2

Coordinate Systems and Coordinate Transformations

The field of mathematics known as topology describes space in a very general sort of way. Many spaces are exotic and have no counterpart in the physical world. Indeed, in the hierarchy of spaces defined within topology, those that can be described by a coordinate system are among the more sophisticated. These are the spaces of interest to the physical scientist because they are reminiscent of the physical space in which we exist. The utility of such spaces is derived from the presence of a coordinate system which allows one to describe phenomena that take place within the space. However, the spaces of interest need not simply be the physical space of the real world. One can imagine the temperature-pressure-density space of thermodynamics or many of the other spaces where the dimensions are physical variables. One of the most important of these spaces for mechanics is phase space. This is a multi-dimensional space that contains the position and momentum coordinates for a collection of particles. Thus physical spaces can have many forms. However, they all have one thing in common. They are described by some coordinate system or frame of reference. Imagine a set of rigid rods or vectors all connected at a point. Such a set of 'rods' is called a frame of reference. If every point in the space can uniquely be projected onto the rods so that a unique collection of rod-points identify the point in space, the reference frame is said to span the space.

2.1 Orthogonal Coordinate Systems

If the vectors that define the coordinate frame are locally perpendicular, the coordinate frame is said to be orthogonal. Imagine a set of unit basis vectors $\hat{e}_i$ that span some space. We can express the condition of orthogonality by

$$\hat{e}_i \bullet \hat{e}_j = \delta_{ij} \quad , \tag{2.1.1}$$

where δ_{ij} is the Kronecker delta that we introduced in the previous chapter. Such a set of basis vectors is said to be orthogonal and will span a space of n-dimensions where n is the number of vectors $\hat{e}_i$. It is worth noting that the space need not be Euclidean. However, if the space is Euclidean and the coordinate frame is orthogonal, then the coordinate frame is said to be a Cartesian frame. The standard xyz coordinate frame is a Cartesian frame. One can imagine such a coordinate frame drawn on a rubber sheet. If the sheet is distorted in such a manner that the local orthogonality conditions are still met, the coordinate frame may remain orthogonal but the space may no longer be a Euclidean space. For example, consider the ordinary coordinates of latitude and longitude on the surface of the earth. These coordinates are indeed orthogonal but the surface is not the Euclidean plane and the coordinates are not Cartesian.

Of the orthogonal coordinate systems, there are several that are in common use for the description of the physical world. Certainly the most common is the Cartesian or rectangular coordinate system (xyz). Probably the second most common and of paramount importance for astronomy is the system of spherical or polar coordinates (r,θ,ϕ). Less common but still very important are the cylindrical coordinates (r,ϑ,z). There are a total of thirteen orthogonal coordinate systems in which Laplace’s equation is separable, and knowledge of their existence (see Morse and Feshback[1]) can be useful for solving problems in potential theory. Recently the dynamics of ellipsoidal galaxies has been understood in a semi-analytic manner by employing ellipsoidal coordinates and some potentials defined therein. While these more exotic coordinates were largely concerns of the nineteenth century mathematical physicists, they still have relevance today. Often the most important part of solving a problem in mathematical physics is the choice of the proper coordinate system in which to do the analysis.

In order to completely define any coordinate system one must do more than just specify the space and coordinate geometry. In addition, the origin of the coordinate system and its orientation must be given. In celestial mechanics there are three important locations for the origin. For observation, the origin can be taken to be the observer (topocentric coordinates). However for interpretation of the observations it is usually necessary to refer the observations to coordinate systems with their origin at the center of the earth (geocentric coordinates) or the center of the sun (heliocentric coordinates) or at the center of mass of the solar system (barycentric coordinates). The orientation is only important when the coordinate frame is to be compared or transformed to another coordinate frame. This is usually done by defining the zero-point of some coordinate with respect to the coordinates of the other frame as well as specifying the relative orientation.

2.2 Astronomical Coordinate Systems

The coordinate systems of astronomical importance are nearly all spherical coordinate systems. The obvious reason for this is that most all astronomical objects are remote from the earth and so appear to move on the backdrop of the celestial sphere. While one may still use a spherical coordinate system for nearby objects, it may be necessary to choose the origin to be the observer to avoid problems with parallax. These orthogonal coordinate frames will differ only in the location of the origin and their relative orientation to one another. Since they have their foundation in observations made from the earth, their relative orientation is related to the orientation of the earth's rotation axis with respect to the stars and the sun. The most important of these coordinate systems is the Right Ascension -Declination coordinate system.

a. The Right Ascension - Declination Coordinate System

This coordinate system is a spherical-polar coordinate system where the polar angle, instead of being measured from the axis of the coordinate system, is measured from the system's equatorial plane. Thus the declination is the angular complement of the polar angle. Simply put, it is the angular distance to the astronomical object measured north or south from the equator of the earth as projected out onto the celestial sphere. For measurements of distant objects made from the earth, the origin of the coordinate system can be taken to be at the center of the earth. At least the 'azimuthal' angle of the coordinate system is measured in the proper fashion. That is, if one points the thumb of his right hand toward the North Pole, then the fingers will point in the direction of increasing Right Ascension. Some remember it by noting that the Right Ascension of rising or ascending stars increases with time. There is a tendency for some to face south and think that the angle should increase to their right as if they were looking at a map. This is exactly the reverse of the true situation and the notion so confused air force navigators during the Second World War that the complementary angle, known as the sidereal hour angle, was invented. This angular coordinate is just 24 hours minus the Right Ascension.

