Lecture Notes on Astronomy

PH2900 – 2007/2008

Glen D. Cowan
Physics Department

Last revised: April 27, 2010

Preface

The main objectives of this course will be to understand how astronomical observations are made and interpreted, to introduce our nearest celestial neighbours in the solar system, and to understand current theories on how these bodies formed and what processes shaped their surfaces.

For 2007/2008 these course notes will still be incomplete. Some supplemental information can be found on the course web page,

`http://www.pp.rhul.ac.uk/~cowan/astro_course.html`

For further information, the books best suited for this course are *Introductory Astronomy and Astrophysics* by Zeilik and Gregory [1] and *Fundamental Astronomy* by Karttunen *et al.* [2]. The book by Carroll and Ostlie [3] is also good but covers a much broader scope than needed for our course. *Observational Astrophysics* by Smith [4] has a good overview of observational techniques.

It would be appreciated if corrections, suggestions and comments on these notes could be communicated to `g.cowan@rhul.ac.uk`.

GDC
October, 2007

Contents

Chapter 1

Introduction and overview

These lecture notes are still in a state of flux. For the first three quarters of the course we will talk about methods of observational astronomy, and for the remaining quarter we will discuss the solar system and have a brief introduction to radio astronomy. Below is an outline of the topics we will cover roughly by week.

1. Introduction and overview, coordinate systems, time keeping.
2. Optical telescopes and detectors.
3. The atmosphere and its effects on observations.
4. Brightness measurement (photometry).
5. Intensity, equation of radiative transfer, solar limb darkening.
6. Colour and spectroscopy.
7. Measuring distance.
8. Extrasolar planets.
9. Introduction to the solar system; origin of the solar system.
10. Planetary atmospheres, tidal forces, planetary geology.
11. Radio astronomy.

Chapter 2

Coordinates and time

Among the most important questions addressed in astronomy is "Where is" a certain planet, star or galaxy. From where we sit on Earth this boils down to two separate questions: "In what direction?" and "How far away?". In this chapter we will look at the issue of direction, which involves determining where and when to point your telescope. In Section 2.1 we will see two important ways of specifying the direction of a star, the horizon (or altitude-azimuth) system and the equatorial system, and we will derive how to relate one to the other. In Section 2.2 we examine time keeping systems so that we can determine when a celestial body will appear at a certain place in the sky.

2.1 Coordinate systems

2.1.1 The horizon system

For an observer on the surface of the earth, the simplest coordinate system for specifying a direction in space is called the *horizon system*. First define the *zenith* as the direction pointing straight up. The opposite direction, called *nadir*, points to the centre of the earth. The plane perpendicular to the zenith defines the *observer's horizon* (or *local horizon)*. Then to specify a direction in space you carry out the steps indicated in Fig. 2.1, namely,

1. face north;
2. turn to the right (i.e., rotate about the vertical axis to the east) an angle A called the *azimuth*;
3. look up from the horizon by an angle a, called the *altitude*. Equivalently we can give the angle from the vertical, called the *zenith distance*, $z = 90^\circ - a$.

Conventions on the definition of azimuth can vary. Some texts measure azimuth starting from the south, others start in the south only if you are in the southern hemisphere. We will always measure azimuth starting from north increasing to the east.

The horizon system is simple to understand and use at a given place and time, but we run into difficulties if we want to tell someone at a different place and time where to see a certain

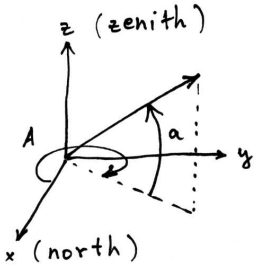


Figure 2.1: The definitions of the altitude a and azimuth A in the horizon (or local) coordinate system.

star. First, the altitude and azimuth of a star will change with time because of the rotation of the earth. Furthermore, the coordinates a and A will depend on the observer's latitude and longitude. For example, the star Polaris will be almost directly overhead for an observer at the north pole, but will be near the horizon for an observer at the equator. So the horizon system coordinates are not numbers that one could use in a star catalogue.

2.1.2 The equatorial system

The view of the ancient astronomers was that the stars were points of light fixed to a distant sphere, which itself rotates about the observer. We can retain this picture to define a the *equatorial system* in which distant stars will have fixed coordinates.

We begin by considering a large *celestial sphere*, centred about the observer and with a radius effectively at infinity (in any case much larger than the radius of the earth), as shown in Fig. 2.2. The points where the axis of the earth's rotation intersects the celestial sphere define the north and south celestial poles, denoted by NCP and SCP. The plane perpendicular to this running through the centre of the earth defines the celestial equator. It is simply the intersection of the earth's equatorial plane with the celestial sphere. In fact we should define this plane to contain the observer, but since we will be using these coordinates to locate distant celestial bodies, we can take the origin equally well to be the centre of the earth.

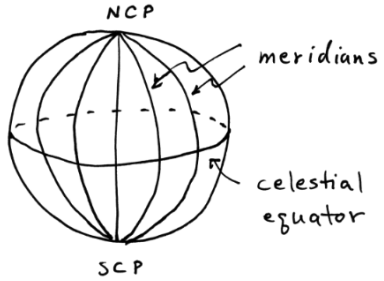


Figure 2.2: The celestial sphere, with its north and south poles and equator.

A circle on a sphere defined by a plane passing through the centre of the sphere is called a *great circle*, and great circles which contain the north and south celestial poles are called *meridians* (also called *hour circles*), as shown in Fig. 2.2. The meridian which passes through the observer's zenith is called the *observer's* (or *local*) meridian. For an observer in Greenwich this is called the *prime meridian*. The meridians correspond to lines of longitude on the earth projected onto the celestial sphere.

In a similar way we we can define for any point on the celestial sphere its angle from the celestial equator, that is, 90° minus the angle it makes with the north celestial pole. This is called the *declination* δ. The NCP has $\delta = 90°$ and SCP has $\delta = -90°$. Declination corresponds to latitude on the surface of the earth, as shown in Fig. 2.3.

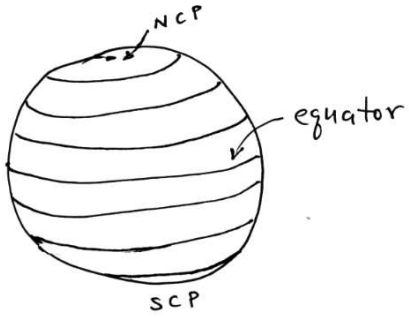


Figure 2.3: Lines of constant declination on the celestial sphere.

Now if we face north and look up at the north celestial pole, we can imagine meridians emanating outwards and circles of constant declination surrounding the NCP, as shown in Fig. 2.4. Remember that now we are looking at the *inside* of the celestial sphere from its centre. Notice that the angle that you look up from the horizon to see the NCP is equal to your local latitude. So as drawn, the figure corresponds to a latitude of around 27°, since the line $\delta = 60°$ disappears below the horizon.

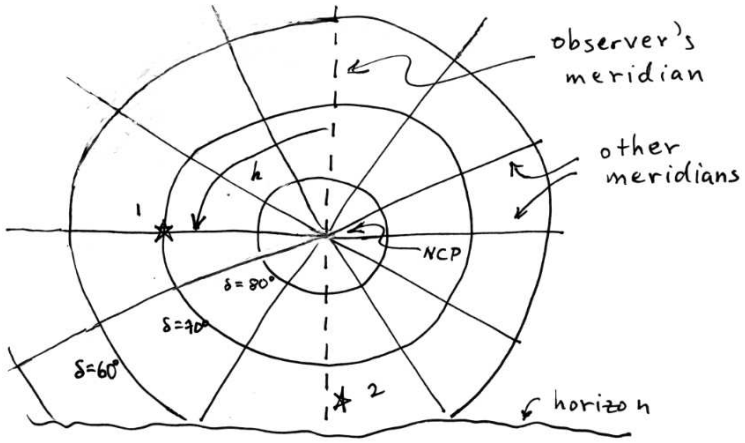


Figure 2.4: The north celestial pole (NCP) shown with meridians and lines of constant declination. The observer's meridian is indicated with a dashed line.

To specify a point at a given time on the celestial sphere, we need the declination δ and we also need to specify the meridian. If we are rotating with the earth and look near the NCP, the stars will appear to follow circles of constant declination rotating anticlockwise as seen from the inside of the celestial sphere. Stars will cross the observer's meridian twice per day, once above the pole, called *upper transit*, and once below it, called *lower transit*. If the lower transit is above the horizon, the star is said to be *circumpolar*, i.e., it never sets.

We can define the *hour angle* h of a star as the angle between the upper part of the observer's meridian to the meridian of the star. Hour angle can be measured in degrees from 0 to 360, but it is more common to measure it in hours, minutes and seconds with 24 hours = 360°, 1 hour = 15°, 1° = 4 minutes, etc. Note that these minutes and seconds are not the same as arc-minutes ($1' = 1/60$ of a degree) or arc-seconds ($1'' = 1/60$ of an arc-minute). In Fig. 2.4, star number 1 has an hour angle $h = 6$h and a declination $\delta = 70°$. Star number two could have $h = 12$h 25m 32s and $\delta = 64°15'32''$.

As the earth rotates, the declination of a star will stay essentially constant, but the hour angle will obviously change. The hour angle basically measures the time since upper transit, although its hours minutes and seconds are *sidereal time*. This differs slightly from the time on your watch for reasons we will see in Section 2.2. So hour angle is still not well suited for something like a star catalogue, since it changes with time.

We can define, however, a set of meridians that rotate with the celestial sphere. The angle of a star's meridian relative to this set of meridians is called the star's *right ascension* α. It is measured starting from some reference meridian that defines $\alpha = 0$ and increasing to the east, i.e., in the opposite direction to that of hour angle, as shown in Fig. 2.5.[1] Like hour angle, right ascension is often measured in hours, minutes and seconds, with 24 h = 360°, etc.

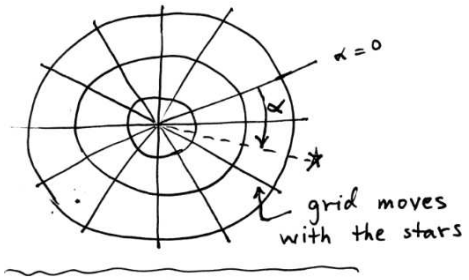


Figure 2.5: Meridians that rotate with the celestial sphere define lines of constant right ascension.

The choice for $\alpha = 0$ is a matter of convention, and you might think that one would simply pick a bright distant star for this purpose. For historical reasons this is not what is done. Rather, one uses the fact that the earth's equator (and hence the celestial equator) is tilted at an angle $\varepsilon \approx 23.5°$ relative to the plane of the *ecliptic*, that is, the plane of the earth's orbit around the sun. To a good approximation the earth's axis stays pointing in the same direction as the earth goes around the sun, so we can use the intersection of the equatorial and ecliptic planes to define a line pointing in a fixed direction. The line connecting the earth and sun coincides with this line twice per year, at the vernal and autumnal equinoxes, that is, around 21 March and 21 September. The direction defining $\alpha = 0$ is taken to be a ray drawn from the earth to the sun at the vernal equinox, also called the *first point of Aries*, as shown in Fig. 2.6. The term 'vernal equinox' is used to refer both to the direction of the first point of Aries as well as the time at which the sun is found in this direction as viewed from the earth.

We can now specify the right ascension and declination of a star, and these values can be

[1] Notice that 'east' on the celestial sphere is meant in the same sense as east on a globe as seen from the outside, although we are viewing the celestial sphere from the inside. Similarly, when saying where to point your telescope on the celestial sphere, 'go further north' means go towards NCP. These are not the same as east or north in the horizon system, which there refer to directions perpendicular to the zenith.

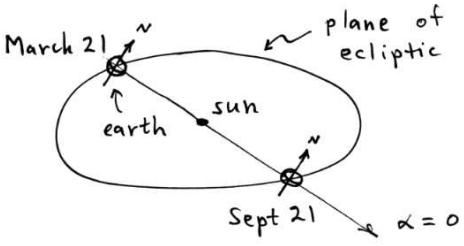


Figure 2.6: Illustration of the vernal equinox, used to define the zero of right ascension.

used by astronomers at different times and different points on the earth. As the earth rotates, the values of α and δ remain almost unchanged. This is 'almost', not 'exactly', for reasons we explore in the next section.

2.1.3 Precession of the equinoxes

Despite what the name suggests, the first point of Aries does not point towards the constellation of Aries, nor is it even a fixed direction in space. This is because the direction of the earth's axis, and hence the orientation of the celestial equator, precesses slowly as the earth goes around the sun. The angle of the earth's axis remains at 23.5° degrees from the ecliptic, but its direction changes slowly making a complete cycle every 25,800 years. This results from the gravitational interaction between the moon, and to a lesser extent the sun, with the equatorial bulge of the earth. The pull of the moon tries to 'straighten out' the tilt of the earth, and as a result the earth's axis precesses, as does a spinning top when it is pulled on by the earth's gravity.

The vernal equinox moves backwards along the ecliptic at a rate of about 50″ per year or one degree every 72 years. Currently the first point of Aries is in the constellation of Pisces. More than 2000 years ago it was in Aries, a period known as the 'age of Aries', and in about 600 years we will enter the age of Aquarius. The precession of the equinoxes obviously also affects the direction of the north celestial pole. Although this now points very close to the star Polaris (0.8° separation), it was three to four degrees away when ships began to sail across the Atlantic. In 2600 BC the north celestial pole was very close to the star Thuban. As Thuban barely moved across the sky and was therefore in some sense undying, it was of special significance to the ancient Egyptians.

So a star catalogue containing right ascension and declination must specify the time at which the coordinates are intended to apply, which is called the *epoch*. Conventionally star catalogues refer to epochs defined every 50 years, e.g., 1950, 2000, etc. In this way the user can if needed make corrections for the precession of the earth's axis to give the right ascension and declination for some other time.

There are further subtle effects that must be taken into account if high accuracy of the coordinates is required. An example is *nutation*, the small wobbling of the earth's axis caused by the precession of the moon's axis relative to its orbital plane. This precession has a period of 18.6 years, and it produces small perturbations in the direction of the earth's axis with the same period. Usually for purposes of this course we will not be concerned with nutation or other similarly small complications.

2.1.4 Relating the horizon and equatorial coordinate systems

If we correct for (or ignore) complications such as precession and nutation, values of right ascension and declination for distant stars will remain essentially constant and therefore we can list them in catalogues. For example, we can find out that the star Altair has α =19h 51m and $\delta = 8°52'$. Suppose we go out tonight to see Altair. Where do we point the telescope, i.e., at what altitude and azimuth? And when do we look? Or we may simply want to know at what time Altair will be at its highest point in the sky, i.e., at upper transit, and what its altitude will be at this point.

First we need to relate the right ascension α to the hour angle h. Recall that the hour angle essentially measures the time that has elapsed since the upper transit of the star. This angle will come back to the same value after the earth has rotated 360° about its axis, which as we mentioned is a bit less than one 'solar day'. Nevertheless we *define* 24 hours of sidereal time to be the time it takes for the earth to rotate 360° relative to the vernal equinox (which is almost equal to a 360° rotation relative to a distant star). Then we make the convention that the *local sidereal time* LST is equal to zero when the vernal equinox ($\alpha = 0$) crosses the observer's meridian.

Consider, for example, the star shown in Fig. 2.7. Its hour angle h is measured from the observer's meridian increasing to the west, and its right ascension α is measured from the vernal equinox starting from the first point of Aires increasing to the east.

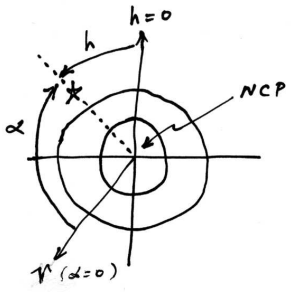


Figure 2.7: Illustration of the relation between LST, α and h (see text).

The *local siderial time* (LST) is defined as the total time since the first point of Aires passed the local meridian, i.e.,

$$\mathrm{LST} = h + \alpha \; . \tag{2.1}$$

The relation between local sidereal time and the time on your watch is not trivial to work out, but is something you can easily obtain from an almanac or from calculators available on the internet (see, e.g., [6]). For the moment let's assume this has been done, so that we only need to relate δ and h to the altitude a and azimuth A. We will return to LST and its relation to other times in Section 2.2.

The mathematics needed to relate (δ, h) to (a, A) is a bit involved but it provides a nice example of how coordinates transform under a rotation, and so we will go through this in some detail. First, consider a 'standard' spherical coordinate system, as shown in Fig. 2.8.

Suppose we consider a point P at a radius $r = 1$. We can specify where it is by giving its Cartesian coordinates x, y and z or by giving the angles θ and ψ.[2] By considering the

[2]Normally we would write ϕ instead of ψ, but we are already using ϕ to denote latitude.

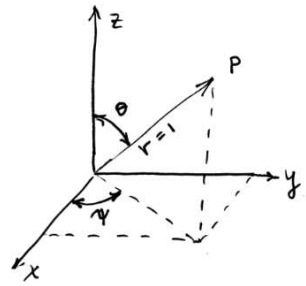


Figure 2.8: A spherical coordinate system. The point P is at unit radius from the origin.

triangles indicated in Fig. 2.8, we can write down the following relations between the Cartesian and spherical coordinates:

$$x = \sin\theta\cos\psi\,, \tag{2.2-a}$$

$$y = \sin\theta\sin\psi\,, \tag{2.2-b}$$

$$z = \cos\theta\,. \tag{2.2-c}$$

We will use the coordinate system in Fig. 2.8 to represent the equatorial system with the z axis pointing towards NCP and with the zenith direction lying in the xz-plane. We can then define another system with coordinates x', y' and z' to represent the horizon system. For it, the z' axis points up (zenith) and the x' axis points north. In this system, NCP lies in the $x'z'$-plane.

The two coordinate systems are related by a rotation about the y axis. The primed system can be obtained from the unprimed one by rotating by an angle γ about the y axis such that the z axis moves from NCP to the zenith. This angle is $\gamma = 90^\circ - \phi$ where ϕ is the local latitude. The new set of axes is shown in Fig. 2.9. As before, the z axis points towards NCP and the z' axis towards the zenith.

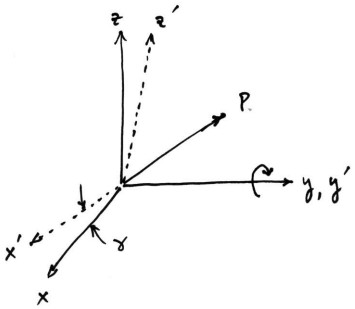


Figure 2.9: The primed coordinate system is defined by a rotation by an angle γ about the y axis.

In the primed frame we can define angles θ' and ψ' as above, and they will have the same relations to x', y' and z' as do the corresponding variables in the unprimed frame as given by equation (2.2-a)–(2.2-c), i.e.,

$$x' = \sin\theta' \cos\psi' \,, \tag{2.3-a}$$

$$y' = \sin\theta' \sin\psi' \,, \tag{2.3-b}$$

$$z' = \cos\theta' \,. \tag{2.3-c}$$

Our first goal will be to relate the angles θ and ψ to θ' and ψ'. The way we do this is to relate (x, y, z) to (x', y', z') and then to use equations (2.2-a)–(2.2-c) and (2.3-a)–(2.3-c).

If we rotate about the y axis, then the value of the y coordinate of the point P does not change, so we have $y' = y$. The values of x and z will be different in the primed frame, as seen in Fig. 2.10, which shows the projection of the point onto the xz-plane.

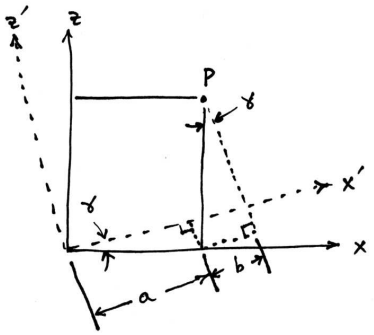


Figure 2.10: The relation between the coordinates (x, z) and (x', z') under a rotation by an angle γ.

By looking at the triangles in Fig. 2.10, we can write $x' = a + b$, where $a/x = \cos\gamma$ and $b/z = \sin\gamma$. We can construct a similar relation for z', and thus we find for all three coordinates

$$x' = x\cos\gamma + z\sin\gamma \,, \tag{2.4-a}$$

$$y' = y \,, \tag{2.4-b}$$

$$z' = -x\sin\gamma + z\cos\gamma \,. \tag{2.4-c}$$

This type of transformation comes up frequently in physics, and one generally finds a relation where the rotated coordinates are linear combinations of the original ones, with the coefficients given by sines and cosines of the angle of rotation. It's usually easy to remember the formula if you consider, say, a very small rotation, so that $\sin\gamma \approx \gamma$ and $\cos\gamma \approx 1$. Then we know that $x' \approx x$ and thus in equation (2.4-a) the cosine term should multiply x. Furthermore we can see that the rotation in the sense indicated makes z' smaller, so the sine term in the equation for z' has a minus sign.

Now we substitute the equations (2.2-a)–(2.2-c) and (2.3-a)–(2.3-c) into (2.4-a)–(2.4-c). This gives the following relations between the angles (θ', ψ') and (θ, ψ):

$$\sin\theta'\cos\psi' = \sin\theta\cos\psi\cos\gamma + \cos\theta\sin\gamma\,, \tag{2.5-a}$$

$$\sin\theta'\sin\psi' = \sin\theta\sin\psi\,, \tag{2.5-b}$$

$$\cos\theta' = -\sin\theta\cos\psi\sin\gamma + \cos\theta\cos\gamma\,. \tag{2.5-c}$$

The angles θ and ψ were defined relative to the (x, y, z) axes in a way consistent with the usual definition of a spherical coordinate system. For historical reasons the angles we use in the horizon and equatorial systems have somewhat different definitions, but they can be related easily to θ, ψ, θ' and ψ'. From Fig. 2.11 we find

$$\theta = 90^\circ - \delta\,, \tag{2.6-a}$$

$$\psi = 180^\circ - h\,, \tag{2.6-b}$$

$$\theta' = 90^\circ - a\,, \tag{2.6-c}$$

$$\psi' = 360^\circ - A\,, \tag{2.6-d}$$

$$\gamma = 90^\circ - \phi\,. \tag{2.6-e}$$

(a) (b)

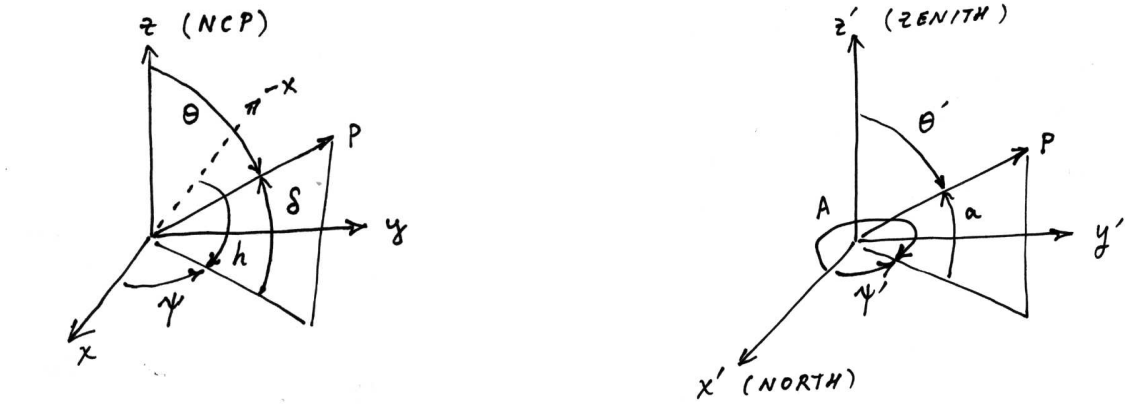


Figure 2.11: (a) The relation between the equatorial system and the unprimed coordinate system. (b) The relation between the horizon system and the primed system.

