5 Matrices

5.1 Introduction

A matrix is a two-dimensional array of numbers with an arbitrary number of rows n and columns m. The standard way to describe a matrix dimensions is $\#rows \times \#columns$, i.e., $n \times m$. The standard way to label the individual elements of a matrix A is with subscripts in the form a_{rc} where r is the row number and c is the column number. For example, an MR image slice is just the patterns of intensities arranged in a 2D matrix, as shown in Figure 5.1. A column vector is really just a $n \times 1$ dimensional matrix, and a row vector is just a $1 \times m$ dimensional matrix. A matrix for which $n = m$ is calle a *square matrix*. The real utility of matrices is in viewing them as a collection of vectors, which then leads to their use in linear algebra. Casting equations in the form of matrices allows general conclusions to be made about the equations (and thus the physical systems they describe) directly from the general properties of the matrices. This can greatly simply analyses and provde a concise description complicated physical systems.

There are many ways to introduce and discuss matrices. For our purposes, this is best done by extending two ideas we defined for vectors - addition and the dot product. But first we define the generalization of the tranpose we defined for vectors.

$$A = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & & \\ \vdots & & \ddots & \\ a_{1n} & & & a_{nm} \end{pmatrix} =$$

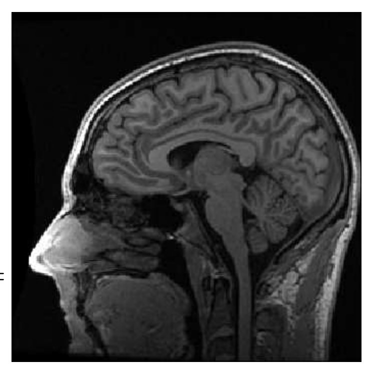
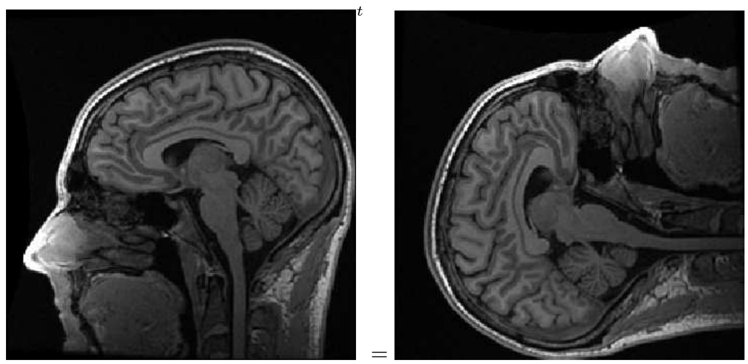


Figure 5.1 A 2D matrix.



(5.3)

Figure 5.2 The matrix transpose.

5.2 The matrix transpose

The matrix *transpose* is obtained by interchanging the rows with the columns and symbolized superscript t: the transpose of $\boldsymbol{A}$ is denoted $\boldsymbol{A}^t$. In component notation:

$$A^t_{ij} \equiv A_{ji} \tag{5.1}$$

For example, an arbitrary 3×2 matrix and its transpose is

$$\boldsymbol{A} = \begin{pmatrix} a & d \\ b & e \\ c & f \end{pmatrix} \quad , \quad \boldsymbol{A}^t = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \tag{5.2}$$

A graphical example of the matrix transpose is shown in Figure 5.2. Note that for a matrix of dimensions $n \times 1$ (i.e., a vector), the matrix transpose reduces to vector transpose Eqn 3.7. The matrix tranpose converts all the column vectors in the matrix to row vectors, and vise versa. The transpose of the transpose of matrix returns the original matrix:

$$(\boldsymbol{A}^t)^t \equiv \boldsymbol{A} \tag{5.4}$$

which is clear from both Eqn 5.2 and Figure 5.2.

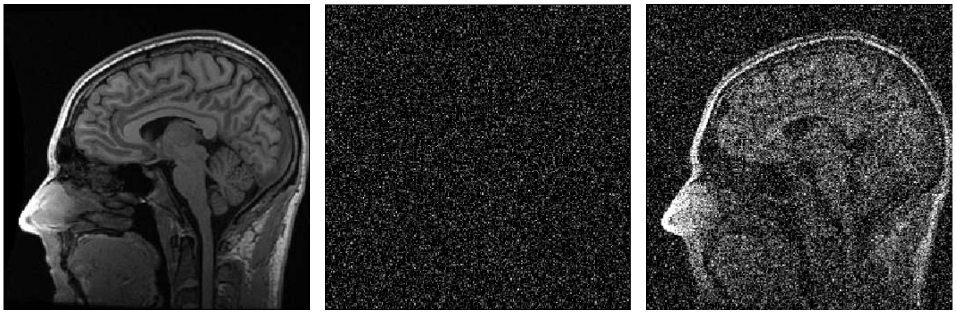
5.3 Matrix addition, subtraction, and scalar multiplication

Multiplying a matrix $\boldsymbol{A}$ by a scalar q just multiplies the individual elements of $\boldsymbol{A}$ by q:

$$q \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} q\,a_{11} & q\,a_{12} \\ q\,a_{21} & q\,a_{22} \end{pmatrix} \tag{5.5}$$

The sum of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ produces a third matrix $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ whose elements are the sum of the corresponding individual elements in $\boldsymbol{A}$ and $\boldsymbol{B}$. For example, for two general 2×2 matrices:

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix} \tag{5.6}$$



(a) The image matrix $\boldsymbol{A}$. (b) The noise matrix $\boldsymbol{B}$. (c) Image + noise matrix $\boldsymbol{A}+\boldsymbol{B}$.

Figure 5.3 Matrix addition. Artificially adding random noise to an image.

Similarly, matrix subtraction $\boldsymbol{C}=\boldsymbol{A}-\boldsymbol{B}$ produces a matrix similar to Eqn 5.6 with the elements subtracted rather than added. Clearly in order to sum (or subtract) like elements, matrix addition or subtraction requires that the two matrices are the same dimensions, and the resulting matrix will also be of the same dimension. A graphical example of matrix addition is shown in Figure 5.3. The transpose the sum of two matrices is the sum of the transpose of the individual matrices:

$$(\boldsymbol{A}+\boldsymbol{B})^t = \boldsymbol{A}^t + \boldsymbol{B}^t \tag{5.7}$$

5.4 The matrix dot or inner product

In Section 3.10 it was shown that the inner product of two vectors was obtained by multiplying the row elements in the first vector with the corresponding column elements in the second vector and summing the result. This required that the number of columns in the first vector was equal to the number of rows in the second vector. The matrix dot product just extends this to the multiple vectors within a matrix: the row vectors of the first (left) matrix multiply the columns of the second (right) matrix. This requires that the number of columns of the left matrix be the same as the number of rows of the right matrix.

For example, the multiplication of a general 3×3 matrix times a 3×1 matrix (i.e., a column vector) is

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + a_{13}v_3 \\ a_{21}v_1 + a_{22}v_2 + a_{23}v_3 \\ a_{31}v_1 + a_{32}v_2 + a_{33}v_3 \end{pmatrix} \tag{5.8}$$

where the components involved in the construction of element f_2 have been color-coded to visualize how it is constructed:

$$f_2 = a_{21}v_1 + a_{22}v_2 + a_{23}v_3 \tag{5.9}$$

Eqn 5.8 shows that the matrix dot product can be used to concisely express a systems of equations. In order for Eqn 5.11 to be true, all *three* equations must be satisfied at the same time. Thus Eqn 5.11 expresses a *simultaneous system of equations*.

A more general example is the inner product of a matrix $\boldsymbol{A}$ of dimension $[n \times m]$ with $\boldsymbol{B}$ of dimension $[m \times p]$. For $n = m = 3$ and $p = 2$

$$\underbrace{\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \\ f_{31} & f_{32} \end{pmatrix}}_{\boldsymbol{F}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}}_{\boldsymbol{B}} \tag{5.10}$$

where the components involved in the construction of element f_{22} have been color-coded to visualize how it is constructed:

$$f_{22} = a_{21}\, b_{12} + a_{22}\, b_{22} + a_{23}\, b_{32} \tag{5.11}$$

The first column of $\boldsymbol{F}$ is just Eqn 5.8 with $a \to b$ and $\{v_1, v_2, v_3\} \to \{b_{11}, b_{21}, b_{31}\}$. The second column of $\boldsymbol{F}$ is just Eqn 5.8 with $a \to b$ and $\{v_1, v_2, v_3\} \to \{b_{12}, b_{22}, b_{32}\}$. Notice the dimension of the final matrix $\boldsymbol{F}$ in relation to the two matrices that are multiplied:

$$\underbrace{\boldsymbol{F}}_{[n\times p]} = \underbrace{\boldsymbol{A}}_{[n\times m]}\, \underbrace{\boldsymbol{B}}_{[m\times p]} \tag{5.12}$$

It is useful to remember that the "inner" dimension m disappears and the "outer" dimensions n and p remain. The transpose the product of two matrices is the product the transpose of the individual matrices *in reverse order*:

$$(\boldsymbol{AB})^t = \boldsymbol{B}^t \boldsymbol{A}^t \tag{5.13}$$

This follows from the requirement for matrix multiplication that the number of columns of the left matrix be the same as the number of rows of the right matrix.

A useful example is the dot product of two-dimensional rotation matrix $\boldsymbol{R}(\theta)$ with a coordinate vector $\boldsymbol{v} = \{x, y\}^t$:

$$\underbrace{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}}_{\boldsymbol{R}(\theta)} \cdot \underbrace{\begin{pmatrix} x \\ y \end{pmatrix}}_{\boldsymbol{v}} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ y\cos\theta + x\sin\theta \end{pmatrix} \tag{5.14}$$

This dot product rotates the vector in the counter-clockwise direction about the origin by an angle θ, as shown in Figure 5.4. Writing the rotation in the form $\boldsymbol{R}\theta \cdot \boldsymbol{v}$ rather than explicitly writing out the components makes expressions not only much cleaner, but ultimately more intuitively clear.

Problems

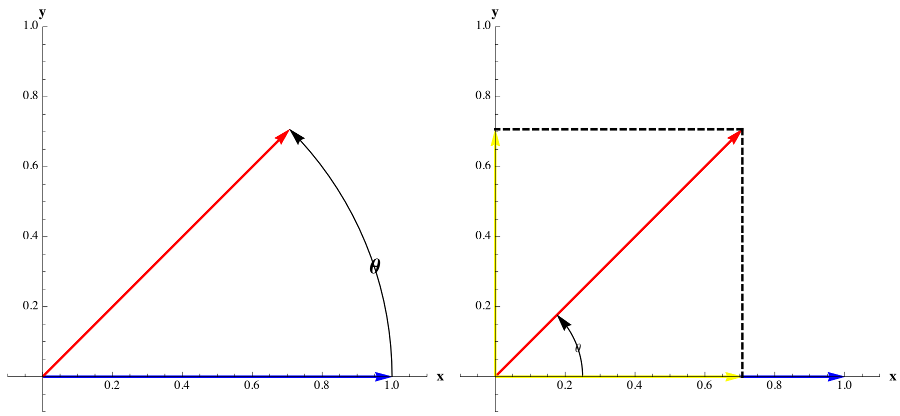
5.1 Write out the components of $\boldsymbol{F}$ in Eqn 5.10.

Problems

5.2 Demonstrate Eqn 5.13 using the two matrices in Eqn 5.2.

5.5 Diagonal matrices

A special subset of matrices that will play an important role in many of the applications we will discuss is one in which only the diagonal elements $\{a_{ii}\}$ of the matrix are non-zero. For example,



(a) Vector is rotated by θ.

(b) The new x and y components are shown in yellow.

Figure 5.4 The dot product of the rotation matrix Eqn 5.14 on a vector originally aligned along the x-axis for $\theta = 45°$.

a diagonal 3×3 matrix has the form:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \tag{5.15}$$

The dot product of a diagonal matrix with a vector has a unique effect, which we see in the case of a general 3×3 system of equations:

$$\boldsymbol{A} \cdot \boldsymbol{v} = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} a_{11}v_1 \\ a_{22}v_2 \\ a_{33}v_3 \end{pmatrix} \tag{5.16}$$

The resulting vector is the original vector scaled by the diagonal elements. The important point is that none of the original vector components are combined - they remain as independent elements in the resulting vector. For example, if the vector $\boldsymbol{v}$ represented orthogonal coordinates such as the Cartesian unit vectors $\boldsymbol{v} = \{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$, then multiplication by a diagonal vector would independently scale the components along these coordinate axes.

An important special case of the diagonal matrices is one in which all the diagonal elements are 1:

$$\boldsymbol{I} = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} \tag{5.17}$$

The matrix $\boldsymbol{I}$ is called the *identity matrix* because any vector multiplied by it returns unchanged:

$\boldsymbol{I}\boldsymbol{v} = \boldsymbol{v}$.

$$\boldsymbol{I} \cdot \boldsymbol{x} = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \boldsymbol{x} \tag{5.18}$$

Problems

5.3 Explicitly demonstrate the truth of Eqn 5.18.

5.6 The matrix trace

The *trace* of an $n \times n$ matrix $\boldsymbol{A}$, denoted $\mathrm{Tr}(\boldsymbol{A})$, is the sum of the diagonal elements:

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i^n a_{ii} \tag{5.19}$$

What might the trace be good for? Consider a special case of section 5.5 where the vector components along the x, y and z directions are scaled by the constants $\{a_1, a_2, a_3\}$:

$$\boldsymbol{A} \cdot \boldsymbol{v} = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_1 x \\ a_2 y \\ a_3 z \end{pmatrix} \tag{5.20}$$

The trace of $\boldsymbol{A}$ is then $\mathrm{Tr}(A) = (a_1 + a_2 + a_3) = 3\langle a \rangle$ where $\langle a \rangle = (a_1 + a_2 + a_3)/3$ is the average value of the scaling coefficients. We will encounter this use of the trace much later when we consider the diffusion tensor and find that this trace is proportional to the average diffusion coefficient.