Another aspect of this Right Ascension that many find confusing is that it is not measured in any common angular measure like degrees or radians. Rather it is measured in hours, minutes, and seconds of time. However, these units are the natural ones as the rotation of the earth on its axis causes any fixed point in the sky to return to the same place after about 24 hours. We still have to define the zero-point from which the Right Ascension angle is measured. This also is inspired by the orientation of the earth. The projection of the orbital plane of the earth on the celestial sphere is described by the path taken by the sun during the year. This path is called the ecliptic. Since the rotation axis of the earth is inclined to the orbital plane, the ecliptic and equator, represented by great circles on the celestial sphere, cross at two points 180° apart. The points are known as equinoxes, for when the sun is at them it will lie in the plane of the equator of the earth and the length of night and day will be equal. The sun will visit each once a year, one when it is headed north along the ecliptic and the other when it is headed south. The former is called the vernal equinox as it marks the beginning of spring in the northern hemisphere while the latter is called the autumnal equinox. The point in the sky known as the vernal equinox is the zero-point of the Right Ascension coordinate, and the Right Ascension of an astronomical object is measured eastward from that point in hours, minutes, and seconds of *time*.

While the origin of the coordinate system can be taken to be the center of the earth, it might also be taken to be the center of the sun. Here the coordinate system can be imagined as simply being shifted without changing its orientation until its origin corresponds with the center of the sun. Such a coordinate system is useful in the studies of stellar kinematics. For some studies in stellar dynamics, it is necessary to refer to a coordinate system with an origin at the center of mass of the earth-moon system. These are known as barycentric coordinates. Indeed, since the term barycenter refers to the center of mass, the term barycentric coordinates may also be used to refer to a coordinate system whose origin is at the center of mass of the solar system. The domination of the sun over the solar system insures that this origin will be very near, but not the same as the origin of the heliocentric coordinate system. Small as the differences of origin between the heliocentric and barycentric coordinates is, it is large enough to be significant for some problems such as the timing of pulsars.

b. Ecliptic Coordinates

The ecliptic coordinate system is used largely for studies involving planets and asteroids as their motion, with some notable exceptions, is confined to the zodiac. Conceptually it is very similar to the Right Ascension-Declination coordinate system. The defining plane is the ecliptic instead of the equator and the "azimuthal" coordinate is measured in the same direction as Right Ascension, but is usually measured in degrees. The polar and azimuthal angles carry the somewhat unfortunate names of celestial latitude and celestial longitude respectively in spite of the fact that these names would be more appropriate for Declination and Right Ascension. Again these coordinates may exist in the topocentric, geocentric, heliocentric, or barycentric forms.

c. Alt-Azimuth Coordinate System

The Altitude-Azimuth coordinate system is the most familiar to the general public. The origin of this coordinate system is the observer and it is rarely shifted to any other point. The fundamental plane of the system contains the observer and the horizon. While the horizon is an intuitively obvious concept, a rigorous definition is needed as the apparent horizon is rarely coincident with the location of the true horizon. To define it, one must first define the zenith. This is the point directly over the observer's head, but is more carefully defined as the extension of the local gravity vector outward through the celestial sphere. This point is known as the astronomical zenith. Except for the oblatness of the earth, this zenith is usually close to the extension of the local radius vector from the center of the earth through the observer to the celestial sphere. The presence of large masses nearby (such as a mountain) could cause the local gravity vector to depart even further from the local radius vector. The horizon is then that line on the celestial sphere which is everywhere 90° from the zenith. The altitude of an object is the angular distance of an object above or below the horizon measured along a great circle passing through the object and the zenith. The azimuthal angle of this coordinate system is then just the azimuth of the object. The only problem here arises from the location of the zero point. Many older books on astronomy will tell you that the azimuth is measured westward from the south point of the horizon. However, only astronomers did this and most of them don't anymore. Surveyors, pilots and navigators, and virtually anyone concerned with local coordinate systems measures the azimuth from the *north* point of the horizon increasing through the east point around to the west. That is the position that I take throughout this book. Thus the azimuth of the cardinal points of the compass are: N(0°), E(90°), S(180°), W(270°).

2.3 Geographic Coordinate Systems

Before leaving the subject of specialized coordinate systems, we should say something about the coordinate systems that measure the surface of the earth. To an excellent approximation the shape of the earth is that of an oblate spheroid. This can cause some problems with the meaning of local vertical.

a. The Astronomical Coordinate System

The traditional coordinate system for locating positions on the surface of the earth is the latitude-longitude coordinate system. Most everyone has a feeling for this system as the latitude is simply the angular distance north or south of the equator measured along the local meridian toward the pole while the longitude is the angular distance measured along the equator to the local meridian from some reference meridian. This reference meridian has historically be taken to be that through a specific instrument (the Airy transit) located in Greenwich England. By a convention recently adopted by the International Astronomical Union, longitudes measured east of Greenwich are considered to be positive and those measured to the west are considered to be negative. Such coordinates provide a proper understanding for a perfectly spherical earth. But for an earth that is not exactly spherical, more care needs to be taken.

b. The Geodetic Coordinate System

In an attempt to allow for a non-spherical earth, a coordinate system has been devised that approximates the shape of the earth by an oblate spheroid. Such a figure can be generated by rotating an ellipse about its minor axis, which then forms the axis of the coordinate system. The plane swept out by the major axis of the ellipse is then its equator. This approximation to the actual shape of the earth is really quite good. The *geodetic latitude* is now given by the angle between the local vertical and the plane of the equator where the local vertical is the normal to the oblate spheroid at the point in question. The *geodetic longitude* is roughly the same as in the astronomical coordinate system and is the angle between the local meridian and the meridian at Greenwich. The difference between the local vertical (i.e. the normal to the local surface) and the astronomical vertical (defined by the local gravity vector) is known as the "*deflection of the vertical*" and is usually less than 20 arc-sec. The oblatness of the earth allows for the introduction of a third coordinate system sometimes called the *geocentric coordinate system*.

c. The Geocentric Coordinate System

Consider the oblate spheroid that best fits the actual figure of the earth. Now consider a radius vector from the center to an arbitrary point on the surface of that spheroid. In general, that radius vector will not be normal to the surface of the oblate spheroid (except at the poles and the equator) so that it will define a different local vertical. This in turn can be used to define a different latitude from either the astronomical or geodetic latitude. For the earth, the maximum difference between the geocentric and geodetic latitudes occurs at about 45° latitude and amounts to about (11' 33"). While this may not seem like much, it amounts to about eleven and a half nautical miles (13.3 miles or 21.4 km.) on the surface of the earth. Thus, if you really want to know where you are you must be careful which coordinate system you are using. Again the *geocentric longitude* is defined in the same manner as the geodetic longitude, namely it is the angle between the local meridian and the meridian at Greenwich.