Note in particular that the zero of hour angle corresponds to the negative x axis, since we have defined it to begin at the upper rather than lower transit.

We can now convert the equations (2.5-a)–(2.5-c) to relate (a, A) to (δ, h). To do this we need the trigonometric relations valid for any angle ω,

$$\begin{aligned}
\sin(90^\circ - \omega) &= \cos\omega\,, && \text{(2.7-a)}\\
\cos(90^\circ - \omega) &= \sin\omega\,, && \text{(2.7-b)}\\
\sin(180^\circ - \omega) &= \sin\omega\,, && \text{(2.7-c)}\\
\cos(180^\circ - \omega) &= -\cos\omega\,, && \text{(2.7-d)}\\
\sin(360^\circ - \omega) &= -\sin\omega\,, && \text{(2.7-e)}\\
\cos(360^\circ - \omega) &= \cos\omega\,. && \text{(2.7-f)}
\end{aligned}$$

Using these together with equations (2.5-a)–(2.5-c) and (2.6-a)–(2.6-c) we find

$$\begin{aligned}
\cos a\cos A &= -\cos\delta\cos h\sin\phi + \sin\delta\cos\phi\,, && \text{(2.8-a)}\\
\cos a\sin A &= -\cos\delta\sin h\,, && \text{(2.8-b)}\\
\sin a &= \cos\delta\cos h\cos\phi + \sin\delta\sin\phi\,. && \text{(2.8-c)}
\end{aligned}$$

Equations (2.8-a)–(2.8-c) are useful for obtaining (a, A) if we know (δ, h). To go the other way, note that we could have called the horizon system the original (unprimed) system. Then we would have rotated by an angle $-\gamma$ to obtain the horizon system. Doing this we would find

$$\begin{aligned}
\cos\delta\cos h &= -\cos a\cos A\sin\phi + \sin a\cos\phi\,, && \text{(2.9-a)}\\
\cos\delta\sin h &= -\cos a\sin A\,, && \text{(2.9-b)}\\
\sin\delta &= \cos a\cos A\cos\phi + \sin a\sin\phi\,. && \text{(2.9-c)}
\end{aligned}$$

If we are given the right ascension and declination, for example, we can use equation (2.8-c) to determine how high in the sky a star is, i.e., we can find the altitude a. Once we have this we can use (2.8-a) and (2.8-b) to find the azimuth A. All three equations are needed to fix which quadrants the angles are in.

We can now work out a simple example of transformations between horizon and equatorial coordinates. We will postpone a more complicated example until after we have discussed the relation between hour angle and sidereal time. Suppose we are told that a supernova has been found at $\delta = 62^\circ$ and $\alpha = 21\text{h}\ 14\text{m}$. We are at a latitude of $\phi = 51^\circ$. What will the be the maximum altitude at which we can observe the supernova?

The maximum altitude will take place at upper transit, and by definition this has an hour angle $h = 0$. So equation (2.8-c) gives

$$\begin{aligned}
a &= \sin^{-1}\left(\cos 62^\circ\cos 51^\circ\cos 0 + \sin 62^\circ\sin 51^\circ\right)\\
&= 79^\circ\,. && \text{(2.10)}
\end{aligned}$$

For this example there is a shortcut which is illustrated in Fig. 2.12. If we face north, we need to look up an angle equal to our latitude ϕ to see the north celestial pole. From there we need to look up further an angle $90^\circ - \delta$ to get to a star at upper transit. So we could have simply concluded

$$a = \phi + (90^\circ - \delta) = 51^\circ + (90^\circ - 62^\circ) = 79^\circ \, . \tag{2.11}$$

In either case we find that it's almost directly overhead, so the view should be good.

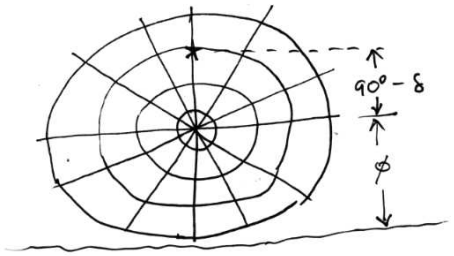


Figure 2.12: A star at upper transit has an altitude of $a = \phi + (90^\circ - \delta)$.

2.2 Timekeeping systems

In addition to knowing where we must also know when to look. We saw in Section 2.1 that this means knowing how long has elapsed since a star's upper transit, but the system of time used to compute this most easily is not the time from an ordinary clock. In Section 2.2.1 we will see the basic difference between *solar* and *sidereal time*. We will discuss the relation of these to *mean time* in Section 2.2.2.

2.2.1 Solar and sidereal time

Historically the basic idea behind the 24-hour clock was to take the time from one noon to the next, i.e., successive upper transits of the sun, and to define this as 24 hours. This is called *solar time*. It is now defined in terms of precise atomic processes, but the basic idea is the same: 24 solar hours gives one average noon-to-noon period.

The celestial sphere appears to rotate once per day, so you might think that a star would return to the same hour angle after 24 hours of solar time. This is of course not quite true, because during those 24 hours the earth has moved about 1° in its orbit around the sun. As a result, the earth's rotation about its axis to go from one noon to the next is slightly more than 360°, as illustrated in Fig. 2.13.

What we need is the time needed for one complete (i.e., 360°) rotation of the celestial sphere. This is called one *sidereal* day, or 24 sidereal hours. In the course of one year we have 365.25 'noon-to-noon' periods, but we get one more rotation by virtue of the earth's revolution about the sun. So we have

$$365.25 \text{ solar days} = 366.25 \text{ sidereal days} \, , \tag{2.12}$$

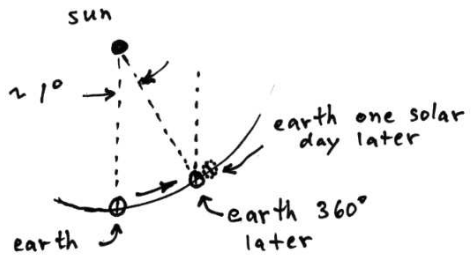


Figure 2.13: The motion of the earth about the sun in the course of one day (scale greatly exaggerated).

or

$$1 \text{ sidereal day} = 0.99727 \text{ solar days} \approx 23 \text{ hours } 56 \text{ minutes} . \tag{2.13}$$

We could have determined the same number by measuring the time it takes between successive upper transits of a distant star.

One of several complications to this picture is that we would like a star with a given right ascension to return to the same hour angle after 24 sidereal hours. But the zero point of right ascension is not defined using the distant stars, but rather with the vernal equinox, which precesses with a period of 25,800 years. So in fact 24 sidereal hours are defined to be the time between successive upper transits of the vernal equinox. The difference is small and for our purposes it will be sufficient to consider sidereal time to be the same as if we had defined it using the distant stars.

The preceding discussion sets the length of a sidereal day, but we also need to fix an arbitrary starting point. As mentioned above, we take $t = 0$ to be the time when the vernal equinox crosses the local meridian. A star's right ascension α and hour angle h are thus related to the local sidereal time by

$$h = \text{LST} - \alpha . \tag{2.14}$$

The LST clearly depends on the observer's longitude (hence 'local'). In order to have a convenient measure of sidereal time one defines the LST for an observer on the prime meridian as *Greenwich Sidereal Time* or GST. If we measure longitude λ in hours, minutes and seconds such that $360^\circ = 24$ hours, etc., with longitude increasing to the east,[3] then we have

$$\text{LST} = \text{GST} + \lambda . \tag{2.15}$$

2.2.2 The mean sun and the equation of time

To proceed we need to find how GST is related to the time that can be determined from an 'ordinary' clock. To first approximation, a clock shows solar time, i.e., 24 hours from noon to noon. But the actual time (as defined, say, by an atomic clock) between successive noons is not constant throughout the year, varying plus or minus around 15 minutes from its average. This

[3]The International Astronomical Union has defined longitude as increasing to the east. The opposite convention, however, is also sometimes found.

is because the orbit of the earth is not quite circular, having an eccentricity of $e = 0.017$. We are thus several percent nearer to the sun at the closest point, called the *perihelion*, in early January, than at the furthest point, the *aphelion*, in early July.

According to Kepler's laws, we move faster when we are closer to the sun, so the angle we move around the sun in one day is greater in January than in July, as shown in Fig. 2.14. So if we start at noon and first rotate 360°, the extra bit further that we need to get to the next noon is longer in January than it is in July. So one *apparent solar day*, that is, the actual time from noon to noon, takes longer in January than it does in July.

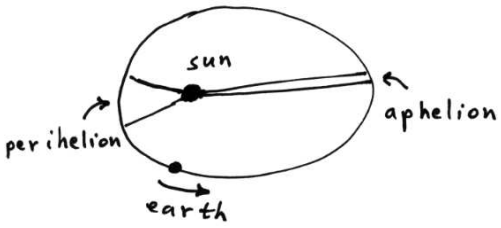


Figure 2.14: The motion of the earth in one day near the perihelion and aphelion (scale greatly exaggerated).

It would be inconvenient to say the least if we needed to adjust the length of one second throughout the year. To get around this we define a point on the celestial sphere called the *mean sun*, which moves at a constant rate around the celestial equator. In the equatorial system the sun also moves around the celestial sphere in the plane of the ecliptic, i.e., in the plane of the earth's orbit around the sun projected onto the celestial sphere. The angular speed of the mean sun is such that it covers one revolution around the celestial equator in the same time it takes for the real sun to move once around the ecliptic. The time between successive passages of the mean sun across our local meridian defines *mean solar time*. For an observer at the prime meridian this is called *universal time* or UT (also called *Greenwich Mean Time* or GMT). Specifically, this is defined using the hour angle of the mean sun as

$$\mathrm{UT} = h_{\text{mean sun}} \pm 12 \text{ solar hours} . \tag{2.16}$$

Here the ± 12 is included so that noon takes place at 12:00 rather than 0:00. If you are in London, then barring complications related to summer time, this is the time on your watch. In other time zones you need to add or subtract the necessary offset, e.g., +3 for Moscow.

The difference ε between mean and apparent solar time is called the *equation of time*. If we measure this in sidereal units then it is simply the difference between the right ascensions of the mean and real sun,

$$\varepsilon = \alpha_{\text{mean sun}} - \alpha_{\text{real sun}} . \tag{2.17}$$

Using equation (2.14) we can write this equivalently as the difference of hour angles

$$\varepsilon = h_{\text{real sun}} - h_{\text{mean sun}} . \tag{2.18}$$

The exact time dependence of ε is complicated further by the fact that the mean sun moves at a constant angular rate around the celestial equator, but the real sun is in the ecliptic. A

1° slice in the equatorial plane is no longer in general 1° when projected onto the ecliptic, and the projected angle depends on whether we are near a solstice or an equinox. So we would have a difference between apparent and mean solar time even if the earth's orbit were circular. The combination of this effect with the elliptical orbit of the earth is the rather complicated equation of time shown in Fig. 2.15.

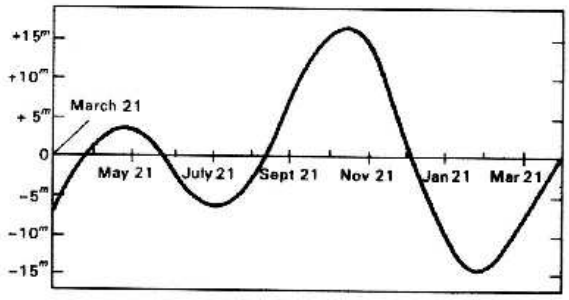


Figure 2.15: The equation of time.

In fact we can largely ignore apparent solar time since our watch gives us UT, or something closely related to it. What we need is the sidereal time. We have seen that they proceed at a different rate, and in general they will be defined to have a different starting point. So we expect a relation of the form

$$\mathrm{GST} = t_0 + \frac{366.25}{365.25}\mathrm{UT}\,. \tag{2.19}$$

We define the starting point of mean solar time to be midnight in Greenwich (i.e,. ± 12 hours from the mean sun's upper transit in Greenwich). The starting point for GST, however, is when the vernal equinox crosses the prime meridian, and this does not have a simple relation to the solar time. So we don't expect the offset t_0 to be a simple number. If we measure GST in degrees (remember 360° is 24 sidereal hours), then we find [7]

$$\mathrm{GST} = 280.46061837^\circ + 360.98564736629^\circ \cdot d\,, \tag{2.20}$$

where

$$d = \text{days (including fractions) since 00:00:00 1 January 2000}\,. \tag{2.21}$$

From the result of (2.20) we need to add or subtract multiples of 360° to bring the value into the range between 0 and 360. Then if needed convert the result to (sidereal) hours, minutes and seconds.

In fact it is much easier to use one of the many calculators available on the internet to compute GST, for example, the U.S. Naval Observatory site at [6]. Obtaining LST is then simply a matter of adding the observer's longitude.[4] These calculators can even take into account effects such as the earth's nutation, which leads to a formula somewhat more complicated than equation (2.20) (see, e.g., [7]).

[4]For Egham, use $\lambda = -0^\circ 34'$.

2.3 Putting it together

Now let's work out a detailed example of how to tell someone on the other side of the world where and when to point their telescope to see a given celestial body. Suppose we want to tell someone in Billings, Montana (latitude $\phi = 45.8^\circ$ N and longitude $\lambda = -108.5^\circ$) where to look to see the globular cluster M15, which is at right ascension $\alpha = 21\mathrm{h}\,30\mathrm{m}$ and declination $\delta = 12^\circ 11'$. At what altitude and azimuth will this be at 9:00 p.m. Montana time on 7 October 2003?

We start by working out the hour angle of M15 at the given date and time. From `www.time.gov` we find that 7 October 9:00 p.m. in Montana is 21:00 + 6:00 = 27:00 = 3:00 UT on 8 October. Then from a sidereal time calculator on the internet we find that that 8 October 2003, 3:00 UT is 4:05 GST. To find the local sidereal time in Billings we use equation (2.15), which tells us we need to add the longitude λ (i.e., subtract the west longitude), so we find

$$\begin{aligned} \mathrm{LST} &= 4\mathrm{h}\,5\mathrm{m} + \left(-108.5^\circ \times \frac{24\mathrm{h}}{360^\circ}\right) \\ &= 4.08\mathrm{h} - 7.23\mathrm{h} = -3.15\mathrm{h} \\ &= 20.85\mathrm{h} = 20\mathrm{h}\,51\mathrm{m}\,, \end{aligned} \tag{2.22}$$

where to get to the last line we added 24 sidereal hours to bring the value into the range between 0 and 24.

Now we use equation (2.14) to obtain the hour angle of M15:

$$h = \mathrm{LST} - \alpha = 20\mathrm{h}\,51\mathrm{m} - 21\mathrm{h}\,30\mathrm{m} = 23\mathrm{h}\,21\mathrm{m}\,, \tag{2.23}$$

where again we added 24 to bring the value into the desired interval.

Now we can use equation (2.8-c) to obtain the altitude a from the declination $\delta = 12^\circ 11'$, the hour angle, $h = 23\mathrm{h}\,21\mathrm{m}$, and the latitude of Billings, $\phi = 45.8^\circ$. Converting all of the angles to decimal degrees we find

$$\begin{aligned} a &= \sin^{-1}\left(\cos\delta\cos\phi\cos h + \sin\delta\sin\phi\right) \\ &= \sin^{-1}\left(\cos 12.2^\circ \cos 45.8^\circ \cos 350.2^\circ + \sin 12.2^\circ \sin 45.8^\circ\right) \\ &= \sin^{-1} 0.821 = 55.2^\circ\,. \end{aligned} \tag{2.24}$$

Then we can use equation (2.8-a) together with the previous result to find

$$\begin{aligned} \sin A &= -\cos\delta\sin h/\cos a \\ &= -\cos 12.2^\circ \sin 350.2^\circ/\cos 55.2^\circ \\ &= 0.290\,, \end{aligned} \tag{2.25}$$

which gives either $A = 16.9^\circ$ or $A = 163.1^\circ$. Now using equation (2.8-b) we find

$$\begin{aligned} \cos A &= (-\cos\delta\cos h\sin\phi + \sin\delta\cos\phi)\,/\cos a \\ &= (-\cos 12.2^\circ\cos 350.2^\circ\sin 45.8^\circ + \sin 12.2^\circ\cos 45.8^\circ)\,/\cos 55.2^\circ \\ &= -0.957\,, \end{aligned} \tag{2.26}$$

so this fixes the azimuth to be $A = 163.1^\circ$. Carrying out these sorts of calculations can be a chore but it's straightforward to write a computer program to do it quickly and easily. A simple java program that takes as input the hour angle and declination and prints out the altitude and azimuth is shown below.

```
/**
 *  Computes altitude and azimuth from latitude, hour angle and declination.
 *  Glen Cowan, RHUL Physics
 *  16 October, 2003
 */

import java.io.*;

public class CoordCalc{
  public static void main (String[] args) throws IOException {

// Get phi, delta and h from console and convert from degrees to radians

    BufferedReader b = new BufferedReader(
                       new InputStreamReader(System.in));

    System.out.println("Enter latitude phi (degrees): ");
    double phi = Double.parseDouble ( b.readLine() ) * Math.PI / 180;

    System.out.println("Enter declination delta (degrees): ");
    double delta = Double.parseDouble ( b.readLine() ) * Math.PI / 180;

    System.out.println("Enter hour angle h (degrees): ");
    double h = Double.parseDouble ( b.readLine() ) * Math.PI / 180;

// Compute altitude

    double alt = Math.asin ( Math.cos(delta) * Math.cos(h) * Math.cos(phi) +
                          Math.sin(delta) * Math.sin(phi) );

// Now get sine and cosine of azimuth and use atan2 to figure out quadrant

    double sinAz =  - Math.cos(delta) * Math.sin(h) / Math.cos(alt);
    double cosAz = ( - Math.cos(delta) * Math.cos(h) * Math.sin(phi)
                     + Math.sin(delta) * Math.cos(phi) ) / Math.cos(alt);
    double Az = Math.atan2(sinAz, cosAz);

// Convert to degrees and output to screen

    System.out.println("altitude = " + alt*180/Math.PI + " degrees");
    System.out.println("Azimuth  = " + Az*180/Math.PI + " degrees");

  }
}
```

Chapter 3

Optical telescopes

Almost all observations in astronomy involve collecting and analyzing electromagnetic radiation. Although some important 'astronomy' has been carried out using charged cosmic rays (electrons, protons, and some heavier nuclei) and neutrinos (e.g., from the sun or from the supernova SN1987A), the photon has been and continues to be the most important carrier of information about the universe around us. In this chapter we will therefore restrict the discussion to detecting electromagnetic radiation.

Photons or electromagnetic waves have an associated wavelength, with that of visible light extending from roughly 400 to 700 nm. This is contained within a somewhat broader range of wavelengths that can penetrate through the Earth's atmosphere, called the *optical window*, from around 300 nm to 1 μm. Much longer wavelengths, from roughly 1 cm to 10 m can also penetrate through the atmosphere, and these can be detected using radio telescopes. Shorter wavelengths corresponding to ultraviolet rays, x-rays and γ-rays are strongly absorbed in the atmosphere and can only be measured by using high-altitude or (better) space-borne devices. The probability for a photon to be absorbed in the Earth's atmosphere (the atmospheric opacity) is shown as a function of wavelength in Fig. 3.1.

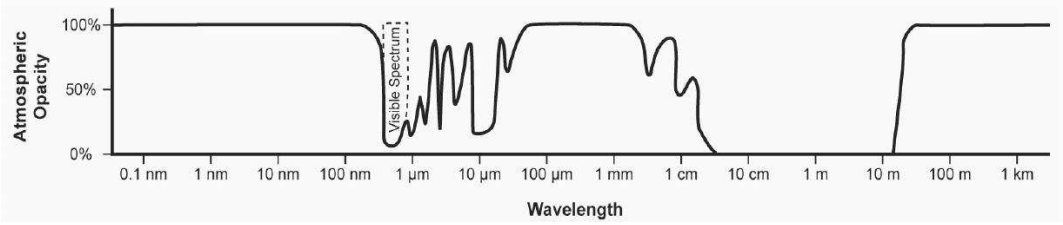


Figure 3.1: Atmospheric opacity as a function of wavelength (from [17]).

Thus many parts of the electromagnetic spectrum tell us about astrophysical objects and in modern astronomy one often tries to combine information from observations at different wavelengths. But historically the optical range was the first to be exploited and it continues to play the most important role. So in this chapter we will cover the basic principles of optical telescopes. We will see what parameters are needed to characterise telescopes and what limits their performance.

3.1 Goals of an optical telescope system

Let us first go over what we want a telescope system to be able to do. Most importantly it should gather light and since the source of the light will often be very dim, it needs to collect as many photons as possible. The quantitative measurement of the light flux, generally within a specified range of wavelengths, is called *photometry*.

Further, we want to know what direction the photons are coming from. There are two related but distinct aspects to this question. We may want to know the direction of a point source of light, such a distant star. That is, we want accurate measurements of its declination and right ascension; this is called *astrometry*. In addition we may want to *resolve* separate sources of light as being indeed separate. For example, we may want to resolve the angular structure of a cluster of stars or perhaps a planetary surface; this is called *imaging*.

In addition we may want to know the colour of the light, or rather, its wavelength spectrum. This branch of measurements is called *spectroscopy*, which we will take up further in Chapter 7. We may also want to measure all of these things as a function of time. Some astrophysical phenomena such as supernova explosions take place on a short time scale but these are rather the exception. In this chapter we will not deal with time measurements in detail.

All of these measurements require not only that the telescope be able to collect the photons but also to detect and record them. First done with the human eye and then photographic plates, this task is now often carried out with solid state detectors called CCDs (Charged Coupled Devices).

3.2 Light gathering power

Telescopes collect light through an aperture which is in general circular, having a diameter D. The *light gathering power* (LGP) of a telescope is a measure of how much light energy is collected per unit time, and this is proportional to the area of the aperture. Therefore we have

$$\text{LGP} \propto \text{ area } \propto D^2 \, . \tag{3.1}$$

The pupil of the dark-adapted human eye has a diameter of around 5 mm (perhaps up to 7 or 8 mm for young eyes). Table 3.1 shows the light gathering power of a number of telescopes relative to a 6 mm eye.