The trace is only defined for square (i.e., $n \times n$) matrices. Since the diagonal elements of a square matrix $\boldsymbol{A}$ do not change when we take its transpose, it is clear that the trace of a matrix is equal to the trace of the transpose of that matrix:

$$\mathrm{Tr}(\boldsymbol{A}) = \mathrm{Tr}(\boldsymbol{A}^t) \tag{5.21}$$

The trace of the sum of two matrices is the sum of the trace of these two matrices, since

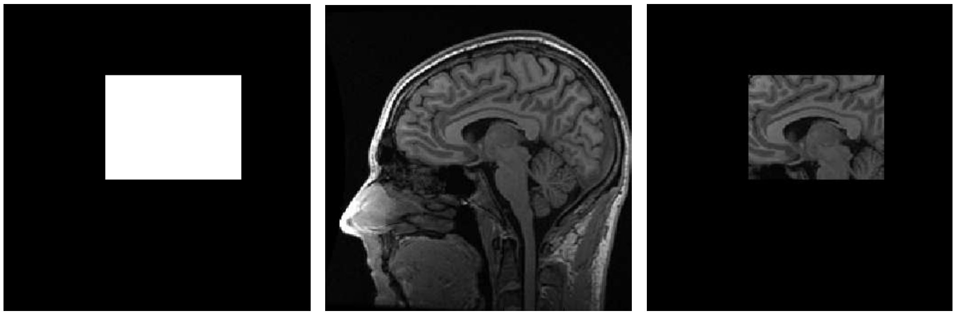
$$\mathrm{Tr}(\boldsymbol{A} + \boldsymbol{B}) = \sum_i (\boldsymbol{A} + \boldsymbol{B})_{ii} = \sum_i a_{ii} + \sum_i b_{ii} = \mathrm{Tr}(\boldsymbol{A}) + \mathrm{Tr}(\boldsymbol{B}) \tag{5.22}$$

And the trace of a matrix $\boldsymbol{A}$ times a constant c is just c times the trace of $\boldsymbol{A}$, since

$$\mathrm{Tr}(c\boldsymbol{A}) = \sum_i c a_{ii} = c \sum_i a_{ii} = c\mathrm{Tr}(\boldsymbol{A}) \tag{5.23}$$

The trace also has an interesting property that will prove very useful in Chapter 8. Consider the trace of the product three $n \times n$ matrices $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}\}$:

$$\begin{aligned} \mathrm{Tr}(\boldsymbol{ABC}) &= \sum_i (\boldsymbol{ABC})_{ii} = \sum_i \sum_j \sum_k a_{ij} b_{jk} c_{ki} \\ &= \sum_i \sum_j \sum_k c_{ki} a_{ij} b_{jk} = \mathrm{Tr}(\boldsymbol{CAB}) \end{aligned} \tag{5.24}$$



(a) $\boldsymbol{A}$ (black = 0, white = 1). (b) $\boldsymbol{B}$. (c) $\boldsymbol{AB} = A_{ij}\boldsymbol{B}$.

Figure 5.5 The Hadamard product.

Thus we see that the trace in invariant to moving the last matrix to the place of the first and moving every other matrix forward on place. This would also have worked by moving the first to the end (i.e., $\mathrm{Tr}(\boldsymbol{BCA})$) and moving every other matrix back a space. The important thing is to keep the matrices in the correct order relative to one another, with the last (or first, depending on which way you move them, right or left, respectively) cycling back to the beginning (or the end). This is called *cyclic permutations* of the matrices. Therefore we have the important quality of the trace that it is *invariant to cyclic permutations of the matrices.*

5.7 Other forms of matrix multiplication

There are actually several ways to multiply matrices. Usually, the context makes the appropriate multiplication clear, but it can sometimes be confusing. Generally, "standard" matrix multiplication means the matrix inner (or dot) product, which is just the multiplication of the row and column vectors, as previously described, for each combination of row and column vectors.

The *Hadamard product*, denoted by the symbol ”$\circ$”, is defined between two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the same size, say $n \times m$, is the matrix formed from the product of the individual elements:

$$\boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{B}$$

or

$$C_{ij} = A_{ij}B_{ij} \tag{5.25}$$

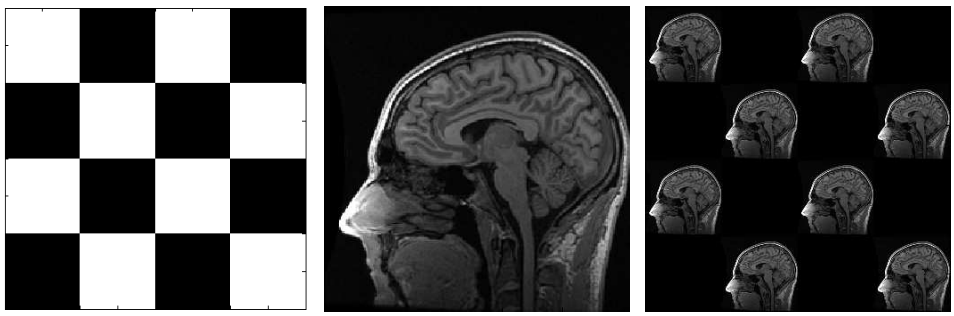
For example,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix} \tag{5.26}$$

Note that both matrices must be of the same size to form the Hadamard product. A graphical example of the Hadamard product is shown in Figure 5.5.

The Hadamard product is the component-wise multiplication of two matrices. If we add up all these products of Eqn 5.25, that is,

$$\boldsymbol{A} : \boldsymbol{B} \equiv \sum_i \sum_j C_{ij} = \sum_i \sum_j A_{ij}B_{ij} \tag{5.27}$$



(a) $\boldsymbol{A}$ (black = 0, white = 1). (b) $\boldsymbol{B}$. (c) $\boldsymbol{A} \otimes \boldsymbol{B} = A_{ij} B_{ij}$.

Figure 5.6 The kronecker product.

we get what is called the *Frobenius inner product*, denoted by the symbol ":". It is the component-wise inner product of two matrices and so is just as if we were treating the matrices as vectors.

The *Kronecker product* of two matrices, denoted by the symbol $\otimes$, is formed from multiplying each element (i.e., the scalar product) of the left matrices with the *entire* right matrix:

$$\begin{aligned} \boldsymbol{C} &= \boldsymbol{A} \otimes \boldsymbol{B} \\ &\text{or} \\ C_{ij} &= A_{ij} \boldsymbol{B} \end{aligned} \tag{5.28}$$

For example

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} & a_{12} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ a_{21} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} & a_{22} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \end{pmatrix} \tag{5.29}$$

$$= \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix} \tag{5.30}$$

The Kronecker product of an $m \times n$ matrix $\boldsymbol{A}$ by of an $o \times p$ matrix $\boldsymbol{B}$ results in a matrix that is of size $mo \times np$. The Kronecker product is actually a special case of a more general product, called the *tensor product*. For the special case that the two matrices are vectors, the tensor product is referred to as the *outer product*. An example of the Kronecker product is shown in Figure 5.6.

5.8 Associative, Distributive, but not necessarily Commutative

Both scalar and vectors obey the three basic properties: they are associative, distributive, and commutative. Similarly, matrix multiplication is *associative*

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}) \tag{5.31}$$

and matrix multiplication is *distributive*

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC} \tag{5.32}$$

However, the important different between matrices and vectors is that matrices do not necessarily commute so that generally multiplication is *non-commutative*:

$$\boldsymbol{AB} \neq \boldsymbol{BA} \qquad \text{usually} \tag{5.33}$$

This actually make sense intuitively, since in $\boldsymbol{A} \cdot \boldsymbol{B}$ the rows of $\boldsymbol{A}$ are multiplied by the columns of $\boldsymbol{B}$ whereas in $\boldsymbol{B} \cdot \boldsymbol{A}$ the rows of $\boldsymbol{B}$ are multiplied by the columns of $\boldsymbol{A}$, and there's no reason to believe those should be equal. A simple example illustrates this:

$$\begin{aligned} \boldsymbol{A} &= \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \quad , \quad \boldsymbol{B} = \begin{pmatrix} 0 & c \\ 0 & d \end{pmatrix} \\ \boldsymbol{A} \cdot \boldsymbol{B} &= \begin{pmatrix} 0 & ac \\ 0 & bc \end{pmatrix} \quad , \quad \boldsymbol{B} \cdot \boldsymbol{A} = \begin{pmatrix} bc & 0 \\ bd & 0 \end{pmatrix} \end{aligned} \tag{5.34}$$

In this case we say that $\boldsymbol{A}$ and $\boldsymbol{B}$ *do not commute*. This has important implications in many applications, perhaps most obviously in the matrix representation of rotations which do not commute and this tells us that the order in which rotations are performed changes the final results (the direction of a rotated vector, for example).

One matrix that does commute with every matrix, however, is the identity matrix. For an arbitrary $n \times n$ matrix $\boldsymbol{A}$ the multiplication with the $n \times n$ identity matrix is

$$\boldsymbol{IA} = \boldsymbol{AI} = \boldsymbol{A} \tag{5.35}$$

So $\boldsymbol{I}$ and $\boldsymbol{A}$ always *commute*.

Problems

5.4 Prove Eqn 5.35.

5.9 The matrix determinant

If a matrix $\boldsymbol{A}$ is square (i.e., of dimensions $n \times n$), then a quantity called the *determinant* can be defined, and is symbolized by either $\det(\boldsymbol{A})$ or $|\boldsymbol{A}|$. There are several very important uses of the determinant that are not readily apparent from its definition. But in order to understand these uses, we must first give the definition, as abstruse as it might be, in order to derive some illustrative examples.

The determinant of matrix $\boldsymbol{A}$ can be computed using any row i from the expression

$$\det(\boldsymbol{A}) = \sum_j a_{ij} C_{ij} \tag{5.36}$$

where the *cofactor* C_{ij} is the determinant of the *minor* M_{ij} multiplied by a sign that is positive or negative depending on whether or not the sum $i+j$ is even or odd:

$$C_{ij} = (-1)^{i+j}\,|M_{ij}| \tag{5.37}$$

The minor M_{ij} is just A_{ij} with the i'th row and j'th column deleted. This all sounds very complicated but visually we can see it is quite easy, because the above formula allows us to build up the determinant of a matrix of size $n \times n$ from the the determinant of a matrix of size $(n-1) \times (n-1)$, so it can be built up in successive steps. We begin with the determinant of a 1×1 matrix $A = a_{11}$, which is just $\det(\boldsymbol{A}) = a_{11}$. From Eqn 5.36, the determinant of a general two-dimensional matrix is (using $i = 1$):

$$\det \boldsymbol{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \tag{5.38}$$

The meaning of this can be made clearer if we consider two 2-dimensional column vectors

$$\boldsymbol{v}_1 = \begin{pmatrix} a \\ c \end{pmatrix} \quad , \quad \boldsymbol{v}_2 = \begin{pmatrix} b \\ d \end{pmatrix} \tag{5.39}$$

shown in Figure 5.7a. The matrix formed with these vectors as its columns is

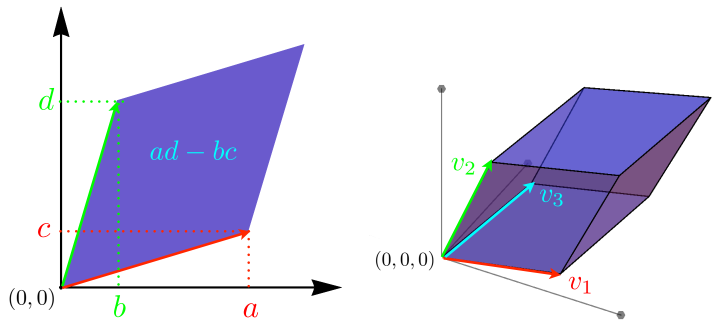
$$\boldsymbol{A} = (\boldsymbol{v}_1, \boldsymbol{v}_2) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{5.40}$$

The determinant of $\boldsymbol{A}$ is $\det(\boldsymbol{A}) = ad - bc$. The absolute value of this is just the area of the parallelogram formed from the two vectors, as seen in Figure 5.7a. We specify that it is the absolute value that is the area because the determinant can be negative. If the angle between the vectors is defined in a clockwise direction, this determinant is positive, but if it turns in a clockwise direction, the determinant is negative. Thus the determinant is coordinate system dependent and *orientation preserving*. The fact that this determinant is an orientation preserving area of a parallelogram suggests a connection between the determinant and the cross product. We make this connection in Section 5.11.

For a 3×3 matrix, the determinant is

$$\begin{aligned} \det \boldsymbol{A} &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix} \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \end{aligned} \tag{5.41}$$

So the 3×3 is just computed in terms of the 2×2 determinant, which we know how to compute from Eqn 5.38. The geometrical interpretation of the determinant of a 3×3 matrix is volume of a parallelopiped defined by the three column vectors of $\boldsymbol{A}$, as shown in Figure 5.7b. It is clear from Eqn 5.41 that the determinant of an $n \times n$ matrix can be written in terms of the determinants of the $(n-1)$-dimensional submatrices. The geometric interpretation for an $n \times n$ dimensional matrix is that the determinant is the volume of an n-dimensional parallelopiped



(a) The absolute value of the determinant of a 2×2 real matrix is the area of the parallelogram defined by the four entries.

(b) The absolute value of the determinant of a 3×3 real matrix is the volume of the parallelopiped defined by the nine entries.

Figure 5.7 The geometry of the matrix determinant.

constructed from, or *spanned*, by the column vectors. The sign of the determinant will depend on the orientation of the column vectors.

The interpretation of the determinant as a volume is quite useful in one particular but very common application - the change of coordinates in integral equations. The change of the volume element (or *measure*) of an integral with the change of coordinate system representation (e.g., from Cartesian to spherical) is computed from the determinant of a matrix (the *Jacobian matrix*) that describes the variations in new coordinates with respect to the old coordinates. This will be discussed in Section **??**.

A few important properties of the determinant of a matrix are that it is equal to the determinant of its transpose:

$$\det \boldsymbol{A} = \det \boldsymbol{A}^t \quad , \tag{5.42}$$

the determinant of the inverse is the inverse of the determinant:

$$\det \boldsymbol{A}^{-1} = (\det \boldsymbol{A})^{-1} \quad , \tag{5.43}$$

that the product of two $n \times n$ matrices is the product of their determinants:

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A}) \det(\boldsymbol{B}) \quad , \tag{5.44}$$

and because $\det(\boldsymbol{A})$ and $\det(\boldsymbol{B})$ are just numbers, then $\det(\boldsymbol{A}) \det(\boldsymbol{B}) = \det(\boldsymbol{B}) \det(\boldsymbol{A})$ from which it can conclude that

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{B}) \det(\boldsymbol{A}) = \det(\boldsymbol{BA}) \tag{5.45}$$

so that the order matrices are multiplied does not matter in computing the determinant of their product.

The determinant also plays an important role in the theory of linear equations. As we saw above, a system of n simultaneous linear equations can be represented by the matrix equation $\boldsymbol{Ax} = \boldsymbol{b}$ where $\boldsymbol{A}$ is an $n \times n$ matrix and $\boldsymbol{x} = \{x_1, \ldots, x_n\}^t$ is the column vector of variables.

