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \begin{pmatrix} p & q \\ r & s \end{pmatrix}$$

to find that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\cdot\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \begin{pmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{pmatrix}. \tag{31}$$

Our simple 2X2 example allows us to verify that matrix multiplication does not commute

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix}\cdot\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} pa+qc & pb+qd \\ ra+sc & rb+sd \end{pmatrix} \neq \begin{pmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{pmatrix} \tag{32}$$

The transpose $\mathbf{P}^T$ of a matrix **P** is defined by

$$\left(\mathbf{P}^T\right)_{ij} = (\mathbf{P})_{ji} \tag{33}$$

if a matrix is equal to its transpose it is said to be symmetric. Covariance matrices formed from a set of real random variables have this property. The transpose of a product of two matrices is the product of the transposes of the individual matrices, but taken in reverse order:

$$(\mathbf{P}\cdot\mathbf{Q})^T = \mathbf{Q}^T \cdot \mathbf{P}^T \tag{34}$$

(You might like to verify this either from the definition (26),(33) or from an explicit example.)

An NXN **P** matrix can form a product with an N dimensional (column) vector **v** to produce another vector **u**

$$\begin{aligned}&\mathbf{P}\cdot\mathbf{v} = \mathbf{u}\\&u_i = P_{ik}v_k\end{aligned} \tag{35}$$

This provides with a very compact way in which to write linear simultaneous equations, with **v** being identified as the vector of solutions and **P** as a matrix of coefficients. If we could find the inverse of the matrix **P**, defined by

$$\mathbf{P}^{-1}\cdot\mathbf{P} = \mathbf{P}\cdot\mathbf{P}^{-1} = 1 \tag{36}$$

we could solve (35) at a stroke

$$\mathbf{P}\cdot\mathbf{v} = \mathbf{u}, \quad \mathbf{P}^{-1}\cdot\mathbf{P}\cdot\mathbf{v} = \mathbf{P}^{-1}\cdot\mathbf{u}, = \mathbf{v}\ . \tag{37}$$

It is possible to construct the inverse of a matrix explicitly, as we shall see shortly. In practice this general result becomes prohibitively unwieldy for even modest N, and matrix inversion is usually carried out numerically. Before we look at matrix inversion we note that the matrix of direction cosines providers us with an example of a matrix whose inverse can be identified without too much trouble. From (22) we see that

$$l_{\alpha\beta}l_{\gamma\beta} = \delta_{\alpha\gamma} \tag{38}$$

which can be written in matrix form as

$$\mathbf{L}^T \cdot \mathbf{L} = \mathbf{1}. \tag{39}$$

This lets us make the identification of the transpose of the matrix of direction cosines with its inverse; a matrix with this property is said to be orthogonal.

$$\mathbf{L}^T = \mathbf{L}^{-1}. \tag{40}$$

To construct the inverse of a NXN matrix we first have to define and look at some of the properties of the determinant of that matrix. The determinant of a matrix **P** with elements $P_{ij}, 1 \le i,j \le N$ is defined by the sum of $N!$ N-fold products of its constituent elements

$$\det(\mathbf{P}) = \varepsilon_{\alpha_1 \quad \alpha_N} P_{1\alpha_1}P_{2\alpha_2}\cdots P_{N\alpha_N} \tag{41}$$

The quantity $\varepsilon_{\alpha_1 \quad \alpha_N}$ vanishes if any of the $\alpha_1\cdots\alpha_N$ are equal, and takes the values 1 and –1 if $\alpha_1\cdots\alpha_N$ is respectively an even or odd permutation of $12\cdots N$. As an example we consider the 2X2 matrix

$$\mathbf{P} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}; \quad P_{11} = a, P_{12} = b, P_{21} = c, P_{22} = d \tag{42}$$

We can form just two twofold products of the form $P_{1\alpha}P_{2\beta}, \quad \alpha \neq \beta$ i.e. $P_{11}P_{22}$ and $P_{12}P_{21}$. In the former $\alpha = 1, \beta = 2$ is an even permutation of 12, the permutation that yields $\alpha = 2, \beta = 1$ in the latter is odd. Thus we can write

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - cb \ . \tag{43}$$

As N gets progressively larger the expression (41) rapidly becomes quite cumbersome. Nonetheless it is possible to derive the salient properties of the determinant from the definition of $\varepsilon_{\alpha_1 \quad \alpha_N}$. For example: $\varepsilon_{\alpha_1 \quad \alpha_N}$ changes sign when any two of its suffices are interchanged. Thus, if **P** has two rows equal, we find that $\det \mathbf{P} = -\det \mathbf{P} = 0$. The determinant can be written in a more symmetrical form as

$$\det(\mathbf{P}) = \frac{1}{N!}\varepsilon_{\beta_1 \quad \beta_N}\varepsilon_{\alpha_1 \quad \alpha_N}P_{\beta_1\alpha_1}P_{\beta_2\alpha_2}\cdots P_{\beta_N\alpha_N} \tag{44}$$

from which we see that the determinant of a matrix and its transpose are equal. The determinant of the product of two matrices is the equal to the product of their separate determinants; the demonstration of this result once again reveals the power of the tensor notation. Essentially by jigging around with the order of the indices, and keeping track of the permutations involved we find that we can write

$$\varepsilon_{\beta_1 \cdots \beta_N} \det \mathbf{A} = \varepsilon_{\alpha_1 \cdots \alpha_N} A_{\beta_1\alpha_1} \cdots A_{\beta_N\alpha_N} \tag{45}$$

Using this result we proceed as follows

$$\begin{aligned} \det \mathbf{A} \det \mathbf{B} &= \det \mathbf{A} \varepsilon_{\alpha_1 \cdots \alpha_N} B_{1\alpha_1} \cdots B_{N\alpha_N} \\ &= \varepsilon_{\gamma_1 \cdots \gamma_N} A_{\alpha_1\gamma_1} \cdots A_{\alpha_N\gamma_N} B_{1\alpha_1} \cdots B_{N\alpha_N} \\ &= \varepsilon_{\gamma_1 \cdots \gamma_N} (\mathbf{B}\cdot\mathbf{A})_{1\gamma_1} \cdots (\mathbf{B}\cdot\mathbf{A})_{N\gamma_N} \\ &= \det(\mathbf{B}\cdot\mathbf{A}) = \det(\mathbf{A}\cdot\mathbf{B}) \end{aligned} \tag{46}$$

These simple properties of the determinant of a matrix can be exploited to construct its inverse. To make things clearer we amplify our notation denoting $\varepsilon_{\alpha_1\alpha_2 \cdots \alpha_N}$ by $\varepsilon^{12\cdots N}_{\alpha_1\alpha_2 \cdots \alpha_N}$, to remind us the $\alpha_1 \cdots \alpha_N$ are obtained by an odd or even number of transpositions from the 'chronological' order 12…N. .The determinant is then written as

$$\det \mathbf{A} = \varepsilon^{12\cdots N}_{\alpha_1 \cdots \alpha_N} A_{1\alpha_1} A_{2\alpha_2} \cdots A_{N\alpha_N} \ . \tag{47}$$