A quick calculation shows why light gather power is such a crucial parameter for the performance of a telescope. Suppose we want to use the Hubble Space Telescope (HST) to observe a Cepheid variable star with a luminosity 10^4 times that of the sun in the galaxy M100 in the Virgo cluster, which is at a distance of 20 Mpc. The sun's luminosity is $L_{\text{sun}} = 3.8 \times 10^{26}$ W. At a distance r from the star, this light is spread over a sphere of surface area $4\pi r^2$. The light flux F (power per unit area) at a distance r from the star is thus

$$F = \frac{L}{4\pi r^2} \, . \tag{3.2}$$

The power P collected by the Hubble telescope ($D = 2.4$ metres) is obtained by integrating this flux over the area of the telescope's aperture. We find

Table 3.1: Light gathering power of some existing and proposed telescopes relative to the dark-adapted human eye ($D = 6$ mm).

telescope	diameter (m)	LGP / LGP(6 mm)
RHUL observatory	0.30	2500
Hubble Space Telescope (HST)	2.4	1.6×10^5
Hale (Mt. Palomar, California)	5	7×10^5
Keck (Mauna Kea, Hawaii)	10	2.8×10^6
CELT (California Extremely Large Telescope)	30	2.5×10^7
OWL (Overwhelmingly Large Telescope)	100	2.8×10^8

$$P = \pi(2.4\text{ m}/2)^2 \times \frac{10^4 \times 3.8 \times 10^{26}\text{W}}{4\pi(20\text{ Mpc})^2} \times \left(\frac{1\text{ Mpc}}{3.1 \times 10^{22}\text{ m}}\right)^2 = 3.6 \times 10^{-18}\text{ W}\,. \tag{3.3}$$

Now let's suppose that the average wavelength of the photons is $\lambda = 500\,\text{nm}$, roughly in the middle of the optical range and also close to the Sun's peak wavelength. The corresponding photon energy is

$$E_\gamma = h\nu = \frac{hc}{\lambda} = \frac{2\pi\hbar c}{\lambda}\,. \tag{3.4}$$

The last form of the equation is more convenient for those who memorize the value $\hbar c = 197.3\,\text{eV-nm}$. For $\lambda = 500\,\text{nm}$ we find $E_\gamma = 2.5\,\text{eV}$. The number of photons per unit time entering the HST is therefore roughly

$$\text{photons/time} = 3.6 \times 10^{-18}\text{ J/s} \times \frac{1\text{ eV}}{1.6 \times 10^{-19}\text{ J}} \times \frac{1\text{ photon}}{2.5\text{ eV}} = 9\,\gamma/\text{s}\,. \tag{3.5}$$

Of course not all photons have the same wavelength but nevertheless this calculation gives us a reasonable estimate. This corresponds to an apparent magnitude of about 26, and Cepheid variable stars at this magnitude have indeed been observed by the HST. The corresponding photon collection rate for a 30 cm telescope would be $0.14\,\gamma/\text{s}$ and for the proposed OWL telescope ($D = 100$ m) one finds almost $16\,000\,\gamma/\text{s}$.

Now the accuracy with which we can determine any property of the object we are observing, e.g., the light flux of the star, will depend on the number n_γ of photons we collect. In general we will find that relative measurement errors (more precisely, the *statistical* or *random* errors) scale in proportion to $1/\sqrt{n_\gamma}$. We will look at this in more detail in Chapter 5 when we discuss the Poisson distribution. The main point is that with a larger diameter we can see fainter objects in a given amount of observing time. The number of photons n_γ and therefore the LGP are proportional to D^2, and for a fixed observing time the statistical accuracy therefore goes as $1/D$. Equivalently, for a desired statistical accuracy (i.e., fixed n_γ), the required observing time goes as $1/D^2$.

3.3 Refracting telescopes

The first telescopes were built in the early 1600s and were based on the refraction of light through glass lenses. They are therefore called *refracting* telescopes or simply refractors. Figure 3.2(a) shows one of the first refractors built by Galileo in 1609. Figure 3.2(b) shows the world's largest refractor: the 1-metre diameter Yerkes telescope operated by the University of Chicago.

(a) (b)

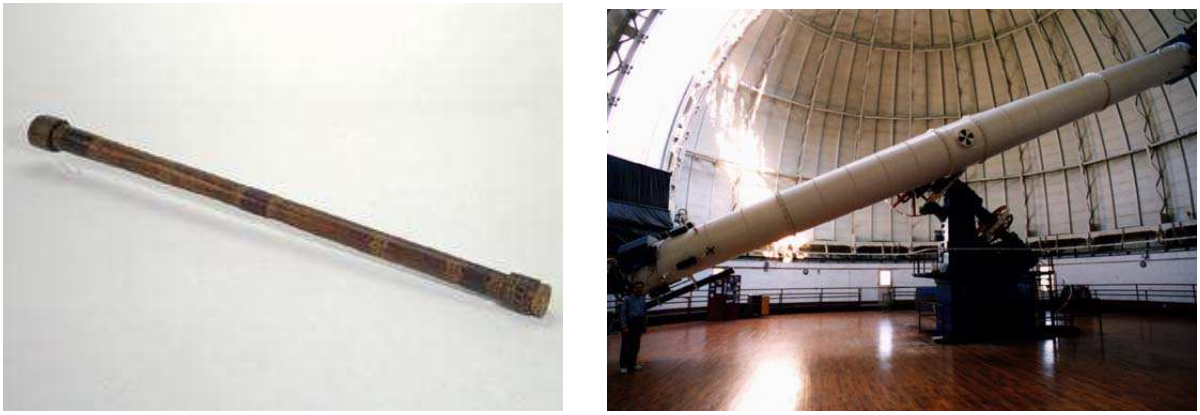


Figure 3.2: (a) Refracting telescope built by Galileo in 1609 [8]. (b) The Yerkes 1-metre refractor [9].

The basic idea of a refracting telescope is to collect light through an aperture and to bring it to a focus using a lens or system of lenses. Here we will review some of the basic formulae connected with lenses. For more details see any good book on optics, e.g., [10]. The distance between the aperture and the focus is called the *focal length*, F, as shown in Fig. 3.3.

(a) (b)

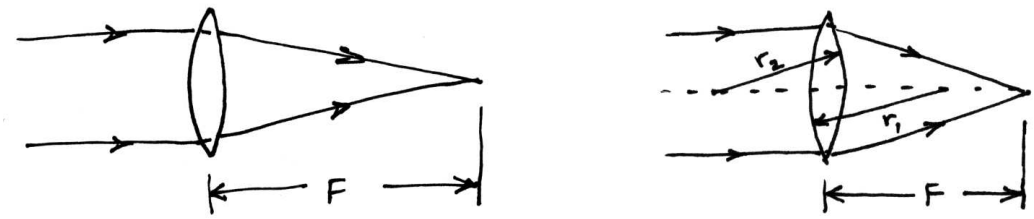


Figure 3.3: (a) Illustration of the focal length for a refracting telescope. (b) Illustration of the lensmaker's formula.

As astronomical targets are all very far away, we can regard the rays of light from a point source incident on different parts of the aperture to be essentially parallel. Of course this does not mean that rays from the centre of an extended object such as the Moon are parallel to those from the edge, merely that rays emerging from a single point on the object are almost parallel, regardless of where they go through the aperture.

We then need to shape the lens in such a way that the incident rays are brought to a focus. This can be achieved approximately by using a lens with spherical surfaces. For the

approximation to hold, the rays must be not too far from parallel to the optical axis (the axis of symmetry of the lens), and the radii of curvature of the two surfaces must both be much greater than the diameter of the lens (the *thin lens approximation*).

Consider a ray of light in a medium with index of refraction $n_{\rm i}$ incident on the surface of with an angle relative to the normal $\theta_{\rm i}$, and then passing into a medium with index of refraction $n_{\rm r}$. These quantities are related to the angle of the outgoing (refracted) ray $\theta_{\rm r}$ by Snell's law,

$$n_{\rm i} \sin\theta_{\rm i} = n_{\rm r} \sin\theta_{\rm r} \ . \tag{3.6}$$

Now suppose we have a lens with spherical surfaces with radii of curvature r_1 and r_2, as shown in Fig. 3.3(b). We define a sign for the radius such that it is positive if the centre of curvature is on the same side as the outgoing light; otherwise it is negative. In Fig. 3.3(b), for example, we have $r_1 > 0$ and $r_2 < 0$. Then using Snell's law we can show that for a thin lens, the focal length is related to the radii of curvature by the *lensmaker's formula*,

$$\frac{1}{F} = \frac{n - n_{\rm air}}{n_{\rm air}} \left(\frac{1}{r_1} - \frac{1}{r_2} \right) \approx (n-1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right) \ . \tag{3.7}$$

Here here n is the index refraction of the lens and the approximation on the right-hand side of (3.7) is generally used since $n_{\rm air} \approx 1.00029$ is very close to unity. For most types of glass one has n in the range of 1.5 to 1.6.

Furthermore we can show that if the light traverses two thin lenses with focal lengths F_1 and F_2, then this behaves as if there were a single lens with inverse focal length

$$\frac{1}{F} = \frac{1}{F_1} + \frac{1}{F_2} \ . \tag{3.8}$$

We can also consider a lens with a negative focal length, as shown in Fig. 3.4. Here we have $r_1 = \infty$ and $r_2 > 0$, and thus

$$\frac{1}{F} = (n-1) \left(\frac{1}{\infty} - \frac{1}{r_2} \right) < 0 \ . \tag{3.9}$$

A negative focal length means that the rays do not converge after passing through the lens but rather diverge from a point behind the lens.

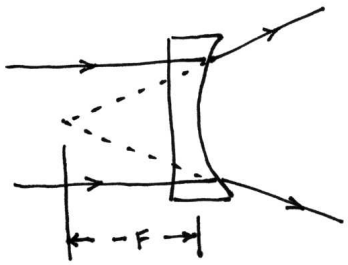


Figure 3.4: Illustration of negative focal length.

We define the *focal ratio* f as the ratio of the focal length to the diameter of the aperture:

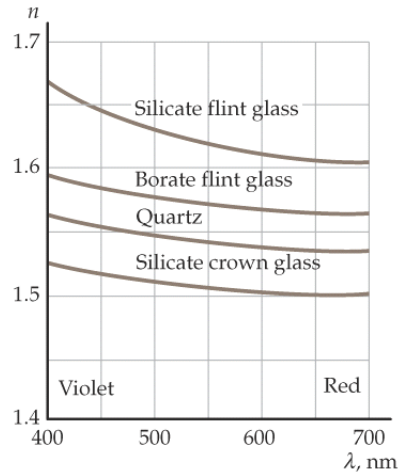
$$f = \frac{F}{D}\,, \tag{3.10}$$

and by tradition one often writes, say, $f/10$ for a lens or mirror with $f = 10$. Photographers will recognize this quantity as the 'f-number' of a lens. It is related to the depth of field that is in focus. This aspect of the focal ratio is not relevant for astronomical observations, since the targets are all very far away, effectively at infinity. The focal ratio is nevertheless often used by astronomers as it relates the angular field of view to the amount of light collected in a given area on the focal plane.

3.3.1 Chromatic aberration

One of the problems with refracting telescopes arises from the fact that the index of refraction n depends in general on the wavelength of the light λ, generally decreasing as shown in Fig. 3.5(a). The focal length F of a lens depends on n and so it therefore also will depend on λ.

(a)



(b)

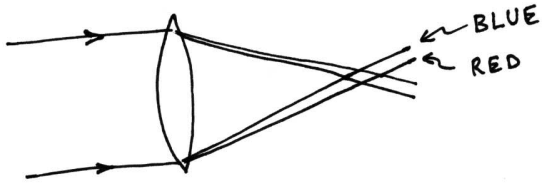


Figure 3.5: (a) The index of refraction versus wavelength for several different types of glass [11]. (b) Illustration of chromatic aberration.

Suppose that parallel rays of light with wavelength λ_1 are brought to a focus at some focal length F_1. If we place a recording device (e.g., a CCD chip) at that distance, then rays of this colour will be in focus. But rays of a different wavelength λ_2 will have in general a different focal length F_2, and thus they will be out of focus, as illustrated in Fig. 3.5(b). The resulting distortion of the image is called *chromatic aberration*.

Chromatic aberration can be partially corrected by constructing lenses out of more than one component. The simplest example is called an *achromatic doublet*, and is illustrated in Fig. 3.6. We will show that with such a doublet lens we can at least arrange for two given wavelengths, λ_1 and λ_2, to have the same focal length. Here we have a diverging lens made of flint glass together with a converging lens made of crown glass. Suppose the radius of curvature for all of the surfaces is r except for the flat face of the diverging lens, which has a radius of infinity.

For the flint or crown glass lenses alone we would have from equation (3.7) inverse focal lengths (note the sign convention for r)

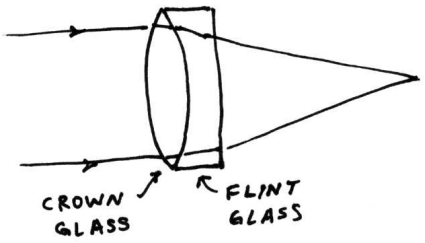


Figure 3.6: Illustration of an achromatic doublet.

$$\frac{1}{F_\mathrm{f}} = (n_\mathrm{f}-1)\left(-\frac{1}{r}-\frac{1}{\infty}\right) = -\frac{n_\mathrm{f}-1}{r}\,, \tag{3.11}$$

$$\frac{1}{F_\mathrm{c}} = (n_\mathrm{c}-1)\left(\frac{1}{r}-\frac{1}{-r}\right) = \frac{2(n_\mathrm{c}-1)}{r}\,. \tag{3.12}$$

Now using equation (3.8) for the focal length of both lenses in combination we find

$$\frac{1}{F} = \frac{1}{F_\mathrm{f}} + \frac{1}{F_\mathrm{c}} = \frac{2n_\mathrm{c}-n_\mathrm{f}-1}{r}\,. \tag{3.13}$$

For glass the index of refraction is around 1.5 to 1.6 so we find that the focal length from (3.13) is positive, i.e., the combination of the two components acts as a converging lens.

From equation (3.13) we can see that the focal length will be the same for two wavelengths λ_1 and λ_2 if the condition

$$2n_\mathrm{c}(\lambda_1) - n_\mathrm{f}(\lambda_1) = 2n_\mathrm{c}(\lambda_2) - n_\mathrm{f}(\lambda_2) \tag{3.14}$$

is satisfied. This is indeed possible by using crown and flint glass, and by a refined choice of glass properties one can arrange for λ_1 and λ_2 to be spaced so as to minimize the effect of chromatic aberration when averaged over the optical range.

An even better correction can for chromatic aberration can be achieved by using more lenses. So-called *apochromats* have equal focal lengths for three wavelengths; *superapochromats* correct for four. A modern telescope eyepiece may contain 8 to 10 individual lenses.

3.3.2 Further limitations of refractors

He we will briefly mention a few further limitations of refracting telescopes. In addition to chromatic aberration there is a further optical distortion that arises when the surfaces of the lenses are spherical (*spherical aberration*). Rays that are further from the optical axis are not brought to a focus at the exactly the same distance as those that pass through the middle of the lens.

The effect of spherical aberration can be lessened by an appropriate choice of the radii of the front and back surfaces of the lens; this is discussed further in [5]. It turns out that in addition to reducing the chromatic aberration, multiple lenses such as achromats also have less spherical

aberration. A further improvement can be achieved by using lenses with aspheric surfaces, which are, however, more expensive to produce.

Another difficulty with refractors is related to the difficulty in producing and supporting very large lenses. A lens needs to be supported by its edges, and if it is too large then it will tend to sag under its own weight, introducing optical distortions. Furthermore lenses are never perfectly transparent, and to reduce light loss (and to keep the mass low), lenses must be made thin. As a result most refractors have very long focal lengths. For example, the Yerkes 1-metre refractor shown in Fig. 3.2(b) has focal ratio of $f/20$, i.e., its tube is 20 m long, requiring a correspondingly large observatory dome.

3.4 Reflecting telescopes

The main idea of reflecting telescopes is to collect the rays of light and bring them to a focus using a mirror, as shown in Fig. 3.7(a). As in the case of refractors, the collecting surface is characterised by a diameter D and the telescope has a focal length F. This first telescope of this type was constructed by Newton in 1668 (Fig. 3.7(b), [12]).

(a) (b)

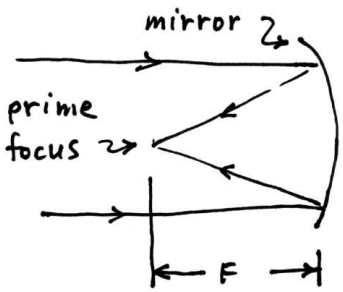


Figure 3.7: (a) Illustration of the focal length for a reflecting telescope. (b) Replica of Newton's original reflector (total length approx. 7 inches) [12].

The law of reflection – angle of incidence equal to angle of reflection – is independent of wavelength and thus the problem of chromatic aberration is avoided completely. Furthermore it is mechanically easier to support a mirror than a lens, and already in the 19th century it became possible to build very large reflecting telescopes. Figure 3.8 shows the 1.8 metre (6 foot) diameter reflector built in the 1840s by William Parsons (Lord Rosse) at Birr Castle, Ireland.

The light gathering power of the Rosse reflector was sufficient to reveal for the first time the spiral structure of galaxies, as seen in the sketch of the galaxy M51 in Fig. 3.8(b). The difficulty in observing this structure stems from the relatively low surface brightness of the galaxy. It is therefore necessary to gather many photons so as to distinguish the light from the dark regions. The angular size of the galaxy (approximately $4'$ between the two core areas) is large enough so that angular resolution is not the issue; even the unaided eye can resolve angles smaller than

(a) (b)

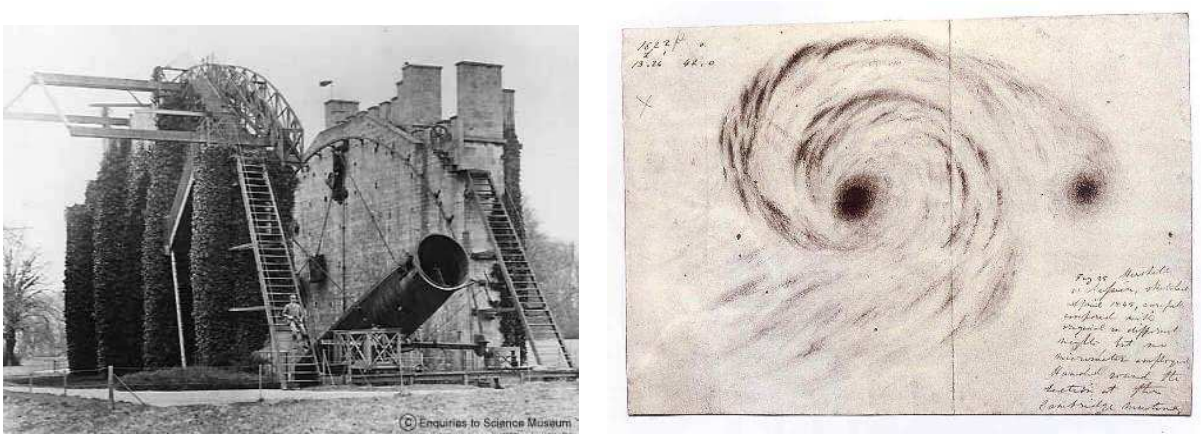


Figure 3.8: (a) The Rosse reflector [13]. (b) Drawing by Lord Rosse of the galaxy M51 [14].

this. The problem is rather the light gathering power needed to reveal the structure of a low contrast object.

Early reflecting telescopes were made using spherical mirrors. As long as the rays are sufficiently close to the mirror's axis of symmetry (the optical axis), the rays will converge at a focal length

$$F = \frac{R}{2}\,, \tag{3.15}$$

where R is the radius of curvature of the spherical surface.

The focal point indicated in Fig. 3.7(a) is called the *prime focus*. If we try to put some measuring device or eyepiece at the prime focus we will block at least some of the incident light. For very large devices the fraction of light blocked may be relatively small and the prime focus can be used. For smaller telescopes – certainly the one used by Newton in Fig. 3.7(b) – the observer would end up blocking the primary mirror and the setup would not work. Newton's solution was to place a flat diagonal mirror before the prime focus to bring the light to a focal point called the *Newtonian focus* on the side of the telescope, as shown in Fig. 3.9(a).

(a) (b)

Figure 3.9: Illustration of (a) the Newtonian focus and (b) the Cassegrain focus.

Alternatively, the light can be reflected off a hyperbolic secondary mirror back through a hole in the primary mirror to the *Cassegrain focus*, as shown in Fig. 3.9(b). This configuration has the advantage of allowing for a relatively short total length. Furthermore it is generally easier to mount large instruments at the Cassegrain focus than would be possible at the Newtonian or prime focus. Large modern telescopes often allow for instruments, depending on their size, to be mounted at either the prime or Cassegrain focus; the Newtonian focus is usually only used in smaller amateur telescopes.

Because reflectors overcome chromatic aberration and also can be much larger than refractors, they have been for most of the last century the primary instrument used in optical astronomy. Reflecting telescopes also have their limitations, however, some of which we will look at here. Spherical mirrors, just like spherical lenses, result in spherical aberration. As shown in Fig. 3.10(a), rays that are further from the optical axis will be brought to a focus closer than those that are more central. One way to cure this problem is to use a parabolic mirror, as shown in Fig. 3.10(b). Already in the 17th century it was shown mathematically that this shape will bring rays to a common focus regardless of their distance from the centre, as long as the rays are parallel to the optical axis. Parabolic mirrors are more difficult to manufacture but nevertheless this is what is used in almost all large reflectors.

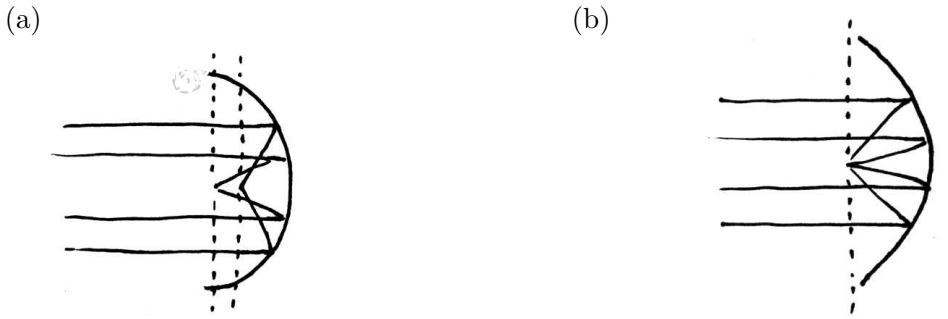


Figure 3.10: Illustration of (a) spherical aberration and (b) how it is overcome by a parabolic mirror.

Parabolic mirrors nevertheless suffer from other types of aberrations, an important example of which is called *coma*. This distortion arises when the rays are not parallel to the optical axis, as shown in Fig. 3.11(a). The distortion is such that a point-like source will be stretched out into the shape of a comet's tail, hence the name 'coma'. The coma-free field of view of a Newtonian telescope with a parabolic mirror may be quite small, say, only a few arcminutes. In sky surveys, however, we want to look at much larger angular regions – perhaps several square degrees.

One solution to the problem of coma was developed by Bernhard Schmidt in 1930. His telescope used a spherical mirror. Because of its symmetry, off-axis rays are focussed at the same point regardless of their angle, as long as they are incident at the same distance from the centre of the mirror. But we are back to having spherical aberration. Schmidt's solution was to cure this with a thin lens – a *Schmidt corrector plate*.[1] A Schmidt corrector plate is often used with the Cassegrain focus; this is called a Schmidt-Cassegrain telescope, as shown in Fig. 3.11(b).