2.4 Coordinate Transformations

A great deal of the practical side of celestial mechanics involves transforming observational quantities from one coordinate system to another. Thus it is appropriate that we discuss the manner in which this is done in general to find the rules that apply to the problems we will encounter in celestial mechanics. While within the framework of mathematics it is possible to define myriads of coordinate transformations, we shall concern ourselves with a special subset called *linear transformations*. Such coordinate transformations relate the coordinates in one frame to those in a second frame by means of a system of linear algebraic equations. Thus if a vector $\vec{X}$ in one coordinate system has components X_j, in a primed-coordinate system a vector $\vec{X}'$ to the same point will have components X_j given by

$$X_i' = \sum_j A_{ij} X_j + B_i \quad . \tag{2.4.1}$$

In vector notation we could write this as

$$\vec{X}' = \mathbf{A}\vec{X} + \vec{B} \quad . \tag{2.4.2}$$

This defines the general class of linear transformation where **A** is some matrix and $\vec{B}$ is a vector. This general linear form may be divided into two constituents, the matrix **A** and the vector $\vec{B}$. It is clear that the vector $\vec{B}$ may be interpreted as a shift in the origin of the coordinate system, while the elements A_{ij} are the cosines of the angles between the axes X_i and X_j and are called the directions cosines (see Figure 2.1). Indeed, the vector $\vec{B}$ is nothing more than a vector from the origin of the un-primed coordinate frame to the origin of the primed coordinate frame. Now if we consider two points that are fixed in space and a vector connecting them, then the length and orientation of that vector will be independent of the origin of the coordinate frame in which the measurements are made. That places an additional constraint on the types of linear transformations that we may consider. For instance, transformations that scaled each coordinate by a constant amount, while linear, would change the length of the vector as measured in the two coordinate systems. Since we are only using the coordinate system as a convenient way to describe the vector, its length must be independent of the coordinate system. Thus we shall restrict our investigations of linear transformations to those that transform orthogonal coordinate systems while preserving the length of the vector.

Thus the matrix **A** must satisfy the following condition

$$\vec{X}' \bullet \vec{X}' = (\mathbf{A}\vec{X}) \bullet (\mathbf{A}\vec{X}) = \vec{X} \bullet \vec{X} \quad , \tag{2.4.3}$$

which in component form becomes

$$\sum_i (\sum_j A_{ij} X_j)(\sum_k A_{ik} X_k) = \sum_j \sum_k (\sum_i A_{ij} A_{ik}) X_j X_k = \sum_i X_i^2 \quad . \tag{2.4.4}$$

This must be true for all vectors in the coordinate system so that

$$\sum_i A_{ij} A_{ik} = \delta_{jk} = \sum_i A_{ji}^{-1} A_{ik} \quad . \tag{2.4.5}$$

Now remember that the Kronecker delta δ_{ij} is the unit matrix and any element of a group that multiplies another and produces that group's unit element is defined as the inverse of that element. Therefore

$$A_{ji} = [A_{ij}]^{-1} \quad . \tag{2.4.6}$$

Interchanging the elements of a matrix produces a new matrix which we have called the transpose of the matrix. Thus orthogonal transformations that preserve the length of vectors have inverses that are simply the transpose of the original matrix so that

$$\mathbf{A}^{-1} = \mathbf{A}^{\mathbf{T}} \quad . \tag{2.4.7}$$

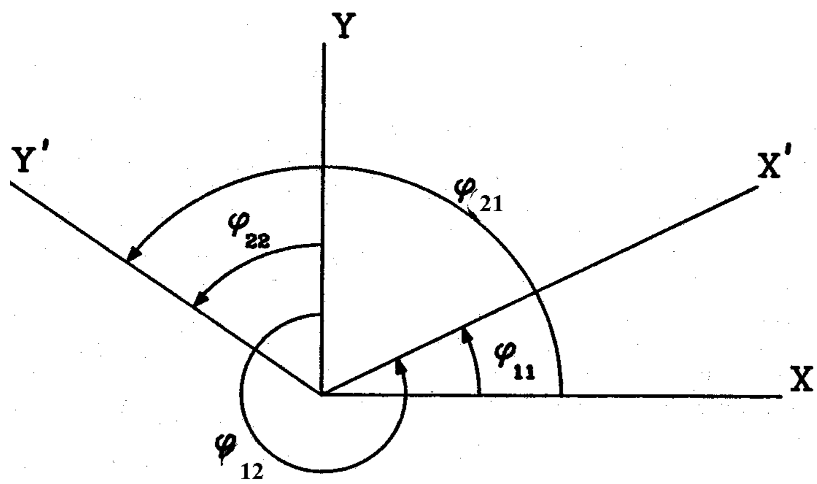

Figure 2.1 shows two coordinate frames related by the transformation angles φ_{ij}. Four coordinates are necessary if the frames are not orthogonal.

This means that given that transformation **A** in the linear system of equations (2.4.2), we may invert the transformation, or solve the linear equations, by multiplying those equations by the *transpose* of the original matrix or

$$\vec{X} = \mathbf{A}^{\mathbf{T}}\vec{X}' - \mathbf{A}^{\mathbf{T}}\vec{B} \quad . \tag{2.4.8}$$

Such transformations are called orthogonal unitary transformations, or orthonormal transformations, and the result given in equation (2.4.8) greatly simplifies the process of carrying out a transformation from one coordinate system to another and back again.