We now look at the coefficient of a single matrix element, obtained by partial differentiation of the determinant with respect to that element. This takes the form

$$\frac{\partial \det \mathbf{A}}{\partial A_{J\alpha_J}} = \varepsilon^{12\cdots N}_{\alpha_1 \cdots \alpha_N} A_{1\alpha_1} A_{2\alpha_2} \cdots A_{J-1\alpha_{J-1}} A_{J+1\alpha_{J+1}} A_{N\alpha_N} \tag{48}$$

This immediately provides us with the raw material from which we construct the matrix inverse. We note that

$$A_{J\alpha_J}\frac{\partial \det \mathbf{A}}{\partial A_{J\alpha_J}} = \det \mathbf{A} \tag{49}$$

while

$$A_{K\alpha_J}\frac{\partial \det \mathbf{A}}{\partial A_{J\alpha_J}} = 0, \quad K \neq J \tag{50}$$

as we are in effect evaluating the determinant of a matrix with two identical rows. Thus we can write the inverse of the matrix **A** as

$$\left(\mathbf{A}^{-1}\right)_{\alpha_J J} = \frac{1}{\det \mathbf{A}}\frac{\partial \det \mathbf{A}}{\partial A_{J\alpha_J}} \tag{51}$$

It remains only to cast (48) in a more familiar form. As it stands we have a sum of products of N-1 matrix elements; the permutation symbol $\varepsilon^{12\cdots N}_{\alpha_1\alpha_2\cdots\alpha_N}$ over N indices relates to that over N-1 indices with J struck out of the index list and α_J from the suffix list by

$$\varepsilon^{12\cdots N}_{\alpha_1\cdots\alpha_N} = (-1)^{J+\alpha_J}\varepsilon^{J12\cdots J-1J+1\cdots N}_{\alpha_J\alpha_1\cdots\alpha_{J-1}\alpha_{J+1}\cdots\alpha_N} \tag{52}$$

Thus we see that

$$\frac{\partial \det \mathbf{A}}{\partial A_{J\alpha_J}} = (-1)^{J+\alpha_J}\det \hat{\mathbf{A}}(J,\alpha_J) \tag{53}$$

where $\hat{\mathbf{A}}(J,\alpha_J)$ is the matrix obtained from **A** by striking out row J and column α_J. This leads us to the standard prescription for the construction of the inverse of **A** as the transpose of the matrix of its signed co-factors, divided by its determinant. A simple 2X2 example should make things clearer

$$\begin{aligned}
&\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \det \mathbf{A} = ad - bc = \Delta \\
&\frac{\partial \Delta}{\partial a} = d, \frac{\partial \Delta}{\partial d} = a, \frac{\partial \Delta}{\partial b} = -c, \frac{\partial \Delta}{\partial c} = -b \\
&\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}
\end{aligned} \tag{54}$$

You might like to fiddle round with a 3X3 example of your own devising. We should stress once more that all this fancy notation obscures a burgeoning algebraic and computational overhead; once matrices get much bigger than 3X3 things get right out of hand. Nonetheless, modern computers are well able to handle big matrices numerically.

The final aspect of matrix theory that we need to mention is the eigenvalue problem

$$\mathbf{A}\cdot\mathbf{u} = \lambda\mathbf{u}\,. \tag{55}$$

Here we merely note that the eigenvalues λ are determined by solving the equation

$$\det(\mathbf{A} - \lambda\mathbf{1}) = 0 \tag{56}$$

that the determinant of **A** is equal to the product of its eiganvalues and that the trace of **A** (the sum of its diagonal elements) is equal to the sum of its eigenvalues.

The non-commutation of matrix multiplication and the non-triviality of the relation between a matrix and its inverse complicate the manipulative details of matrix algebra. Here we mention a couple of tricks that are not usually included in introductory accounts. To motivate the first of these we look at the matrix version of the geometric series. We set up an equation that defines $(\mathbf{1} - \mathbf{A})^{-1}$ implicitly

$$\begin{aligned} \mathbf{1} &= (\mathbf{1} - \mathbf{A}) + \mathbf{A} \\ (\mathbf{1} - \mathbf{A})^{-1} &= \mathbf{1} + \mathbf{A}(\mathbf{1} - \mathbf{A})^{-1} \end{aligned} \tag{57}$$

This can now be iterated N times to give

$$(\mathbf{1} - \mathbf{A})^{-1} = \mathbf{1} + \mathbf{A} + \mathbf{A}^2 \cdots + \mathbf{A}^N + \mathbf{A}^{N+!}(\mathbf{1} - \mathbf{A})^{-1} \tag{58}$$

which can be rearranged to yield

$$\sum_{n=0}^{N} \mathbf{A}^n = (\mathbf{1} - \mathbf{A}^{N+1})(\mathbf{1} - \mathbf{A})^{-1}. \tag{59}$$

In much the same vein we have

$$\begin{aligned} \mathbf{A} &= (\mathbf{A} + \mathbf{B}) - \mathbf{B} \\ \mathbf{A} \cdot (\mathbf{A} + \mathbf{B})^{-1} &= \mathbf{1} - \mathbf{B} \cdot (\mathbf{A} + \mathbf{B})^{-1} \\ (\mathbf{A} + \mathbf{B})^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1} \cdot \mathbf{B} \cdot (\mathbf{A} + \mathbf{B})^{-1} \end{aligned} \tag{60}$$

This again can be iterated to yield

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{A}^{-1} + \mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{A}^{-1} - \mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{A}^{-1} \cdot \mathbf{B} \cdot \mathbf{A}^{-1} \cdots \tag{61}$$

Implicit in this are all sorts of ‘theoretical physics’ nuggets: perturbation expansions, Dyson’s equation, Mori’s equation etc.

Our second trick, which is very useful in dealing with exponential functions of non-commuting matrices, is that of differentiation with respect to a scalar parameter. Let us consider

$$\mathbf{F}(\lambda) = \exp(\lambda\mathbf{A}) \cdot \mathbf{B} \cdot \exp(-\lambda\mathbf{A}) \tag{62}$$

We differentiate this with respect to the parameter λ to obtain

$$\frac{d\mathbf{F}(\lambda)}{d\lambda} = \mathbf{A} \cdot \mathbf{F} - \mathbf{F} \cdot \mathbf{A} = [\mathbf{F}, \mathbf{A}]. \tag{63}$$

Higher order derivatives can be evaluated by the iteration of this result

$$\frac{d^2\mathbf{F}(\lambda)}{d\lambda^2} = \mathbf{A}\cdot\frac{d\mathbf{F}(\lambda)}{d\lambda} - \frac{d\mathbf{F}(\lambda)}{d\lambda}\cdot\mathbf{A} = [\mathbf{A},[\mathbf{A},\mathbf{F}]]$$
$$\frac{d^3\mathbf{F}(\lambda)}{d\lambda^3} = [\mathbf{A},[\mathbf{A},[\mathbf{A},\mathbf{F}]]] \tag{64}$$