[1]See [15] for a history of the Schmidt telescope as well as the Mt. Palomar 200-inch reflector.

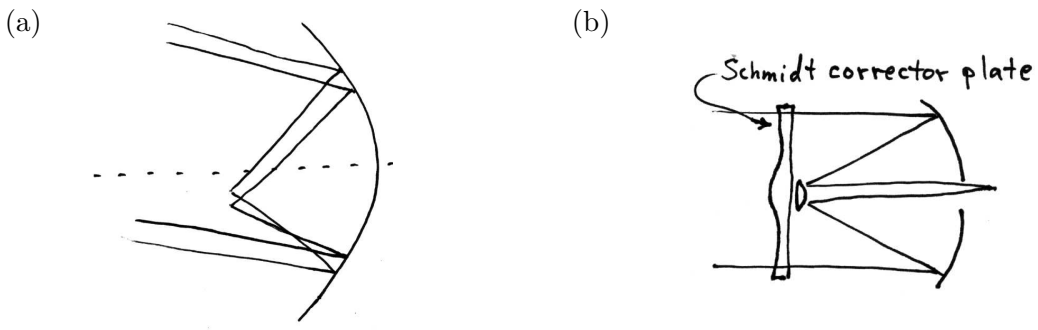


Figure 3.11: (a) Illustration of coma. (b) Schematic illustration of a Schmidt-Cassegrain telescope.

Because of the difficulties in manufacturing and supporting large lenses, Schmidt telescopes cannot be as large as simple reflectors. The world's largest is the Alfred-Jensch Telescope in Tautenburg, Germany, which has a 2 metre primary mirror with a 1.34 metre corrector plate, giving a field of view of $3.3° \times 3.3°$ [16].

3.5 Image formation

Up to now we have discussed how a telescope gathers light and focusses parallel rays to a point. This is of course only half the story – we also want the telescope to provide an image of the target object. The basic ideas of image formation are essentially the same for refracting and reflecting telescopes, so the illustrations below with mirrors could equally well be drawn with lenses and *vice versa*.

3.5.1 Plate scale

Consider a small, distant object, traditionally represented as an arrow, as shown in Fig. 3.12(a). The solid lines represent rays coming from the arrow's tail, and the dashed lines come from its head. As the arrow is very far away, the rays from any given point on the object are almost parallel regardless of where they hit the mirror.

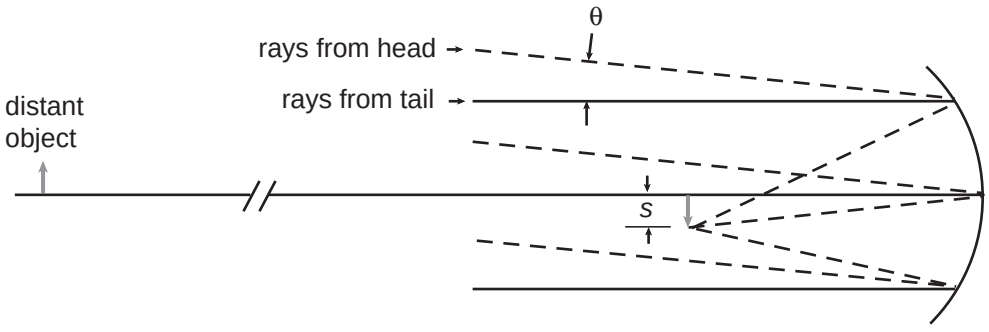


Figure 3.12: Illustration of image formation.

Because of the law of reflection, if we increase the angle of incidence of a ray by an angle θ, then its angle of reflection is increased by the same amount. This is illustrated using the angle subtended by the arrow in Fig. 3.12. The rays from the head of the arrow are brought to a focus at the same focal length F but some distance s below the image of the tail.

The distance s subtended by the image is something we can measure directly by, say, placing a piece of photographic film in the focal plane. We then want to relate this to the angular size of the real target. From the figure we have

$$\frac{s}{F} = \tan\theta \approx \theta \ , \tag{3.16}$$

where the approximation holds for small angles (almost always valid). Using this approximation we therefore have the change in angle with respect to image size,

$$\frac{d\theta}{ds} = \frac{1}{F} \equiv P \ , \tag{3.17}$$

which is called the *plate scale*, P.

Equation (3.17) gives the plate scale in radians per unit length. In order to express this in, say, arcseconds per mm we can use

$$P = \frac{1\,\mathrm{rad}}{F\,[\mathrm{mm}]} \times \frac{180^\circ}{\pi\,\mathrm{rad}} \times \frac{3600''}{1^\circ} = \frac{206\,265''}{F\,[\mathrm{mm}]} \ . \tag{3.18}$$

If we have the image in digital form as a two-dimensional array of pixels, then we need to know the distance per pixel, or equivalently we need the plate scale in arcseconds per pixel. Suppose, for example that two stars are found to have a separation of 50 pixels, and that the distance between pixels is $8.5\,\mu$m and the focal length of the telescope is $3\,048$ mm. The plate scale is thus

$$P = \frac{206\,265''}{3\,048\,\mathrm{mm}} \times \frac{0.0085\,\mathrm{mm}}{\mathrm{pixel}} = 0.575''/\mathrm{pixel} \ . \tag{3.19}$$

Therefore the angular separation of the stars is

$$\theta = sP = 50\,\mathrm{pixels} \times 0.575''/\mathrm{pixel} = 29'' \ . \tag{3.20}$$

In practice the plate scale is often calibrated using a pair of stars whose angular separation has been previously determined.

3.5.2 Magnification

If we place a piece of photographic film in the focal plane, then we could record an image. If we placed our eye at the telescope's focus, however, then we would see a blur. This is because lens of the human eye is arranged so that it can focus incoming parallel rays onto the retina, but the rays arriving at the focal plane are converging. To view the image with our eye we need to pass the rays first through another lens to bring them parallel. This is the role of the eyepiece.

Consider the setup in Fig. 3.13. The primary (or objective) lens brings the incoming rays to a focus at a focal length F_o. If an object, represented by the arrow, subtends an angle α_o, then this will form an image of size s at the focal plane. The rays continue from the focal plane through an eyepiece lens with focal length F_e. If it is placed a distance F_e from the focal plane of the objective, then the rays emerging from the eyepiece from a given point on the object, say, the head, will be parallel, and the image can then be viewed by a human eye.

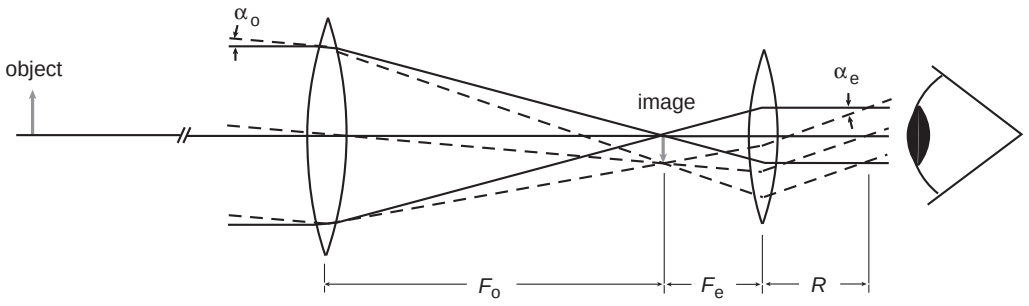


Figure 3.13: Illustration of angular magnification.

The off-axis rays, indicated by dashed lines in Fig. 3.13, make an angle α_o relative to the optical axis when they arrive from the object, and an angle α_e when the emerge from the eyepiece. With the same reasoning used to derive equation (3.16), these angles follow the relations

$$\alpha_\mathrm{o} = \frac{s}{F_\mathrm{o}} , \tag{3.21}$$

$$\alpha_\mathrm{e} = \frac{s}{F_\mathrm{e}} . \tag{3.22}$$

The ratio of these angles gives the *angular magnification*, M,

$$M = \frac{\alpha_\mathrm{e}}{\alpha_\mathrm{o}} = \frac{F_\mathrm{o}}{F_\mathrm{e}} . \tag{3.23}$$

For the RHUL 12-inch telescope, for example, we have $F_\mathrm{o} = 3048\,\mathrm{mm}$ and we may use an eyepiece with $F_\mathrm{e} = 26\,\mathrm{mm}$, giving $M = 117$.

It is important to note that the angular magnification is a property of both the objective mirror (or lens) and the eyepiece. Furthermore since one can easily make the eyepiece focal length very small, it is a simple matter to obtain a very high magnification. But this is not really a meaningful measure of the telescope's ability to resolve small angular structures. The *angular resolution* of the telescope, i.e., the smallest distinguishable angular separation, depends on the diameter of the objective and on the atmospheric conditions. We will return to angular resolution in Section 3.6.1.

A few more features in Fig. 3.13 are worth noting. At a certain distance from the eyepiece, the emerging bundle of rays has its smallest cross section, called the *exit pupil*. By looking at Fig. 3.13 we can see that the diameter D_e of the exit pupil is given by

$$D_\mathrm{e} = D_\mathrm{o}\frac{F_\mathrm{e}}{F_\mathrm{o}} = \frac{F_\mathrm{e}}{f} = \frac{D_\mathrm{o}}{M} , \tag{3.24}$$

where D_o is the diameter of the objective lens or mirror and f is the focal ratio. If the diameter of the exit pupil is any larger than the pupil of the human eye, say 6 mm, then some of the light is wasted. The amount of light entering the eye is maximized by placing it at the exit pupil.

We can easily show that the distance R in Fig. 3.13 is given by

$$R = F_e + \frac{F_e^2}{F_o} \approx F_e \ , \tag{3.25}$$

where the approximation holds in the usual case in which $F_e \ll F_o$. In practice, eyepieces are compound lenses and the effective position of the equivalent single lens will be somewhere in the middle of the eyepiece. This means that the actual distance between the last lens (the eye lens) and the exit pupil, called the *eye relief*, will in general be less than F_e. Having an eye relief of at least 15 mm or so is important for eyeglass wearers.

Finally, notice that if the diameter of the eyepiece is any smaller than the value given by equation (3.24), then rays that hit the objective lens near its outer edge will miss the eyepiece and be lost. Even if the eyepiece's diameter is equal to this value, some of the off-axis rays will miss the eyepiece. This then results in a reduction of intensity near the edge of the field of view called *vignetting*.

3.6 Angular measurement

In this section we will take up two different but related aspects of angular measurement. First we will look at angular resolution, i.e., the ability to resolve two or more sources with a given angular separation. Then we will consider astrometry, the measurement of the angular position of an individual source.

3.6.1 Angular resolution and the diffraction limit

The *angular resolution* of a telescope refers to the minimum angular separation between two light sources that can be seen as separate, rather than appearing as merged into a single smeared out blob. High angular resolution (i.e., small minimum resolvable angle) is crucial for investigating the structure of extended objects such as clusters, galaxies or planetary surfaces.

The angular resolution will depend on a number of factors, including imperfections of the optical system and the distorting effect of the Earth's atmosphere, which we will examine in Chapter 4. Even if we had perfect optics and could eliminate atmospheric effects completely (which of course we can with a space-based telescope), there is still a unbeatable limit to angular resolution due to diffraction.

The diffraction limit is a fundamental consequence of the wave nature of light. Diffraction is an interference effect that results in a spreading out of light waves when they are constrained to pass around or through a structure, in our case through the aperture of the telescope.

As astrophysical sources are generally very far away we can regard the incident radiation to be plane waves. For now consider a single wavelength λ, although of course the light from a real star will be a mixture of wavelengths. Suppose plane waves with a wavelength λ pass through a circular aperture with diameter D and then are projected onto a screen at a distance r. In

principle we can work out the intensity of light at the screen by solving Maxwell's equations for the given boundary conditions. In practice we can write down the solution in closed form only for a few special cases, such as the *Fraunhofer diffraction* limit of $r \gg D^2/\lambda$, i.e., far away from the aperture.

For a circular aperture, Fraunhofer diffraction results in a central peak surrounded by a series of concentric rings as shown in Fig. 3.14. This is called an Airy diffraction pattern, first calculated by the Astronomer Royal Sir George Airy. He showed that the intensity I as a function of the angle θ from the optical axis is given by (see e.g. [10])

$$I(\theta) = I(0) \left[\frac{2J_1(ka \sin\theta)}{ka \sin\theta} \right]^2 \tag{3.26}$$

where $a = D/2$ is the radius of the aperture, the wavenumber k is defined as

$$k = \frac{2\pi}{\lambda} \,, \tag{3.27}$$

and the function J_1 is the first order *Bessel function* (of the first kind). The intensity pattern is shown in Fig. 3.14.

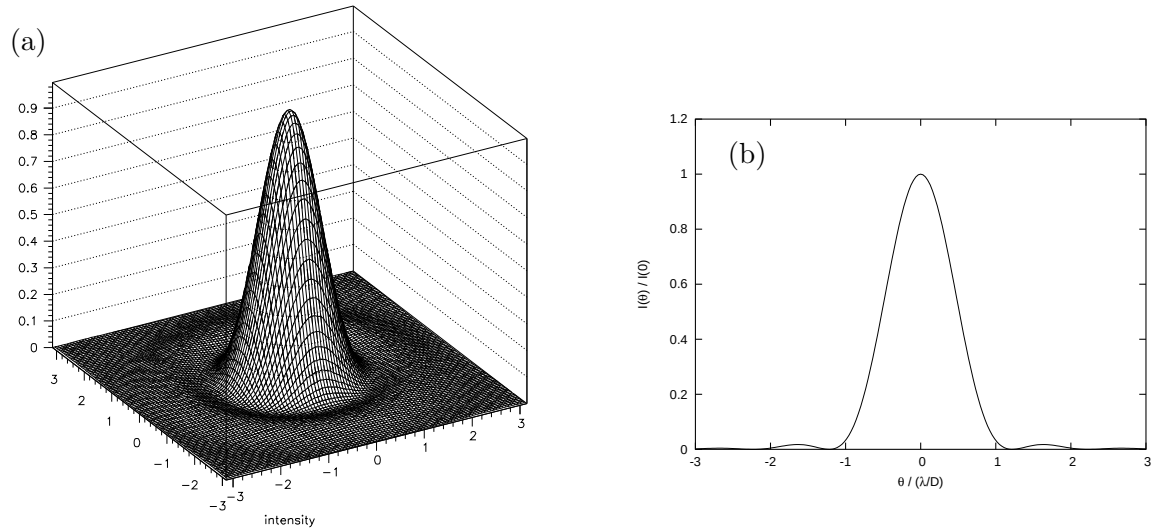


Figure 3.14: (a) The Airy diffraction pattern shown in two-dimensions; (b) the intensity shown as a function of angle from the centre in units of λ/D.

The Bessel function J_1 is one of a family of functions that are solutions to Bessel's equation (see e.g. [18]). They arise in a number of physical problems where there is a cylindrical symmetry. The Bessel function of (integer) order n can be written

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos(n\theta - x \sin\theta)\, d\theta \tag{3.28}$$

The integrals can be computed numerically; the results for $n = 0, 1, 2$ are shown in Fig. 3.15. The zeros of the function J_1 are found to be at $x = 3.8317, 7.0156, 10.1735, ...$

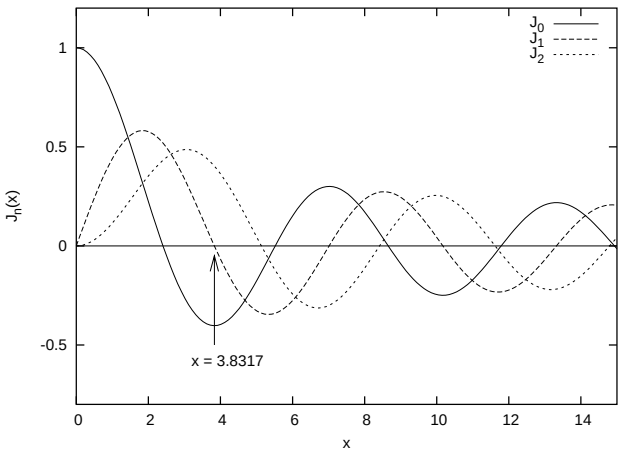


Figure 3.15: The Bessel functions J_0, J_1 and J_2.

The first dark ring around the central peak in the Airy diffraction pattern where the intensity goes to zero corresponds to the first zero of J_1, which means that for that angle we have

$$\frac{2\pi a \sin\theta}{\lambda} = 3.8317 \,, \tag{3.29}$$

or solving for θ,

$$\theta = 1.2197 \frac{\lambda}{D} \text{ radians} \,. \tag{3.30}$$

In a telescope the light does not simply pass through the aperture, it is focussed such that if the rays were initially parallel to the optical axis, they would converge to a point. But the incident parallel rays diverge relative to what they would have done without diffraction, and instead of converging to a point they form an Airy diffraction pattern on the telescope's focal plane. The central disc of the pattern (the Airy disc) contains 84% of the total flux.

Now if we look at two point sources of light (e.g., two stars), then in the absence of other distortions we will see two Airy patterns on the telescope's focal plane. If these are too close together then we will not be able to distinguish easily the resulting blob from that of a single star. Roughly speaking, we can resolve the two stars if their angular separation is large compared to the size of their Airy discs. This is the basis of the *Rayleigh criterion*, in which one deems the smallest resolvable separation to be such that the peak from one star is in the position of the first minimum of the other (and *vice versa*) as shown in Fig. 3.16.

If the peak of one source falls in the first minimum of the other, then their angular separation is simply given by equation (3.30) and one can use this to define numerically the diffraction limit. The Rayleigh criterion is somewhat arbitrary in that one could have chosen some other measure of the size of the Airy disc other than the position of its first minimum, e.g., the full width at half maximum. But any other reasonable measure of angular resolution would scale as λ/D and the coefficient would be some number of order unity.

As an example consider again observing light with $\lambda = 500\,\text{nm}$ using the Hubble Space Telescope, having a diameter $D = 2.4\,\text{m}$. Using equation (3.30) we find the diffraction limit for the angular resolution to be

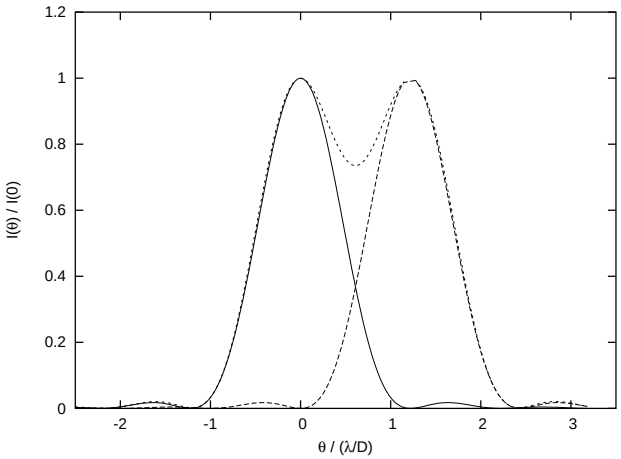


Figure 3.16: Illustration of the Rayleigh criterion.

$$\theta = 1.22 \times \frac{500 \times 10^{-9}\,\mathrm{m}}{2.4\,\mathrm{m}} \times \frac{180^\circ}{\pi\,\mathrm{radians}} \times \frac{3600''}{1^\circ} = 0.05'' \; . \tag{3.31}$$

The corresponding value for a 30-cm diameter is $0.4''$. In fact, the distorting effects of atmospheric turbulence (called *seeing*) limit the achievable angular resolution to typically $1''$. So most Earth-based telescopes are not 'diffraction limited' but rather 'seeing limited'. We will discuss atmospheric effects further in Chapter 4.

The intensity profile one observes for a point-like source is called the *Point Spread Function* (PSF) of the telescope. It is characterised by a certain angular width, often quoted as the full width at half maximum. Alternatively one can give the radius inside which a specified fraction of the light, e.g., 50% or 90%, is contained. The angular resolution from equation (3.30) (the Rayleigh criterion) is the 84% containment radius for a diffraction limited telescope. In general the PSF is a combined property of the telescope and the atmospheric conditions, with the latter often being dominant.

As a final note on angular resolution, we can now easily see that individual stars, with the exception of the Sun, are almost always too small and too far away to be resolved as anything but point-like sources of light. Suppose we consider a star with a diameter equal to that of the sun, $D = 1.4 \times 10^9\,\mathrm{m}$, located at a distance of $r = 10\,\mathrm{pc}$. Viewed from Earth, the star would subtend an angle

$$\theta = \frac{1.4 \times 10^9\,\mathrm{m}}{10\,\mathrm{pc}} \times \frac{1\,\mathrm{pc}}{3.1 \times 10^{16}\,\mathrm{m}} \times \frac{180^\circ}{\pi\,\mathrm{radians}} \times \frac{3600''}{1^\circ} \approx 0.001'' \; , \tag{3.32}$$

i.e., fifty times smaller than the angular resolution of the HST. Stellar diameters can however be measured by interferometric techniques, which are described further in [4].

As you can see from the numbers above, the Hubble Space Telescope's abilities are not too far away from what is required to resolve the angular size of a very large, nearby star. This was achieved for the first time in 1996 with the red giant star Betelgeuse (also called Alpha Orionis, the left shoulder of Orion) and is shown in Fig. 3.17. The image was recorded using ultraviolet light with a wavelength around 255 nm, giving an angular resolution of $0.038''$. This

false colour image shows UV emission from the star's chromosphere extending to at least 0.125″, significantly beyond the 0.055″ optical disc [19].

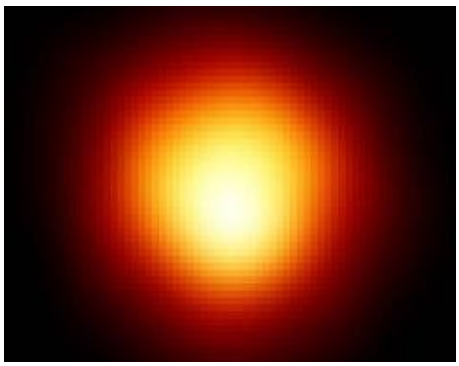


Figure 3.17: Image of the star Betelgeuse by the Hubble Space Telescope [19, 20].

3.6.2 Astrometric accuracy

The second aspect of angular measurement we will look at is the accurate determination of the direction of an individual point source of light, e.g., the right ascension and declination of a star. An important goal of astrometry is to determine the distances to stars using stellar parallax, which we will look at further in Chapter 8.

We should note right away that the astrometric accuracy is not the same as the angular resolution. Suppose we look at a distant (point-like) star with a diffraction-limited telescope. The image we see would have an intensity profile (the Point Spread Function) similar to Fig. 3.14. If we include other effects such as seeing we would see an even broader PSF generally having some sort of bell-shaped form.