We can further divide orthonormal transformations into two categories. These are most easily described by visualizing the relative orientation between the two coordinate systems. Consider a transformation that carries one coordinate into the negative of its counterpart in the new coordinate system while leaving the others unchanged. If the changed coordinate is, say, the x-coordinate, the transformation matrix would be

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad , \tag{2.4.9}$$

which is equivalent to viewing the first coordinate system in a mirror. Such transformations are known as *reflection transformations* and will take a right handed coordinate system into a left handed coordinate system. The length of any vectors will remain unchanged. The x-component of these vectors will simply be replaced by its negative in the new coordinate system. However, this will not be true of "vectors" that result from the vector cross product. The values of the components of such a vector will remain unchanged implying that a reflection transformation of such a vector will result in the orientation of that vector being changed. If you will, this is the origin of the "right hand rule" for vector cross products. A left hand rule results in a vector pointing in the opposite direction. Thus such vectors are not invariant to reflection transformations because their orientation changes and this is the reason for putting them in a separate class, namely the axial (pseudo) vectors. Since the Levi-Civita tensor generates the vector cross product from the elements of ordinary (polar) vectors, it must share this strange transformation property. Tensors that share this transformation property are, in general, known as tensor densities or pseudo-tensors. Therefore we should call ε_{ijk} defined in equation (1.2.7) the Levi-Civita tensor density.

Indeed, it is the invariance of tensors, vectors, and scalars to orthonormal transformations that is most correctly used to define the elements of the group called tensors. Finally, it is worth noting that an orthonormal reflection transformation will have a determinant of -1. The unitary magnitude of the determinant is a result of the magnitude of the vector being unchanged by the transformation, while the sign shows that some combination of coordinates has undergone a reflection.

As one might expect, the elements of the second class of orthonormal transformations have determinants of +1. These represent transformations that can be viewed as a rotation of the coordinate system about some axis. Consider a transformation between the two coordinate systems displayed in Figure 2.1. The components of any vector $\vec{C}$ in the primed coordinate system will be given by

$$\begin{pmatrix} C_{x'} \\ C_{y'} \\ C_{z'} \end{pmatrix} = \begin{pmatrix} \cos\phi_{11} & \cos\phi_{12} & 0 \\ \cos\phi_{21} & \cos\phi_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix} \quad . \tag{2.4.10}$$

If we require the transformation to be orthonormal, then the direction cosines of the transformation will not be linearly independent since the angles between the axes must be $\pi/2$ in both coordinate systems. Thus the angles must be related by

$$\left.\begin{aligned} &\phi_{11} = \phi_{22} = \phi \\ &\phi_{21} = \phi_{11} + \pi/2 = \phi + \pi/2 \\ &(2\pi - \phi_{12}) = (\pi/2) - \phi_{22}, \ \Rightarrow \phi_{12} = (\phi + \pi/2) + \pi \end{aligned}\right\} \quad . \tag{2.4.11}$$

Using the addition identities for trigonometric functions, equation (2.4.10) can be given in terms of the single angle ϕ by

$$\begin{pmatrix} C_{x'} \\ C_{y'} \\ C_{z'} \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix} \quad . \tag{2.4.12}$$

This transformation can be viewed simple rotation of the coordinate system about the Z-axis through an angle ϕ. Thus, as a

$$\mathrm{Det} \begin{vmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{vmatrix} = \cos^2\phi + \sin^2\phi = +1 \quad . \tag{2.4.13}$$

In general, the rotation of any Cartesian coordinate system about one of its principal axes can be written in terms of a matrix whose elements can be expressed in terms of the rotation angle. Since these transformations are about one of the coordinate axes, the components along that axis remain unchanged. The rotation matrices for each of the three axes are

$$
\left.
\begin{aligned}
\mathbf{P}_x(\phi) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix} \\[2ex]
\mathbf{P}_y(\phi) &= \begin{pmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{pmatrix} \\[2ex]
\mathbf{P}_z(\phi) &= \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned}
\right\} \quad . \tag{2.4.14}
$$

It is relatively easy to remember the form of these matrices for the row and column of the matrix corresponding to the rotation axis always contains the elements of the unit matrix since that component are not affected by the transformation. The diagonal elements always contain the cosine of the rotation angle while the remaining off diagonal elements always contains the sine of the angle modulo a sign. For rotations about the X- or Z-axes, the sign of the upper right off diagonal element is positive and the other negative. The situation is just reversed for rotations about the Y-axis. So important are these rotation matrices that it is worth remembering their form so that they need not be re-derived every time they are needed.

One can show that it is possible to get from any given orthogonal coordinate system to another through a series of three successive coordinate rotations. Thus a general orthonormal transformation can always be written as the product of three coordinate rotations about the orthogonal axes of the coordinate systems. It is important to remember that the matrix product is not commutative so that the order of the rotations is important. So important is this result, that the angles used for such a series of transformations have a specific name.

2.5 The Eulerian Angles

Leonard Euler proved that the general motion of a rigid body when one point is held fixed corresponds to a series of three rotations about three orthogonal coordinate axes. Unfortunately the definition of the Eulerian angles in the literature is not always the same (see Goldstein[2] p.108). We shall use the definitions of Goldstein and generally follow them throughout this book. The order of the rotations is as follows. One begins with a rotation about the Z-axis. This is followed by a rotation about the new X-axis. This, in turn, is followed by a rotation about the resulting Z"-axis. The three successive rotation angles are $[\phi, \theta, \psi]$.