Cramer's Rule ((**?**)) states that the the components x_j of the solution vector $\boldsymbol{x}$ are given by

$$x_i = \frac{\det \boldsymbol{B}_i}{\det \boldsymbol{A}} \tag{5.46}$$

where $\boldsymbol{B}_i$ is $\boldsymbol{A}$ with the i'th column replaced by $\boldsymbol{b}$. Notice that the denominator in Eqn 5.46 contains $\det \boldsymbol{A}$ which therefore must be non-zero in order for a solution to exist. This is perhaps the most important property of the determinant: it is a test to see if there is a solution to a system of equations. A matrix whose determinant is zero is said to be *singular*. A non-homogeneous system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has a unique solution if and only if $\det \boldsymbol{A}$ is non-singular (i.e., not zero).

5.10 Interlude: The Jacobian determinant

Because there are different coordinate systems to describe the same situation, we need to know how to transform between them. As we saw in Section 2, a physical situation can be described in different coordinate systems, and it is often necessary to transform from one to another. This is often motivated because there is a "natural" coordinate system for the problem. Describing the weather on the surface of the Earth suggests a spherical coordinate system, for example. Now, it might seem pretty straightforward to just pick a coordinate system and then go about describing the system within it. For example, in three-dimesions, you pick the Cartesian system $\{x, y, z\}$ or you pick the spherical system $\{r, \vartheta, \varphi\}$, and be done with it. Transformation between coordinate systems can be done by representing the parameters of one in terms of the other. For example, the Cartesian coordinates in terms of the spherical coordinates are:

$$x = r \sin\vartheta \cos\varphi \tag{5.47a}$$
$$y = r \sin\vartheta \sin\varphi \tag{5.47b}$$
$$z = r \cos\varphi \tag{5.47c}$$

whereas the spherical coordinates in terms of the Cartesian coordinates are

$$r = \sqrt{x^2 + y^2 + z^2} \tag{5.48a}$$
$$\vartheta = \arctan\left(\sqrt{x^2 + y^2}/z\right) \tag{5.48b}$$
$$\varphi = \arctan(y/x) \tag{5.48c}$$

However, it turns out that there is a subtle but very important thing that happens when you go from one coordinate system to another: the density of points changes. And this turns out to be of profound importance to MRI. We'll encounter this several times, in analyzing the spatial distortions from curving flow (Section 20.8), in moving from the Cartesian coordinates used standard formulation of the diffusion tensor to a spherical desription (Section **??**), in designing spiral trajectories for efficient diffusion imaging (Section 22.8), and in assessing the image distortion caused by field inhomogeneities in echo planar imaging (Section 30.1). This subtle issue arises in the following way. Imagine that we have a problem that involves the following integral, described in terms of the n coordinates $\boldsymbol{x} = \{x_1, \ldots, x_n\}$:

$$I = \int \boldsymbol{f}(\boldsymbol{x}) d\boldsymbol{x} \tag{5.49}$$

And now we want to transform this to another coordinate system described by the n new coordinates $\boldsymbol{u} = \{u_1, \ldots, u_n\}$. Our original coordinates $\boldsymbol{x}$ are a function of the new coordinates: $\boldsymbol{x}(\boldsymbol{u})$. Thus to go to the new coordinates we write Eqn 5.49 in the new coordinates is

$$I = \int \boldsymbol{f}\left[\boldsymbol{x}(\boldsymbol{u})\right] \left|\frac{d\boldsymbol{x}}{d\boldsymbol{u}}\right| d\boldsymbol{u} \tag{5.50}$$

where

$$\frac{d\boldsymbol{x}}{d\boldsymbol{u}} \equiv \boldsymbol{J}(\boldsymbol{u}) \equiv \begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_n} \end{pmatrix} \tag{5.51}$$

is called the *Jacobian* of the transformation. This is usually denoted by the shorthand notation

$$\boldsymbol{J} \equiv \frac{\partial\left(x_1, \ldots, x_n\right)}{\partial\left(u_1, \ldots, u_n\right)} \tag{5.52}$$

Eqn 5.50 is called the *Change of Variables Theorem* (**?**) and involves the determinant (cf Section 5.9) of the Jacobian of the transformation from $\boldsymbol{x}$ to $\boldsymbol{u}$, which is called *Jacobian determinant*. The interpretation of $|\boldsymbol{J}|$ is the following. The symbol where $d\boldsymbol{x}$ in Eqn 5.49 can be written as

$$d\boldsymbol{x} = dx_1 \ldots dx_n \tag{5.53}$$

where dx_i symbolizes an infinitesimal line element along the i'th coordinate direction. The symbol $d\boldsymbol{x}$ is called the *measure* and is the volume of an infinitesimal element in that space, created from infinitesimal elements along all the coordinate axes (think of three dimensions - an infinitesimal cube is Cartesian coordinates is just $dx\, dy\, dz$). Now, if we want to switch to new coordinates $\boldsymbol{u} = \{u_i\}, i = 1, \ldots, n$, the integral becomes

$$I = \int \boldsymbol{f}(\boldsymbol{u})\, |\boldsymbol{J}|\, d\boldsymbol{u} \tag{5.54}$$

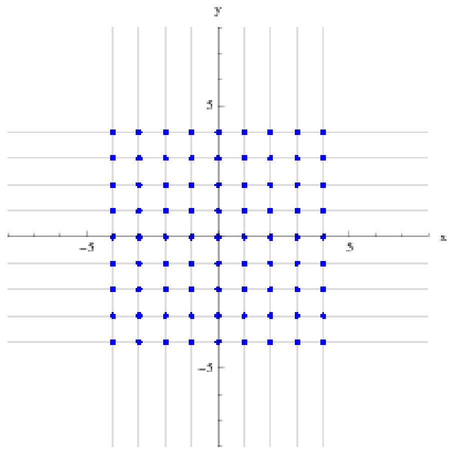
The new volume element $|\boldsymbol{J}|\, d\boldsymbol{u}$ is the infinitesimal volume element $d\boldsymbol{u} = du_1 \ldots du_n$ of the new coordinates $\boldsymbol{u}$, times a scaling factor $|\boldsymbol{J}|$.

The fact that the Jacobian determinant arises as a volume element is a consequence of the fact that the determinant of an $n \times n$ matrix is the volume of an n-dimensional parallelepiped (**?**). Changes in variables stretch the infinitesimal volume element and the Jacobian determinant takes into account the subsequent change in volume this stretching produces. An example is in the comparison of the Cartesian grid with the two-dimensional slice through a spherical grid, also called the *polar grid*, shown in Eqn 5.8. It is clear that the density of points in the Cartesian grid is constant, whereas there is a greater concentration of points near the origin of the polar grid. The Jacobian determinant reflects this density variation.

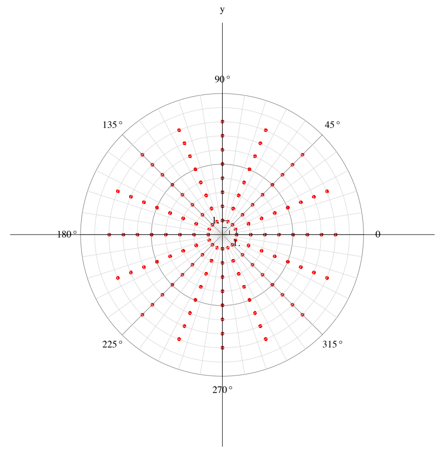
An important case that will occur frequently in this book is when we want to go from the Cartesian coordinate system with coordinate variables $\{x, y, z\}$ to the Spherical coordinate system with coordinate variables $\{r, \vartheta, \varphi\}$. We will then encounter integrals for which Eqn 5.54 looks like:

$$\int \boldsymbol{f}(x, y, z) dx\, dy\, dz = \int \boldsymbol{f}(r, \vartheta, \varphi)\, |\boldsymbol{J}|\; dr\, d\vartheta\, d\varphi \tag{5.55}$$

This is an important case, so let's go ahead and calculate the Jacobian determinant for this



(a) Cartesian grid.



(b) Polar grid.

Figure 5.8 Points on a Cartesian grid are shown in blue, and those on a (planar) sphereical grid are shown in red.

transformation. The Jacobian in Eqn 5.51 becomes

$$\boldsymbol{J}(r,\vartheta,\varphi) \equiv \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r} \\ \frac{\partial x}{\partial \vartheta} & \frac{\partial y}{\partial \vartheta} & \frac{\partial z}{\partial \vartheta} \\ \frac{\partial x}{\partial \varphi} & \frac{\partial y}{\partial \varphi} & \frac{\partial z}{\partial \varphi} \end{pmatrix} \tag{5.56}$$

Writing $\{x, y, z\}$ in spherical coordinates $\{r\sin\vartheta\cos\varphi, r\sin\vartheta\sin\varphi, r\cos\vartheta\}$, and taking the appropriate derivatives, the Jacobian becomes

$$\boldsymbol{J}(r,\vartheta,\varphi) \equiv \begin{pmatrix} \sin\vartheta\cos\varphi & r\cos\vartheta\cos\varphi & -r\sin\vartheta\sin\varphi \\ \sin\vartheta\sin\varphi & r\cos\vartheta\sin\varphi & r\sin\vartheta\cos\varphi \\ \cos\vartheta & -r\cos\vartheta & 0 \end{pmatrix} \tag{5.57}$$

Computing the determinant of Eqn 5.57 we get the Jacobian determinant of the transformation from Cartesian coordinates to Spherical coordinates:

$$|\boldsymbol{J}(r,\vartheta,\varphi)| = r^2 \sin\vartheta \tag{5.58}$$

This tells us that an infinitesimal volume elements in Cartesian coordinates is related to an infinitesimal volume in spherical coordinates by

$$dx\,dy\,dz = r^2 \sin\vartheta\, dr\, d\vartheta\, d\varphi \tag{5.59}$$

You will see this factor repeatedly in our discussion of the analysis of high angular resolution DTI data when it will become advantageous to switch from a Cartesian to a Spherical coordinate system.

One very practical manifestation of the change in density of points occurs when curved flow is present in MR images which causes artifactually bright regions in vessels (such as the carotid artery) to appear in images. An example of this effect is shown in the numerical simulation in Figure 5.9, and is discussed in greater detail in Section 20.8

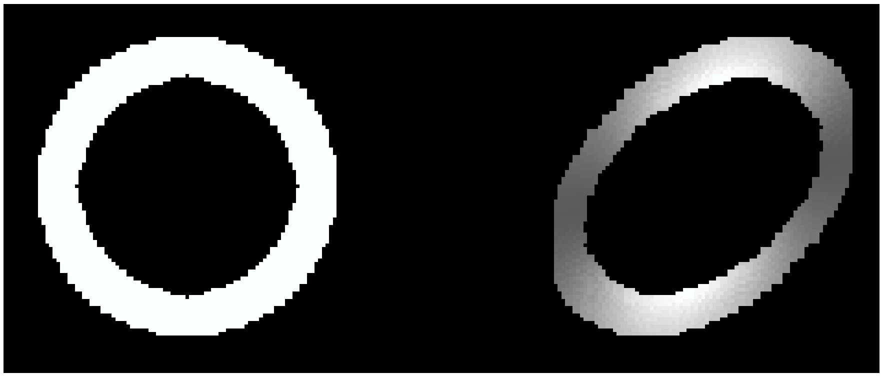


Figure 5.9 Simulated flow in a helical tube. On the left is stationary (non-flowing) water. On the right the water is flowing. This causes the circular object to distort in shape into an ellipse, and the intensities to appear spatially non-uniform. These distortions can be described by the Jacobian determinant.

5.11 Interlude: The matrix form of the cross product

The cross product introduced in Section 3.14 appears frequently in MRI theory and it has very useful matrix representations that facilitate compact expressions of several important equations (such as the Bloch equations).

For two vectors $\{\boldsymbol{u}, \boldsymbol{v}\} \in \mathbb{R}^3$

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \quad , \quad \boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{5.60}$$

the cross product of $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix} \tag{5.61}$$

This can also be written in terms of the matrix multiplication

$$\boldsymbol{u} \times \boldsymbol{v} = \boldsymbol{A} \cdot \boldsymbol{v} \tag{5.62}$$

where

$$\boldsymbol{A} = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \tag{5.63}$$

In the previous section we suggested a connection between the matrix determinant and the vector cross product introduced in Section 3.14. Making this connection gives us another matrix representation of the cross product, this one involving the matrix determinant. With the unit vectors along the Cartesian axes $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$, we form the matrix

$$\boldsymbol{M} = \begin{pmatrix} \hat{\boldsymbol{x}} & \hat{\boldsymbol{y}} & \hat{\boldsymbol{z}} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix} \tag{5.64}$$

Then the cross product is just the determinant:

$$\boldsymbol{u} \times \boldsymbol{v} = |\boldsymbol{M}| \tag{5.65}$$

Note that in this case the determinant $|\boldsymbol{M}|$ is not a number (scalar) but a vector. Intuitively, the cross product suggests a connection with rotations, which will be explored in Section **??**.

5.12 The matrix inverse

For the simple equation $x = ay$ where a is a scalar constant, the solution for y is just $y = x/a$. A more general notation for this is $y = a^{-1}x$ where the $^{-1}$ represents the *inverse* of a for which $a^{-1} = 1/a$ and $a^{-1}a = 1$ defines the inverse operation. Similarly, matrix *inverse*, denoted $\boldsymbol{A}^{-1}$, is the matrix that, when multiplied by the original matrix, produces the identity matrix: $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.

In one special kind of matrix the inverse *is* very similar to the scalar inverse, and that is the diagonal matrix, where the inverse is

$$\boldsymbol{A}^{-1} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}^{-1} = \begin{pmatrix} 1/a & 0 \\ 0 & 1/b \end{pmatrix} \tag{5.66}$$

That is, the inverse is just found by taking the scalar inverse of each element. But this is a special case. For a general matrix $\boldsymbol{A}$ (with elements a_{ij}) the computation of $\boldsymbol{A}^{-1}$ is *not* found by taking $1/a_{ij}$. For example, the inverse of the general 2×2 matrix is

$$\boldsymbol{A}^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\underbrace{ad - bc}_{|\boldsymbol{A}|}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{5.67}$$

Note that the determinant of $\boldsymbol{A}$ is the denominator in Eqn 5.67. This suggests that if the determinant of $\boldsymbol{A}$ is zero, it is not possible to calculate the inverse. Indeed, there are two very important properties of the matrix determinant:

1. If $|\boldsymbol{A}| \neq 0$ then A is invertible.
2. If $|\boldsymbol{A}| = 0$ then A is not invertible.

A matrix that is not invertible is called a *singular* or *degenerate* matrix. The invertibility of a matrix is a critical issue in many application, the most ubiquitous perhaps being the solution to a system of linear equation $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$. The solution is $\boldsymbol{x} = (\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{b}$ which involves the inverse of the matrix $\boldsymbol{B} \equiv \boldsymbol{A}^t\boldsymbol{A}$.