Suppose the PSF width is characterized by some value $\sigma_{\rm PSF}$. If we were to detect only a single photon from the star, then $\sigma_{\rm PSF}$ would tell us the accuracy of the photon's measured direction. But we observe in general many photons, say, n_γ, and our angular measurement is based on their average. Under fairly unrestrictive assumptions we can show that the statistical accuracy in the determination of the mean angle follows

$$\sigma_{\rm star} = \frac{\sigma_{\rm PSF}}{\sqrt{n_\gamma}} \, . \tag{3.33}$$

Often n_γ is so large that the total measurement uncertainty is dominated not by this statistical component but by other systematic effects. The important point is that diffraction does not impose a fundamental limit to the accuracy of the angular measurement. For a sufficiently long exposure time, the statistical accuracy can be made arbitrarily small.

As examples of the attainable accuracy, the Carlsberg Meridian Telescope quotes a measurement uncertainty of 36 milli-arcseconds (mas) [21]. The Hipparcos mission [22], a space-based telescope, achieved around 1 mas, and future projects such as GAIA [23] are attempting to obtain angular accuracies in the micro-arcsecond range. This will be required, for example, in order to observe the wobble in a star's position due to the presence of an orbiting planet.

3.7 Telescope mounts

A telescope generally needs to be mounted in such a way that it can track a star as the Earth turns. One way to do this is to mount the telescope in a structure that rotates about an axis parallel to that of the Earth, at the same rate (360° per sidereal day) but in the opposite direction. In this way the telescope stays pointing at a fixed direction in space. This is called an *equatorial mount*, an example of which is shown in Fig. 3.18. This shows the 3-metre Shane Telescope at the Lick Observatory in California. The telescope's 'tube', actually an open 'Serrurier' truss tube, is held in a massive fork, which rotates about a central axis (like the axis of a tuning fork). This axis points toward the North Celestial Pole, so the angle between it and the floor is equal to the local latitude, in this case 37°.

Figure 3.18: An example of an equatorial mount: the 3-metre diameter Shane Telescope at the Lick Observatory, California [24].

A mechanically simpler way to support a telescope is using what is called an altitude-azimuth (or alt-az) mount. Here the telescope can move about a vertical axis, which changes the azimuth and about a horizontal one, which adjusts the altitude. In order to track a star while the Earth rotates, one needs to adjust continuously both the altitude and azimuth. Figure 3.19 shows an example of this type of mount: the 6-metre diameter Bol'shoi Teleskop Azimutal'ny (BTA) or 'Large Azimuthal Telescope' on Mt. Pastukhov, Russia.

An alt-az mount entails a further complication. As the Earth turns, the orientation of the telescope's field of view rotates. This can be understood by considering, say, the constellation Cassiopeia, with its characteristic 'W' shape. Suppose in the early evening it appears in the East with the prongs of its W pointing to the right. As the Earth turns, the W moves about the North Celestial Pole, always with the prongs of the W pointing away from NCP. Some hours later it will appear in the west with the W pointing in the opposite direction. This *field rotation* must be compensated by rotating the camera. So with an alt-az mount one must simultaneously

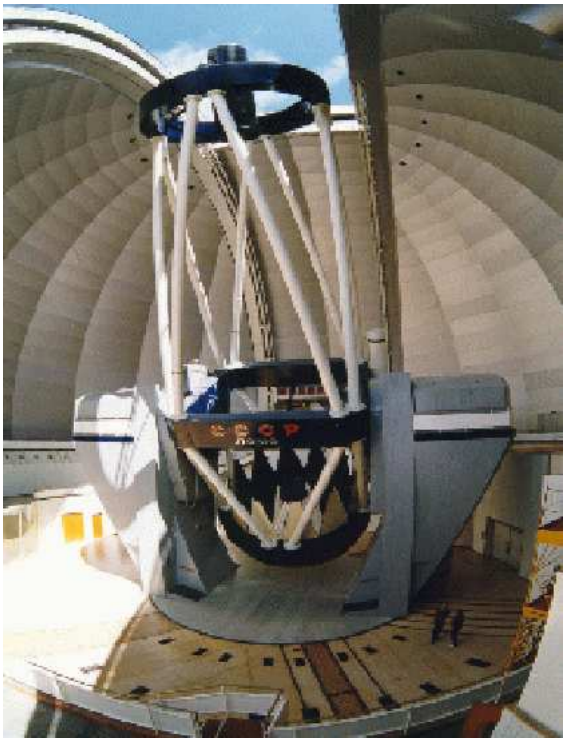


Figure 3.19: The 6-metre BTA (Bol'shoi Teleskop Azimutal'ny) or 'Large Azimuthal Telescope' on Mt. Pastukhov, Russia. This was the first large telescope to employ an alt-az mount [25].

rotate three things, all at adjustable rates, compared to only one rotation at a constant rate with the equatorial mount.

The technology needed to continuously compute and adjust the altitude, azimuth and field orientation did not become available until the 1970s with the advent of sufficiently fast computers. The BTA in Fig. 3.19 was the first large telescope to use an alt-az mount. Today, of course, computing technology has far surpassed what one needs for this task. Because it is mechanically much easier to support a telescope with an alt-az mount, this is what is used today for essentially all large telescopes.

As a final type of telescope mount we will look at the *meridian telescope*, also called a *transit telescope* or *transit circle*. This is a rather specialized device used for astrometry, i.e., making accurate measurements of the coordinates of stars. An example of this type of telescope is the Carlsberg Meridian Telescope shown in Fig. 3.20 [26].

A meridian telescope must be very accurately mounted so that it moves only north and south, always pointing along the local meridian. It is thus similar to an alt-az mount, except the azimuth is permanently fixed to zero. One then measures the time when a star crosses the telescope's field of view, i.e., the time of its transit. The time difference between the upper transit of the first point of Aries and that of a star gives the star's right ascension. In addition the angle of the telescope from the horizontal gives the altitude.

Because one is only interested in measuring the star's direction, chromatic aberration is not of great concern. The primary science goal is to measure the positions of nearby stars in order to investigate how they move relative to more distant ones. Therefore the requirements on light gathering power are not too great and there is no need for a very large diameter aperture. For smaller telescopes it is generally easier to achieve the required rigidity with lenses rather than mirrors, so meridian telescopes are generally refractors.

Figure 3.20: The Carlsberg Meridian Telescope on La Palma, Canary Islands [26].

A meridian telescope also serves as the basis for an accurate system of time, since the upper transit of the first point of Aries gives the starting point of the Local Sidereal Time. So the honour of defining the world's timekeeping system, as well as providing the meridian from which longitude is measured, clearly goes to the observatory with the best meridian telescope. Since the 1850s this was the Airy transit circle at the Royal Greenwich Observatory, and since 1884 its position has defined the position of 0° longitude. The upper transit of the Sun at this meridian is used to define Universal Time (UT, equivalent to Greenwich Mean Time, GMT).

3.8 Detectors

In preparation.

Chapter 4

Seeing through the atmosphere

In preparation.

Chapter 5

Measuring brightness

In this section we'll take a brief look at *stellar photometry*, that is, how to measure the brightness of a star. This will bring together a number of areas that we have discussed in previous chapters as well as introducing some new concepts.

5.1 Background subtraction

Suppose we have used a CCD detector and telescope to record an image of a star field as shown in Fig. 5.1. The image is displayed using the program ImageJ and the brightness and contrast have been adjusted so that a number of stars are visible.

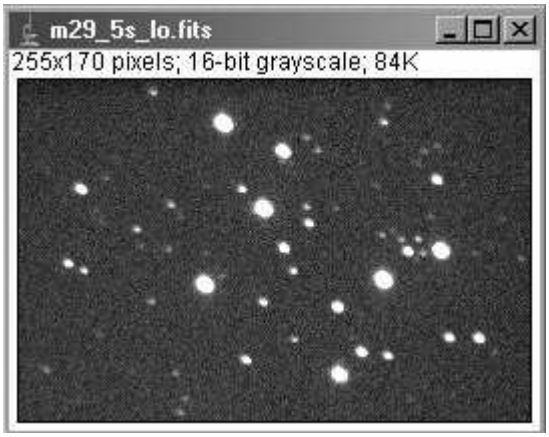


Figure 5.1: The open cluster M29.

Suppose we want to estimate the brightness of the stars in this image. The CCD camera will store the pixel values using a certain number of bits, which in this case was 16, so the brightness is given by a value in the range from 0 to $2^{16} - 1 = 65,535$. Using ImageJ you can see the pixel value by placing the cursor over a given location. Using this tool we see that the dark regions around the stars do not have zero counts, but rather have a background level of around 300 ADU. (Often the unit for "number of counts" is called the ADU or "analogue to digital unit".) The central pixels of the brighter stars have more than 20,000 ADU, and the dimmer stars have at their peaks less than 500 ADU.

What we would like is an estimate of the total brightness of a star. If we consider a narrow slice of a image passing through a star, it might show an intensity profile something like that shown in Fig. 5.2. There is a peak corresponding to the star, but in addition there is roughly constant background level from reflected light (light pollution), skyglow, etc. In order to estimate the brightness of a star, we need to subtract off the contribution from the background.

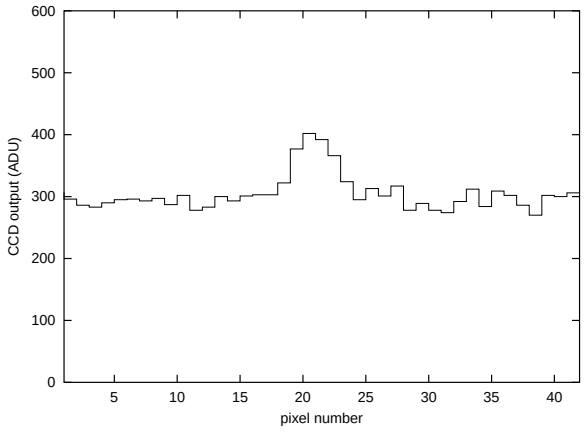


Figure 5.2: A profile of the intensity in a row of pixels passing through a star.

We can do this by first measuring the total number of counts in a region around the star, which we will call the *signal region*. For the moment we will choose the signal region to be sufficiently large to contain all of the light of the star, i.e., we will not try to correct for light which scatters outside this region. Using ImageJ we can specify a *Region of Interest* (ROI) as shown in Fig. 5.3(a) and centre it around the star we want to measure. The program will then tell us (using the 'measure' command) the area A of the region (measured in pixels) and the total number of counts C.

(a) (b)

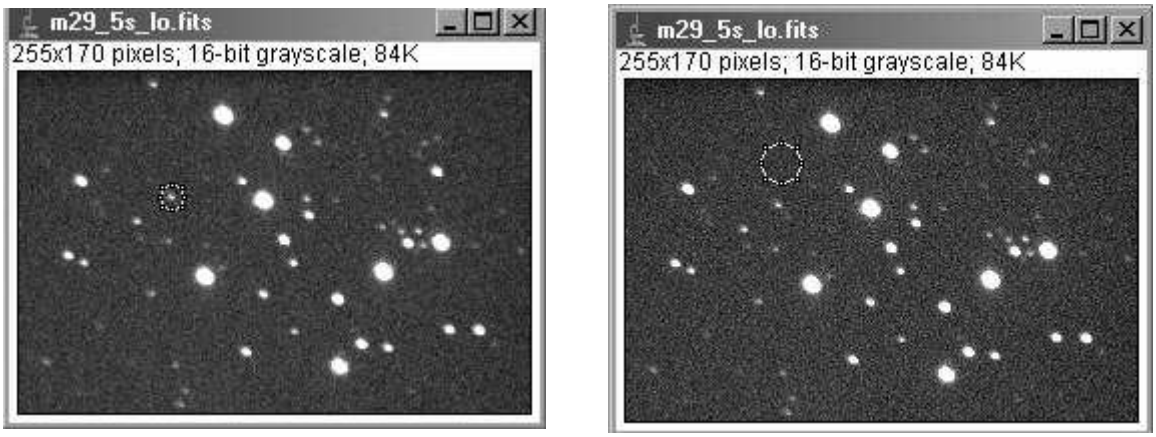


Figure 5.3: (a) A signal region around a star whose brightness is to be measured. (b) A background region.

From the regions shown in Fig. 5.3 we find for the signal region an area $A_\mathrm{s} = 104$ pixels and an output value of $C_\mathrm{s} = 32,471$ ADU. For the background region we find $A_\mathrm{b} = 300$ pixels and

a total output of $C_\mathrm{b} = 88,227$ ADU.

If we were simply interested in the relative intensity of stars, it would be sufficient to use the output values in ADU. We are going to want to estimate the uncertainty in our measurements, however, and one of the sources of uncertainty will be the random fluctuations due to the finite number of photons collected. For this we will need an estimate of the actual number of photoelectrons accumulated in each of the two regions. To convert the CCD output C to a number of photoelectrons N we need to multiply by the gain G:

$$N(\mathrm{e}^-) = C(\mathrm{ADU}) \cdot G(\mathrm{e}^-/\mathrm{ADU}) \; . \tag{5.1}$$

This step is needed because it is the number of photoelectrons N and not the CCD output C which we can treat as a Poisson random variable. The CCD used for the image of M29 above has a gain of $G = 2.3\,\mathrm{e}^-/\mathrm{ADU}$, so the numbers of photoelectrons in the signal and background regions are therefore

$$N_\mathrm{s} = 32,471\,\mathrm{ADU} \times 2.3\,\mathrm{e}^-/\mathrm{ADU} = 74,683\,\mathrm{e}^- \; , \tag{5.2-a}$$

$$N_\mathrm{b} = 88,227\,\mathrm{ADU} \times 2.3\,\mathrm{e}^-/\mathrm{ADU} = 202,922\,\mathrm{e}^- \; . \tag{5.2-b}$$

To correct for the background present in the signal region we need to account for the fact that the two regions do not have the same areas. This is done simply by scaling the number of background counts by the ratio of areas. Our background-subtracted estimate for the number of photoelectrons due only from the star is therefore

$$\begin{aligned} N_* &= N_\mathrm{s} - \frac{A_\mathrm{s}}{A_\mathrm{b}} N_\mathrm{b} \\ &= 74,683\,\mathrm{e}^- - \frac{104}{300}\,202,922\,\mathrm{e}^- = 4,337.0\,\mathrm{e}^- \; . \end{aligned} \tag{5.3}$$

Suppose we do this for two stars in an image and obtain background-subtracted estimates for the number of photoelectrons N_1 and N_2. These numbers are directly proportional to the brightness of the star, i.e., to the flux of light from each star. So we can obtain the difference in their magnitudes by using

$$m_1 - m_2 = -2.5 \log_{10} \frac{N_1}{N_2} \; . \tag{5.4}$$

Note that it would not matter here if we had left the numbers in ADU rather than converting to photoelectrons, as the gain would cancel in the ratio, i.e., $C_1/C_2 = N_1/N_2$. Nevertheless we will need the numbers of photoelectrons to estimate the random error in our measurements, which we will investigate in the next section.

5.2 Error analysis

In general measurements are subject to uncertainties of various sorts and it is an important part of experimental science to quantify them.[1] This is absolutely necessary if one is to use the measurements to try to distinguish between differing interpretations of the observed phenomenon. There are in general many sorts of experimental error in a measurement, and with enough care and money we can in some cases reduce them to arbitrarily low levels. Sometimes, however, experimental uncertainties have an essentially unbeatable lower limit. In our case, a fundamental limit comes from the finite number of photons collected by our detector.

5.2.1 The Poisson distribution

The number of photoelectrons collected in a finite time is an example of a random variable following the *Poisson probability distribution*. If we were to repeat the entire measurement many times under the identical conditions, we would not obtain the same number n each time. Rather, we would in general find different values and we could determine the probability $P(n)$ for finding n. What we would find in the case of the number of photons from a star in a fixed time t (and also for the number of photoelectrons found in a detector) is the Poisson distribution,

$$P(n;\nu) = \frac{\nu^n}{n!} e^{-\nu} \,. \tag{5.5}$$

Here n is a *random variable*, i.e., its value characterizes the output of the repeatable experiment; ν is a *parameter*, i.e., a constant value, which characterizes the probability distribution. We will often use the notation of equation (5.5) and list random variables to the left and parameters to the right of a semi-colon when specifying a probability function. In this case the parameter ν is equal to the mean value of n, i.e., the average one would obtain from an infinite number of observations of n. The Poisson distribution is shown in Fig. 5.4 for several values of ν.

We will not derive the Poisson distribution here but will only state what conditions must be fulfilled in order for a random variable to follow it. Consider a length of time t in which a certain number of events can occur. In our case these events are the creation of photoelectrons. Suppose that in each small element of time dt that one has a probability for a single event of

$$P(\text{one event in } dt) = \lambda \, dt \,, \tag{5.6}$$

where λ is a constant. Furthermore suppose that the events are *independent*. This means that the occurrence of one event should have no influence on the probability for the occurrence of another. We can then show that the probability to obtain n events in a time t will be given by the Poisson distribution (5.5) with the mean value given by

$$\nu = \lambda t \,. \tag{5.7}$$

The number of photoelectrons found in the signal and background regions, N_s and N_b can both be treated as Poisson random variables. If we were to repeat the experiment under

[1] Following long established but perhaps unfortunate practise, we will often use the word 'error' not to mean 'mistake', but rather 'uncertainty'.

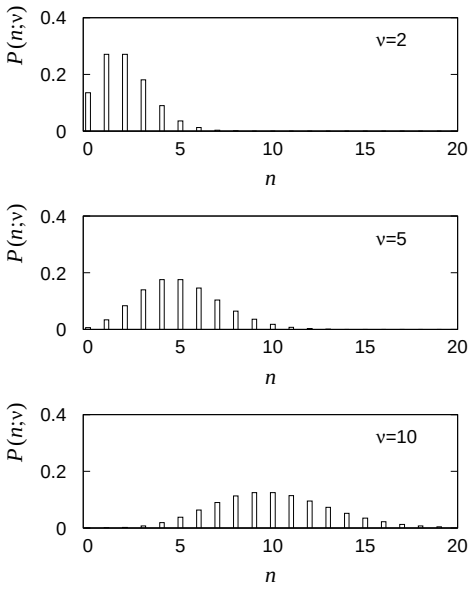


Figure 5.4: The Poisson distribution $P(n;\nu)$ for several values of the mean ν.

identical conditions many times, we would in general find different values for N_s and N_b, and the probabilities for the different values would follow Poisson distribution with some mean values ν_s and ν_b. We don't actually know ν_s and ν_b, but we can use the observed values, N_s and N_b, as estimates.

The 'random error' or 'statistical error' can be estimated by using the width of the corresponding Poisson distribution. A convenient measure of the width of a probability distribution is given by the standard deviation σ. This is defined by the relation

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 \,, \tag{5.8}$$

where the angled brackets indicate mean values. We have already mentioned that the mean value of n is $\langle n \rangle = \nu$. We could in also work out $\langle n^2 \rangle$ and use this to find that the standard deviation of a Poisson distribution is given by

$$\sigma = \sqrt{\nu} \,. \tag{5.9}$$

Although we do not (and may never) know ν, we can estimate it with the observed number n, and thus we can estimate the standard deviation with $\sqrt{n}$. We will use this result to estimate the measurement uncertainty of a Poisson distributed quantity such as N_s or N_b.

5.2.2 Error propagation

Using the results from the previous section we can estimate the random error in the number of photoelectrons N_s and N_b. What we want, however, is an estimate of the uncertainty in N_*, the background-subtracted measurement of the brightness of a star. And although N_s and N_b are both Poisson variables, the combination given by equation (5.3) is not, so we cannot simply take its square root to obtain the random error. We need to 'propagate' the uncertainties in N_s and N_b into the final quantity N_*.

To make the discussion more general, suppose we have measured quantities x_1, x_2, ..., with standard deviations σ_1, σ_2, ..., which we use as a measure of their random errors. Suppose that what we are interested in, however, is some function $f(x_1, x_2, \ldots)$. What is the experimental error in f? We can find this with the *error propagation* formula. If the quantities x_1 and x_2 are independent in the sense discussed above, that is, knowing the value of one does not affect the probability of where one will find the other, then the square of the standard deviation of the function f can be approximated by

$$\sigma_f^2 \approx \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_1^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_2^2 + \ldots \ . \tag{5.10}$$

For reference you should know that if the quantities x_1, x_2, ...are not independent, then the error propagation formula must include additional terms. Neglecting these terms can render the result completely invalid. The formula for the general case is

$$\sigma_f^2 \approx \sum_{i,j} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j} V_{ij} \ , \tag{5.11}$$

where the sums extend over all of the measured variables and V_{ij} is an element of the *covariance matrix*. The diagonal terms of this matrix are given by $V_{ii} = \sigma_i^2$, and if all of the off-diagonal terms are zero then one recovers the previous equation (5.10). If the x_i are not independent, however, (strictly speaking, if they are *correlated*), then the off-diagonal elements of the covariance matrix are nonzero and the two equations will give different results, the second of the two being the correct one.

5.2.3 Application to stellar brightness

We can now apply the results of Sections (5.2.1) and (5.2.2) to our problem of determining the uncertainty in N_* as given by equation (5.3). Taking the derivatives required by the error propagation formula (5.10) gives

$$\begin{aligned} \sigma_{N_*}^2 &= \left(\frac{\partial N_*}{\partial N_\mathrm{s}}\right)^2 \sigma_{N_\mathrm{s}}^2 + \left(\frac{\partial N_*}{\partial N_\mathrm{b}}\right)^2 \sigma_{N_\mathrm{b}}^2 \\ &= 1^2(\sqrt{N_\mathrm{s}})^2 + (-A_\mathrm{s}/A_\mathrm{b})^2(\sqrt{N_\mathrm{b}})^2 \\ &= N_\mathrm{s} + \left(\frac{A_\mathrm{s}}{A_\mathrm{b}}\right)^2 N_\mathrm{b} \ . \end{aligned} \tag{5.12}$$

or equivalently as our final estimate of the random error in N_*,

$$\sigma_{N_*} = \sqrt{N_\mathrm{s} + \left(\frac{A_\mathrm{s}}{A_\mathrm{b}}\right)^2 N_\mathrm{b}} \ . \tag{5.13}$$

We should emphasize that this formula for the random error only includes the contributions from the Poisson fluctuations in N_s and N_b. There could in general be many different sources of

measurement error, and equation (5.13) only covers one area. It will turn out to be the dominant source of uncertainty in the limit of a very weak source.

Often one wants the *relative* error in N_*, which is simply the ratio of equations (5.3) and (5.13),

$$\frac{\sigma_{N_*}}{N_*} = \frac{\sqrt{N_\mathrm{s} + (A_\mathrm{s}/A_\mathrm{b})^2 N_\mathrm{b}}}{N_\mathrm{s} - (A_\mathrm{s}/A_\mathrm{b}) N_\mathrm{b}} = \frac{\sqrt{C_\mathrm{s} + (A_\mathrm{s}/A_\mathrm{b})^2 C_\mathrm{b}}}{C_\mathrm{s} - (A_\mathrm{s}/A_\mathrm{b}) C_\mathrm{b}} \frac{1}{\sqrt{G}} \,. \tag{5.14}$$

Here the second expression uses the number of counts in ADU and the gain G rather than the number of photoelectrons. Sometimes one quotes the equivalent *signal to noise ratio*, usually written S/N, which is simply the ratio of N_* (the signal) to σ_{N_*} (the noise).