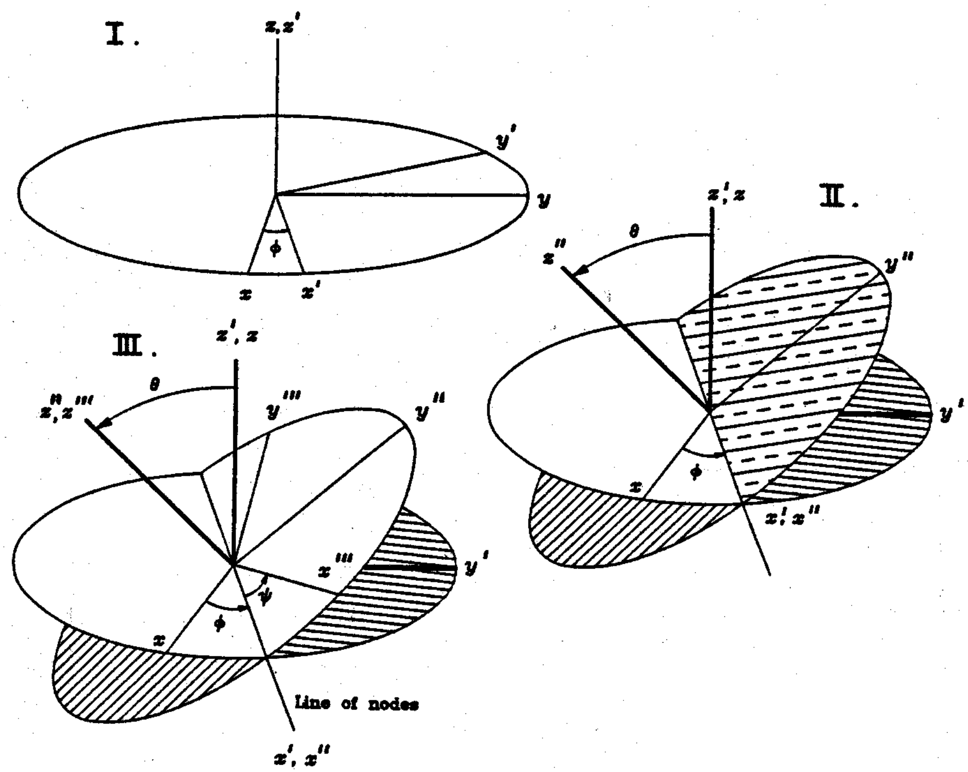

Figure 2.2 shows the three successive rotational transformations corresponding to the three Euler Angles (ϕ, θ, ψ) transformation from one orthogonal coordinate frame to another that bears an arbitrary orientation with respect to the first.

This series of rotations is shown in Figure 2.2. Each of these rotational transformations is represented by a transformation matrix of the type given in equation (2.4.14) so that the complete set of Eulerian transformation matrices is

$$\left.\begin{aligned}
\mathbf{P}_z(\phi) &= \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
\mathbf{P}_{x'}(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \\
\mathbf{P}_{z''}(\psi) &= \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned}\right\} \quad , \tag{2.5.1}$$

and the complete single matrix that describes these transformations is

$$\mathbf{A}(\phi,\theta,\psi) = \mathbf{P}_{z''}(\psi)\mathbf{P}_{x'}(\theta)\mathbf{P}_z(\phi) \quad . \tag{2.5.2}$$

Thus the components of any vector $\vec{X}$ can be found in any other coordinate system as the components of $\vec{X}'$ from

$$\vec{X}' = \mathbf{A}\vec{X} \quad . \tag{2.5.3}$$

Since the inverse of orthonormal transformations has such a simple form, the inverse of the operation can easily be found from

$$\vec{X} = \mathbf{A}^{-1}\vec{X}' = \mathbf{A}^{\mathbf{T}}\vec{X}' = [\mathbf{P}_z^{\mathbf{T}}(\phi)\mathbf{P}_{x'}^{\mathbf{T}}(\theta)\mathbf{P}_{z''}^{\mathbf{T}}(\psi)]\,\vec{X}' \quad . \tag{2.5.4}$$

2.6 The Astronomical Triangle

The rotational transformations described in the previous section enable simple and speedy representations of the vector components of one Cartesian system in terms of those of another. However, most of the astronomical coordinate systems are spherical coordinate systems where the coordinates are measured in arc lengths and angles. The transformation from one of these coordinate frames to another is less obvious. One of the classical problems in astronomy is relating the defining coordinates of some point in the sky (say representing a star or planet), to the local coordinates of the observer at any given time. This is usually accomplished by means of the Astronomical Triangle which relates one system of coordinates to the other through the use of a spherical triangle. The solution of that triangle is usually quoted *ex cathedra* as resulting from spherical trigonometry. Instead of this approach, we shall show how the result (and many additional results) may be generated from the rotational transformations that have just been described.

Since the celestial sphere rotates about the north celestial pole due to the rotation of the earth, a great circle through the north celestial pole and the object (a meridian) appears to move across the sky with the object. That meridian will make some angle at the pole with the observer’s local prime meridian (i.e. the great circle through the north celestial pole and the observer's zenith). This angle is known as the local hour angle and may be calculated knowing the object's right ascension and the sidereal time. This latter quantity is obtained from the local time (including date) and the observer's longitude. Thus, given the local time, the observer's location on the surface of the earth (i.e. the latitude and longitude), and the coordinates of the object (i.e. its Right Ascension and declination), two sides and an included angle of the spherical triangle shown in Figure 2.3 may be considered known. The problem then becomes finding the remaining two angles and the included side. This will yield the local azimuth A, the zenith distance z which is the complement of the altitude, and a quantity known as the parallactic angle η. While this latter quantity is not necessary for locating the object approximately in the sky, it is useful for correcting for atmospheric refraction which will cause the image to be slightly displaced along the vertical circle from its true location. This will then enter into the correction for atmospheric extinction and is therefore useful for photometry.