There is a distributive property of the matrix inverse like that for the transpose (Eqn **??**): the inverse of the product of two matrices, is the product of the inverse of the individual matrices, rearranged in reverse order.

$$(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \tag{5.68}$$

From this we can also see that if two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are invertible, then their product $\boldsymbol{AB}$ is also invertible. Eqn 5.44 leads to the useful property for an invertible matrix $\boldsymbol{A}$ that

$$\det(\boldsymbol{A})\det(\boldsymbol{A^{-1}}) = \det(\boldsymbol{AA^{-1}}) = \det \boldsymbol{I} = 1 \tag{5.69}$$

5.13 Calculating the matrix inverse*

The analytic solution for the matrix inverse is

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|} \underbrace{\begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & & & \vdots \\ \vdots & & & \\ C_{1n} & & \ldots & C_{nn} \end{pmatrix}}_{\mathfrak{C}^t} \tag{5.70}$$

The inverse involved the *transpose* of the *cofactor matrix* $\boldsymbol{\mathfrak{C}}^t$, which is called the *adjugate matrix*, and contains the cofactors

$$\mathfrak{C}_{ij} = (-1)^{i+j}\,|M_{ij}| \tag{5.71}$$

where the *minor* $|M_{ij}|$ of an $n \times n$ matrix $\boldsymbol{A}$ is defined as the determinant of the $(n-1) \times (n-1)$ matrix formed by removing the i'th row and the j'th column from $\boldsymbol{A}$. For example, for the matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \tag{5.72}$$

the cofactor elements $\boldsymbol{\mathfrak{C}}_{32} = (-1)^{3+2} M_{32}$ where

$$M_{32} = \begin{vmatrix} 1 & \square & 3 \\ 4 & \square & 6 \\ \square & \square & \square \end{vmatrix} = \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} = (1)(6) - (4)(3) = -6 \tag{5.73}$$

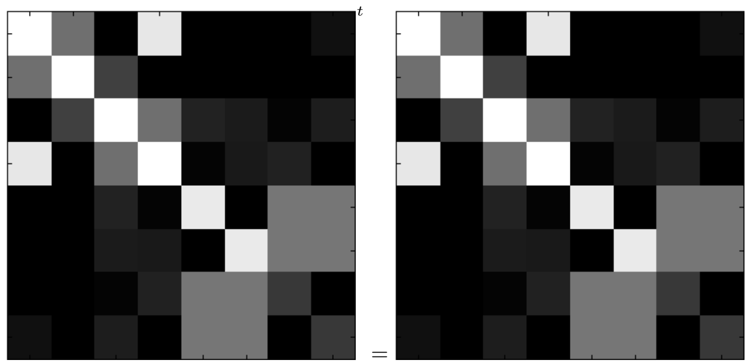
so that $\boldsymbol{\mathfrak{C}}_{32} = 6$. In practice, however, the matrix inverse is usually just calculated numerically, as there are fast algorithms to do so.

5.14 The matrix rank

If a matrix $\boldsymbol{A}$ is singular, then the set of simultaneous equations $\boldsymbol{Ax} = \boldsymbol{b}$ can be decomposed in an important way. Since $\boldsymbol{A}$ is singular, then there is a subspace of $\boldsymbol{A}$, called the *nullspace* of $\boldsymbol{A}$, for which $\boldsymbol{Ax} = \boldsymbol{0}$. The dimension of the nullspace is called the *nullity* of $\boldsymbol{A}$. The remainder of the space of $\boldsymbol{A}$ that is not zero, the subspace of $\boldsymbol{b}$ that is not zeros, is called the *range* of $\boldsymbol{A}$. Its dimension is called the *rank* of $\boldsymbol{A}$. The dimension of $\boldsymbol{A}$ is the dimension of its range and its nullspace , i.e.,

$$\text{Dimension}\,[\boldsymbol{A}] = \text{Rank}\,[\boldsymbol{A}] + \text{Nullity}\,[\boldsymbol{A}] \tag{5.74}$$

An important concept related to the rank concerns how many "unique" vectors there are in a system of equations. If an equation can be constructed from the sum of a set of vectors



(5.79)

Figure 5.10 A symmetric matrix and its transpose.

$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ with coefficients $\{a_1, \ldots, a_k\}$ such that the sum is not zero unless all the coefficients are zero, i.e.

$$\sum_{i=1}^{k} a_i \boldsymbol{v}_i \neq 0 \qquad \text{unless} \quad a_i = 0 \quad (\text{for all } i) \tag{5.75}$$

then the vectors are said to be *linearly independent*. For example, the vectors that make up the Cartesian axes (Eqn 3.8) are linearly independent. Moreover, as we saw, *any* Cartesian vector in the three-dimensional space (e.g., the baseball field in Figure 3.1) can be expressed as a combination of these vectors. These vectors are thus said to *span* the *vector space* (i.e., the baseball field) V. We somewhat prematurely called these basis vectors in Section 3.5, but now we can provide the formal definition that *basis vectors* are a set of vectors that are (1) linearly independent and (2) span the space, which indeed these are.

5.15 Symmetric Matrices

There is a special type of matrix, called a *symmetric* matrix, which is equal to its tranpose:

$$\boldsymbol{A}^t = \boldsymbol{A} \qquad \text{symmetric matrix} \tag{5.76}$$

Consider general 2×2 matrix that is symmetric

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \boldsymbol{A}^t = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \tag{5.77}$$

For this to be true, it must be that $b = c$, and in general, from Eqn 5.1

$$A_{ij} \equiv A_{ji} \tag{5.78}$$

A moment's thought will convince you also that a symmetric matrix must be a square matrix. An example of a symmetric matrix is shown in Figure 5.10.

Problems

5.5 Try equating the matrix and its transpose in Eqn 5.2 to convince yourself that a symmetric matrix must be a square matrix.

This case of the symmetric matrix is a good example of how knowing the *structure* of a matrix tells us something about the solutions to equations involving it, even if the specific values of the elements are not known. For, if we had a problem in which we wanted to determine the elements of the matrix $\boldsymbol{M}$ in Eqn 5.77, the fact that it is symmetric means that the number of unknowns is not 4, but rather 3, since $b = c$. For a general $n \times n$ symmetric matrix, which has n^2 matrix elements, the number of *unique* matrix elements is $n(n+1)/2$.

There is a very important theorem: For any $m \times n$ matrix $\boldsymbol{A}$ of rank r, the matrix formed from inner product $\boldsymbol{M} = \boldsymbol{A}^t\boldsymbol{A}$ is a symmetric matrix that is also of rank r. It is easy to see that it is symmetric because the transpose of M is $\boldsymbol{M}^t = \left(\boldsymbol{A}^t\boldsymbol{A}\right)^t = \boldsymbol{A}^t\left(\boldsymbol{A}^t\right)^t = \boldsymbol{A}^t\boldsymbol{A} = \boldsymbol{M}$ where we have used the rule Eqn **??** and the fact that $\boldsymbol{A} = \boldsymbol{A}^t$, because it is symmetric. Since $\boldsymbol{M}$ equals its transpose, it is symmetric.

The symmetric matrix will be important in DTI because the diffusion tensor *is* a symmetric matrix that looks like this:

$$\boldsymbol{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix} = \boldsymbol{D}^t \tag{5.80}$$

where $D_{ij} = D_{ji}$ (i.e., $D_{xy} = D_{yx}$, etc).

Problems

5.6 Prove that the number of unique elements in an $n \times n$ symmetric matrix is $n(n+1)/2$.

5.16 Orthogonal Matrices

There is a special type of matrix, called an *orthogonal* matrix, whose transpose is equal to its inverse:

$$\boldsymbol{Q}^t = \boldsymbol{Q}^{-1} \qquad \text{Orthogonal matrix} \tag{5.81}$$

so that $\boldsymbol{Q}^t\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^t = \boldsymbol{I}$. The name should be *orthonormal* matrix, but historical usage has overruled precision in this case. Orthogonal matrices also have the important property that their determinant is $\det(Q) = \pm 1$. A very important property of orthonormal properties is that they preserve length. This is easy to show if we consider a vector $\boldsymbol{v}$ transformed by an orthogonal matrix $\boldsymbol{Q}$: $\boldsymbol{u} = \boldsymbol{Q}\boldsymbol{v}$. Then the length of $\boldsymbol{u}$ is

$$\boldsymbol{u}^t\boldsymbol{u} = (\boldsymbol{Q}\boldsymbol{v})^t(\boldsymbol{v}\boldsymbol{Q}) = \boldsymbol{v}^t\underbrace{\boldsymbol{Q}^t\boldsymbol{Q}}_{\boldsymbol{I}}\boldsymbol{v} = \boldsymbol{v}^t\boldsymbol{v} \tag{5.82}$$

An important example of an orthogonal matrix is the rotation matrix Eqn 5.83:

$$\boldsymbol{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{5.83}$$

Problems

5.7 Show that the 2D rotation matrix Eqn 5.83 is an orthogonal matrix, i.e., that it satisfies $\boldsymbol{R}^t(\theta)\boldsymbol{R}(\theta) = \boldsymbol{I}$.

5.8 Show that a rotated 2D vector does not change its length, i.e., that $|R(\theta)\cdot\boldsymbol{v}| = |\boldsymbol{v}|$. This is apparent from Figure 5.4.

Example 5.1 Show that the product $\boldsymbol{C} = \boldsymbol{AB}$ of orthogonal matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is also an orthogonal matrix.

Solution

Compute the product

$$\boldsymbol{C}^t\boldsymbol{C} = (\boldsymbol{AB})^t\,(\boldsymbol{AB}) = \boldsymbol{B}^t\boldsymbol{A}^t\boldsymbol{AB} = \boldsymbol{B}^t\boldsymbol{B} = \boldsymbol{I} \tag{5.84}$$

where Eqn 5.13 has been used.

5.17 Complex matrices

Matrices with complex elements are called *complex* matrices. The analoque of the transpose of a complex matrix $\boldsymbol{H}$ is called the *Hermitian conjugate*

$$\boldsymbol{H}^\dagger \equiv (\boldsymbol{H}^*)^t \;\text{ complex matrix transpose (Hermitian conjugate) :} \tag{5.85}$$

and the complex analog of the matrix inner product utilizes the Hermitian conjugate, rather than the tranpose:

$$\boldsymbol{H}^\dagger\boldsymbol{H} \;\text{ complex matrix inner product :} \tag{5.86}$$

The transposition rule for the product of real matrices (Eqn 5.4) becomes for complex matrices

$$(\boldsymbol{H}^\dagger\boldsymbol{B})^\dagger = \boldsymbol{B}^\dagger\boldsymbol{H} \tag{5.87}$$

H very important type of matrix is the complex analogue of the symmetric matrix, a complex matrix that equals its Hermitian conjugate:

$$\boldsymbol{H}^\dagger = \boldsymbol{H} \tag{5.88}$$

Such a matrix is called a *Hermitian* matrix. The diagonal elements of a Hermitian matrix must be real because each must equal its complex conjugate. Hermitian matrices have some interesting properties that allow general conclusions to be drawn about results involving them. For example, consider the complex number c generated by multiplying an arbitrary Hermitian matrix $\boldsymbol{H}$ on both the right and the left by an arbitrary complex vector $\boldsymbol{v}$.

$$c = \boldsymbol{v}^\dagger\boldsymbol{H}\boldsymbol{v} = \boldsymbol{v}^\dagger\boldsymbol{H}^\dagger\boldsymbol{v} = \left(\boldsymbol{v}\boldsymbol{H}^\dagger\boldsymbol{v}\right)^\dagger = c^\dagger \tag{5.89}$$

Note the since c is a just a number, its Hermitian conjugate is just its conjugate $c^\dagger = c^*$. But in order for a number to be equal to its complex conjugate, i.e., $c = c^*$, the number must be real. Therefore, it can be concluded that if $\boldsymbol{H}$ is Hermitian, the number $c = \boldsymbol{v}^\dagger\boldsymbol{H}\boldsymbol{v}$ formed from any complex vector $\boldsymbol{v}$ is real. This situation arises frequently in least squares problems where $\boldsymbol{H}$ is the correlation matrix and $\boldsymbol{v}$ is the variable vector.

There is also a complex analog of the orthogonal matrix. The transpose of an orthonormal matrix is its inverse, and the conjugate transpose (i.e., the Hermitian conjugate) of a *unitary* matrix $\boldsymbol{U}$ is its inverse:

$$\boldsymbol{U}^\dagger = \boldsymbol{U}^{-1} \qquad \text{Unitary matrix} \tag{5.90}$$

Unitary matrices $\boldsymbol{U}$ satisfy the complex generalization of Eqn 5.84

$$\boldsymbol{U}^\dagger \boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^\dagger = \boldsymbol{I} \tag{5.91}$$

and, like orthogonal matrices, preserve the length of the vectors they act upon. That is, a vector $\boldsymbol{u} = \boldsymbol{U}\boldsymbol{v}$ formed by the action of a unitary matrix $\boldsymbol{U}$ on the complex vector $\boldsymbol{v}$ has length squared

$$\boldsymbol{u}^t \boldsymbol{u} = (\boldsymbol{U}\boldsymbol{v})^\dagger \boldsymbol{U}\boldsymbol{v} = \boldsymbol{v}^\dagger \boldsymbol{U}^\dagger \boldsymbol{U}\boldsymbol{v} = \boldsymbol{v}^\dagger \boldsymbol{v} \tag{5.92}$$

which is just the length squared of the original vector $\boldsymbol{v}$.

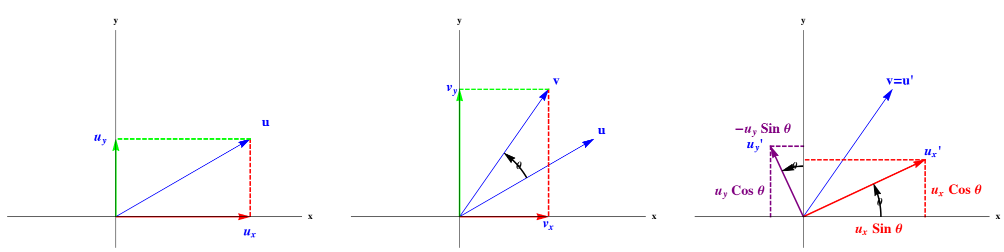
6 Rotations and Affine Transformations

6.1 Motivation

A situation that is encountered often in MRI is that in which some object, such as a vector (e.g., the magnetization) or a shape (e.g., a cubic voxel) has its location or geometry altered in some way by a *transformation*. The most common of these transformations are translations, rotations, scaling, and shearing. These are called the *affine transformations*. The matrix contruction of the previous chapter is particularly well suited to succintly describe and apply such transformations, and we will see those benefits in this chapter. Among these affine transformations, the rotations are of particular importance in MRI, and arise in a wide range of situations, in a variety of forms, and are by far the most complicated of the affine transformations. They therefore consume the bulk of this chapter. A brief section below with show that the rotations, along with the rest of the affine transformations, can be combined into a single transformation matrix.