Finally we can construct a Taylor series representation of **F**:

$$\mathbf{F}(\lambda) = \mathbf{F}(0) + \lambda\left.\frac{d\mathbf{F}(\lambda)}{d\lambda}\right|_{\lambda=0} + \frac{\lambda^2}{2}\left.\frac{d^2\mathbf{F}(\lambda)}{d\lambda^2}\right|_{\lambda=0} + \frac{\lambda^3}{3!}\left.\frac{d^3\mathbf{F}(\lambda)}{d\lambda^3}\right|_{\lambda=0}\cdots$$
$$= \mathbf{B} + \lambda[\mathbf{A},\mathbf{B}] + \frac{\lambda^2}{2}[\mathbf{A},[\mathbf{A},\mathbf{B}]] + \frac{\lambda^3}{3!}[\mathbf{A},[\mathbf{A},[\mathbf{A},\mathbf{B}]]]\cdots \tag{65}$$

This result is particularly useful if the repeated commutators of **A** with **B** vanish at some order and the series terminates: if

$$[\mathbf{A},[\mathbf{A},\mathbf{B}]] = \mathbf{0} \tag{66}$$

then the series has only two terms

$$\exp(\mathbf{A})\cdot\mathbf{B}\cdot\exp(-\mathbf{A}) = \mathbf{B} + [\mathbf{A},\mathbf{B}]\ . \tag{67}$$

Some practice in manipulations of this type is provided in the exercises. Should you be interested in seeing how far this simple approach can be taken you might want to look at R.M. Wilcox, J. Math. Phys., **8**, 962, 1968.

<u>Vector products</u>

Having made this lengthy excursion into determinants and matrices, we return to Cartesian tensors. The anti-symmetric symbol ε_{ijk} plays a central role in the construction of determinants of 3X3 matrices; is there any chance that it might also be a Cartesian tensor? To see if this is the case we look at how it transforms subject to (21) and form

$$\varepsilon_{\alpha\beta\gamma}' = l_{\eta\alpha} l_{\zeta\beta}\ l_{\xi\gamma}\varepsilon_{\eta\zeta\xi} \tag{68}$$

$\varepsilon_{\alpha\beta\gamma}'$ changes sign when we interchange any two of its indices, and so vanishes when any two indices are the same. Now it remains to evaluate ε_{123}'; a moment's contemplation should convince us that

$$\varepsilon_{123}' = \det(\mathbf{L}) \tag{69}$$

the determinant of the direction cosine matrix. Taking the determinant of each side of

$$\mathbf{L}^T\cdot\mathbf{L} = \mathbf{1}$$

tells us that

$$(\det(\mathbf{L}))^2 = 1;\quad \det(\mathbf{L}) = \pm 1. \tag{70}$$

For rotational transformations of the basis vectors $\det(\mathbf{L}) = 1$; $\det(\mathbf{L}) = -1$ corresponds to transformations in which the coordinate system is rotated and reflected in some plane including the origin. So, as long as we confine our attention to rotational transformations, the anti-symmetric symbol does behave like a Cartesian tensor. Using ε_{ijk} we can form the vector product of two vectors $\mathbf{a}, \mathbf{b}$ as

$$(\mathbf{a} \wedge \mathbf{b})_i = \varepsilon_{ijk} a_j b_k \; ; \tag{71}$$

this can be written as a determinant in the form

$$\mathbf{a} \wedge \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} \tag{72}$$

Armed with our knowledge of determinants we can easily show that

$$\mathbf{a} \wedge \mathbf{b} = -\mathbf{b} \wedge \mathbf{a} \tag{74}$$

and that

$$\mathbf{a} \cdot (\mathbf{a} \wedge \mathbf{b}) = \mathbf{b} \cdot (\mathbf{a} \wedge \mathbf{b}) = 0 \, . \tag{75}$$

Thus $\mathbf{a} \wedge \mathbf{b}$ is perpendicular to both **a** and **b**. To determine the magnitude of the vector product we can (just this once) bash things out in component form. In outline we find that

$$\begin{aligned} (\mathbf{a} \wedge \mathbf{b}) \cdot (\mathbf{a} \wedge \mathbf{b}) &= (a_y b_z - a_z b_y)^2 + (a_z b_x - a_x b_z)^2 + (a_x b_y - a_y b_x)^2 \\ &= (a_x^2 + a_y^2 + a_z^2)(b_x^2 + b_y^2 + b_z^2) - (a_x b_x + a_y b_y + a_z b_z)^2 \\ &= (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) - (\mathbf{a} \cdot \mathbf{b})^2 \end{aligned} \tag{76}$$

so that, remembering (7), we have

$$|\mathbf{a} \wedge \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin \Phi \, . \tag{77}$$

All the labour expended in working through (76) can be summarised in the tensor identity

$$\varepsilon_{ijk} \varepsilon_{ilm} = \delta_{jl} \delta_{km} - \delta_{jm} \delta_{lk} \, . \tag{78}$$

This little fellow is easy to remember and, once one has had a bit of practice, renders manipulations of vector products almost trivial – a triumph of notation in fact. The vector product crops up physically in the specification of a torque or angular momentum; perhaps it should be no surprise when it turns out to be very useful in the characterisation of the rotations on a rigid body.

Vector calculus

In the previous session we described how the derivative of a function of a single variable could be calculated; this idea can be extended to a function of several variables as follows. If we have a function of several variables, a partial derivative with respect to any one of these variables is generated by differentiation with respect to that variable, all others being held constant. For example

$$\begin{aligned} &f(x,y) = x^2 + 3y \\ &\frac{\partial f}{\partial x} = 2x, \quad \frac{\partial f}{\partial y} = 3, \quad \frac{\partial^2 f}{\partial x^2} = 2, \quad \frac{\partial^2 f}{\partial x \partial y} = 0 \end{aligned} \tag{79}$$

All the usual rules for constructing derivatives carry over to the partial case; there are some examples in the exercises. If *x,y* are themselves functions of a variable *t* we have

$$f \equiv f\big(x(t), y(t)\big); \quad \frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \tag{80}$$

in an obvious generalisation of the chain rule. There is nothing new in principle in partial differentiation; most of the problem is a matter of book-keeping and presentation. The vector and tensor formalism we introduced earlier provide us with a way of doing this. Partial differential operators with respect to Cartesian coordinates have particularly simple transformation properties that allow us to treat them essentially as components of a vector. To this end we consider the position vector of a point, with components x_1, x_2, x_3; the components of the position vector of the same point, referred to a rotated frame of reference are given by

$$x_\alpha' = l_{\beta\alpha} x_\beta \tag{81}$$

which can be inverted to give us

$$x_\beta = l_{\beta\alpha} x_\alpha' . \tag{82}$$

Using the chain rule we now form the partial derivative with respect to this transformed coordinate, in terms of those with respect to the un-transformed coordinates. In this way be find that

$$\begin{aligned} \frac{\partial}{\partial x_\alpha'} &= \frac{\partial x_\beta}{\partial x_\alpha'}\frac{\partial}{\partial x_\beta} \\ &= l_{\beta\alpha}\frac{\partial}{\partial x_\beta} \end{aligned} \tag{83}$$

Thus we see that the partial derivative with respect to a Cartesian coordinate transforms in just the same way as the coordinate itself. Using this we can construct a vector differential operator (often referred to as ‘grad’) of the form

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z} \equiv \mathbf{e}_\alpha \frac{\partial}{\partial x_\alpha} . \tag{84}$$