In order to solve this problem, we will solve a separate problem. Consider a Cartesian coordinate system with a z-axis pointing along the radius vector from the origin of both astronomical coordinate systems (i.e. equatorial and alt-azimuth) to the point Q. Let the y-axis lie in the meridian plane containing Q and be pointed toward the north celestial pole. The x-axis will then simply be orthogonal to the y- and z-axes. Now consider the components of any vector in this coordinate system. By means of rotational transformations we may calculate the components of that vector in any other coordinate frame. Therefore consider a series of rotational transformations that would carry us through the sides and angles of the astronomical triangle so that we return to exactly the initial xyz coordinate system. Since the series of transformations that accomplish this must exactly reproduce the components of the initial arbitrary vector, the transformation matrix must be the unit matrix with elements δ_{ij}. If we proceed from point Q to the north celestial pole and then on to the zenith, the rotational transformations will involve only quantities related to the given part of our problem [i.e. $(\pi/2-\delta)$, h, $(\pi/2-\phi)$] .Completing the trip from the zenith to Q will involve the three local quantities [i.e. A, $(\pi/2-H)$, η] . The total transformation matrix will then involve six rotational matrices, the first three of which involve given angles and the last three of which involve unknowns and it is this total matrix which is equal to the unit matrix. Since all of the transformation matrices represent orthonormal transformations, their inverse is simply their transpose. Thus we can generate a matrix equation, one side of which involves matrices of known quantities and the other side of which contains matrices of the unknown quantities.

Let us now follow this program and see where it leads. The first rotation of our initial coordinate system will be through the angle $[-(\pi/2 - \delta)]$. This will carry us through the complement of the declination and align the z-axis with the rotation axis of the earth. Since the rotation will be about the x-axis, the appropriate rotation matrix from equation (2.4.14) will be

$$\mathbf{P}_x\left[-\left(\frac{\pi}{2}-\delta\right)\right]=\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin\delta & -\cos\delta \\ 0 & \cos\delta & \sin\delta \end{pmatrix} \quad . \qquad (2.6.1)$$

Now rotate about the new z-axis that is aligned with the polar axis through a counterclockwise or positive rotation of (h) so that the new y-axis lies in the local prime meridian plane pointing away from the zenith. The rotation matrix for this transformation involves the hour angle so that

$$\mathbf{P}_z(h)=\begin{pmatrix} \cos h & \sin h & 0 \\ -\sin h & \cos h & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \qquad (2.6.2)$$

Continue the trip by rotating through [+(π/2 - ϕ)] so that the z-axis of the coordinate system aligns with a radius vector through the zenith. This will require a positive rotation about the x-axis so that the appropriate transformation matrix is

$$\mathbf{P}_x\left[-\left(\frac{\pi}{2}-\phi\right)\right]=\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin\phi & \cos\phi \\ 0 & -\cos\phi & \sin\phi \end{pmatrix} \quad . \tag{2.6.3}$$

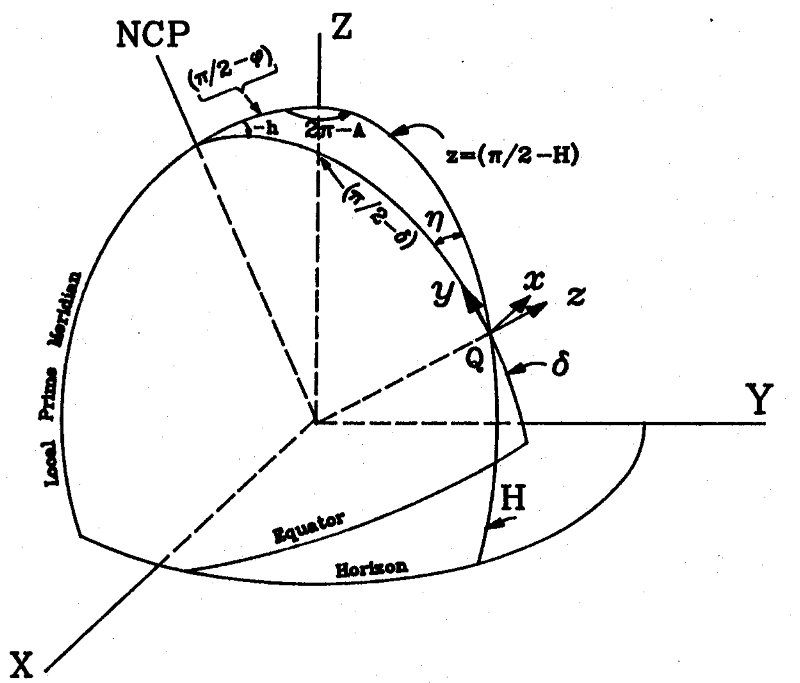

Figure 2.3 shows the Astronomical Triangle with the zenith in the Z-direction. The solution of this triangle is necessary for transformations between the Alt-Azimuth coordinate system and the Right Ascension-Declination coordinate system. The latter coordinates are found from the hour angle h and the distance from the North Celestial Pole.

Now rotate about the z-axis through the azimuth [2π-A] so that the y-axis will now be directed toward the point in question Q. This is another z-rotation so that the appropriate transformation matrix is

$$\mathbf{P}_z[2\pi - A] = \begin{pmatrix} \cos A & -\sin A & 0 \\ \sin A & \cos A & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{2.6.4}$$

We may return the z-axis to the original z-axis by a negative rotation about the x-axis through the zenith distance (π/2-H) which yields a transformation matrix

$$\mathbf{P}_x\left[-\left(\frac{\pi}{2} - H\right)\right] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin H & -\cos H \\ 0 & \cos H & \sin H \end{pmatrix} . \tag{2.6.5}$$

Finally the coordinate frame may be aligned with the starting frame by a rotation about the z-axis through an angle [π+η] yielding the final transformation matrix

$$\mathbf{P}_z[\pi + \eta] = \begin{pmatrix} -\cos\eta & -\sin\eta & 0 \\ +\sin\eta & -\cos\eta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{2.6.6}$$