In the example of the star in Section 5.1 with $N_* = 4,337\,\mathrm{e}^-$ we find

$$\sigma_{N_*} = \sqrt{74,683\,\mathrm{e}^- + (104/300)^2 202,922} = 315\,\mathrm{e}^- \,, \tag{5.15}$$

or a relative error of

$$\frac{\sigma_{N_*}}{N_*} = 7.3\% \,, \tag{5.16}$$

or equivalently a signal-to-noise ratio of $1/0.073 = 14$.

The number of photoelectrons N_s and N_b are both proportional to the exposure time t, and so equation (5.14) says that the relative error follows

$$\frac{\sigma_{N_*}}{N_*} \propto \frac{1}{\sqrt{t}} \,. \tag{5.17}$$

The image in Fig. 5.1 was obtained with a 5 s exposure. We can see from equation (5.17) that with an exposure twice as long we would have a relative error of $7.3\%/\sqrt{2} = 5.2\%$. Of course if we allowed the exposure time to become arbitrarily long, then the actual relative uncertainty will not go to zero, but rather at some point another source of experimental error will begin to dominate over that from Poisson fluctuations in N_s and N_b.

5.2.4 Application to apparent magnitude

Often rather than leaving the result in units of photoelectrons we would like to calibrate this brightness level relative to that of a known star. Suppose we have measured N_* for our star of interest and we have calibrated our camera such that a reference star of magnitude m_ref will give a brightness of N_ref photoelectrons for a background-subtracted exposure of the same duration. We can then estimate the magnitude of our star as

$$m_* = m_\mathrm{ref} - 2.5 \log_{10} \frac{N_*}{N_\mathrm{ref}} \,. \tag{5.18}$$

Let's assume for now that m_ref and N_ref are known sufficiently well that their uncertainty can be neglected. The only quantity on the right-hand side of (5.18) which is uncertain is then N_*, and thus our error propagation formula becomes

$$\sigma_{m_*} = \left|\frac{\partial m_*}{\partial N_*}\right| \sigma_{N_*} \ . \tag{5.19}$$

Using $\log_{10} x = \ln x / \ln 10$, we find

$$\sigma_{m_*} = \left|\frac{-2.5}{\ln 10}\right| \frac{\sigma_{N_*}}{N_*} \approx 1.086 \frac{\sigma_{N_*}}{N_*} \ . \tag{5.20}$$

This convenient relation says that the uncertainty in the apparent magnitude is approximately equal to the relative uncertainty in the brightness. For our example above with $\sigma_{N_*}/N_* = 0.073$ we would find $\sigma_{m_*} = 0.079$.

Chapter 6

Radiative transfer and solar limb darkening

In this chapter we extend our study of brightness to cover objects whose size can be resolved, such as the sun. We will introduce the important concept of the *intensity* of a radiation field and we will see how this changes as the radiation passes through matter. As a result we will be able to explain why the *limb* or edge of the sun appears darker than the centre of its disc, a phenomenon called *limb darkening*. We will then be able to relate the observed intensity of the sun to quantities that can be predicted by solar models, such as the temperature as a function of depth.

6.1 Intensity

To describe limb darkening quantitatively, we need to define the *intensity* of the sun's radiation. Consider an area element dA with a normal unit vector $\mathbf{n}$, as shown in Fig. 6.1. This area element could be anywhere in the radiation field, e.g., inside the sun, although we will finally be interested in the case where dA is on the sun's surface. Suppose radiation with energy dE_λ passes through dA into a cone of solid angle $d\omega$, which has an angle θ from the vector $\mathbf{n}$, in a wavelength interval from λ to $\lambda + d\lambda$ in a time dt. The figure should not imply that the radiation has to pass through the surface exactly at the tip of the cone. The cone only specifies the direction, i.e., a range of angles. It can be regarded as originating anywhere on dA. The intensity I_λ is defined by

$$dE_\lambda = I_\lambda \cos\theta \, d\lambda \, dA \, d\omega \, dt \; . \tag{6.1}$$

The factor of $\cos\theta$ is included in the definition because the area element dA has an projected area of $dA\cos\theta$ when viewed along the direction of the cone $d\omega$. In this way the intensity of a radiation field depends on the position and on the direction into which the radiation goes (i.e., the direction of $d\omega$), but it is actually independent of our choice of the unit vector $\mathbf{n}$.

The intensity I_λ has units of W m^{-3} sr^{-1}. Note that one of the powers of length in the m^{-3} is related to the wavelength $d\lambda$ and the other two are from the area dA. Equivalently, the

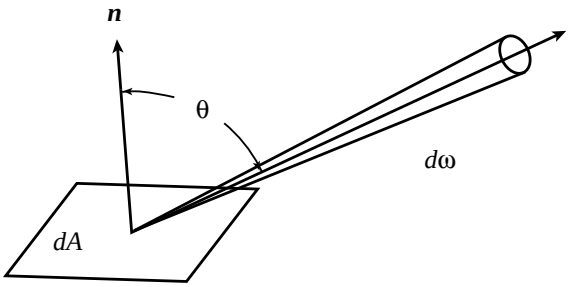


Figure 6.1: Illustration of the definition of intensity (see text).

intensity I_ν can be defined using the energy dE_ν in a frequency interval from ν to $\nu + d\nu$, which has units of W m^{-2} sr^{-1} s.

The intensity is what we can measure when we observe the surface of a resolved object such as the sun, i.e., when we measure the amount of energy coming from a given area of its surface projected along our line of sight. It is related to but should not be confused with the flux or the luminosity. If we consider the energy passing per unit time through the area element dA and integrate over all solid angle, then we get the *radiative flux*,

$$F_\lambda = \int I_\lambda \cos\theta \, d\omega \; . \tag{6.2}$$

If we integrate the flux over the surface of a sphere of radius r centred about a star, we obtain the *monochromatic luminosity*,

$$L_\lambda = \int_S F_\lambda \, dA \; = \; 4\pi r^2 F_\lambda \; , \tag{6.3}$$

where the second equality holds if the flux is independent of direction. The total luminosity is obtained by integrating L_λ over all wavelengths; this is what determines the absolute magnitude of a star.

6.2 The equation of radiative transfer

We will now derive the *equation of radiative transfer*, which describes how the intensity of a radiation field changes as it traverses matter. Consider an element of matter of thickness ds which could be, for example, somewhere in the middle of the photosphere. As radiation with a certain incident intensity I_λ traverses ds it can be scattered or absorbed. The resulting reduction in intensity is proportional to ds, to the density of matter ρ and to the incident I_λ:

$$dI_\lambda = -\kappa_\lambda \rho I_\lambda \, ds \; . \tag{6.4}$$

The constant of proportionality, κ_λ, is called the *absorption coefficient*, which depends in general on the wavelength λ.

The same element of matter can emit photons which will increase the intensity. The change in I_λ due to emission is proportional to $\rho\, ds$ but is assumed independent of the incident intensity, i.e.,

$$dI_\lambda = j_\lambda \rho\, ds \ . \tag{6.5}$$

The constant of proportionality, j_λ, is called the *emission coefficient*, and it also depends in general on the wavelength λ.

Putting together equations (6.4) and (6.5) gives the rate of change of the intensity,

$$\frac{1}{\rho}\frac{dI_\lambda}{ds} = j_\lambda - \kappa_\lambda I_\lambda \ . \tag{6.6}$$

We now define the *source function* S_λ as the ratio

$$S_\lambda = \frac{j_\lambda}{\kappa_\lambda} \ , \tag{6.7}$$

which has the same units as the intensity. Equation (6.6) then becomes

$$\frac{1}{\kappa_\lambda \rho}\frac{dI_\lambda}{ds} + I_\lambda = S_\lambda \ , \tag{6.8}$$

which is called the equation of radiative transfer or simply the *transfer equation*.

It is convenient to define the *optical thickness* τ_λ of a region of matter by the relation

$$d\tau_\lambda = \kappa_\lambda \rho\, ds \ . \tag{6.9}$$

That is, the total optical thickness is found by integrating (6.9) over the region in question,

$$\tau_\lambda = \int \kappa_\lambda \rho\, ds \ , \tag{6.10}$$

where both the density ρ and the absorption coefficient κ_λ can depend in general on the position. The optical thickness is a dimensionless quantity. If a region of matter has an optical thickness $\tau_\lambda \ll 1$ it is said to be *optically thin*; if $\tau_\lambda \gg 1$ it is *optically thick*.

Rewriting the equation of radiative transfer in terms of the optical thickness τ gives

$$\frac{dI}{d\tau} + I = S \ . \tag{6.11}$$

Here for simplicity we have suppresed the subscript λ, but we should keep in mind that I, S and τ all depend in general on the wavelength of the radiation.

6.3 Solving the equation of radiative transfer

We can solve the equation of radiative transfer (6.11) with an integrating factor f. We want to find a function f such that when we multiply both sides of (6.11) by it, we find a total derivative on the left-hand side. We can easily guess (and check) that

$$f = e^{\tau} \tag{6.12}$$

has the desired property:

$$e^{\tau}\left(\frac{dI}{d\tau} + I\right) = \frac{d}{d\tau}\left(e^{\tau} I\right) \ . \tag{6.13}$$

Multiplying (6.11) by f therefore gives

$$\frac{d}{d\tau}\left(e^{\tau} I\right) = e^{\tau} S \ . \tag{6.14}$$

Temporarily renaming the variable from τ to τ', we can integrate both sides of equation (6.14) from 0 to τ,

$$\int_0^{\tau} \frac{d}{d\tau'}\left(e^{\tau'} I(\tau')\right)\, d\tau' = \int_0^{\tau} e^{\tau'} S(\tau')\, d\tau' \ . \tag{6.15}$$

The left-hand side of (6.15) is simply $e^{\tau} I(\tau) - I(0)$. Solving the equation for $I(\tau)$ gives

$$I(\tau) = I(0) e^{-\tau} + \int_0^{\tau} S(\tau') e^{-(\tau - \tau')}\, d\tau' \ . \tag{6.16}$$

where the factor of $e^{-\tau}$ was moved inside the integral on the right-hand side.

Equation (6.16) really just rewrites the differential equation (6.11) as an integral, which we still need to solve. Let us consider first the simple case where there is no emission from the medium. That is, suppose the emission coefficient j_λ is zero and therefore also the source function S is zero. The solution (6.16) becomes

$$I(\tau) = I(0) e^{-\tau} \ . \tag{6.17}$$

This shows that the effect of an absorbing medium is to exponentially suppress the intensity of the incident radiation.

As a second example, suppose the source function S is a constant, i.e., it does not depend on the position within the medium. The solution (6.16) becomes

$$\begin{aligned} I(\tau) &= I(0) e^{-\tau} + S \int_0^{\tau} e^{-(\tau - \tau')}\, d\tau' \\ &= I(0) e^{-\tau} + S(1 - e^{-\tau}) \ . \end{aligned} \tag{6.18}$$

The two terms in (6.18) can be easily interpreted. The first shows the incident radiation, which is exponentially suppressed by absorption. The second term corresponds to the radiation emitted by the medium itself.

If the medium is optically thick ($\tau \gg 1$), then the exponential terms in (6.18) go to zero and we find $I = S$. This provides a simple physical interpretation of the source function. It is what the intensity becomes after the incident radiation has died away. That is, it is what one sees coming from the medium itself.

6.4 Solar limb darkening

We will now apply the equation of radiative transfer to the photosphere of a star such as the sun. The limb darkening of the sun can be seen easily in Fig. 6.2 The key to understanding limb darkening is that not all of the photons that escape the sun are emitted at the same radius. Rather, they are emitted from a layer called the *photosphere*, which is several hundred km thick (small compared to the 700,000 km radius of the sun). The temperature increases as one goes deeper into the photosphere, and according to laws of blackbody radiation, the light from the hotter regions is more intense.

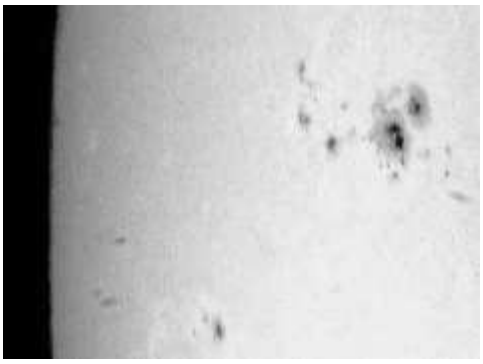


Figure 6.2: The solar limb (photo by Bevan Scott, RHUL Physics Deptartment).

The light from the interior of the photosphere must, however, pass through the outer layers in order to escape. In doing so there is given probability that it will be absorbed or scattered. There is a certain typical distance over which a photon has a significant probability of being absorbed, namely, the distance that corresponds to one optical thickness, τ_λ. Photons that traverse one optical thickness coming from the centre of the disc are originating at a vertical depth equal to τ_λ. Photons coming from closer to the limb that travel the same distance through the photosphere, however, originate closer to the surface of the sun, where it is cooler. Therefore, the intensity of the centre of the sun's disc is brighter than at the edge. This is illustrated schematically in Fig. 6.3.

It is convenient to introduce the *optical depth* $\tau_{\lambda s}$ of a region in the photosphere by defining

$$d\tau_{\lambda s} = -\kappa_\lambda \rho \, ds \ . \tag{6.19}$$

Notice the minus sign since here optical depth is defined to increase moving away from the observer, whereas s increases as one moves towards the observer. In contrast, the optical

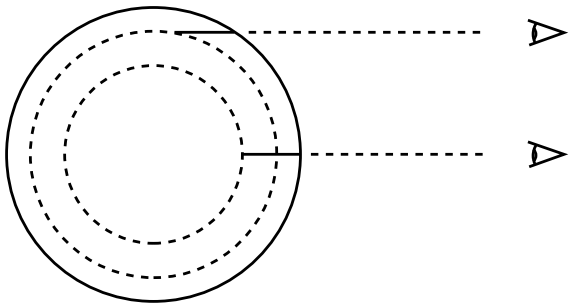


Figure 6.3: A schematic illustration of limb darkening (see text).

thickness of a region of matter is by definition positive.

The optical depth $\tau_{\lambda s}$ describes the thickness of matter that a ray of light passes through from somewhere inside the sun to the observer. Except at the centre of the sun's disc, this thickness does not correspond directly to the depth below the surface. It is therefore useful to define the *vertical* optical depth $\tau_{\lambda v}$ as the optical depth of a given region inside the sun, measured vertically downward from the sun's surface. If the normal to the sun's surface directly above the region in question makes an angle θ relative to our line of sight, then $d\tau_{\lambda v}$ is given by

$$d\tau_{\lambda v} = \cos\theta \, d\tau_{\lambda s} = -\rho\kappa_\lambda \cos\theta \, ds \ . \tag{6.20}$$

This relation holds as long as one can regard the photosphere as a slab between parallel planes. Since the photosphere is only several hundred km thick, which is small compared to the sun's 700,000 km radius, this is a good approximation except very close to the edge of the disc. In terms of the vertical optical depth, the transfer equation becomes

$$\mu \frac{dI_\lambda(\tau_\lambda, \mu)}{d\tau_\lambda} = I_\lambda(\tau_\lambda, \mu) - S_\lambda(\tau_\lambda) \ , \tag{6.21}$$

where we will use the abbreviation $\mu = \cos\theta$. Here following well established tradition, we have dropped the subscript v from the vertical optical depth. This has the unfortunate consequence of making it look just like the optical thickness, which has the opposite sign. Nevertheless, this is the conventional definition (and notation) when describing stellar atmospheres. Keep in mind that both I_λ and S_λ depend on position, and in addition, I_λ depends on the direction θ, or equivalently on μ.

What we can measure is the intensity of the sun's radiation as it leaves the photosphere, i.e., at $\tau_\lambda = 0$, and we would like to be able to solve the transfer equation in order to relate $I_\lambda(0, \mu)$ to the source function S_λ. Just as we did in Section 6.2, we can write down a formal solution which in effect only converts the differential equation (6.21) into an integral equation. To simplify the notation we will temporarily drop the subscript λ from I, S, and τ, but keep in mind that these quantities all depend in general on the wavelength. Multiplying both sides of (6.21) by $e^{-\tau/\mu}$ and rearranging terms gives

$$\mu \frac{dI}{d\tau} e^{-\tau/\mu} - I e^{-\tau/\mu} = -S e^{-\tau/\mu} \ , \tag{6.22}$$

which can be written as

$$\frac{d}{d\tau}\left(I\mu e^{-\tau/\mu}\right) = -Se^{-\tau/\mu} \,. \tag{6.23}$$

Integrating both sides from a depth 0 (the surface) down to τ and solving for the surface intensity $I(0,\mu)$ gives

$$I(0,\mu) = I(\tau,\mu)e^{-\tau/\mu} + \frac{1}{\mu}\int_0^\tau S(\tau')e^{-\tau'/\mu}\,d\tau' \,. \tag{6.24}$$

If we now consider that the sun extends many optical depths below the surface, we can let τ go to infinity. Reinserting the subscript λ, we obtain

$$I_\lambda(0,\mu) = \int_0^\infty S_\lambda(\tau_\lambda)e^{-\tau_\lambda/\mu}\,\frac{d\tau_\lambda}{\mu} \,. \tag{6.25}$$

Equation (6.25) with $\mu = \cos\theta$ gives the intensity we see from the sun where the normal to surface makes an angle θ relative to our line of sight. This angle is directly related to the position that we view on the sun's disc, as illustrated in Fig. 6.4. If the sun's radius is R and we look at a radius r on the disc, we have

$$\mu = \cos\theta = \sqrt{1 - \left(\frac{r}{R}\right)^2} \,. \tag{6.26}$$

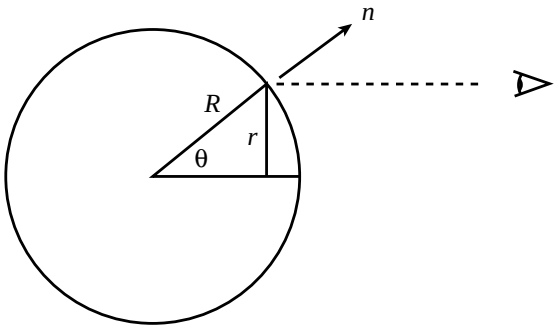


Figure 6.4: Illustration of the relation between $\cos\theta$, the radius on the disc r, and the sun's radius R (see text).

6.5 Measuring the source function

We would like to solve equation (6.25) for the source function $S_\lambda(\tau_\lambda)$ as a function of the vertical optical depth. If we can then measure $I_\lambda(0,\mu)$, the intensity emerging from the surface at different angles (i.e., at different positions on the disc), we can use this to determine $S_\lambda(\tau_\lambda)$.

In practice, we do not solve (6.25) for $S_\lambda(\tau_\lambda)$ explicitly. We can, however, suppose a reasonably general functional form for the source function which depends on certain adjustable parameters, and we can use the measured intensity to determine the values of the parameters.

For example, suppose that the source function is a power series in τ_λ times the surface intensity at the centre of the disc ($\mu = 1$),

$$S_\lambda(\tau_\lambda) = I_\lambda(0,1) \sum_{n=0}^{m} a_{\lambda n} \tau_\lambda^n \, . \tag{6.27}$$

Substituting this into (6.25) and using

$$\int_0^\infty x^n e^{-x} \, dx = n! \, , \tag{6.28}$$

we find

$$I_\lambda(0,\mu) = I_\lambda(0,1) \sum_{n=0}^{m} a_{\lambda n} n! \mu^n \, . \tag{6.29}$$

It turns out that a good description of the intensity can often be obtained by using only three terms ($m = 2$) in the power series. We then determine the coefficients $a_{\lambda 0}$, $a_{\lambda 1}$ and $a_{\lambda 2}$ from the measured intensity at different values of μ relative to the centre of the disc, $I_\lambda(0,\mu)/I_\lambda(0,1)$. This can be done using statistical techniques such as the *method of least squares*.

From equation (6.27) we see that in addition to the parameters $a_{\lambda 0}$, $a_{\lambda 1}$ and $a_{\lambda 2}$, we need the surface intensity at the centre of the sun's disc $I_\lambda(0,1)$, to determine $S_\lambda(\tau_\lambda)$. Some measured values for $I_\lambda(0,1)$ are shown in Table 6.1 (from [27] p. 233).

Table 6.1: Values of the sun's surface intensity at the centre of the disc $I_\lambda(0,1)$ for different wavelengths λ (from [27] p. 233).

λ (nm)	$I_\lambda(0,1)$ (W m^{-3} sr^{-1})
373.7	4.20×10^{13}
426.0	4.49×10^{13}
501.0	4.03×10^{13}
699.0	2.50×10^{13}

6.6 Determining the temperature as a function of depth

Once we have found the source function $S_\lambda(\tau_\lambda)$, we can use this to determine the temperature of the sun as a function of optical depth. To do this we need to introduce the concept of *local thermal equilibrium* (LTE).

Consider first a situation where a radiation field is in complete thermodynamic equilibrium with a blackbody of temperature T. The intensity is then given by the Planck function, $B_\lambda(T)$,

$$I_\lambda = B_\lambda(T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1} \, . \tag{6.30}$$

Furthermore, if we consider an element of matter in thermal equilibrium, then the intensity of the radiation passing through it does not change, since if it did, the element would either heat up or cool down. This means that in thermal equilibrium, $dI_\lambda/ds = 0$, and therefore from the radiative transfer equation (6.8) we have

$$I_\lambda = S_\lambda \ . \tag{6.31}$$

Both of these quantities are then equal to the Planck function $B_\lambda(T)$ given by equation (6.30).

The photosphere is not, however, in thermal equilibrium, since there is a temperature gradient of several thousand degrees over a distance of several hundred km and a net flow of radiation outwards. However we will assume that these changes are sufficiently gradual that there is a well-defined temperature at any given position. The assumption of *local thermal equilibrium* means that we can find a temperature at each depth such that

$$S_\lambda(\tau_\lambda) = B_\lambda(T) \tag{6.32}$$

holds. Further discussion on the validity of LTE can be found in the book by Böhm-Vitense [27].

Assuming LTE, i.e., setting $S_\lambda(\tau_\lambda) = B_\lambda(T)$, we can solve for the temperature to obtain

$$T(\tau_\lambda) = \frac{hc/k\lambda}{\ln\left(1 + \frac{2hc^2}{\lambda^5 S_\lambda(\tau_\lambda)}\right)} \ . \tag{6.33}$$

By substituting the source function derived from the measured intensity, we can determine the temperature of the sun as a function of the vertical optical depth.

6.7 Observation of solar limb darkening

Solar limb darkening can be observed and measured using the set-up in the RHUL observatory. This is done using a neutral density filter with a transmission of about 10^{-5} together with the LX200 telescope. The data are recorded with a CCD camera (currently we use the cheap QuickCam-based CCD).

The sun subtends an angle of about 0.5°, but the field of view obtained on the CCD chip is less than 0.1°, so it is not possible in a single frame to see the intensity variation from centre to limb. One solution would be to use a focal reducer to decrease the effective focal length and hence increase the field of view. The problem with the focal reducers is that the image brightness becomes significantly non-uniform, with decreased response near the edges. This is an effect called vignetting, and it would interfere with our attempt to measure the actual variation in the sun's intensity.

Rather than having to correct for vignetting, it is easier to work without the focal reducer, but to use only the central region of the CCD's image, e.g., the central 4×4 pixels. A measurement of the centre-to-limb variation can be then made by using the *drift scan* method.