When applied to a scalar function of position this generates a vector; for example the application of -grad to a potential energy function gives us a force, which is a vector:

$$-\nabla\phi(x,y,z) = -\left(\mathbf{i}\frac{\partial\phi(x,y,z)}{\partial x} + \mathbf{j}\frac{\partial\phi(x,y,z)}{\partial y} + \mathbf{k}\frac{\partial\phi(x,y,z)}{\partial z}\right) = \mathbf{F}(x,y,z) \quad (85)$$

Similarly the rate of change of a scalar function along the direction specified by the unit vector **a** is given by

$$\mathbf{a}\cdot\nabla\phi(x,y,z). \quad (86)$$

Successive application of partial derivatives yields quantities that transform as higher rank Cartesian tensors e.g.

$$F_{\alpha\beta} = \frac{\partial}{\partial x_\alpha}\frac{\partial}{\partial x_\beta}\phi \quad (87)$$

is a second rank tensor; contracting this with the Kronecker delta gives us the familiar Laplacian operator

$$F_{\alpha\alpha} = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = \nabla^2\phi \quad (88)$$

The operator grad can also be applied to functions that are themselves vectors (e.g. an electric or magnetic field). Analogous to the scalar product we have the divergence or div operator

$$\mathrm{div}\mathbf{F}(x,y,z) = \nabla\cdot\mathbf{F}(x,y,z) = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \equiv \frac{\partial F_\alpha}{\partial x_\alpha} \quad (89)$$

while the ‘curl’ operator resembles the vector product

$$\mathrm{curl}\mathbf{F}(x,y,z) = \nabla\wedge\mathbf{F}(x,y,z) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ F_x & F_y & F_z \end{vmatrix}. \quad (90)$$
$$\left(\mathrm{curl}\mathbf{F}(x,y,z)\right)_\alpha = \varepsilon_{\alpha\beta\gamma}\frac{\partial}{\partial x_b}F_\gamma$$

The names given to these operators give us some idea of their physical interpretation, which will become clearer when we consider vector integration shortly. A great many potentially useful identities can be constructed using these operators; these are probably best derived using Cartesian tensor notation, (especially (78)), rather than bashing things out a component at a time. An exercise provides you with some practice in doing this.

Integration in two and more dimensions

In the previous session we confined ourselves to integrals along the real line *x*. This provides us with a particularly simple path of integration. In a plane, characterised by Cartesian coordinates *x*,*y*, we have two options for extending the concept of integration: along a path defined by $y = y(x)$ or over a given area (i.e. a two dimensional integral) in the plane. Thus we might have

$$I = \int_{x_1}^{x_2} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \tag{91}$$

as a length of arc providing an example of the former and

$$\iint_{x^2+y^2 \le a^2} dxdy \tag{92}$$

as an example of the latter. The techniques for integral evaluation we developed in the one dimensional case carry over; one has to be careful if the limit of integration in one variable depends on the other. We will take the circle as a simple example

$$y = \sqrt{a^2 - x^2} \quad \frac{dy}{dx} = -\frac{x}{\sqrt{a^2 - x^2}} \tag{93}$$

Therefore the length of arc in the first quadrant is given by

$$\int_o^a \sqrt{1 + \frac{x^2}{a^2 - x^2}}\, dx = a\int_0^a \frac{dx}{\sqrt{a^2 - x^2}} = a\int_0^1 \frac{dx}{\sqrt{1 - x^2}} = a\left(\sin^{-1}(1) - \sin^{-1}(0)\right) = \frac{a\pi}{2} \tag{94}$$

while the area of the first quadrant is given by

$$\int_0^a dy \int_0^{\sqrt{a^2-y^2}} dx = \int_0^a dy\sqrt{a^2 - y^2} = a^2\int_0^1 dy\sqrt{1 - y^2} = \frac{\pi a^2}{4} \tag{95}$$

It is of course possible to evaluate integrals of functions of *x,y* along a path or over a region of space, in just the same way. We have seen in previous weeks how coordinate sytems other than the Cartesian can be used to parameterise points on a plane. Thus, if we adopt some other coordinate system, our integration will be carried out with respect to these new coordinates. In doing this we must (i) make sure that the limts on the ranges of integration are properly taken care of and (ii) line and area elements are correctly expressed in terms of the new coordinates. All this can be done in terms of general tensor notation; rather than do this we will consider a specific example to see how it goes:

Plane polar coordinates

$$x = r\cos\theta; \quad y = r\sin\theta \tag{96}$$

Line element

$$ds^2 = (dr)^2 + (rd\theta)^2 \tag{97}$$

Area element

$$r dr d\theta \tag{98}$$

In this coordinate system the evaluation of the arc length and area in the first quadrant is particularly straight forward.

If we take new coordinates *p,q* so that

$$x = x(p,q), \quad y = y(p,q) \tag{99}$$

then the area element in these new coordinate is given by

$$\begin{vmatrix} \partial x/\partial p & \partial x/\partial q \\ \partial y/\partial p & \partial y/\partial q \end{vmatrix} dp dq \tag{100}$$

This determinant of derivatives is referred to as the Jacobian of the coordinate transformation; what do you think the line element is in terms of these coordinates?

A more general form of line integral is

$$\int_{y=y(x)} \left[P(x,y)dx + Q(x,y)dy \right] \tag{101}$$

The integrand is a total derivative if

$$P(x,y) = \frac{\partial F(x,y)}{\partial x}; \quad Q(x,y) = \frac{\partial F(x,y)}{\partial y} \tag{102}$$

so that

$$\frac{\partial P(x,y)}{\partial y} = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial Q(x,y)}{\partial x} . \tag{103}$$

If this condition is satisfied then the value of the path integral is determined by the endpoints of the path; in particular the integral round a close path vanishes.

When we move up to three dimensions the vector notation lets us express things compactly. We now have integrals along paths

$$\int_C \mathbf{F} \cdot d\mathbf{l}; \quad d\mathbf{l} = \mathbf{i}dx + \mathbf{j}dy + \mathbf{k}dz \tag{104}$$

over surfaces in the three space

$$\int_S \mathbf{F}\cdot d\mathbf{S}; \quad d\mathbf{S} = \mathbf{n}dS \tag{105}$$

and over volumes

$$\int_V \phi \, dV; \quad \int_V \mathbf{F} \, dV \, . \tag{106}$$

Surface and volume integrals are connected through Gauss' theorem

$$\int_S \mathbf{F}\cdot d\mathbf{S} = \int_V \nabla\cdot\mathbf{F}dV \tag{107}$$

where *V* is the volume enclosed within the surface *S*. Surface and line integrals are connected through Stoke's theorem

$$\int_C \mathbf{F}\cdot d\mathbf{l} = \int_S \left(\nabla \wedge \mathbf{F}\right)\cdot d\mathbf{S} \tag{108}$$

Here *S* is a surface circumscribed by the closed path *C.*

These identities allow us to carry out analogues of partial integration e.g.