Since the end result of all these transformations is to return to the starting coordinate frame, the product of all the transformations yields the identity matrix or

$$\mathbf{P}_z[+(\pi+\eta)]\mathbf{P}_x(-z)\mathbf{P}_z(2\pi - A)\mathbf{P}_x[+(\pi/2-\phi)]\mathbf{P}_z(h)\mathbf{P}_x[\delta - \pi/2] = \mathbf{1} \quad . \tag{2.6.7}$$

We may separate the knowns from the unknowns by remembering that the inverse of an orthonormal transformation matrix is its transpose so that

$$\mathbf{P}_z[+(\pi+\eta)]\mathbf{P}_x(-z)\mathbf{P}_z(2\pi - A) = \mathbf{P}_x^{\mathbf{T}}(\delta - \pi/2)\mathbf{P}_z^{\mathbf{T}}(h)\mathbf{P}_x^{\mathbf{T}}[(\pi/2) - \phi] \quad . \tag{2.6.8}$$

We must now explicitly perform the matrix products implied by equation (2.6.8) and the nine elements of the left hand side must equal the nine elements of the right hand side. These nine relations provide in a natural way all of the relations possible for the spherical triangle. These, of course, include the usual relations quoted for the solution to the astronomical triangle. These nine relations are

$$\left.\begin{aligned}
&\sin H = \cos h \cos\phi \cos\delta + \sin\delta \sin\phi \\
&\cos H \cos A = \sin\delta \cos\phi - \cos h \cos\delta \sin\phi \\
&\cos H \sin A = -\cos\delta \sin h \\
&\cos H \cos\eta = \cos h \sin\delta \cos\phi - \cos\delta \sin\phi \\
&\cos H \sin\eta = \sin h \cos\phi \\
&\sin A \sin\eta + \sin H \cos A \cos\eta = -(\cos\delta \cos\phi + \cos h \sin\phi \sin\delta) \\
&\sin A \cos\eta + \sin H \cos A \sin\eta = -\sin\phi \sin h \\
&\cos A \sin\eta + \sin H \sin A \cos\eta = +\sin\delta \sin h \\
&\cos A \cos\eta + \sin H \sin A \sin\eta = -\cos h
\end{aligned}\right\} . \qquad (2.6.9)$$

Since the altitude is defined to lie in the first or fourth quadrants, the first of these relations uniquely specifies H. The next two will then uniquely give the azimuth A and the following two allow for the unique specification of the parallactic angle. Thus these relations are sufficient to effect the coordinate transformation from either the defining coordinate frame to the observer's frame or vice versa. However, the more traditional solution of the astronomical triangle can be found from

$$\mathbf{P}_z[+(\pi+\eta)]\mathbf{P}_x(H-\pi/2)\mathbf{P}_z(2\pi-A)\mathbf{P}_x[(\pi/2)-\phi] = \mathbf{P}_x^{\mathbf{T}}(\delta-\pi/2)\mathbf{P}_z^{\mathbf{T}}(h) \; , \qquad (2.6.10)$$

where only the last row of the matrices is considered. These elements yield

$$\left.\begin{aligned}
&\sin h \cos\delta = -\sin A \cos H \\
&\cos h \cos\delta = -\cos A \cos H \sin\phi + \sin H \cos\phi \\
&\sin\delta = +\cos A \cos H \cos\phi + \sin H \sin\phi
\end{aligned}\right\} \quad . \qquad (2.6.11)$$

These results differ from those found in some astronomical textbooks as we have defined the azimuth from the north point. So to get the traditional results we would have to replace A by (π-A). Having discussed how we locate objects in the sky in different coordinate frames and how to relate those frames, we will now turn to a brief discussion of how they are to be located in time.

2.7 Time

The independent variable of Newtonian mechanics is time and thus far we have said little about it. Newton viewed time as absolute and 'flowing' uniformly throughout all space. This intuitively reasonable view was shown to be incorrect in 1905 by Albert Einstein in the development of what has become known as the Special Theory of Relativity. However, the problems introduced by special relativity are generally small for objects moving in the solar system. What does complicate the concept of time is the less sophisticated notion of how it is measured. As with other tangled definitions of science, historical developments have served to complicate immensely the definition of what ought to be a simple concept. We will choose to call the units of time seconds, minutes, hours, days, years, and centuries (there are others, but we will ignore them for this book). The relationships between these units are not simple and have been dictated by history.

In some very broad sense, time can be defined in terms of an interval between two events. The difficulty arises when one tries to decide what events should be chosen for all to use. In other words, what "clock" shall we use to define time? Clocks run in response to physical forces so we are stuck with an engineering problem of finding the most accurate clock. Currently, the most accurate clocks are those that measure the interval between atomic processes and have an accuracy of the order of 1 part in 10^{11} to 1 part in 10^{15}. Clocks such as these form the basis for measuring time and time kept by them is known as international atomic time (TAl for short). However, the world for centuries has kept time by clocks that mimic the rising and setting of the sun or rotation of the earth. Certainly prehistoric man realized that all days were not of equal length and therefore could not serve to define a unit of time. However, the interval between two successive transits (crossings of the local meridian) of the sun is a more nearly constant interval. If the orbit of the earth were perfectly circular, then the motion of the sun along the ecliptic would be uniform in time. Therefore, it could not also be uniform along the equator. This non-uniformity of motion along the equator will lead to differences in successive transits of the sun. To make matters worse, the orbit of the earth is elliptical so that the motion along the ecliptic is not even uniform. One could correct for this or choose to keep time by the stars.

Time that is tied to the apparent motion of the stars is called sidereal time and the local sidereal time is of importance to astronomers as it defines the location of the origin of the Right Ascension-Declination coordinate frame as seen by a local observer. It therefore determines where things are in the sky. Local sidereal time is basically defined as the hour angle of the vernal equinox as seen by the observer.