First, we adjust the declination of the telescope to be equal to that of the sun. Turn off the telescope's tracking motors and then, without changing the telescope's declination, move the

telescope past the western limb of the sun and simply let it drift all the way through the field of view. We can set up the CCD camera to automatically take and record pictures at a set rate, say, one per second. Some of the images from such a drift scan are shown in Fig. 6.5.

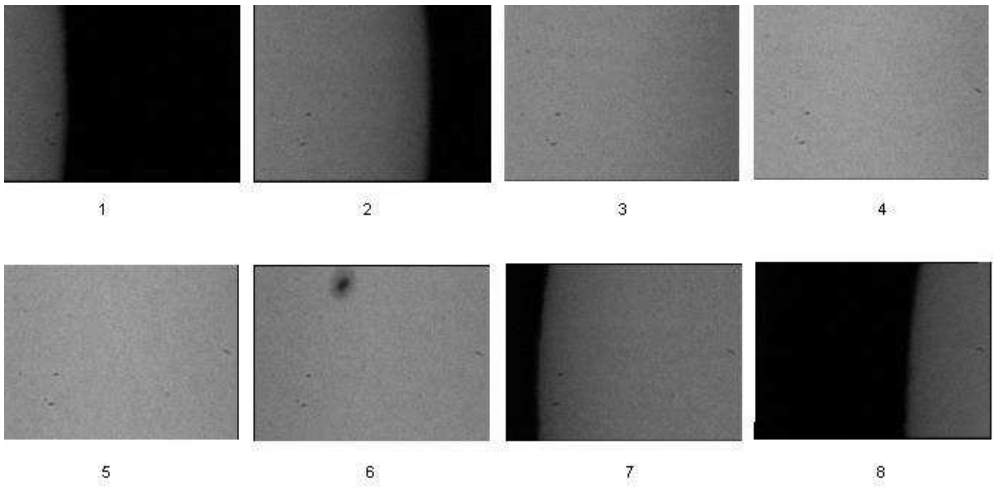


Figure 6.5: A sequence of images of the sun taken from a drift scan. Note the sunspot in frame 6.

As part of third-year projects we have had RHUL students develop software to analyze the images and measure the variation of intensity as a function of position on the disc. Figure 6.6 shows the intensity measured from a series of drift scans using B (blue), V (visual) and R (red) filters. These correspond to wavelengths of around 425, 525 and 650 nm, respectively. The horizontal axis represents r/R, i.e., the radius of the image divided by the radius of the sun's disc, with -1 corresponding to the western limb and $+1$ to the eastern one. The vertical axis shows the intensity normalized such that $I = 1$ at the centre of the disc.

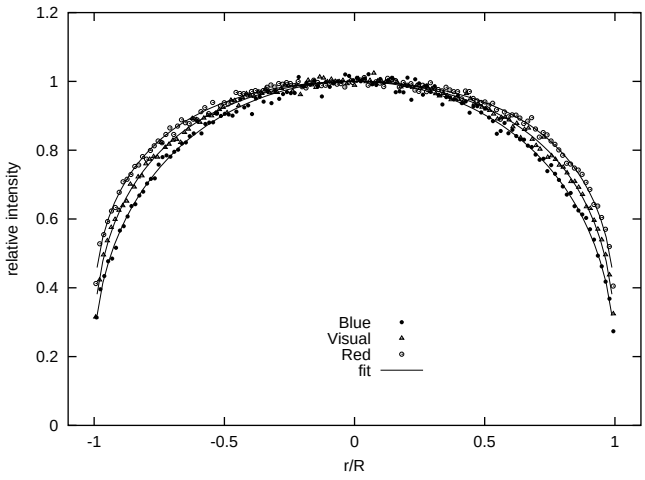


Figure 6.6: Normalized intensity versus r/R for solar drift scans (see text).

The curves in Fig. 6.6 have the functional form

$$f(\mu; \mathbf{a}) = a_0 + a_1\mu + 2a_2\mu^2 \tag{6.34}$$

where $\mu = \cos\theta$ is related to r/R by equation (6.26). This is simply the polynomial form from

equation (6.29) with $m = 2$. The values of the parameters a_0, a_1 and a_2 were determined separately for the B, V and R data sets using the method of least squares. The resulting values can then be plugged into equation (6.27) for the source function. The results are shown in Fig. 6.7.

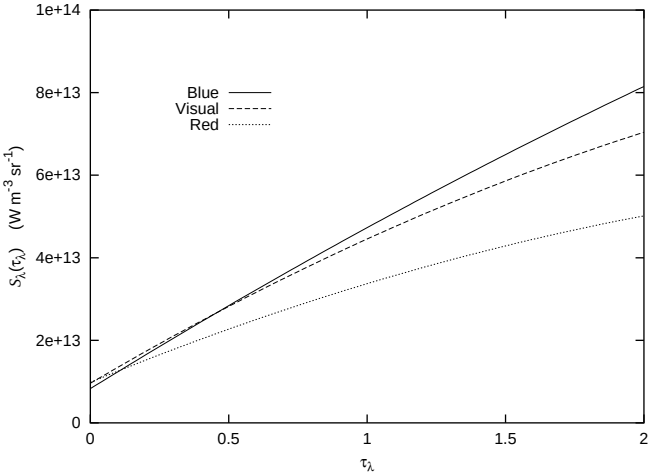


Figure 6.7: The source function $S_\lambda(\tau_\lambda)$ as a function of optical depth.

The source functions can then be used in equation (6.33) with their corresponding wavelengths to determine the temperature of the photosphere as a function of optical depth; this is shown in Fig. 6.8.

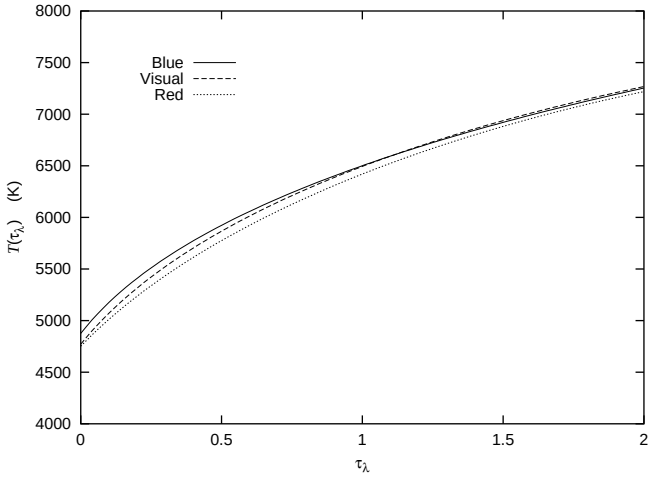


Figure 6.8: The temperature of the sun as a function of optical depth from limb darkening measurements using B, V and R filters.

One can define an effective temperature for the sun (or any star) through the relation

$$L = 4\pi R^2 \sigma T^4 \ , \tag{6.35}$$

where L is the total luminosity, R is the radius, and $\sigma = 5.670 \times 10^{-8}\,\mathrm{W\,K^{-4}m^{-2}}$ is the Stefan-Boltzmann constant. This is of course what one would obtain if the sun were a blackbody with a given radius and temperature. It can be shown that the radius and temperature in this relation correspond to the values in the photosphere at an optical depth of $\tau = 2/3$, which from the curves is seen to give $T \approx 5,800$ K.

The results of this type of measurement can be compared to the calculations of solar models, which is something you will consider in greater depth next term. In particular the temperature profile for different wavelengths provides information on the composition of the sun. It was through measurements of solar limb darkening that the H^- ion was determined to be the primary agent responsible for absorption in the solar photosphere.

Chapter 7

Colour and spectroscopy

In this chapter we will discuss spectroscopy, which answers the basic astronomical question "What colour is it?". We will of course do this in a quantitative way and so we will have to investigate in some detail how to measure the mixture of colours or *spectrum* of an astrophysical source. We will then discuss how the detailed properties of spectral features can provide information on the source's temperature, pressure, composition, the presence of magnetic fields, and so forth.

7.1 Photometry with filters

One way to determine how much of a given colour is emitted by an astrophysical source is to measure its flux through a set of filters. By tradition if the measurement is made with filters then it is referred to as *photometry* rather than spectroscopy, i.e., the emphasis is on learning how bright the source is in some number of wavelength bands. In this section we will take a brief look at the widely used UBVRI filters and we will examine some examples of the type of information one can obtain with them about astrophysical sources.

7.1.1 Filters

A filter is characterized by a certain transmission $T(\lambda)$, defined as the probability that a photon of wavelength λ will pass through. The transmission is in general characterized by a curve like the one shown in Fig. 7.1. It has a certain central wavelength λ_0 and a full width at half maximum (FWHM) $\Delta\lambda$, also called the *bandwidth*. Depending on the bandwidth $\Delta\lambda$, a filter may be called 'wide band' ($\Delta\lambda$ around 100 nm or more), 'intermediate band' ($\Delta\lambda \sim 10$ nm) or 'narrow band' ($\Delta\lambda \sim 1$ nm).

Standard filters have been developed for astronomy, with the Johnson UBVRI system being the most widely used. The letters stand for: ultraviolet, blue, visual, red and infrared. Here "visual" refers to roughly the middle of the optical part of the spectrum, i.e., green. Thus RVB is similar to the RGB (red, green blue) filters that one would use on, say, a colour television or digital camera, but in contrast, the astronomical filters have accurately defined transmisison characteristics. The transmission versus wavelength for a commercially available set of UBVRI filters is shown in Fig. 7.2. The central wavelengths and bandwidths are shown in Table 7.1.

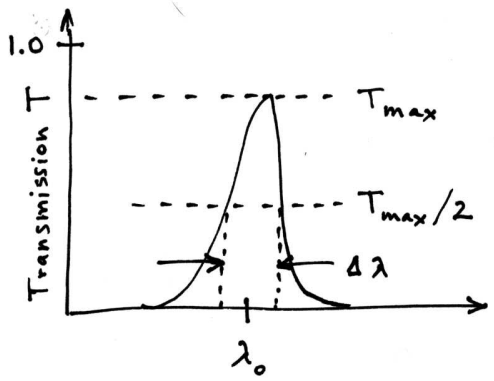


Figure 7.1: Transmission of a filter, illustrating the central wavelength λ_0 and bandwidth $\Delta\lambda$, given by the full width at half maximum.

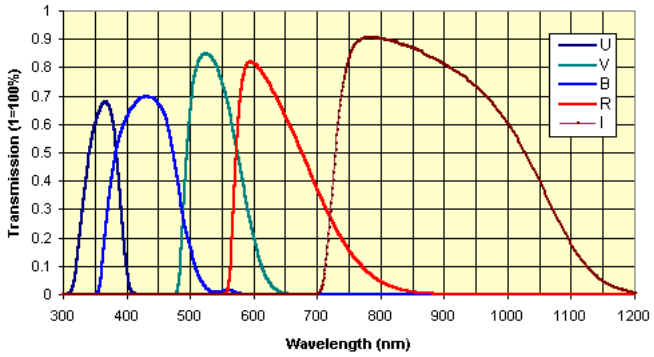


Figure 7.2: Transmission versus wavelength for a set of UBVRI filters available from the Santa Barbara Instruments Group [28].

Table 7.1: Central wavelength λ_0 and bandwidth $\Delta\lambda$ values for UBVRI filters.

Filter	λ_0 (nm)	$\Delta\lambda$ (nm)
U	365	70
B	440	100
V	550	90
R	650	150
I	790	150

7.1.2 Photometric magnitude

Let us first recall the magnitude system for quantifying the relative flux values F_1 and F_2 one measures from two sources. One defines the difference between the magnitudes m_1 and m_2 of the two sources by the relation

$$m_1 - m_2 = -2.5 \log_{10} \frac{F_1}{F_2} \, . \tag{7.1}$$

Implicit in the relation above is that the flux measurements are made with a device sensitive to a particular set of wavelengths. Usually we measure the flux using, say, one of the UBVRI filters, and then the magnitude is written with the corresponding subscript, e.g., m_B, m_V, etc., or sometimes one also abbreviates $m_B = B$, $m_V = V$, and so forth.

To define the numerical value of the apparent magnitude one chooses a standard star, and in effect the magnitude values one reports for all other stars are simply comparing their flux to this chosen standard. One conventional choice is the star Vega, and by definition one assigns Vega to have $m_U = m_B = m_V = m_R = m_I = 0$. Other stars can be used as well, and therefor one may encounter published magnitudes for Vega that are slightly different from zero, but the differences between different standards are relatively small and for purposes of this course can be ignored.

The initial definition of the UBVRI system was made with a particular combination of telescope, filter and instrument (detector), and this was used to determine the magnitude differences for a specific set of standard stars. A different telescope/filter/instrument combination can then be calibrated using these stars.

Notice that to define a magnitude in a certain passband, equation (7.1) needs to be used with two different sources, but with the *same* set of wavelengths, i.e., the same filter.

If the flux values included all of the light at all wavelengths, then this is what we would call the *bolometric* magnitude $m_{\rm bol}$. This can be estimated if we have a photon detector sensitive to a broad range of wavelengths (called a bolometer). If we define the bolometric correction as

$$\mathrm{BC} = m_{\rm bol} - m_V \, , \tag{7.2}$$

then we find that this is a relatively small correction that depends on the temperature of the star. For hot stars, say, $T = 9\,000$ K, the bolometric correction is -0.5 for a cooler star of $T = 4\,200$ K it is $+0.5$.

7.1.3 Colour indices

7.2 Spectrographs

general

7.2.1 Prisms

7.2.2 Gratings

Blazed

Echelle

7.3 General features of spectra

continuum

absorption lines

emission lines

7.4 Doppler shift and the Lyman-α forest

7.5 Spectral lineshapes

intro blurb

absorptance

equivalent width

7.5.1 Instrumental lineshape

7.5.2 Natural lineshape

7.5.3 Thermal broadening

7.5.4 Rotational broadening

7.5.5 Collisional (pressure) broadening

7.5.6 The Zeeman effect

7.5.7 The Voigt lineshape

Chapter 8

Measuring distance

In preparation.

Chapter 9

Extrasolar planets

In preparation.

Chapter 10

Introduction to the Solar System

In this chapter we give a brief overview of the solar system. Its cast of characters includes:

- the Sun;
- the terrestrial (rocky, or Earth-like) planets: Mercury, Venus, Earth, Mars;
- the Jovian (gas giant, or Jupiter-like) planets: Jupiter, Saturn, Uranus and Neptune;
- asteroids and comets;
- the planet Pluto and other Kuiper belt objects, the Oort cloud;
- moons.

In the remainder of the chapter we will look more closely at these ingredients. We will not, however, treat in detail its most important player, the Sun, and we will only touch briefly on planetary geology, as both of these topics are treated in separate courses.

10.1 Planets and their orbits

Table 10.1 gives some basic information on the planets of the solar system. With one exception (Pluto) they divide neatly into two main categories: terrestrial (earth-like) and Jovian (or gas giant). The terrestrial planets are all relatively close to the Sun, and are mainly composed of heavy elements, e.g., Fe, Si, O, Mg, Al. The Jovian planets are further away from the Sun and are much larger than their terrestial counterparts. For example, Jupiter is roughly one eleventh the size of the Sun or around 10 times larger than the Earth. The gas giants are mostly composed of hydrogen and helium. The exception is Pluto, which is smaller than any of the terrestrial planets and yet further from the Sun than the gas giants.

Some of the values from Table 10.1, particularly the distances, are perhaps counter to our intuition. For example, suppose the Sun, with its diameter of 1 400 000 km were the size of a beach ball 60 cm in diameter. On this scale, the Earth would be roughly the size of a pea, orbiting at a distance of 64 m. Jupiter would be the size of a tennis ball, some 360 m away from

Table 10.1: Properties of planets and their orbits (from [29]). (Note Pluto has been reclassified as a *dwarf planet*; see Section 10.5.)

Planet	Diameter (km)	a (AU)	Orbital eccentricity	Inclination of orbital plane (deg.)	Obliquity (tilt of axis, deg.)
Mercury	4 879	0.387	0.206	7.0	0.0
Venus	12 104	0.723	0.007	3.39	178
Earth	12 756	1.0	0.017	0	23.45
Mars	6 787	1.524	0.093	1.85	25
Jupiter	142 800	5.203	0.048	1.3	3.08
Saturn	120 660	9.537	0.056	2.49	26.7
Uranus	51 118	19.19	0.047	0.77	97.9
Neptune	49 528	30.07	0.009	1.77	29.6
Pluto	2 300	39.48	0.248	17.15	122.5

the Sun. The nearest star is Proxima Centauri at a distance of 1.3 pc (1 pc = 206 265 AU). In the imaginary world with the Sun as a beach ball, Proxima Centauri would be sitting at a distance of 35 000 km.

From Table 10.1 we can see that most of the planets have orbits that lie within a few degrees of the ecliptic, with only Pluto having a relatively large inclination angle of 17°. All of the planets orbit in the same direction, and most of the planets revolve about their own axes also in the same direction. Here the exceptions are Uranus and Pluto. Uranus's axis of rotation almost lies in the plane of its orbit. It has been proposed that Uranus suffered a major collision early in the history of the solar system that caused its rotational axis to tip over by more than 90°. Pluto's axis has a tilt angle of 122.5°, which also is presumably related to interactions with other bodies during the formation of the solar system.

10.2 Asteroids and the Titius-Bode Law

From Table 10.1 we also see that the sizes of the orbits show a curious regularity. The semi-major axis values, a, can be approximately described by the forumla

$$a = [0.4 + 0.3 \times 2^n] \text{ AU, with } n = -\infty, 0, 1, 2, 3, \ldots \ . \tag{10.1}$$

This emperical rule is called the Titius-Bode Law, and was found independently by Titius and Bode in the 18th century to describe the six planets known at the time (out to Uranus). Table 10.2 shows the predictions of this rule along with the observed orbital sizes. The agreement is seen to be relatively good, at least out to Uranus, with one important exception: for $n = 3$ one predicts an object orbiting at $a = 2.8$ AU, and this did not appear to correspond to any known planet.

In 1801 Giuseppe Piazzi discovered an object orbiting the Sun at a distance very close to 2.8 AU, and this was interpreted by Bode as the missing planet. Piazzi's planet was actually the

Table 10.2: Values of the semi-major axis of planets as observed and as predicted by the Titius-Bode Law.

Planet	n	a predicted (AU)	a observed (AU)
Mercury	$-\infty$	0.4	0.39
Venus	0	0.7	0.72
Earth	1	1.0	1.0
Mars	2	1.6	1.52
Ceres	3	2.8	2.77
Jupiter	4	5.2	5.20
Saturn	5	10.0	9.54
Uranus	6	19.6	19.2
Neptune	7	38.8	30.07
Pluto	8	77.2	39.48

asteroid Ceres, one of thousands of rocky objects found in the *asteroid belt* between the orbits of Mars and Jupiter. Asteroids, also called 'minor planets', are small rocky bodies that are left over from the formation of the solar system. They are prevented from fusing together by the gravitational perturbations caused by Jupiter, which is by far the most massive planet in the solar system. Ceres is the largest asteroid, having a diameter of 940 km and containing a third of the mass of the entire asteroid belt. Some properties of some of the larger asteroids are given in Table 10.3.

Table 10.3: Properties of some of the asteroids [30].

Asteroid	Size (km)	Semimajor axis (AU)	Orbital eccentricity
Ceres	960×932	2.767	0.0789
Pallas	$570 \times 525 \times 482$	2.774	0.2299
Juno	240	2.669	0.2579
Vesta	530	2.362	0.0895
Eugenia	226	2.721	0.0831
Siwa	103	2.734	0.2157
Kleopatra	217×94	2.793	0.2535
Ida	58×23	2.861	0.0451
Mathilde	$66 \times 48 \times 46$	2.646	0.2660
Eros	$33 \times 13 \times 13$	1.458	0.2229
Gaspra	$19 \times 12 \times 11$	2.209	0.1738

Although the discovery of Ceres appeared to be a stunning victory for the Titius-Bode Law, its success is to a large extent regarded as a coincidence. Simulations of the formation of a planetary system from the collapse of a gas cloud can indeed result in regularities in the spacing

of the orbits, but these turn out to be difficult to predict in advance and are highly sensitive to the initial conditions. This is typical of complex systems governed by nonlinear force laws such as gravity.

Asteroids larger than around 500 km diameter contain enough mass to be pulled into a roughly spherical shape. Smaller asteroids such as Kleopatra, Ida, etc., are often irregularly shaped boulders. Figure 10.1 shows images of the asteroids Ida and Eros.

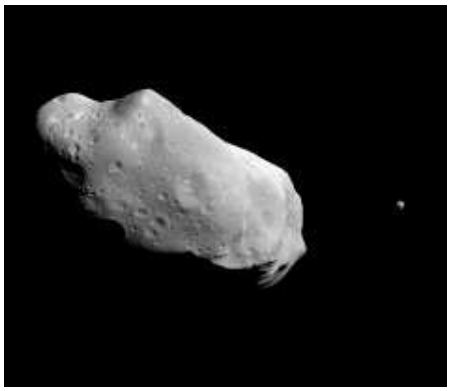
(a)  (b) 

Figure 10.1: (a) Image of the asteroid Ida and its moon Dactyl taken by the Galileo spacecraft in 1993. [31]. (b) Image of the asteroid Eros taken by the NEAR spacecraft in 2001 [32]. NEAR proceeded to land on the surface of Eros returning images from as low as 120 m [33].

10.3 Comets and trans-Neptunian objects

Comets are small bodies composed of dust and ices (mainly water, but also C0 and $C0_2$) that reside in the outer solar system. According to Kepler's laws, bodies orbiting at a very large radius move very slowly, and thus it does not take a very large perturbation to bring their angular motion to a halt. If this happens, they then fall in toward the inner solar system, often falling into the Sun but also possibly passing around it in a highly eccentric orbit. As a comet approaches the Sun its ices sublimate giving its nucleus an atmosphere of gas and dust called the *coma*.

The coma is swept by the Sun's light into a visible *tail*, which can be up to an AU in length. Often comets have two distinct tails, one from ionized gas and one from dust. The ion tail extends opposite the direction of the Sun and often appears blue, since the most commonly found ion, $C0^+$, preferentially scatters blue light. The dust tail also extends away from the Sun as dust particles are pushed away from the Sun's radiation pressure. As the comet swings around the Sun, the particles find themselves in orbits slightly further away, and therefore with slower orbital speeds. The dust tail is therefore bent into a curved shape that lags behind the gas tail. Figure 10.2(a) shows the comet Hale-Bopp, which was clearly visible in the spring of 1997. The lower tail is from ionized gas, and the upper one from dust.

Comets can be classified as either *short-period* or *long-period*. The long-period comets (period greater than 200 years) enter the solar system isotropically and have orbits that originate around 50 000 AU from the Sun. It has been postulated that at this distance there is a spherical

(a)  (b) 

Figure 10.2: (a) Image of Comet Hale-Bopp, showing the ion and dust tails [34]. In the spring of 1997 this comet was clearly visible to the unaided eye. (b) Nucleus of Comet Halley, approximately 15 km long by 8 km wide, imaged in 1986 by the Giotto spacecraft [35].

distribution of cometary nuclei called the *Oort Cloud*. The gravitational perturbation that causes an Oort Cloud object to fall in toward the Sun could come from a nearby star.