$$\begin{aligned}
&\int_V \nabla\cdot\left(\psi\mathbf{A}\right)dV = \int_S \psi\mathbf{A}\cdot d\mathbf{S} \\
&\int_V \left(\psi\nabla\cdot\mathbf{A} + \mathbf{A}\cdot\nabla\psi\right)dV = \int_S \psi\mathbf{A}\cdot d\mathbf{S} \\
&\int_V \left(\mathbf{A}\cdot\nabla\psi\right)dV = \int_S \psi\mathbf{A}\cdot d\mathbf{S} - \int_V \left(\psi\nabla\cdot\mathbf{A}\right)dV
\end{aligned} \tag{109}$$

A special case of this result is Green's theorem:

$$\int_V \left(\phi\nabla^2\psi - \psi\nabla^2\phi\right)dV = \int_S \left(\phi\nabla\psi - \psi\nabla\phi\right)\cdot d\mathbf{S} \tag{110}$$

<u>Calculus with a complex variable</u>

When we introduced the concept of differentiating a real function of a real variable we formed the quotient

$$\frac{\Delta y}{\Delta x} = \frac{y\left(x+\Delta x\right) - y\left(x\right)}{\Delta x} \tag{111}$$

when we developed the numerator in this expression appropriately a term linear in Δx crops up that cancels out with the bottom line. If we were to form something similar for the function *f* of a complex variable *z*

$$\frac{\Delta f}{\Delta z} = \frac{f(z+\Delta z) - f(z)}{\Delta z} \tag{112}$$

and this quantity tended to a limit that depended only on *z* and not on the direction of Δz in the complex plane then this limit would be identified as the derivative of *f* with respect to *z.* So what, you might say. However the existence of this derivative places very significant constraints on the real and imaginary parts of *f*, considered as functions of *x,y* the real and imaginary parts of *z*. So, we start off with

$$\begin{aligned} z &= x + iy \\ f(z) &\equiv u(x,y) + iv(x,y) \end{aligned} \tag{113}$$

If the derivative of *f* exists, its value will be the same, whether *z* undergoes an increment in its real or imaginary part. Thus we can write

$$\begin{aligned} z \equiv x; &\quad \frac{df}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \\ z \equiv iy; &\quad \frac{df}{dz} = \frac{1}{i}\left(\frac{\partial u}{\partial y} + i\frac{\partial v}{\partial y}\right) \end{aligned} \tag{114}$$

On equating the real and imaginary parts of these two expressions for the derivative, which must be the same, we find that

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial y} &= -\frac{\partial v}{\partial x} \end{aligned} \tag{115}$$

These are known as the Cauchy Reimann conditions; they imply that the real and imaginary parts of an analytic function satisfy Laplace's equation in two dimensions
i.e.

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0\,. \tag{116}$$

Thus the condition that a function of a complex variable is differentiable imposes considerable smoothness onto the behaviour of its real and imaginary parts. This idea carries over to integrals in the complex plane. Thus we construct an integral around a closed contour in the complex plane as

$$\int_C f(z)dz \equiv \int_C \left(u(x,y)+iv(x,y)\right)[dx+idy]$$
$$= \int_C \left(\left(u(x,y)+iv(x,y)\right)dx + \left(iu(x,y)-v(x,y)\right)dy\right) \tag{117}$$

The Cauchy Reimann conditions then ensure that, if *f* is analytic in and on *C*, this integral vanishes. Contributions to such integrals in the complex plane are derived only from those points within the path where *f* is not analytic; you can also take the most tremendous liberties when it comes to manipulating the path of integration without changing the value of the integral. These two observations form the basis of the method of contour integration, to which an exercise provides a simple introduction.

<u>The description of rotations</u>

Rotational transformations and their evolution in time are of considerable practical importance. The analysis of imaging methods that depend for their azimuthal resolution on the relative motion of the target and radar platform (SAR, ISAR) requires us to characterise rotational motion, both in the description of the imaging process (e.g. Walker's theory of ISAR and its extensions) and in the exploitation of inertial navigation systems in motion compensation. The vector and matrix methods we have discussed this session provide us with a very useful tool for the analysis of these problems. Our specification of the rotation of a rigid body is in terms of an axis, which remains unchanged in the rotation, and the angle that measures the extent of the rotation. This seemingly natural parameterisation of rotations was introduced in the eighteenth century by Rodrigues, but has been sorely neglected in favour of Euler angles etc. It is only recently that Rodrigues' methods have been recognised and rehabilitated.

To start off, we consider the transformation of a general vector **v** under a rotation in a little more detail. To this end we identify the unit vector (the axis) that is left unchanged by the rotation; this we denote by **n.** The extent of the rotation about this axis is characterized by an angle Θ; this is taken to be positive for an anti-clock wise rotation. The vector **v** is then decomposed into its components parallel and perpendicular to **n**

$$\mathbf{v} = (\mathbf{v}\cdot\mathbf{n})\mathbf{n} + (\mathbf{v} - (\mathbf{v}\cdot\mathbf{n})\mathbf{n}) = (\mathbf{v}\cdot\mathbf{n})\mathbf{n} - \mathbf{n}\wedge(\mathbf{n}\wedge\mathbf{v}) \tag{118}$$

The component parallel to **n** is unchanged by the rotation, while that perpendicular to **n** is rotated in a plane normal to **n**. Consequently we can write (c.f.(9))

$$\mathbf{v}' = (\mathbf{n}\cdot\mathbf{v})\mathbf{v}(1-\cos\Theta) + \mathbf{v}\cos\Theta + \mathbf{n}\wedge\mathbf{v}\sin\Theta \tag{119}$$

This result can be written in a particularly compact form if we recall the simple vector identities

$$(\mathbf{n}\wedge)^{2r}\mathbf{v} = (-1)^r(\mathbf{v} - (\mathbf{n}\cdot\mathbf{v})\mathbf{n}), \quad r > 0$$
$$(\mathbf{n}\wedge)^{2r+1}\mathbf{v} = (-1)^r(\mathbf{n}\wedge\mathbf{v}) \tag{120}$$

so that

$$\mathbf{v}' = \exp(\Theta\mathbf{n}\wedge)\mathbf{v} \tag{121}$$

In explicit component form the vector product $\mathbf{n}\wedge\mathbf{v}$ can also be represented as multiplication of column vector by a matrix

$$\begin{pmatrix} n_y v_z - n_z v_y \\ n_z v_x - n_x v_z \\ n_x v_y - n_y v_x \end{pmatrix} = \begin{pmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \tag{122}$$

The representation of the rotation operator (121) is thus rather reminiscent of the simple exponential function in (29).

The direction cosines corresponding to a rotation specified by an axis and angle are derived as follows

$$\begin{aligned} l_{jk} &= \mathbf{e}_j \cdot \mathbf{e}_k' = \mathbf{e}_j \cdot \exp(\Theta \mathbf{n} \wedge)\mathbf{e}_k \\ &= n_j n_k (1 - \cos\Theta) + \delta_{jk} \cos\Theta + \sin\Theta (\mathbf{e}_k \wedge \mathbf{e}_j) \cdot \mathbf{n} \end{aligned} \tag{123}$$

It is reassuring to note that the orthogonality result (22) can be verified directly from this expression.