6.2 Rotation of a vector in 2 dimensions

Consider the simple problem of a vector $\boldsymbol{u}$ (Figure 6.1a) that is rotate by an angle θ to form a new vector $\boldsymbol{v}$ shown in Figure 6.1b. The two-dimensional vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ can be written in terms of their components along the two basis vectors $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}$ as $\boldsymbol{u} = \{u_x, u_y\}$ and $\boldsymbol{v} = \{v_x, v_y\}$. We want to relate the components $\{v_x, v_y\}$ of the rotated vector $\boldsymbol{v}$ to the components $\{u_x, u_y\}$ of the original vector $\boldsymbol{u}$. This is shown geometrically in Figure 6.1c where the rotate all the vectors ($\boldsymbol{u}$, $\boldsymbol{u}_x$, and $\boldsymbol{u}_y$) in Figure 6.1a by an angle θ, and denote these new vectors by a $\boldsymbol{u}'$, $\boldsymbol{u}'_x$, and



(a) Original vector $\boldsymbol{u}$ and its components (projections) u_x and u_y along the x and y axes.

(b) Rotation of $\boldsymbol{u}$ by θ to give a new vector $\boldsymbol{v}$ with components v_x and v_y along the x and y axes.

(c) The x projections of the rotated projections in (a) sum to v_x and the y projections sum to v_y.

Figure 6.1 Rotation of a vector in two dimensions.

$\boldsymbol{u}'_y$, where $\boldsymbol{u}' = \boldsymbol{v}$. From this we see that the components of the rotated vector $\boldsymbol{v}$ in terms of the components of the original vector $\boldsymbol{u}$ are

$$v_x = u_x \cos\theta - u_y \sin\theta \tag{6.1a}$$

$$v_y = u_x \sin\theta + u_y \cos\theta \tag{6.1b}$$

But notice that this can be written in the form

$$\underbrace{\begin{pmatrix} v_x \\ v_y \end{pmatrix}}_{\boldsymbol{v}} = \underbrace{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}}_{\boldsymbol{R}(\theta)} \underbrace{\begin{pmatrix} u_x \\ u_y \end{pmatrix}}_{\boldsymbol{u}} \tag{6.2}$$

The matrix

$$\boldsymbol{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{6.3}$$

is called a *rotation matrix* because multiplying a vector by $\boldsymbol{R}(\theta)$ rotates that vector through the angle θ. The rotation depicted in Figure 6.1b can then be written simply $\boldsymbol{v} = \boldsymbol{R}_\theta \boldsymbol{u}$ (where we use the shorthand notation $\boldsymbol{R}_\theta = \boldsymbol{R}(\theta)$)

6.3 Rotation of an ellipse in 2 dimensions

The rotation of a vector involved the simplest rotational transformation: Multiplication of the original vector $\boldsymbol{u}$ by a two-dimensional rotation matrix $\boldsymbol{R}_\theta$ to produce a rotated vector $\boldsymbol{v} = \boldsymbol{R}_\theta \boldsymbol{u}$. However, *how* an object (e.g., a vector) is rotated depends upon *what* that object is. In fact, mathematical objects are *defined* by how they transform. That is, a vector can be *defined* as a quantity that transforms according to this formula. To demonstrate this dependence explicitly, let's now consider the rotation of a *shape*. A simple example is an ellipse, shown in Figure 6.2a, whose equation is

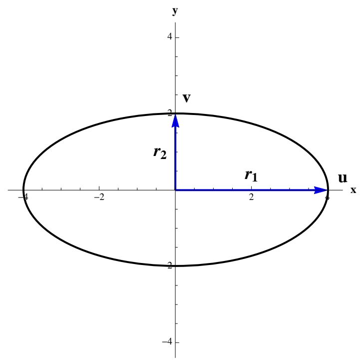
$$\frac{x^2}{r_1^2} + \frac{y^2}{r_2^2} = 1 \tag{6.4}$$

where r_1 is called the *semi-major axis* and r_2 is the *semi-minor axis*, and $r_1 > r_2$. Together are called the *principal axes* of the ellipse. At the point the ellipse intersects the y-axis (i.e., $y = 0$) $1 = x^2/r_1^2$, that is $x = r_1$. Similarly, at the point it intersects the x-axis (i.e., $x = 0$) $1 = y^2/r_2^2$, that is $y = r_2$. So we see that r_1 and r_2 are the lengths of the principal axes. Eqn 6.4 is an example of a quadratic equation because of the powers to which the variables x and y are raised. To facilitate the use of our rotation matrix, let's write this equation in matrix form by defining $\boldsymbol{x}^t = (x, y)$ and $\lambda_i = 1/r_i^2$ to get

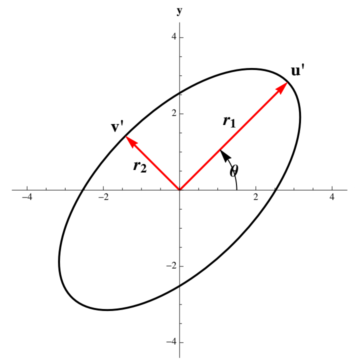
$$\boldsymbol{x}^t \boldsymbol{\Lambda} \boldsymbol{x} = 1 \tag{6.5a}$$

$$\text{where} \quad \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \tag{6.5b}$$

Eqn 6.6, being a matrix version of the quadratic equation Eqn 6.4, is an example of a *quadratic form*. We can think of every point on the ellipse as being a vector from the origin to that point,



(a) Unrotated ellipse.



(b) Rotated ellipse.

Figure 6.2 Rotation of an ellipse by an angle θ.

so to rotate the shape means to rotate all the points at once: $\boldsymbol{\xi} = \boldsymbol{R}_\theta \boldsymbol{x}$. So a rotated ellipse will just be

$$\boldsymbol{\xi}^t \boldsymbol{\Lambda} \boldsymbol{\xi} = \boldsymbol{x}^t \boldsymbol{R}^t \boldsymbol{\Lambda} \boldsymbol{R} \boldsymbol{x} = 1 \tag{6.6}$$

where we've used Eqn 5.13. But this looks just like Eqn 6.5a if we write it as

$$\boldsymbol{x}^t \boldsymbol{Q} \boldsymbol{x} = 1 \tag{6.7a}$$

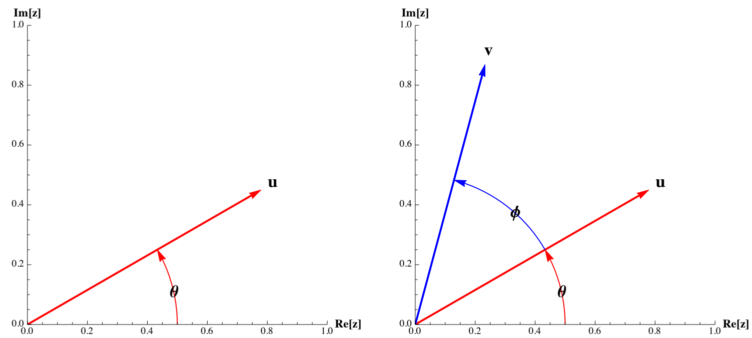
$$\text{where} \quad \boldsymbol{Q} = \boldsymbol{R}^t \boldsymbol{\Lambda} \boldsymbol{R} \tag{6.7b}$$

Note that the matrix $\boldsymbol{Q}$, unlike the matrix $\boldsymbol{\Lambda}$, is *not* diagonal:

$$\boldsymbol{Q} = \begin{pmatrix} \lambda_2 \sin^2\theta + \lambda_1 \cos^2\theta & \lambda_2 \sin\theta\cos\theta - \lambda_1 \sin\theta\cos\theta \\ \lambda_2 \sin\theta\cos\theta - \lambda_1 \sin\theta\cos\theta & \lambda_1 \sin^2\theta + \lambda_2 \cos^2\theta \end{pmatrix} \tag{6.8}$$

The diagonal quality of Λ expresses the fact that the semi-major and semi-minor axes point along the x and y axes. The non-diagonal quality of Q expresses the fact that the semi-major and semi-minor axes have been rotated relative to the x and y axes. The transformation Eqn 6.7 that takes $\boldsymbol{\Lambda}$ to $\boldsymbol{Q}$ is a special case of a *similarity transformation*. Now, we have constructed this rotated ellipse by rotating all of the points $\boldsymbol{x}^t$ simultaneously through the angle θ. But a very important point must be made here. The ellipse is defined by the parameters $\{r_1, r_2\}$ and thus by the vectors $\{\boldsymbol{u}, \boldsymbol{v}\}$. The rotation of the points in the ellipse results in the rotation of $\{\boldsymbol{u}, \boldsymbol{v}\}$ to their rotated version $\{\boldsymbol{u'}, \boldsymbol{v'}\}$. Therefore we make an interesting, and ultimately very practically important, observation. If we had *started* with the rotated ellipse Figure 6.2b and rotated this by an angle $-\theta$ so that it was aligned as in Figure 6.2a, then the matrix $\boldsymbol{Q}$ would become diagonal (i.e., reduce to $\boldsymbol{\Lambda}$) and we could just read off the parameters $\{r_1, r_2\} = \{1/\lambda_1, 1/\lambda_2\}$ that define the ellipse. This relationship between rotations, shape parameters, and the principle axes will become a central issue when we discuss eigenvectors and eigenvalues in Chapter 8.

In passing, it is worth noting that since the λ's are real numbers (as are the trigonometric functions of θ), the matrix $\boldsymbol{Q}$ is also real. Also notice that the off-diagonal terms of $\boldsymbol{Q}$ are equal, so that $\boldsymbol{Q}$ is a *real, symmetric* matrix.



(a) Complex vector $z = re^{i\theta}$.

(b) Complex vector in (a) rotated by an angle ϕ: $z = re^{i(\theta+\phi)}$

Figure 6.3 Rotation of a complex vector of radius $r = .9$ and phase angle θ further rotated by an angle ϕ.

6.4 Rotations in 2 dimensions: Complex Representation

It was probably already clear to you after reading Chapter 4 that complex numbers provide an efficient way to describe rotations in two-dimensions since the rotation angle of a vector in the complex plane is precisely the phase in exponent of Euler's relations (Eqn 4.9) and that the representation of a complex number in terms of its amplitude and phase (Eqn 4.10) therefore provides a simple means of describing an arbitrary two-dimension vector pointing out from the origin. An example is shown in Eqn 6.3 where a complex vector initially oriented at an angle θ is further rotated through an angle ϕ. This example serves to emphasize that the complex representation, using Euler's relation, which decouples the amplitude and phase, provides a very simple way to induce a rotation, without changing the length of the vector, merely by adding an additional phase in the exponent.

6.5 Rotations in 3 dimensions

An arbitrary rotation in three dimensions can be described by three angles so that any rotation can be performed by the succession of three separate rotations about the three Cartesian axes $\{x, y, z\}$. This is known as *Euler's rotation theorem* (**?**). That is

$$\boldsymbol{R}(\alpha, \beta, \gamma) = \boldsymbol{R}_x(\alpha)\boldsymbol{R}_y(\beta)\boldsymbol{R}_z(\gamma) \tag{6.9}$$

where the rotations about the x, y, and z axes in a clockwise direction (when looking towards the origin) are represented in matrix form, respectively, as the following three matrices in $\mathbb{R}^3$:

$$\boldsymbol{R}_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \quad , \quad \text{Rotation about } x \text{ by } \alpha \tag{6.10a}$$

$$\boldsymbol{R}_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \quad , \quad \text{Rotation about } y \text{ by } \beta \tag{6.10b}$$

$$\boldsymbol{R}_z(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad , \quad \text{Rotation about } z \text{ by } \gamma \tag{6.10c}$$

Notice that the order of application of these operations is from right to left: the rotations take place along z, then y, and then x.

The structure of the rotation matrices in Eqn 6.10 is actually quite simple: They are the two-dimensional rotation matrix Eqn 6.3 embedded within a 3×3 matrix in the locations of the plane of rotation, with a 1 in the location that depends only on the rotation axis (the axis *about* which the rotation occurs), and zeros in the elements that couple the plane with the rotation axis. For example, $R_z(\gamma)$ (Eqn 6.10c) has a 1 in the location that depends only on the rotation axis z, about which the rotation is occuring, zeros in the elements that couple the plane (x or y) with the rotation axis z, and the elements of the 2×2 rotation matrix are in components that depend on x and y, since the rotation is in the $x - y$ plane. Thus the element of a vector along the rotation axis remains unchanged, while the elements in the plane are rotated. We can see this explicitly by creating the dot product of each of these matrices with an arbitrary vector of cartesian components $\boldsymbol{v}^t = (x, y, z)$:

$$\boldsymbol{R}_x(\alpha) \cdot \boldsymbol{v} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y\cos\alpha - z\sin\alpha \\ z\cos\alpha + y\sin\alpha \end{pmatrix} \tag{6.11a}$$

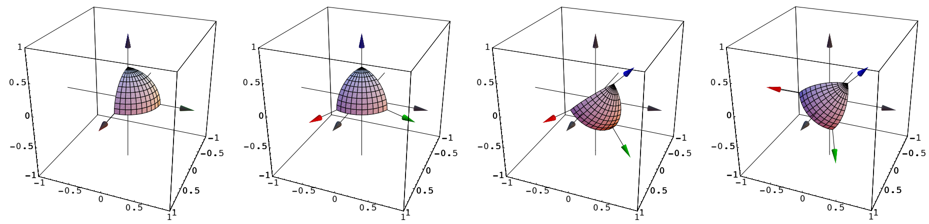
$$\boldsymbol{R}_y(\beta) \cdot \boldsymbol{v} = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x\cos\beta + z\sin\beta \\ y \\ z\cos\beta - x\sin\beta \end{pmatrix} \tag{6.11b}$$

$$\boldsymbol{R}_z(\gamma) \cdot \boldsymbol{v} = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x\cos\gamma - y\sin\gamma \\ y\cos\gamma + x\sin\gamma \\ z \end{pmatrix} \tag{6.11c}$$

Alternatively, an arbitrary rotation can also be performed by the following sequence of rotations that are defined *in the successively rotated coordinate systems*: a rotation of γ is performed about the z-axis in the original coordinate system. Then a rotation of β is performed about the x-axis in this *rotated* coordinate system (let's call it x') and then a rotation of α is performed about the z-axis again in this twice rotated system (let's call it z''). That is,

$$\boldsymbol{R}(\alpha, \beta, \gamma) = \boldsymbol{R}_{z''}(\gamma)\boldsymbol{R}_{x'}(\beta)\boldsymbol{R}_z(\alpha) \tag{6.12}$$

So defined, the angles $\{\alpha, \beta, \gamma\}$ are called the *Euler angles*. This is shown in Figure 6.4. Each of the rotation matrices $\boldsymbol{R}_x(\alpha), \boldsymbol{R}_y(\beta), \boldsymbol{R}_z(\gamma)$ are orthogonal matrices and thus from Section 5.16 the



(a) No rotation. (b) $R_z(\alpha)$ (c) $R_{x'}(\beta)R_z(\alpha)$ (d) $R_{z''}(\gamma)R_{y'}(\beta)R_z(\alpha)$

Figure 6.4 The Euler Angles $\{\alpha, \beta, \gamma\}$. The $\{x, y, z\}$ axes in the rotated frame are colored *red*, *green*, and *blue*. Rotations angles are $\{\alpha, \beta, \gamma\} = \{\pi/8, \pi/4, \pi/6\}$. **Need to put arcs and angle labels in these figures!**. This demonstrates Euler's theorem that a general rotation can be written as the composition of rotations about the three axes.

general rotation matrix $R(\alpha, \beta, \gamma)$, as the product of orthogonal matrices, is also an orthogonal matrix. From the discussion in Section 5.16 we can conclude that the rotation of a vector does not change the length of the vector, which we know intuitively. We will use this fact in Section 29.6, among other places, in our DTI calculations. As discussed in Section **??**, on 3D rotation matrices do not commute, which means that the order of their application matters, as demonstrated in Figure 6.5.