However, as our ability to measure intervals of time became more precise, it became clear that the earth did not rotate at a constant rate. While a spinning object would seem to provide a perfect clock as it appears to be independent of all of the forces of nature, other objects acting through those forces cause irregularities in the spin rate. In fact, the earth makes a lousy clock. Not only does the rotation rate vary, but the location of the intersection of the north polar axis with the surface of the earth changes by small amounts during the year. In addition long term precession resulting from torques generated by the sun and moon acting on the equatorial bulge of the earth, cause the polar axis, and hence the vernal equinox, to change its location among the stars. This, in turn, will influence the interval of time between successive transits of any given star. Time scales based on the rotation of the earth do not correspond to the uniformly running time envisioned by Newton. Thus we have need for another type of time, a dynamical time suitable for expressing the solution to the Newtonian equations of motion for objects in the solar system. Such time is called terrestrial dynamical time (TDT) and is an extension of what was once known as ephemeris time (ET), abandoned in 1984. Since it is to be the smoothly flowing time of Newton, it can be related directly to atomic time (TAl) with an additive constant to bring about agreement with the historical ephemeris time of 1984. Thus we have

$$\mathrm{TDT} = \mathrm{TAI} + 32.184 \ \mathrm{seconds} \quad . \tag{2.7.1}$$

Unfortunately, we and the atomic clocks are located on a moving body with a gravitational field and both these properties will affect the rate at which clocks run compared with similar clocks located in an inertial frame free of the influence of gravity and accelerative motion. Thus to define a time that is appropriate for the navigation of spacecraft in the solar system, we must correct for the effects of special and general relativity and find an inertial coordinate frame in which to keep track of the time. The origin of such a system can be taken to be the barycenter (center of mass) of the solar system and we can define barycentric dynamical Time (TDB) to be that time. The relativistic terms are small indeed so that the difference between TDT and TDB is less than .002 sec. A specific formula for calculating it is given in the Astronomical Almanac[3] .Essentially, terrestrial dynamical time is the time used to calculate the motion of objects in the solar system. However, it is only approximately correct for observers who would locate objects in the sky. For this we need another time scale that accounts for the irregular rotation of the earth.

Historically, such time was known as Greenwich Mean Time, but this term has been supplanted by the more grandiose sounding universal time (UT). The fundamental form of universal time (UT1) is used to determine the civil time standards and is determined from the transits of stars. Thus it is related to Greenwich mean sidereal time and contains non-uniformities due to variations in the rotation rate of the earth. This is what is needed to find an object in the sky. Differences between universal time and terrestrial dynamical time are given in the Astronomical Almanac and currently (1988) amount to nearly one full minute as the earth is "running slow". Of course determination of the difference between the dynamical time of theory and the observed time dictated by the rotation of the earth must be made after the fact, but the past behavior is used to estimate the present.

Finally, the time that serves as the world time standard and is broadcast by WWV and other radio stations is called coordinated universal time (UTC) and is arranged so that

$$|\mathrm{UT1} - \mathrm{UTC}| < 1\ \mathrm{second} \quad . \tag{2.7.2}$$

Coordinated universal time flows at essentially the rate of atomic time (give or take the relativistic corrections), but is adjusted by an integral number of seconds so that it remains close to UT1. This adjustment could take place as often as twice a year (on December 31 and June 30) and results in a systematic difference between UTC and TAl. This difference amounted to 10 sec. in 1972. From then to the present (1988), corrections amounting to an additional 14 sec. have had to be made to maintain approximate agreement between the heavens and the earth.

Coordinated universal time is close enough to UTl to locate objects in the sky and its conversion to local sidereal time in place of UTl can be effectively arrived at by scaling by the ratio of the sidereal to solar day. Due to a recent adoption by the International Astronomical Union that terrestrial Longitude will be defined as increasing positively to the east, local mean solar time will just be

$$(\mathrm{LMST}) = \mathrm{UTC} + \lambda \quad , \tag{2.7.3}$$

where A is the longitude of the observer. The same will hold true for sidereal time so that

$$(\mathrm{LST}) = (\mathrm{GST}) + \lambda \qquad , \tag{2.7.4}$$

where the Greenwich sidereal time (GST) can be obtained from UTl and the date. The local sidereal time is just the local hour angle of the vernal equinox by definition so that the hour angle of an object is

$$\mathrm{h} = (\mathrm{LST}) - (\mathrm{R.A.}) \qquad . \tag{2.7.5}$$

Here we have taken hour angles measured west of the prime meridian as increasing.

Given the appropriate time scale, we can measure the motion of objects in the solar system (TDT) and find their location in the sky (UTl and LST). There are numerous additional small corrections including the barycentric motion of the earth (its motion about the center of mass of the earth-moon system) and so forth. For those interested in time to better than a millisecond all of these corrections are important and constitute a study in and of themselves. For the simple acquisition of celestial objects in a telescope, knowledge of the local sidereal time as determined from UTC will generally suffice.

Chapter 2: Exercises

1. Transform from the Right Ascension-Declination coordinate system (α,δ) to Ecliptic coordinates (λ , β) by a rotation matrix. Show all angles required and give the transformation explicitly.

2. Given two n-dimensional coordinate systems X_i and X_j and an orthonormal transformation between the two $\mathbf{A}_{ij}$, prove that exactly n(n-1)/2 terms are required to completely specify the transformation.

3. Find the transformation matrix appropriate for a transformation from the Right Ascension-Declination coordinate system to the Alt-Azimuth coordinate system.

4. Consider a space shuttle experiment in which the pilot is required to orient the spacecraft so that its major axis is pointing at a particular point in the sky (i.e. α,δ). Unfortunately his yaw thrusters have failed and he can only roll and pitch the spacecraft. Given that the spacecraft has an initial orientation (α_1, $\delta_{1)}$, and (α_2, δ_2) of the long and short axes of the spacecraft) and must first roll and then pitch to achieve the desired orientation, find the roll and pitch angles (ξ,η) that the pilot must move the craft through in order to carry out the maneuver.