A succession of rotations, in general about different axes, is equivalent to a single rotation. The identification of the axis of this rotation, and the corresponding angle, for the resultant of two rotations about non-parallel axes is therefore a key step in application of the axis angle formalism in practice. Let $\mathbf{n}, \Theta$ parameterise the composite rotation generated by first rotating through θ_1 about $\mathbf{n}_1$ then through θ_2 about $\mathbf{n}_2$ i.e.

$$\exp(\Theta \mathbf{n} \wedge) \equiv \exp(\theta_2 \mathbf{n}_2 \wedge)\exp(\theta_1 \mathbf{n}_1 \wedge). \tag{124}$$

To find $\mathbf{n}, \Theta$ we consider the eigenvalue problem

$$\exp(\Theta \mathbf{n} \wedge)\mathbf{v} \equiv \lambda \mathbf{v} ; \tag{125}$$

$\mathbf{n}$ is the solution corresponding to the unit eigenvalue. To facilitate the solution of this problem we write $\mathbf{v}$ as a linear combination of $\mathbf{n}_1, \mathbf{n}_2$ and the vector product $\mathbf{n}_2 \wedge \mathbf{n}_1$

$$\mathbf{v} = \alpha \mathbf{n}_1 + \beta \mathbf{n}_2 + \gamma \mathbf{n}_2 \wedge \mathbf{n}_1. \tag{126}$$

The matrices representing these successive rotations in this basis are

$$\begin{pmatrix} 1 & (1-c_1)\Omega & -\Omega s_1 \\ 0 & c_1 & s_1 \\ 0 & -s_1 & c_1 \end{pmatrix} \text{and} \begin{pmatrix} c_2 & 0 & -s_2 \\ (1-c_2)\Omega & 1 & \Omega s_2 \\ s_2 & 0 & c_2 \end{pmatrix} \tag{127}$$

where

$$\begin{aligned} c_k &= \cos(\theta_k), \quad s_k = \sin(\theta_k) \quad k = 1,2 \\ \Omega &= \mathbf{n}_1 \cdot \mathbf{n}_2 \end{aligned} \tag{128}$$

and we have used the vector identity $\mathbf{n}_1 \wedge (\mathbf{n}_2 \wedge \mathbf{n}_1) = \mathbf{n}_2 - \Omega \mathbf{n}_1$

Their product can be formed

$$
\begin{pmatrix} c_2 & 0 & -s_2 \\ (1-c_2)\Omega & 1 & \Omega s_2 \\ s_2 & 0 & c_2 \end{pmatrix} \cdot \begin{pmatrix} 1 & (1-c_1)\Omega & -\Omega s_1 \\ 0 & c_1 & s_1 \\ 0 & -s_1 & c_1 \end{pmatrix}
$$

$$
= \begin{pmatrix} c_2 & c_2\Omega(1-c_1)+s_2 s_1 & -(s_2 c_1 + s_1 c_2 \Omega) \\ (1-c_2)\Omega & (1-c_1)(1-c_2)\Omega^2 + c_1 - s_1 s_2 \Omega & s_1 + \Omega s_2 c_1 - (1-c_1) s_2 \Omega^2 \\ s_2 & \Omega s_2 (1-c_1) - s_1 c_2 & c_2 c_1 - s_1 s_2 \Omega \end{pmatrix}
$$

This can now be used to construct the eigenvalue problem

$$
\begin{pmatrix} c_2 & c_2\Omega(1-c_1)+s_2 s_1 & -(s_2 c_1 + s_1 c_2 \Omega) \\ (1-c_2)\Omega & (1-c_1)(1-c_2)\Omega^2 + c_1 - s_1 s_2 \Omega & s_1 + \Omega s_2 c_1 - (1-c_1) s_2 \Omega^2 \\ s_2 & \Omega s_2 (1-c_1) - s_1 c_2 & c_2 c_1 - s_1 s_2 \Omega \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \lambda\alpha \\ \lambda\beta \\ \lambda\gamma \end{pmatrix} \qquad (129)
$$

The matrix occurring in (129) has a unit determinant (as rotations preserve length); thus we would expect its three eigenvalues to be of the form $1, \exp(i\Theta), \exp(-i\Theta)$. The trace of the matrix is the sum of these three eigenvalues, from which we deduce that

$$
\cos\Theta = \frac{1}{2}\left((1-c_1)(1-c_2)\Omega^2 + c_1 + c_2 + c_1 c_2 - 2\Omega s_1 s_2 - 1\right) \qquad (130)
$$

To identify **n** we solve (129) with $\lambda = 1$, subject to the normalization condition $n^2 = 1$.. Now the equation

$$
\begin{pmatrix} c_2 - 1 & c_2\Omega(1-c_1)+s_2 s_1 & -(s_2 c_1 + s_1 c_2 \Omega) \\ (1-c_2)\Omega & (1-c_1)(1-c_2)\Omega^2 + c_1 - s_1 s_2 \Omega - 1 & s_1 + \Omega s_2 c_1 - (1-c_1) s_2 \Omega^2 \\ s_2 & \Omega s_2 (1-c_1) - s_1 c_2 & c_2 c_1 - s_1 s_2 \Omega - 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad (131)
$$

is indeterminate; nonetheless it provides us with a soluble set of linear equations for the ratios

$$
\hat{\beta} = \beta/\alpha; \quad \hat{\gamma} = \gamma/\alpha
$$

which take the form

$$
\begin{pmatrix} c_2\Omega(1-c_1)+s_2 s_1 & -(s_2 c_1 + s_1 c_2 \Omega) \\ \Omega s_2(1-c_1) - s_1 c_2 & c_1 c_2 - \Omega s_1 s_2 - 1 \end{pmatrix} \cdot \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 1 - c_2 \\ -s_2 \end{pmatrix} \qquad (132)
$$

The determinant of the 2X2 matrix in this set of equations has a relatively simple form, when account is taken of some elementary trigonometric identities; it boils down to

$$
s_1 s_2 + \left(s_1^2 + (c_1 - 1)(1-c_2)\right)\Omega = s_1 s_2 + \Omega(c_1 + c_2)(1 - c_1) \qquad (133)
$$

so that

$$\hat{\beta} = \frac{s_1 s_2 + (1+c_1)(1-c_2)}{s_1 s_2 + \Omega(c_1+c_2)(1-c_1)}, \quad \hat{\gamma} = \frac{\Omega s_2(1-c_1) + s_1(1-c_2)}{s_1 s_2 + \Omega(c_1+c_2)(1-c_1)}. \tag{134}$$

From this it follows that

$$\mathbf{n} = C\left[\left[s_1 s_2 + \Omega(1-c_1)(1+c_2)\right]\mathbf{n}_1 + \left[s_1 s_2 \Omega + (1+c_1)(1-c_2)\right]\mathbf{n}_2 + \left[\Omega s_2(1-c_1) + s_1(1-c_2)\right]\mathbf{n}_2 \wedge \mathbf{n}_1\right] \tag{135}$$

where

$$\frac{1}{C^2} = \left[s_1 s_2 + \Omega(1-c_1)(1+c_2)\right]^2 + \left[\Omega s_1 s_2 + (1+c_1)(1-c_2)\right]^2 + 2\Omega\left[s_1 s_2 + \Omega(1-c_1)(1+c_2)\right]\left[\Omega s_1 s_2 + (1+c_1)(1-c_2)\right] + \left(1-\Omega^2\right)\left[\Omega s_2(1-c_1) + s_1(1-c_2)\right]^2 \tag{136}$$

This expression can be simplified, but not in any dramatic or revealing way. At this point we have shown how two arbitrary rotations, combined in a specific order, can be identified with a single rotation about a unique axis. The results just presented show explicitly that rotations do not in general commute.