The ability to compactly represent rotation in matrix form as $R_i(\varphi)$ to represent a rotation by angle φ about the i'th axis, and the ability to combine these matrices from multiple rotations into a single matrix will provide very useful. For example, the application of radio-frequency (RF) pulses can be represent by such matrices and multiple pulses by the combined matrices. This notation allows you to represent such phenomenon without writing all the components which adds no new intuitive insight and just clutters up the picture.

6.6 Commutation relations

Now, what if we follow the rotation of a vector $\boldsymbol{u}$ by an angle θ to a new vector $\boldsymbol{u}' = \boldsymbol{R}(\theta)\boldsymbol{u}$ by a rotation by $-\theta$? The vector returns to its original position pointing along the y-axis - we have undone the first rotation. Let's call this twice-rotated vector $\boldsymbol{u}''$, and express it both as the rotation by $-\theta$ of the once-rotated vector $\boldsymbol{u}'$, and as the original vector $\boldsymbol{u}$ to which it returns:

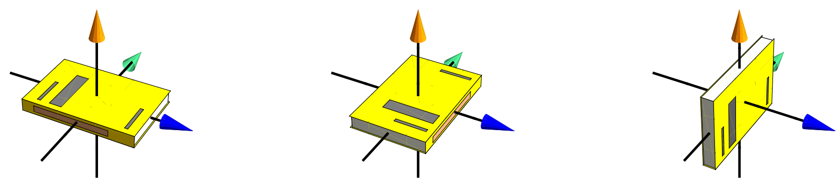
$$\boldsymbol{u}'' = \boldsymbol{R}(-\theta)\boldsymbol{u}' = \boldsymbol{u} \tag{6.13}$$

If we substitute Eqn **??** into Eqn 6.13 we see that

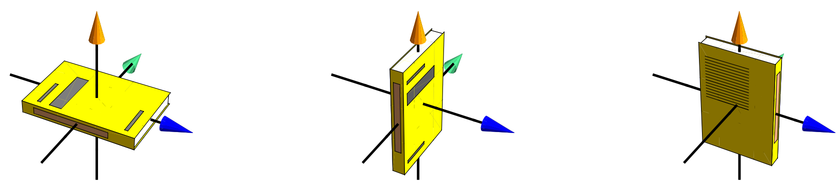
$$\boldsymbol{u} = \boldsymbol{R}(-\theta)\boldsymbol{R}(\theta)\boldsymbol{u} \tag{6.14}$$

from which we see that the following must be true:

$$\boldsymbol{R}(-\theta)\boldsymbol{R}(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{I}_2 \tag{6.15}$$



(a) No rotation. (b) Rotation of 90° about $\hat{\boldsymbol{z}}$. (c) Rotation of 90° about $\hat{\boldsymbol{y}}$.



(d) No rotation. (e) Rotation of 90° about $\hat{\boldsymbol{y}}$. (f) Rotation of 90° about $\hat{\boldsymbol{z}}$.

Figure 6.5 An example of non-commuting matrices. Three-dimensional rotation matrices do not commute. That is, the order of their application matters. Rotation about $\hat{\boldsymbol{z}}$ then $\hat{\boldsymbol{y}}$ (top) does *not* equal rotation about $\hat{\boldsymbol{y}}$ then $\hat{\boldsymbol{z}}$ (bottom).

where $\boldsymbol{I}_2$ is the identity matrix in 2-dimensions which returns unchanged the vector it multiplies: $\boldsymbol{v} = \boldsymbol{I}\boldsymbol{v}$. From Eqn 6.15 we conclude

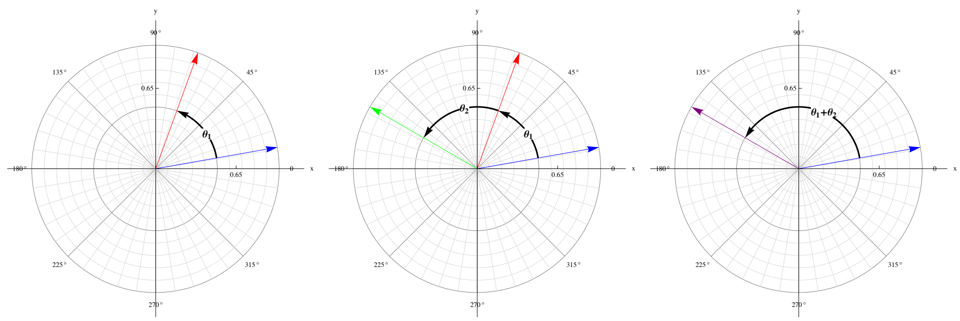
$$\boldsymbol{R}(-\theta) = \boldsymbol{R}^{-1}(\theta) \tag{6.16}$$

That is, $\boldsymbol{R}(-\theta)$ is the *inverse* denoted by $\boldsymbol{R}^{-1}$) of $\boldsymbol{R}(\theta)$. That is, it undoes the effect of $\boldsymbol{R}(\theta)$. Now consider the effect of two successive rotations, by angles θ_1 and θ_2, respectively, as shown in Figure 6.6. As above, the first rotation brings the vector $\boldsymbol{u}$ to $\boldsymbol{u}'$ and the second bring $\boldsymbol{u}'$ to $\boldsymbol{u}''$. Using the shorthand $\boldsymbol{R}_i = \boldsymbol{R}(\theta_i)$ This we write as

$$\boldsymbol{u}'' = \boldsymbol{R}_2\boldsymbol{u}' = \boldsymbol{R}_2\boldsymbol{R}_1\boldsymbol{u} \tag{6.17}$$

Since we can multiply matrices together let's do it and write Eqn 6.17 as

$$\boldsymbol{u}'' = \boldsymbol{R}(\theta_1, \theta_2)\boldsymbol{u} \tag{6.18}$$



(a) Rotation by θ_1. (b) ... followed by a rotation by θ_2. (c) ... is the same as a single rotation by $\theta_1 + \theta_2$.

Figure 6.6 Successive rotations in two dimensions.

where

$$\boldsymbol{R}(\theta_1, \theta_2) = \underbrace{\begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix}}_{\boldsymbol{R}_1(\theta_1)} \underbrace{\begin{pmatrix} \cos\theta_2 & \sin\theta_2 \\ -\sin\theta_2 & \cos\theta_2 \end{pmatrix}}_{\boldsymbol{R}_2(\theta_2)}$$
$$= \begin{pmatrix} \cos(\theta_1+\theta_2) & \sin(\theta_1+\theta_2) \\ -\sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{pmatrix} \tag{6.19}$$

But this is just a rotation of the original vector $\boldsymbol{u}$ by the angle $\theta_1+\theta_2$. Thus the multiple rotations affected by the rotation matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ is equivalent to the operation of a single rotation matrix $\boldsymbol{R}(\theta_1, \theta_2)$. Notice that as a consequence of Eqn 6.17, Eqn 6.18, and Eqn 6.19, that the order that we apply two different rotations does not matter. That is, since $\boldsymbol{R}(\theta_1, \theta_2) = \boldsymbol{R}(\theta_2, \theta_1)$

$$\boldsymbol{R}_1\boldsymbol{R}_2\boldsymbol{u} = \boldsymbol{R}_2\boldsymbol{R}_1\boldsymbol{u} = (\boldsymbol{R}_1\boldsymbol{R}_2 - \boldsymbol{R}_2\boldsymbol{R}_1) = [\boldsymbol{R}_1, \boldsymbol{R}_2]\boldsymbol{u} \tag{6.20}$$

where we have defined

$$[A, B] = AB - BA \tag{6.21}$$

This is called the *commutator*, which is important because, as we saw in Section 5.8, matrices do not necessarily commute. If matrices do not commute, that means that the order in which they are applied makes a difference. The commutator thus tells us specifically how this difference in the order of operation is manifest. As we will see in the next section, two dimensions is a lot simpler than three dimensions, because while rotations in two dimensions *can* be applied in any order, since the two-dimensional rotation matrices commute, this is *not* true in three-dimensions. This so-called non-commutative property of rotations in 3-dimensions has profound physical consequences that we shall encounter in later chapters. The fact that two-dimensional rotations commute is easily seen in Figure 6.7

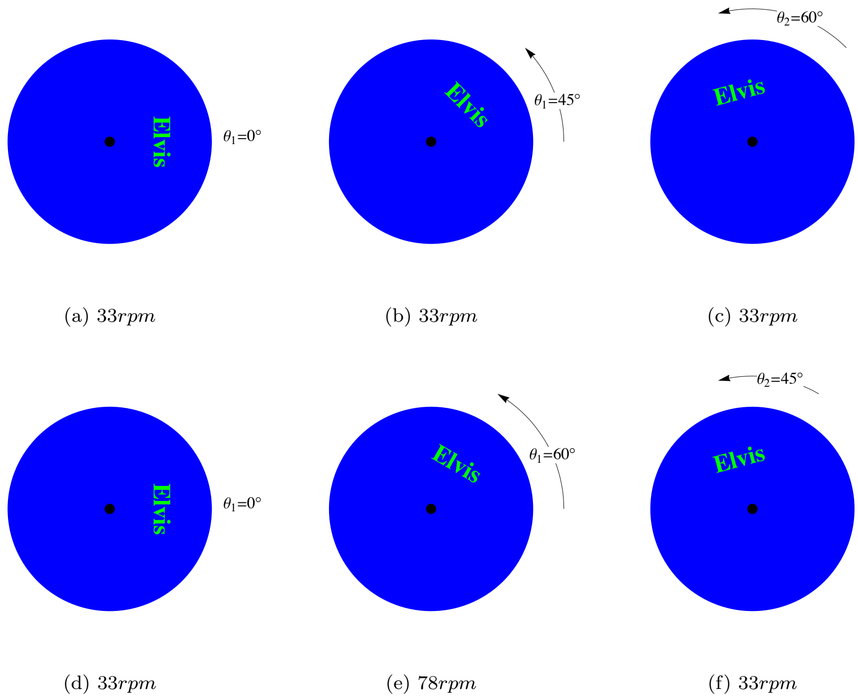


Figure 6.7 Rotations in 2-dimensions commute. Rotation by 45° then 90° is the same as rotation first by 60° then by 45°.

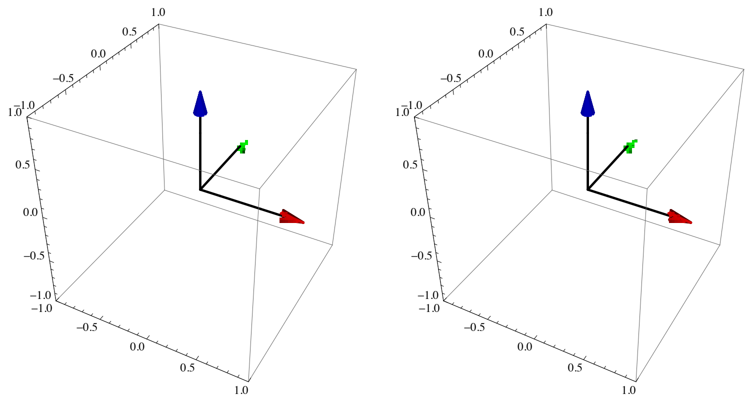
6.7 Interlude: Coordinate rotations

An important example of a 3D rotation that will arise frequently is the rotation of coordinates. A very simple example is illustrated in Figure 6.8 where the Cartesian coordinate system is rotated about the z-axis. This transformation can be simply affected by multiplying the matrix of basis vectors $\boldsymbol{M}$ (Eqn **??**) by the z rotation matrix $R_z(\varphi)$:

$$\boldsymbol{R}_z(\varphi)\cdot\boldsymbol{M} = \begin{pmatrix}\cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} = \begin{pmatrix}\cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{pmatrix} \tag{6.22}$$

6.8 Rotations in 3 dimensions: Complex Representation*

As we saw in Chapter 4 and Section **??**, there is a close relationship between the complex representation of numbers and rotations. So it is should not be surprising that there should be a way to represent rotations in 3D using complex numbers. Indeed, there is, and we will find this representation to be important in several areas of MRI and DTI, such the Bloch equations, the



(a) The Cartesian coordinate system.

(b) The Cartesian coordinate system rotated by an angle θ about the z-axis. (**Figure just placeholder - put in rotated version!**)

Figure 6.8 Rotation of the Cartesian coordinate system.

description of RF pulses and coherence pathways, and the manipulation of generalized tensors using spherical harmonics (these terms will become familiar to you.) However, while this reformulation of the rotations in terms of complex numbers is an exceedingly useful way to represent rotations, it turns out to be quite a subtle business.

So let's motivate our discussion by giving the punchline. Just as a rotation can be described by the product of three 3×3 real, orthogonal matrices $\{\boldsymbol{R}_x(\alpha), \boldsymbol{R}_y(\beta), \boldsymbol{R}_z(\gamma)\}$, describing rotations about the $\{x, y, z\}$ axes respectively (Eqn 6.9) by angles $\{\alpha, \beta, \gamma\}$, so to can the same rotation be described by the product of three 2×2 unitary matrices (Section **??**)

$$\boldsymbol{U}(\alpha, \beta, \gamma) = \boldsymbol{U}_x(\alpha/2)\boldsymbol{U}_y(\beta/2)\boldsymbol{U}_z(\gamma/2) \tag{6.23}$$

with angles (defined below) $\{\xi, \eta, \chi\}$. Recall that a unitary matrix is in some ways the complex analogue of the orthogonal matrix in that it preserves the length of a complex vector (Section **??**), just as an orthogonal matrix preserves the length of a real vector (Section 5.16). The preservation of length is one of our requirements of the transformation being just a rotation.