So far we have seen how the axis and angle parameterising a rotation equivalent to two successive, non-commuting, rotations are found. This is essentially a ‘static’ problem. We now consider a ‘dynamic’ problem: if a rotation develops in time described by $\mathbf{n}(t), \Theta(t)$ how do the rates of change of the axis and angle parameters relate to an instantaneous angular velocity $\boldsymbol{\omega}(t)$?. To answer this question introduce

$$\begin{aligned} &\theta_1 = \Theta(t) \quad \mathbf{n}_1 = \mathbf{n}(t) \\ &\theta_2 = \omega(t)\delta t, \quad \mathbf{n}_2 = \frac{\boldsymbol{\omega}(t)}{|\boldsymbol{\omega}(t)|}, \quad \Omega = \mathbf{n}(t)\cdot\frac{\boldsymbol{\omega}(t)}{|\boldsymbol{\omega}(t)|} \end{aligned} \tag{137}$$

into the results we have just derived, and note that

$$\frac{d\Theta(t)}{dt} = \mathbf{n}(t)\cdot\boldsymbol{\omega}(t) \tag{138}$$

The expression for the normalisation constant C can be expanded to first order in δt as

$$C = \frac{1}{2(1-\cos\Theta)\Omega + \left(1+\Omega^2\right)\sin\Theta\,\omega(t)\delta t}. \tag{139}$$

If we bring all these results together we have

$$\delta\,\mathbf{n}(t) = \delta t\,\frac{1}{2}\left(\frac{\sin\Theta(t)}{1-\cos\Theta(t)}\boldsymbol{\omega}(t)\cdot\left(\mathbf{1}-\mathbf{n}(t)\mathbf{n}(t)\right) + \boldsymbol{\omega}(t)\wedge\mathbf{n}(t)\right) + 0\left(\delta t^2\right) \tag{140}$$

from which we deduce the differential equation

$$\frac{d\mathbf{n}(t)}{dt} = \frac{1}{2}\left(\frac{\sin\Theta(t)}{1-\cos\Theta(t)}\boldsymbol{\omega}(t)\cdot\left(\mathbf{1}-\mathbf{n}(t)\mathbf{n}(t)\right) + \boldsymbol{\omega}(t)\wedge\mathbf{n}(t)\right) \tag{141}$$

Note that this preserves the normalization of **n** as

$$\frac{d|\mathbf{n}(t)|^2}{dt} = 2\mathbf{n}(t)\cdot\frac{d\mathbf{n}(t)}{dt} = 0 \qquad (142)$$

In some circumstances it is helpful to define a vector $\mathbf{q} = \Theta\mathbf{n}$ whose equation of motion follows from the results just derived as

$$\frac{d\mathbf{q}}{dt} = \mathbf{q}\frac{\mathbf{q}\cdot\boldsymbol{\omega}}{q^2} + \frac{1}{2}\left(\frac{q\sin q}{1-\cos q}\boldsymbol{\omega}\cdot\left(\mathbf{1}-\frac{\mathbf{qq}}{q^2}\right)+\boldsymbol{\omega}\wedge\mathbf{q}\right) \qquad (143)$$

The final topic we look at here is the derivation of the equation of motion for the direction cosine matrix. This plays a fundamental role in the interpretation of inertial measurements, their incorporation into a Kalman filter and their subsequent exploitation in several MoCo solutions (e.g. Holomax and Thales). To make closer contact with this work adopt a notation similar used in the nav literature (G. Minkler and J. Minkler, '*Aerospace Co-ordinate Systems and Transformations*', Magellan, 1990): we label the basis vectors **e** with a superscript denoting the frame of reference (A or B) they define. Thus the vector **v**, can be projected onto two different sets of basis vectors

$$\begin{aligned}\mathbf{v} &= \mathbf{v}\cdot\mathbf{e}_k^A\mathbf{e}_k^A = v_k^A\mathbf{e}_k^A \\ &= \mathbf{v}\cdot\mathbf{e}_k^B\mathbf{e}_k^B = v_k^B\mathbf{e}_k^B\end{aligned} \qquad (144)$$

As before we identify the direction cosines transforming from frame A to frame B as

$$\left(\mathbf{C}_A^B\right)_{jl} = \mathbf{e}_l^A\cdot\mathbf{e}_j^B \qquad (145)$$

and see straight away that

$$\left(\mathbf{C}_A^B\right)_{jl} = \left(\mathbf{C}_B^A\right)_{lj} \qquad (146)$$

If the frame A, at time zero, is rotated into frame B by time *t* we can regard

$$\left(\mathbf{C}_A^B(t)\right)_{jl} = \mathbf{e}_l^A\cdot\mathbf{e}_j^B(t) \qquad (147)$$

as a function of time and investigate its rate of change. The following identity lies at the heart of this equation of motion of the direction cosine matrix :

$$\left(\dot{\mathbf{C}}_A^B\cdot\mathbf{C}_B^A\right)_{lk} = \dot{\mathbf{e}}_l^B\cdot\mathbf{e}_j^A\ \mathbf{e}_j^A\cdot\mathbf{e}_k^B = \dot{\mathbf{e}}_l^B\cdot\mathbf{e}_k^B; \qquad (148)$$

here we have identified the identity operator in the form

$$\mathbf{e}_j^A\ \mathbf{e}_j^A = \mathbf{1}. \qquad (149)$$

The rate of change in the basis vector $\mathbf{e}_j^B$ as a result of its rotation with an angular momentum $\boldsymbol{\omega}$ is given by

$$\dot{\mathbf{e}}_j^B = \boldsymbol{\omega} \wedge \mathbf{e}_j^B \tag{150}$$

so that

$$\left(\dot{\mathbf{C}}_A^B \cdot \mathbf{C}_B^A\right)_{lk} = \left(\boldsymbol{\omega} \wedge \mathbf{e}_j^B\right) \cdot \mathbf{e}_k^B = \boldsymbol{\omega} \cdot \left(\mathbf{e}_j^B \wedge \mathbf{e}_k^B\right) \tag{151}$$

This is the required equation of motion of the direction cosine matrix. (Published derivations of the result can go on a bit. The aforementioned Minklers expend eight pages achieving as much as (148)-(151)) We should note that the basis vectors in frame *B* are used to establish the anti-symmetry in the matrix. Should we wish to refer things to the basis vectors in frame *A* we first recall that the scalar triple product of three vectors is invariant under a rotational transformation so that

$$\left(\mathbf{e}_j^B \wedge \mathbf{e}_k^B\right) = \left(\left(\exp(\mathbf{n}\theta \wedge)\mathbf{e}_j^A\right) \wedge \left(\exp(\mathbf{n}\theta \wedge)\mathbf{e}_k^A\right)\right) = \exp(\mathbf{n}\theta \wedge)\left(\mathbf{e}_j^A \wedge \mathbf{e}_k^A\right) \tag{152}$$

Thus we have

$$\boldsymbol{\omega} \cdot \left(\mathbf{e}_j^B \wedge \mathbf{e}_k^B\right) = \boldsymbol{\omega} \cdot \left(\exp(\mathbf{n}\theta \wedge)\left(\mathbf{e}_j^A \wedge \mathbf{e}_k^A\right)\right) = \left(\exp(-\mathbf{n}\theta \wedge)\boldsymbol{\omega}\right) \cdot \left(\mathbf{e}_j^A \wedge \mathbf{e}_k^A\right) \tag{153}$$

You are encouraged to derive this result more directly, using the equations of motion for $\mathbf{n}$ and θ obtained earlier.

Exercises

1. The following exercise is based on

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\cdot\mathbf{B}\cdot(\mathbf{A}+\mathbf{B})^{-1}$$

and should give some practice in manipulating matrices.

Prove that

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\cdot(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\cdot\mathbf{A}^{-1}$$

S and Σ are *N*X*N* matrices; **M** is an *N* dimensional vector, related by

$$\mathbf{S} = \mathbf{\Sigma} + \mathbf{M}\mathbf{M}^T .$$

Show that

$$\mathbf{S}^{-1} = \mathbf{\Sigma}^{-1} - \frac{\mathbf{\Sigma}^{-1}\mathbf{M}\mathbf{M}^T\mathbf{\Sigma}^{-1}}{\left(1+\mathbf{M}^T\mathbf{\Sigma}^{-1}\mathbf{M}\right)}$$

and deduce that

$$\mathbf{M}^T\mathbf{S}^{-1}\mathbf{M} = \frac{\mathbf{M}^T\mathbf{\Sigma}^{-1}\mathbf{M}}{1+\mathbf{M}^T\mathbf{\Sigma}^{-1}\mathbf{M}}$$
$$\mathbf{M}^T\mathbf{\Sigma}^{-1}\mathbf{M} = \frac{\mathbf{M}^T\mathbf{S}^{-1}\mathbf{M}}{1-\mathbf{M}^T\mathbf{S}^{-1}\mathbf{M}}$$

The second part to this question provides us with an example of the so-called matrix inversion lemma, which crops up a lot in the discussion of Kalman filters.

2 Confirm the following vector identities:

$$\nabla\cdot(\nabla\wedge\mathbf{A}) = 0$$
$$\nabla\wedge(\nabla\phi) = 0$$
$$\nabla(\phi\psi) = \phi\nabla\psi + \psi\nabla\phi$$
$$\nabla\cdot(\phi\mathbf{A}) = \mathbf{A}\cdot\nabla\phi + \phi\nabla\cdot\mathbf{A}$$
$$\nabla\wedge(\phi\mathbf{A}) = \phi\nabla\wedge\mathbf{A} + (\nabla\phi)\wedge\mathbf{A}$$
$$\nabla\cdot(\mathbf{A}\wedge\mathbf{B}) = \mathbf{B}\cdot(\nabla\wedge\mathbf{A}) - \mathbf{A}\cdot(\nabla\wedge\mathbf{B})$$
$$\nabla(\mathbf{A}\cdot\mathbf{B}) = (\mathbf{A}\cdot\nabla)\mathbf{B} + (\mathbf{B}\cdot\nabla)\mathbf{A} + \mathbf{A}\wedge(\nabla\wedge\mathbf{B}) + \mathbf{B}\wedge(\nabla\wedge\mathbf{A})$$
$$\nabla\wedge(\nabla\wedge\mathbf{A}) = \nabla(\nabla\cdot\mathbf{A}) - \nabla^2\mathbf{A}$$
$$\nabla\cdot(\phi\nabla\psi - \psi\nabla\phi) = \phi\nabla^2\psi - \psi\nabla^2\phi$$
$$\nabla\cdot\left(\mathbf{A}\wedge(\nabla\wedge\mathbf{B}) - \mathbf{B}\wedge(\nabla\wedge\mathbf{A})\right) = \mathbf{B}\cdot(\nabla\wedge\nabla\wedge\mathbf{A}) - \mathbf{A}\cdot(\nabla\wedge\nabla\wedge\mathbf{B})$$

These should provide practice in the use of the Cartesian tensor notation; flogging them out a component at a time would take far too long.

3. We noted that in the session that the integral of a function of a complex variable around a closed contour in the complex plane vanishes if the function is well behaved (analytic) inside the contour. Use this idea to evaluate

$$\int_C \frac{dz}{z}$$

when the contour of integration C (i) does not and (ii) does contain the origin. When might you expect

$$\frac{1}{2\pi i}\int_C \frac{f(z)}{z-z_0}dz = f(z_0)?$$

This result is known as Cauchy's theorem, and forms the basis of some very neat methods for evaluating definite integrals. For example the integral

$$\int_{-\infty}^{\infty} f(x)dx$$

can be attacked by considering the integral of $f(z)$ around a contour in the complex plane consisting of the real axis (this will give the answer we want) and a curve (e.g. a semi-circle in the complex plane) chosen to close the contour and to make a simple (hopefully zero) contribution to the total contour integral. This can be evaluated using Cauchy's theorem, picking up contributions from points inside the closed contour where $f(z)$ misbehaves. Fill out these sketchy details and evaluate

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2}, \quad \int_{-\infty}^{\infty} \frac{\cos(ax)}{b^2+x^2}dx, \quad \int_{-\infty}^{\infty} \frac{x\sin(ax)}{\left(b^2+x^2\right)^2}dx$$

The first of these should be familiar, and provide a check on what you've done, the second is a bit more taxing and the third is for masochists.

4. Show that the two matrices $\mathbf{S}^{-1}\mathbf{AS}$ and $\mathbf{A}$ have the same eigenvalues. Such a relationship between matrices with identical eigenvalues is often referred to as a similarity transformation.

Two matrices **A** and **B** do not in general commute i.e.

$$\mathbf{AB} \neq \mathbf{BA}.$$

Carry out a few simple numerical experiments to see if the eigenvalues of these two different products might be the same. If they are, can you find an apposite similarity transformation? How are your results related to those obtained in our discussion of the combination of rotations?

5. Assume that the matrices **A** and **B** do not commute but that their commutator nonetheless commutes with each of them i.e.

$$[\mathbf{A},\mathbf{B}] \neq \mathbf{0}; \quad [[\mathbf{A},\mathbf{B}],\mathbf{A}] = [[\mathbf{A},\mathbf{B}],\mathbf{B}] = \mathbf{0}$$

Show that

$$\exp(\mathbf{A})\exp(\mathbf{B}) = \exp(\mathbf{A} + \mathbf{B} + [\mathbf{A},\mathbf{B}]/2).$$

How is this result relevant to our discussion of rotations?

Hint: Render the hoped for result plausible by expanding each side up to quadratic terms and observing what the commutator in the exponent of the right hand side does. With one's confidence suitably boosted, prove that

$$[\exp(\mathbf{A}),\mathbf{B}] = \exp(\mathbf{A})[\mathbf{A},\mathbf{B}]$$

then find and solve a matrix differential equation satisfied by

$$\mathbf{G}(\lambda) = \exp(\lambda\mathbf{A})\exp(\lambda\mathbf{B}).$$