A general element of the group $SU(2)$ is

$$\boldsymbol{U} = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \tag{6.24}$$

where a and b are called the *Cayley-Klein parameters*. Since $\boldsymbol{U}$ is unitary, $|\boldsymbol{U}| = a^*a + b^*b = 1$. The matrix U transforms a complex two component vector w, called a *spinor*:

$$\underbrace{\begin{pmatrix} u' \\ v' \end{pmatrix}}_{w'} = \underbrace{\begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} u \\ v \end{pmatrix}}_{w} \tag{6.25}$$

The Cayley-Klein parameters depend upon the representation of the rotation. Now we need to find an explicit form for $\boldsymbol{U}$. First, we recall ((**?**)) that any 2×2 matrix can be expressed in terms of the *Pauli matrices*:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{6.26}$$

which, along with the 2×2 identity matrix,

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{6.27}$$

form an orthogonal basis for the complex Hilbert space of all 2×2 matrices. The Pauli matrices satisfy (Exercise 6.1).

$$\sigma_i^2 = 1 \tag{6.28a}$$

$$\sigma_i \sigma_j = \sigma_k \quad , \quad \text{cyclic permutations} \tag{6.28b}$$

$$\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij}\sigma_0 \tag{6.28c}$$

Thus we can write

$$\boldsymbol{U} = \sum_{i=1}^{4} c_i \sigma_i \tag{6.29}$$

where the c_i are constants. For example, if $c_i = 0, x, y, z$, then the matrix

$$\boldsymbol{A} = 0\sigma_0 + x\sigma_1 + x\sigma_2 + z\sigma_3 = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix} \tag{6.30}$$

which is a matrix that can be seen as converting our cartesian representation $\{x, y, z\}$ into a representation now still involving z but having convert the $\{x, y\}$ components into two complex vectors, $\{x + iy, x - iy\}$ rotating in the $x - y$ plane in opposite directions. But it can also shown that rotations about the separate axes $\{x, y, z\}$ can be written in terms of separate matrix $\boldsymbol{U}_k$, where $k = \{x, y, z\}$:

$$\boldsymbol{U}_k(\varphi) = \boldsymbol{\sigma}_o \cos(\varphi) + i\boldsymbol{\sigma}_k \sin(\varphi) \quad , \quad k = \{x, y, z\} \tag{6.31}$$

From the transformations of Eqn 6.30 using Eqn 6.31, it can be deduced (**?**) that there is a correspondence between the matrices unitary matrices $\boldsymbol{U}$ and the orthogonal rotation matrices $\boldsymbol{R}$:

$$\boldsymbol{U}_x(\varphi/2) = \begin{pmatrix} \cos\varphi/2 & \sin\varphi/2 \\ -\sin\varphi/2 & \cos\varphi/2 \end{pmatrix} \leftrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix} = R_x(\varphi) \tag{6.32a}$$

$$\boldsymbol{U}_y(\varphi/2) = \begin{pmatrix} \cos\varphi/2 & i\sin\varphi/2 \\ -i\sin\varphi/2 & \cos\varphi/2 \end{pmatrix} \leftrightarrow \begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix} = R_y(\varphi) \tag{6.32b}$$

$$\boldsymbol{U}_z(\varphi/2) = \begin{pmatrix} e^{i\varphi/2} & 0 \\ 0 & e^{-i\varphi/2} \end{pmatrix} \leftrightarrow \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} = R_z(\varphi) \tag{6.32c}$$

Therefore the general rotation in terms of the product of 3×3 orthogonal rotation matrices $\boldsymbol{R}$ can be also be represented as the product of 2×2 unitary $\boldsymbol{U}$:

$$\boldsymbol{R}(\alpha, \beta, \gamma) = \boldsymbol{R}_x(\alpha)\boldsymbol{R}_y(\beta)\boldsymbol{R}_z(\gamma) = \boldsymbol{U}_x(\alpha/2)\boldsymbol{U}_y(\beta/2)\boldsymbol{U}_z(\gamma/2) = \boldsymbol{U}(\alpha/2, \beta/2, \gamma/2) \tag{6.33}$$

Or, in terms of Euler angles,

$$\boldsymbol{R}(\alpha,\beta,\gamma) = \boldsymbol{R}_z(\alpha)\boldsymbol{R}_y(\beta)\boldsymbol{R}_z(\gamma) = \boldsymbol{U}_z(\alpha/2)\boldsymbol{U}_y(\beta/2)\boldsymbol{U}_z(\gamma/2) = \boldsymbol{U}(\alpha/2,\beta/2,\gamma/2) \tag{6.34}$$

Substituting in Eqns 6.32a- 6.32c, gives

$$\boldsymbol{U}(\alpha/2,\beta/2,\gamma/2) = \begin{pmatrix} e^{i\xi}\cos\eta & e^{i\chi}\sin\eta \\ -e^{-i\chi}\sin\eta & e^{-i\xi}\cos\eta \end{pmatrix} \tag{6.35}$$

where

$$\xi = (\gamma+\alpha)/2 \quad , \quad \eta = \beta/2 \quad , \quad \chi = (\gamma-\alpha)/2 \quad , \tag{6.36}$$

in terms of the real angles as $\{\xi,\eta,\chi\}$ so the Cayley-Klein parameters are

$$a = e^{i\xi}\cos\eta \tag{6.37a}$$
$$b = e^{i\chi}\sin\eta \tag{6.37b}$$

The matrix $\boldsymbol{U}$ can actually be written in higher dimensions whereupon it is written in terms of the half-integer indeces $\{m,m'\}$ as $\boldsymbol{U}^j_{mm'}$ where $m = \{-j,\ldots,j\}$ the above representations correspond to $j = 1/2$ so that $\{m,m'\} = \{\pm1/2,\pm1/2\}$. The $\boldsymbol{U}_{mm'}$ represent rotations of the coordinate system. One can also write these in a form appropriate for rotations of functions. These matrices are

$$\boldsymbol{D}^j_{mm'} = \boldsymbol{U}^{j*}_{mm'} \tag{6.38}$$

The matrices $\boldsymbol{D}^j_{mm'}$ are called the *Wigner matrices*. These will be used later to rotate spherical tensors.[1]

One can also work in the opposite direction (from 2D unitary to 3D orthogonal) and express the general 3D rotation matrix $\boldsymbol{R}$ in terms of the Cayley-Klein parameters as

$$\boldsymbol{R} = \begin{pmatrix} (a^*)^2 & -(b^*)^2 & -2a^*b^* \\ -b^2 & a^2 & -2ab \\ ba^* & ab^* & aa^*-bb^* \end{pmatrix} \tag{6.39}$$

Problems

6.1 Prove Eqn 6.28

6.9 Generators of the elements of SU(2)*

Consider the complex exponential of the Pauli matrices:

$$\exp(i\varphi\sigma_k) = \sigma_0\cos\varphi + i\sigma_k\cos\varphi \tag{6.40a}$$

where the approximation follows from expanding the exponential as a MacLaurin series (Exercise 6.2). The generators of the elements of SU(2) are

$$\exp(ic_k\sigma_k/2) \quad , \quad k = \{1,2,3\} \tag{6.41}$$

where c_k are real. Thus the rotation in terms of the Euler angles Eqn 6.34 can be written

$$\boldsymbol{U} = e^{-i\boldsymbol{\sigma}_z\gamma/2}e^{-i\boldsymbol{\sigma}_x\beta/2}e^{-i\boldsymbol{\sigma}_z\alpha/2} \tag{6.42}$$

[1] Need to check this Wigner section.

Problems

6.2 Prove Eqn 6.41 Hint: Expand the exponential as a MacLaurin series, collect terms in different powers of the Pauli matrices, and use Eqn 6.28.

6.10 Rotations and the Cross Product: Generators*

Let's begin by considering how the rotation matrices in Eqn 6.10, which are valid for any angle of rotation, look for the special case of small rotations, i.e., $\{\alpha, \beta, \gamma\} \ll 1$. Using the approximation $\cos\theta \approx 1$ and $\sin\theta \approx \theta$ for $\theta \ll 1$, the rotation matrices can be approximated to first order as

$$\boldsymbol{R}_x(\alpha) \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & -\alpha & 1 \end{pmatrix} = \boldsymbol{I} + \alpha \boldsymbol{A}_x \tag{6.43a}$$

$$\boldsymbol{R}_y(\beta) \approx \begin{pmatrix} 1 & 0 & \beta \\ 0 & 1 & 0 \\ -\beta & 0 & 1 \end{pmatrix} = \boldsymbol{I} + \beta \boldsymbol{A}_y \tag{6.43b}$$

$$\boldsymbol{R}_z(\gamma) \approx \begin{pmatrix} 1 & \gamma & 0 \\ -\gamma & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \boldsymbol{I} + \gamma \boldsymbol{A}_z \tag{6.43c}$$

where $\boldsymbol{I}$ is the identity matrix in $\mathbb{R}^3$ and we have defined the real matrices

$$\boldsymbol{A}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{A}_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{A}_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.44}$$

Note that these matrices have the property that $\boldsymbol{A}^t = -\boldsymbol{A}$, which is called *anti-symmetric*. The $\boldsymbol{A}$ matrices contain all the information necessary to generate infinitesimal rotations, and are thus call the *generators* of the rotations. Let's look a bit closer at the actions of the $\boldsymbol{A}$'s by considering how they operate on a general vector $\boldsymbol{v} = \{v_x, v_y, v_z\}$:

$$\boldsymbol{A}_x \cdot \boldsymbol{v} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} 0 \\ -v_z \\ v_y \end{pmatrix} = \hat{\boldsymbol{e}}_x \times \boldsymbol{v} \tag{6.45a}$$

$$\boldsymbol{A}_y \cdot \boldsymbol{v} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} v_z \\ 0 \\ -v_x \end{pmatrix} = \hat{\boldsymbol{e}}_y \times \boldsymbol{v} \tag{6.45b}$$

$$\boldsymbol{A}_z \cdot \boldsymbol{v} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} -v_y \\ v_x \\ 0 \end{pmatrix} = \hat{\boldsymbol{e}}_z \times \boldsymbol{v} \tag{6.45c}$$

where $\hat{\boldsymbol{e}}_k$ is just the unit vector along the $\hat{k}$-direction. Thus, in general

$$\boldsymbol{A}_k \cdot \boldsymbol{v} = \hat{\boldsymbol{e}}_k \times \boldsymbol{v} \tag{6.46}$$

These *dot* product of these matrices with a vector produces a *cross* product with the vector. Thus these equations allow us to construct a general matrix representation for the cross product between two vectors

$$\boldsymbol{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \quad , \quad \boldsymbol{B} = \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} \tag{6.47}$$

as a matrix equation:

$$\boldsymbol{v} \times \boldsymbol{B} = \boldsymbol{\beta} \cdot \boldsymbol{v} \tag{6.48}$$

where [2]

$$\boldsymbol{\beta} = B_x \boldsymbol{A}_x + B_y \boldsymbol{A}_y + B_z \boldsymbol{A}_z = \begin{pmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{pmatrix} \tag{6.49}$$

The relationship between the infinitesimal generators and the cross product also allows a nice construction of time dependent equations that involve rotations. This will become very useful when we study the motion of the magnetization using the Bloch equation Chapter 14. From Eqn 6.43, any rotation $\boldsymbol{R}(\boldsymbol{\epsilon})$ of the vector $\boldsymbol{v}$ by $\boldsymbol{\epsilon} = \epsilon\, \hat{\boldsymbol{e}}$, i.e., a small angle $\epsilon \ll 1$ about the axis $\hat{\boldsymbol{e}}$, can thus be written,

$$\boldsymbol{R}(\epsilon)\, \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{\epsilon} \times \boldsymbol{v} \tag{6.50}$$

From this we see that the incremental change $\delta \boldsymbol{v}$ in the vector $\boldsymbol{v}$ can be written

$$\delta \boldsymbol{v} = \boldsymbol{R}(\epsilon)\, \boldsymbol{v} - \boldsymbol{v} = \boldsymbol{\epsilon} \times \boldsymbol{v} \tag{6.51}$$

Here we see that there is indeed a connection between the cross product and rotations, as advertised in Section 3.14: the incremental change in the vector when rotated an infinitesimal amount about an axis of rotation $\hat{\boldsymbol{\epsilon}}$ is the cross product of that vector with $\hat{\boldsymbol{\epsilon}}$. [3]

It is worth noting here that the components of the rotated vector are often written using the Levi-Civita symbol ε_{ijk} used to express the components of the cross product (Eqn 3.38):

$$v_i^{'} = v_i + \sum_{j,k=1}^{3} \varepsilon_{ijk} \epsilon_j v_k \tag{6.52}$$

Eqn 6.52 appears in problems of rotation where the angle of rotation is assumed to be small, such as in certain approaches to motion correction for (?).

6.11 Example: The Matrix Bloch Equation

The evolution of the magnetization in a gradient G_x in the x direction and a complex RF pulse of the form [4]

$$B_1 = B_{1,x} + iB_{1,y} \tag{6.53}$$

[2] signs in β are reversed from Liu
[3] show explicitly here how the time rate of change of v is then constructed - see Goldstein (?)
[4] from MJT slides.

is

$$\begin{pmatrix} \dot{M}_x \\ \dot{M}_y \\ \dot{M}_z \end{pmatrix} = \begin{pmatrix} 0 & G_x x & -B_{1,y} \\ -G_x x & 0 & B_{1,x} \\ B_{1,y} & -B_{1,x} x & 0 \end{pmatrix} \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} \tag{6.54}$$

which can be written

$$\dot{\boldsymbol{M}} = -\boldsymbol{R}\boldsymbol{M} \tag{6.55}$$

where

$$\boldsymbol{R} = B_{1,x}\boldsymbol{A}_x + B_{1,y}\boldsymbol{A}_y + G_x x \boldsymbol{A}_z \tag{6.56}$$

6.12 Rotating Frames of Reference

The way in which we perceive a physical phenomenon can depend a great deal on where we are when we observe it, or our *frame of reference*. Very often we are concerned with two reference frames that move relative to one another in a very simple and predictable way. Imagine that we are in outer space, not moving, and looking down on the spinning Earth at Mt Everest. As the Earth spins below us, we see Mt Everest rotating. Now imagine that we wanted to measure its height. Making such a measurement from that vantage point is complicated by the fact that our target (Mt Everest) is moving. However, if we were instead standing on Earth (and in view of Mt Everest, of course) such a measurement would be much easier because we have eliminated the motion between us and Mt Everest. What we have done is move from the *fixed frame of reference* of the Universe, to the *rotating frame of reference* of Mt Everest. It turns out that this is a very close analogy with the situation in MRI, as we shall see in Chapter 14: The quantity that is measured (the magnetization) is a vector that is always precessing (rotating) at a rate that depends upon the strength of the magnet. Since all of our descriptions, manipulations, and measurements take place in the presence of this main field, and thus in the presence of this rotation, it is often very convenient to move to a reference frame that rotates at this rate (the *angular frequency*).

An every-day example of a rotating coordinate system that is a very useful analogue of the situation we find in NMR is the Merry-Go-Round, a platform that rotates in two-dimensions about a vertical axis (a pole) (Figure 6.9). Let's consider two boys who are trying to talk to a girl sitting on one of the horses as the Merry-Go-Round rotates. The first boy is not on the Merry-Go-Round, but standing on the grass outside it. Let's call this the *fixed frame* (of reference) (Figure 6.9a), and designate him as b_{fix}, and designate his Cartesian coordinate system $\{x_{fix}, y_{fix}, z_{fix}\}$. The second boy is riding on the Merry-Go-Round. Let's call his frame of reference the *rotating frame* (Figure 6.9b), designate him as b_{rot}, and designate his Cartesian coordinate system $\{x_{rot}, y_{rot}, z_{rot}\}$. Let's notice the obvious stuff first: only the x and y components of the two frames are different - the z-axis (the pole) is the same in both frames: $z_{fix} = z_{rot}$. Furthermore, the boy on the Merry-Go-Round, b_{rot}, has the advantage in talking with the girl because she is *stationary in his frame*. On the contrary, the boy in the fix frame, b_{fix}, seeing the girl rotating with respect to him, and this makes it difficult to speak with her.

Now let's put this into a mathematical language. Consider a the time rate of change of a vector in some fixed frame, $\boldsymbol{V}$, written in terms of its components $\boldsymbol{v} = \{v_x, v_y, v_z\}$ along the Cartesian


(a) Fixed frame of merry-go-round.


(b) Rotating frame of merry-go-round.

Figure 6.9 The rotating frame and the Merry-Go-Round.

axes $\hat{\boldsymbol{r}} = \{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ in the fixed frame:

$$\boldsymbol{V} = v_x\hat{\boldsymbol{x}} + v_y\hat{\boldsymbol{y}} + v_z\hat{\boldsymbol{z}} = \boldsymbol{v}\cdot\hat{\boldsymbol{r}} \tag{6.57}$$

The time rate of change of $\boldsymbol{V}$ in the fixe frame is, by the chain rule,

$$\left(\frac{d\boldsymbol{V}}{dt}\right)_{fix} = \left(\frac{\partial \boldsymbol{v}}{\partial t}\right)_{rot}\cdot\hat{\boldsymbol{r}} + \boldsymbol{v}\cdot\frac{\partial\hat{\boldsymbol{r}}}{\partial t} \tag{6.58}$$

The first term on the right hand side depends on how much $\boldsymbol{v}$ is changing with time, while the second term depends on how the axes $\hat{\boldsymbol{x}}$ are changing with time. The unit vectors are of fixed length so their time rate of change can only involve a rotation, which, from the last section, is affected by the cross product:

$$\frac{\partial\hat{\boldsymbol{r}}_i}{\partial t} = \boldsymbol{\omega}\times\hat{\boldsymbol{r}}_i \tag{6.59}$$

where the magnitude of $\boldsymbol{\omega}$ is the angular frequency of rotation of the unit vector and the direction of $\boldsymbol{\omega}$ is the axis about which rotation the rotation occurs. Substituting Eqn 6.59 into Eqn 6.58 shows that the rate of change of the vector in the fixed frame is related to the rate of change in the rotating frame by the relationship

$$\left(\frac{d\boldsymbol{V}}{dt}\right)_{fix} = \left(\frac{\partial \boldsymbol{V}}{\partial t}\right)_{rot} + \boldsymbol{\omega}\times\boldsymbol{V} \tag{6.60}$$

The total derivative $d\boldsymbol{V}/dt$ expresses the motion of $\boldsymbol{V}$ in the fixed frame and the partial derivative $\partial\boldsymbol{V}/\partial t$ is the explicit dependence on time of $\boldsymbol{V}$ in the rotating frame. Notice that we can transform to the rotating frame by rearrangement of Eqn 6.61:

$$\left(\frac{\partial\boldsymbol{V}}{\partial t}\right)_{rot} = \left(\frac{d\boldsymbol{V}}{dt}\right)_{fix} - \boldsymbol{\omega}\times\boldsymbol{V} \tag{6.61}$$

This relationship will play an important role in our representation of the motion of the magnetization vector in MRI (Section 14.3).

Notice also that the transformation to the rotating frame can also be affected in a simple and intuitively clear way using the matrix representation of the cross product introduced in

Section 6.10. Putting Eqn 6.49 into Eqn 6.61 gives

$$\left(\frac{\partial \boldsymbol{V}}{\partial t}\right)_{rot} = \left(\frac{d\boldsymbol{V}}{dt}\right)_{fix} - \boldsymbol{\beta}\boldsymbol{V} \quad , \quad \boldsymbol{\beta} = \begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix} \tag{6.62}$$

A very nice feature of this representation is that $\boldsymbol{\beta}$ is decomposible into its components along the different axes (by virtue of Eqn 6.49) which makes transformation to rotating frames *along particular axes* easy to do. For example, consider a set of axes that is rotating only along the $\hat{z}$ axis and we want to transform to the rotating frame. In this case $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^t$ and so from Eqn 6.49

$$\boldsymbol{\beta} = \omega_z \boldsymbol{A}_z \tag{6.63}$$

Then the transformation to the rotating frame is Eqn 6.62 with

$$\boldsymbol{\beta} = \begin{pmatrix} 0 & -\omega_z & 0 \\ \omega_z & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.64}$$

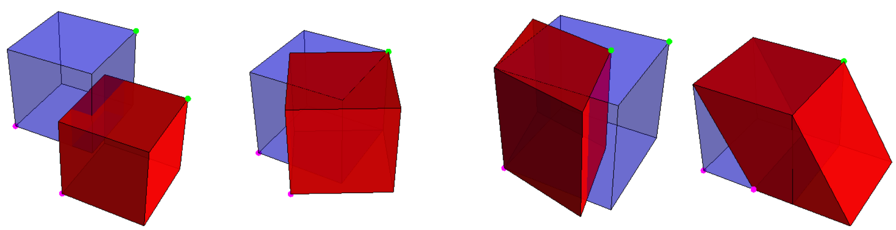
This is actually just the case we will encounter in the standard representation of the signal in MRI, as we will see in Chapter 14.

6.13 Affine Transformations

Because MRI is creating a spatial map of the body, there arise a class of spatial transformations that play an important role in improving the fidelity of MR images whose quality has been compromised by subject motion or imperfections in the magnetic fields used to create the images. Imagine that a patient moves their head in between the acquisition of two images. If the head is considered to be a solid object, this motion is called *rigid body motion*, and the movement can be generally considered to be some mixture of *translational* motion (shifting the center of the head from point x to x') and *rotational* motion (rotating about the center point of the head). Together, this motion is called *rigid body* motion. In later chapters we will see how well control magnetic fields are applied to create spatial maps (images) of water in the body. We will also see that if there are *relative* changes between the spatial encoding fields ("gradients"), the size of the image along that direction will by *scaled* can be different along different spatial directions, resulting in the image appearing stretched or contracted along certain directions. More complicated field variations can produce an effect that is called *shearing*, where all points on one axis are left unaffected but points away from the axis are shifted parallel to the axis by a distance proportional to their perpendicular distance from the axis. This shearing transformation does not change the volume of the region it is transforming.

In Figure 6.10 is shown an example of each of these transformations: translation (Figure 6.10a), rotation (Figure 6.10b), scaling (Figure 6.10c), and shearing Figure 6.10d). These four transformations: translation, rotation, scaling, and shearing, together make up the class of transformations called *affine* transformations. The beauty of the matrix methods presented in Chapter 5 becomes apparent in the study of affine transformations, as each of these can be represented by a 4×4 matrix in *homogeneous* coordinates (Section **??**).[5]

[5] Explain why homogeneous coordinates are used.



(a) Translation. (b) Rotation. (c) Scaling. (d) Shearing.

Figure 6.10 Affine transformations. (a) Translation by $\{\Delta x, \Delta y, \Delta z\} = \{-.75, -.25, -.5\}$. (b) Rotation by $\theta = \pi/6$ about the z-axis through the green point. (c) Scaling by .65 along the direction $\{x, y, z\} = \{1, 1, 0\}$ (i.e, along the line $x = y$). (d) Shearing by an amount $\phi = -\pi/6$ along the x-direction $\{1, 0, 0\}$, normal to the y-direction $\{0, 1, 0\}$ with the green point $\{0, 0, 1\}$ along the z-axis held fixed.

The general translation transformation is represented by the matrix

$$T = \begin{pmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.65}$$

The general scaling transformation is represented by the matrix

$$S = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.66}$$

Rotations can be constructed from the rotation matrices in Section 6.2 placed in the upper left hand 3×3 matrix elements, with zeros in the other elements except for a 1 in the position a_{44}. For example, a rotation around z by θ is, using Eqn 6.10c,

$$R = \begin{pmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.67}$$

Shearing along x-axis by ϕ can be represented as

$$H = \begin{pmatrix} 1 & 0 & \tan\phi & -\tan\phi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.68}$$

[6]

And there is one more remarkable matrix fact that makes affine transformations very efficient to use practically: All four of the separate transformations represented by matrices above can

[6] Say something about general shear transformations.

be combined (or *composed*) into a single transformation matrix. Representing the process of composition with the circle symbol, the general affine transformation can be written:

$$A = T \circ R \circ S \circ H = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.69}$$

We will use these affine transformations in the correction methods in Section 38.3.

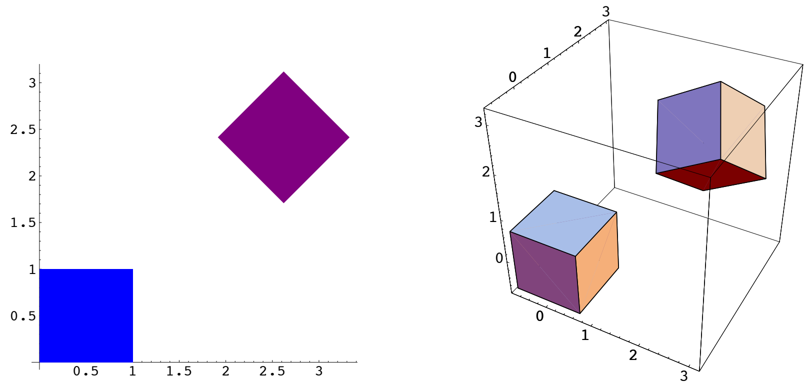
6.14 Interlude: Rigid body motion

Sometimes it is useful to consider a single coordinate system, for example the Cartesian coordinates (Section 2.2), but allow the possibility that its origin and orientation change. This type of transformation that occurs repeatedly in the context of MRI in the characterization of the motion of the subject (e.g., a human head) during the scan process. In order to estimate and correct for the effects due to such motion, it is necessary to have a model for the motion that can then be used to estimated the parameters of the motion and remove its effects from the images. A simple, and tractable, description of the motion begins with the assumptions that the head is a *solid object*. By this we mean that points in the object do not change relative to one another. Imagine a marble bust, for instance. This is not the case, for example, with a piece of Jello, in which the internal points *do* change locations relative to one another. We say that Jello *deforms*. Unfortunately, the brain is more like Jello than marble, since not only does the head position move, but the brain pulsates (via the CSF pressure changes with heart beat) and thus deforms. This is a much more complicated problem, and we will not address it. But let's look at the effect of the motion of a solid object, which is called *rigid body motion*.

Pick up any solid object near you and extend your arm and twist your wrist, and watch the position of the object. It should be immediately clear that the motion of the object is comprised two types of motion. The first is the movement of its center from one location to another as your arm extends. This is called *translation*. The second is the motion about its center as your wrist turns. This is our familiar *rotation*. The order of these two transformations is independent. That is, we can first rotate the object, the translating it, or vise versa, and the final position and orientation are the same. But notice that the rotation of the object occurs *relative to its internal coordinates*. That is, we rotate the object about its center (or center of mass if it is an asymmetric object). The translation, on the other hand, occurs relative to the origin located at, say, the shoulder. This problem thus has two natural coordinate systems: One with an origin at our shoulder (so that our arms length is the distance from the origin, and thus the translation distance) and a coordinate system that sits within the object and is the one about which it rotates. The translation and rotation of the object can thus also be seen as the translation and rotation of two Cartesian coordinate system relative to one another. These are critical distinctions to be made if one is to describe these mathematically, which we do now. To put this on a firm mathematical footing, the location of a point on the rigid body can be written

$$\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{R}(\alpha, \beta, \gamma)\boldsymbol{x}_o \tag{6.70}$$

where $\boldsymbol{x}_r$ is a reference point on the body (say, its center of mass), $\boldsymbol{x}_o$ is a point on the body with respect to the reference point $\boldsymbol{x}_r$, and $\boldsymbol{R}(\alpha, \beta, \gamma)$ is the 3×3 rotation matrix. Thus $\boldsymbol{x}_r$ is



(a) Rigid body motion of a solid square whose center is at the location $\boldsymbol{x}_0 \equiv \{x_0, y_0\}$ (which in this case is $\boldsymbol{x} = \{.5, .5, .5\}$). The cube is rotated about its center $\{.5, .5\}$ by 45° and then translated from $\boldsymbol{x} \equiv \{x_0, y_0\}$ to $\boldsymbol{x} + \boldsymbol{dx}$ where $\boldsymbol{dx} \equiv \{dx, dy\} = \{2, 2\}$.

(b) Rigid body motion of a solid cube whose center is at the location $\boldsymbol{x}_0 \equiv \{x_0, y_0, z_0\}$ (which in this case is $\boldsymbol{x} = \{.5, .5, .5\}$). The cube is rotated about its center by the Euler angles (Section **??**) $\{\phi, \theta, \psi\} = \{\pi/4, \pi/3, \pi/2\}$ and then translated from $\boldsymbol{x} \equiv \{x_0, y_0, z_0\}$ to $\boldsymbol{x} + \boldsymbol{dx}$ where $\boldsymbol{dx} \equiv \{dx, dy, dz\} = \{2, 2, 2\}$.

Figure 6.11 Rigid body motion in (a) 2D of a square (b) 3D of a cube.

a translation of the object from the origin, while $\boldsymbol{R}(\alpha, \beta, \gamma)\boldsymbol{x}_o$ rotates the object *with respect to the coordinates of the body.* This is illustratied in Figure 6.11.

Suggested Reading

1. *book 1*
   This book is good.