The short-period comets (period less than around 200 years) come from a reservoir called the *Kuiper Belt*, which starts roughly at the orbit of Neptune (30 AU) and extends out to 50 AU or more. The Kuiper Belt lies in the ecliptic plane and contains more than 70 000 objects with diameters larger than 100 km. The short-period comets therefore also have orbits close to the ecliptic and move around the Sun in the same direction do the rest of the planets. Comet Halley, having a period of 76 years, falls into this category.

Cometary nuclei are believed to be objects left over from the formation of the solar system. The Kuiper Belt objects were never accreted into planets, but their orbits still lie roughly in the plane of the solar system. Any such objects that were at smaller radii near the orbits of Neptune and Uranus would have undergone close encounters with these planets whereby they were ejected out to much further distances and in random directions; these are the present Oort Cloud objects. Cometary nuclei that had close encounters with Jupiter or Saturn would most likely be ejected from the solar system entirely.

The planet Pluto has an orbital semi-major axis of 39.5 AU. Although its diameter of 1200 km perhaps gives it a certain similarity to the terrestrial planets, its orbit puts it in the middle of the Kuiper Belt. One therefore expects Pluto's origin to be more directly connected to that of comets than to the inner planets. It is perhaps best classified as a large Kuiper Belt object.

Recently a 10th 'planet' has been discovered at an orbital radius of 97 AU, placing it in the outer reaches of the Kuiper Belt [36]. From its apparent brightness we can infer that it must be at least as large as Pluto. The object's provisional name is 2003 UB 313, based on the date of the first recorded images. Three of these images, taken 90 minutes apart, are shown in Fig. 10.3. The planet, indicated by the circle, is clearly moving relative to the background stars. Further images have even shown that this body is orbited by a smaller moon.

10.4 Moons

Many of the planets in the solar system have in orbit around them smaller bodies or moons. Mercury and Venus do not have moons, the Earth of course has a relatively large moon, (usually

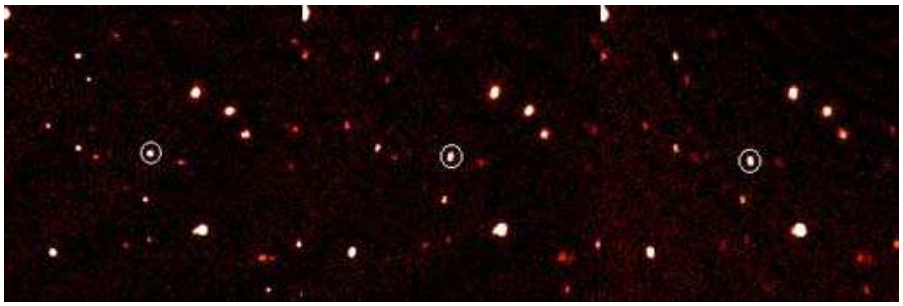


Figure 10.3: The tenth planet, 2003 UB 313, now renamed the dwarf planet Eris [36].

written with a capital letter, the *Moon*). Mars has two small moons ($\sim$ 10 km diameter), Phobos and Deimos, which may be captured asteroids. All of the Gas Giants have many moons. Jupiter, for example, has more than 60, the largest four of which are called the Galilean moons, having been seen by Galileo in 1610. His discovery counts as one of the earliest to be made with a telescope and revealed the first example of celestial motion not apparently centred about the Earth [37]. Figure 10.4 shows the four Galilean moons: Io, Europa, Ganymede and Callisto.

Figure 10.4: A composite image of Jupiter and its four largest moons, from top to bottom (also in order of increasing distance from Jupiter): Io, Europa, Ganymede and Callisto [38].

The moons of the Jovian planets have sizes typically smaller than but roughly the same order of magnitude as the terrestrial planets. The largest moon, Ganymede has a diameter of 5 262 km and is actually larger than Mercury or Pluto. The moons of the Jovian planets differ significantly from the inner planets in composition, however, with significantly greater quantities of ices (water, $C0_2$, amonia, methane).

10.5 2006 update: Pluto demoted to dwarf

In August 2006 the International Astronomical Union voted at its 26th General Assembly Meeting in Prague on a new classification of the members of the solar system. The outcome was the creation of a new type of celestial body called *dwarf planet*. Pluto was reclassified as

a dwarf planet along with the asteroid Ceres and the previously mentioned '10th planet', 2003 UB 313. Several weeks later, 2003 UB 313 was renamed (136199) Eris (pronounced ee′-ris, the Greek goddess of discord and strife), or Eris for short. Its moon, formerly known as S/2005 (2003 UB 313) 1 and now known technically as (136199) Eris I, has been named Dysnomia (the daughter of Eris).

The new classification scheme defines a planet to be a body that orbits the Sun and is massive enough so that its gravity pulls it into an approximately spherical shape. Furthermore it must be massive enough to have ejected or accreted other objects in its path. To be a dwarf planet, it is sufficient to orbit the Sun and be roughly round; it is not required that the body clear away other nearby objects. Both of these categories exclude moons, which are bound to planets or dwarf planets that are themselves orbiting around the Sun.

Chapter 11

Origin of the Solar System

In this chapter we examine the origin of the solar system. A successful model for solar system formation must account for a large number of observed features, such as the nature of planetary orbits, the age of the Sun and the planets and their composition. Furthermore we should be able to apply this model to other stars and planetary systems currently being born.

The current view is that the solar system originated from the collapse of a large gas cloud some 4.6 billion years ago. This *nebular hypothesis* was first put forward by Kant in 1755 and later developed by Laplace and others.

Other models that have since been excluded include so-called catastrophe theories, which were widely discussed in the 19th century. In these models the initially isolated Sun had a close encounter with another star. Material from this star would be pulled into orbit around the Sun and then condense into planets. But given the observed number density of stars in the galaxy, the probability of such a close encounter is very low. Furthermore one would not expect such an event to leave enough matter orbiting around the Sun to form the planets we observe today.

Another possibility is that the Sun and planets formed separately but that the planets at some point wandered close to the Sun and were captured. An important difficulty with this theory is that it does not explain why the planets are all found orbiting in roughly the same plane, and all in the same direction.

We will not go into the details of any of the alternative models here but rather will concentrate on the nebular hypothesis.

11.1 Overview of solar system formation

Below we outline some of the main events in the formation of the solar system.

- A cold gas cloud is perturbed, perhaps by a nearby supernova explosion, and collapses under its own gravity.
- As the cloud collapses it heats up, and energy is radiated away as light. The loss of energy together with conservation of angular momentum force the material of the solar system into a rotating disc. The time scale for the initial collapse is roughly 10^6 years.

- The gas cools enough for condensation of metals, dust particles.
- Dust particles collide, form larger particles up to boulder size.
- Some boulders grow large enough for their own gravity to play a significant role in attracting more material. This accelerates their growth (runaway growth) and planetesimals are formed.
- At around 10^7 years after the initial collapse, the sun ignites and the solar wind expels the remaining gas.
- Planetesimals collide and coalesce, forming protoplanets.
- Some 10^8 years after the start the solar system has roughly its present form. The planets often carry a characteristic of their last large collision.
- The accretion process continues as asteroids and comets collide with planets leaving impact craters. This was particularly intense for the first 500 million years, called the heavy bombardment era, but asteroid impacts continue today.

We will now look at these steps in greater detail starting with the nature of the gas clouds from which the solar system is thought to have formed.

11.2 The solar nebula

We see many thousands of clouds of gas and dust in the galaxy showing up as dark patches blocking light from a field of more distant stars. An example is the molecular cloud Barnard 68, shown in Fig. 11.1(a). The temperature of the clouds can be estimated from the thermal broadening of spectral lines. Typical values are in the range from 10 to 50 K, mostly consisting of molecular hydrogen (H_2), and a typical size is on the order of 10^5 AU. The density of the gas can be estimated from the strength of the cloud's radio emission, with typical values of 10^{10} molecules per m^3.

One of the strongest pieces of evidence supporting the nebular hypothesis is the observation of stars currently being formed from the collapse of gas clouds. An example is the Eagle Nebula, part of the open cluster M16, shown in Fig. 11.1(b). The long pillars are clouds of gas and dust, and the bright spots at the tips are newly formed stars.

Although the overall composition of our solar system is dominated by hydrogen and helium, it contains significant amounts of heavier elements, such as silicon, iron, etc. In the first several minutes after the Big Bang, however, only hydrogen and helium were formed, with very small quantities of slightly heavier elements like lithium (see e.g. [41]). Essentially all of the heavier elements formed much later in the interiors of stars. This means that the solar nebula must have contained products from earlier generations of stars.

We will now discuss the conditions under which a cloud is expected to collapse. We will use an important result called the *virial theorem*, which relates the time average kinetic and potential energies, $\langle K \rangle$ and $\langle U \rangle$, of a gravitationally bound system in equilibrium. The theorem states that for such a system,

(a) (b)

Figure 11.1: (a) The molecular cloud Banard 68 [39]. (b) Star formation in the Eagle nebula [40].

$$\langle K \rangle = -\frac{1}{2}\langle U \rangle \ . \tag{11.1}$$

A derivation of the virial theorem is given in Appendix B. Here we are using the convention that the potential energy of a system is zero when its constituents are separated to infinite distance, so for a bound system the potential energy is negative.

For a cloud of a given mass M, temperature T, and composition we can work out the kinetic and potential energies, K and U. Here the time averaging is not important since there are so many particles in the system that both K and U are individually very close to constant. For the cloud to be gravitationally stable, equation (11.1) must hold. Ignoring the time averaging means we will have $K = -\frac{1}{2}U$. If we find that the kinetic energy is smaller than $-\frac{1}{2}U$, then this means the cloud cannot be in equilibrium, but rather it will collapse to a smaller size.

Consider a spherical cloud of radius R and constant density ρ. Of course a real cloud will not have a uniform density but this model will nevertheless reveal the main issues involved. The mass of the cloud is

$$M = \frac{4}{3}\pi R^3 \rho \ . \tag{11.2}$$

From dimensional arguments we know that the gravitational potential energy of the cloud must be $U \sim -GM^2/R$. The exact value will depend on how the mass is distributed inside the volume. If we assume a spherical cloud with uniform density we can show

$$U = -\frac{3}{5}\frac{GM^2}{R} \ . \tag{11.3}$$

Now to find the kinetic energy for a cloud of a given composition and temperature we use the equipartition theorem. If we ignore rotational and vibrational motion of the molecules we have

$$K = \frac{3}{2} kTN \ , \tag{11.4}$$

where k is Boltzmann's constant and N is the number of molecules. This can be written

$$N = \frac{M}{m_{\rm molecule}} = \frac{M}{\mu m_{\rm H}} \ , \tag{11.5}$$

where here M is the mass of the cloud and the average molecular mass $m_{\rm molecule}$ can be expressed as the mean molecular weight μ times the mass of a hydrogen atom $m_{\rm H}$.

Substituting these ingredients into the virial theorem, equation (11.1), we find

$$\frac{3}{2} \frac{MkT}{\mu m_{\rm H}} = \frac{3}{10} \frac{GM^2}{R} \ . \tag{11.6}$$

The kinetic term on the left is proportional to $M \propto R^3$, while the potential term on the right goes as $M^2/R \propto R^5$. Thus for a cloud of a given density and a sufficiently large size, the potential term will dominate and the cloud will collapse. The radius at which the equilibrium condition (11.6) just holds is called the *Jeans radius* $R_{\rm J}$. Solving equation (11.6) for the radius and using also (11.2) for the mass we find

$$R_{\rm J} = \sqrt{\frac{15kT}{4\pi G \mu m_{\rm H} \rho}} \ . \tag{11.7}$$

Equivalently we can use equation (11.2) to solve for the mass at which the equilibrium condition holds, which gives the *Jeans mass*,

$$M_{\rm J} = \sqrt{\frac{3}{4\pi\rho}} \left(\frac{5kT}{\mu m_{\rm H} G} \right)^{3/2} \ . \tag{11.8}$$

By making different assumptions about the initial shape and mass distribution of the cloud we would find expressions for $R_{\rm J}$ and $M_{\rm J}$ similar to these but which could contain slightly different numerical factors.

We can now plug in some numbers and estimate the Jeans radius for a typical gas cloud. Let us take a temperature of $T = 20\,{\rm K}$ and a number density of $n = 10^{10}$ molecules per $\rm m^3$. For molecular hydrogen we have a molecular weight of $\mu = 2$, so the density is

$$\rho = n\mu m_{\rm H} = 10^{10}\,{\rm m}^{-3} \ \times 2 \times 1.67 \times 10^{-27}\,{\rm kg} = 3.3 \times 10^{-17}\,{\rm kg/m}^3 \ . \tag{11.9}$$

Using this in equation (11.7), we find

$$R_{\rm J} = 6.7 \times 10^{15}\,{\rm m} = 0.22\,{\rm pc} = 4.5 \times 10^4\,{\rm AU} \ , \tag{11.10}$$

and from (11.8) we have the corresponding Jeans Mass,

$$M_{\rm J} = 4 \times 10^{31}\,{\rm kg} \approx 20 M_{\rm sun} \ . \tag{11.11}$$

Only some fraction of the cloud's mass will wind up in the star, so $20M_{\rm sun}$ for the initial mass of a collapsing gas cloud is not unreasonable as an order of magnitude estimate.

The value of $R_{\rm J}$ from (11.10) also seems reasonable in that we see many thousands of clouds with sizes in this range in our galaxy. This then presents a problem in that one would naively expect them to collapse in a time of roughly 10^6 years (see Section 11.3 below). Given that our galaxy formed more than 10^{10} years ago it seems that these clouds should have collapsed long ago. It is generally assumed that other important effects come into play, in particular that of magnetic fields. These can act to stabilize a cloud that otherwise would collapse under its own gravity.

In many galaxies, however, we find a very high rate of formation of new stars, perhaps 10^6 stars created in a timescale of 10^7 to 10^8 years. This rate is so high that it could not possibly be sustained throughout the galaxy's lifetime and therefore it is assumed to be a transient phenomenon. This is called a *starburst galaxy*, where the collapse of a very large region of gas is triggered by an encounter with another galaxy.

So although we have a partially coherent picture of how a gas cloud can begin to collapse, the details remain a bit murky. The situation improves when we refine the model and include magnetic effects but still it is difficult to make accurate predictions for the rate of star formation. The Jeans radius (or mass) nevertheless remains the fundamental quantity to consider, even though the naive calculation above fails to explain the long lifetimes of large, cold clouds. These quantities are also used when we try to understand the collapse of the gas in the early universe into galaxies (see, for example, [42]).

11.3 Contraction of the solar nebula

If the gas cloud were to consist of noninteracting particles with zero initial kinetic energy, they would all simply fall toward the centre. The time that it takes to do so is called the *free-fall time*, $t_{\rm ff}$. A particle at the outer edge of the cloud at a radius R would fall as if all of the rest of the cloud's mass M were concentrated at the centre. The trajectory would be the limiting case of the elliptical Kepler orbit, here with zero angular momentum. The distance to the centre, R, is simply twice the semi-major axis a of an ellipse with an eccentricity of unity and the centre-of-mass (the eventual star) at one focus. The time to cover this distance is one half of the period, P. Kepler's third law tells us $P^2 = \frac{4\pi^2}{GM}a^3$, and we can also use $M = (4/3)\pi R^3\rho$, where here R and ρ are the initial radius and density. We therefore have for the free-fall time

$$t_{\rm ff} = \frac{P}{2} = \frac{1}{2}\cdot\frac{2\pi}{\sqrt{GM}}a^{3/2} = \frac{\pi}{\sqrt{GM}}\left(\frac{R}{2}\right)^{3/2} = \left(\frac{3\pi}{32G\rho}\right)^{1/2} . \tag{11.12}$$

Notice that for a given density, the free-fall time is independent of the mass or radius of the cloud. Substituting numbers from the example above gives $t_{\rm ff} \approx 400\,000$ years.

The free-fall time characterizes the *dynamical timescale*, the timescale for motion in the absence of pressure. But we know the interior temperature and pressure must eventually rise, and this slows the time for collapse. Initially, dust particles collide and radiate in the infrared, and for sufficiently low densities this radiation can escape and the cloud remains cool. Eventually, however, the optical thickness of the cloud exceeds unity and the radiation cannot escape without further interaction, and thus the cloud heats up.

After roughly one million years, the solar nebula had collapsed to protostar. The core temperature was still too low to allow nuclear burning, and the energy source was still based on the conversion of gravitational potential radiation into thermal energy, which is then radiated away as light. This light and the solar wind expelled gas and dust from the surrounding region and the Sun became optically visible. This is what is called the T-Tauri phase of stellar formation.

If gravitational energy were to be the only source of light from the sun, then the timescale on which the Sun would radiate away all of the initial potential energy, called the *Kelvin-Helmholtz timescale*, would be

$$t_{\rm KH} \approx \frac{U}{L} \,, \tag{11.13}$$

where U is the available potential energy and L is the average luminosity, i.e., the power output of the Sun. If we plug in numbers based on the current size and luminosity of the Sun, we find $t_{\rm KH} \approx 3\times10^7$ years. When this was first worked out by Kelvin in the 19th century, however, there was already geological evidence that showed the age of the earth to be significantly older than this. So one was confronted with a sort of 'age crisis', whose resolution required an understanding of nuclear physics that came only many decades later.

Eventually the temperature in the core exceeded the nuclear ignition temperature of around 10^7 K and nuclear fusion took over as the energy source for the Sun's light. The timescale for our Sun to reach this stage was on the order of 10^7 years. The timescale for the Sun to exhaust its supply of hydrogen through nuclear burning is more like 10^{10} years, which is compatible with its currently estimated age of 4.6×10^9 years.

11.4 Disc formation

During the collapse of the molecular cloud the solar system flattened to a disc. Initially the cloud will have had some small rotational motion, e.g., from the turbulence of the interstellar medium. At the very least some rotation of the gas cloud would arise from the fact that the parts of the cloud closer to the galaxy's centre have a shorter period of galactic rotation. As the cloud contracts, angular momentum is conserved and so therefore the rate of angular rotation must increase. If particles move towards the midplane (perpendicular to the cloud's axis of rotation), then they can collide inelastically whereby kinetic energy is converted to heat and eventually released as blackbody radiation. Such a process, shown in Fig. 11.2(a), involves no change in the total angular momentum of the system.

Let us now view the gas cloud along its axis of rotation as in Fig. 11.2(b). If two particles collide inelastically, kinetic energy is converted to heat, but the total angular momentum must be conserved. For this to happen the particles will move into orbits with the largest angular momentum for a given kinetic energy — this is a circular orbit. So the net result of angular momentum conservation and inelastic collisions is to move the particles into circular orbits lying in a central plane.

Notice that if there was no energy loss mechanism, then the flattening would not happen. If the collision in Fig. 11.2(a) were to be elastic, then the outgoing momentum components along the axis of rotation would be equal to those incoming, and the cloud would remain spherical.

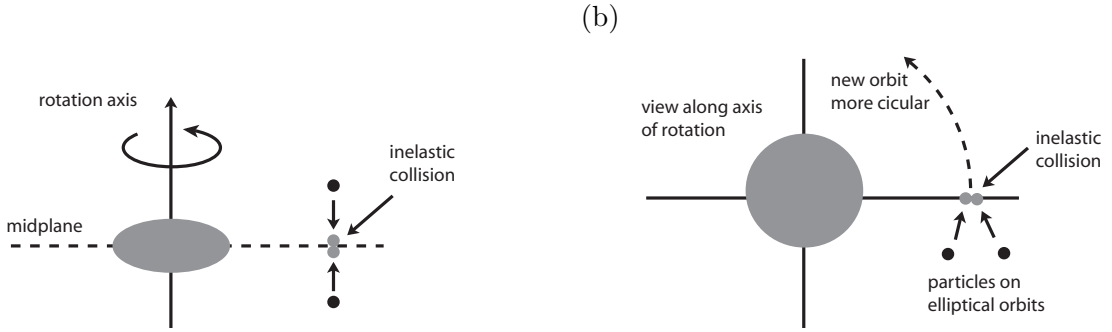


Figure 11.2: Two views of the contracting solar nebula, (a) side view and (b) top view, along the rotational axis. (See text.)

This is similar to what 'dark matter' particles (called WIMPS) are believed to do in galaxy formation.

11.5 Formation of planets

In preparation

Chapter 12

Further topics in planetary science

In preparation

12.1 Planetary atmospheres

12.2 Tidal forces in the solar system

12.3 Planetary geology

Chapter 13

Introduction to radio astronomy

In preparation

Appendix A

The Poisson distribution

Here we will derive the functional form of the Poisson distribution and we will investigates some of its properties. Consider a time t in which some number n of events may occur. Examples are the number of photons collected by a telescope or the number of decays of a large sample of radioactive nuclei. Suppose that the events are *independent*, i.e., the occurrence of one event has no influence on the probability for the occurrence of another. Furthermore, suppose that the probability of a single event in any short time interval δt is

$$P(1;\delta t) = \lambda\,\delta t\;, \tag{A.1}$$

where λ is a constant. In Section A.1 we will show that the probability for n events in the time t is given by

$$P(n;\nu) = \frac{\nu^n}{n!}e^{-\nu}\;, \tag{A.2}$$

where the parameter ν is related to λ by

$$\nu = \lambda t\,. \tag{A.3}$$

We will follow the convention that arguments in a probability distribution to the left of the semi-colon are random variables, that is, outcomes of a repeatable experiment, such as the number of events n. Arguments to the right of the semi-colon are parameters, i.e., constants.

The Poisson distribution is shown in Fig. A.1 for several values of the parameter ν. In Section A.2 we will show that the mean value $\langle n\rangle$ of the Poisson distribution is given by

$$\langle n\rangle = \nu\;, \tag{A.4}$$

and that the standard deviation σ is

$$\sigma = \sqrt{\nu}\;. \tag{A.5}$$

The mean ν roughly indicates the central region of the distribution, but this is not the same as the most probable value of n. Indeed n is an integer but ν in general is not. The standard deviation is a measure of the width of the distribution.

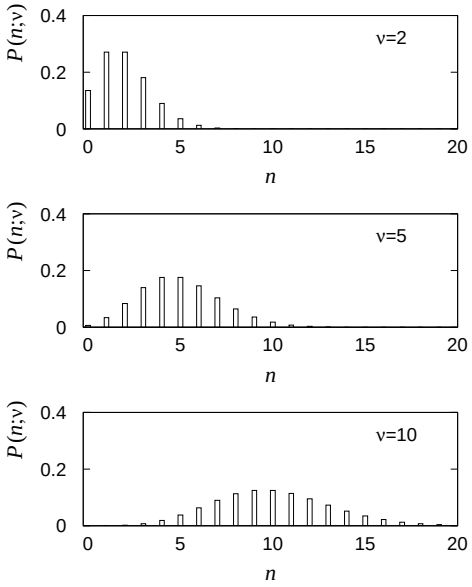


Figure A.1: The Poisson distribution $P(n;\nu)$ for several values of the mean ν.

A.1 Derivation of the Poisson distribution

Consider the time interval t broken into small subintervals of length δt. If δt is sufficiently short then we can neglect the probability that two events will occur in it. We will find one event with probability

$$P(1;\delta t) = \lambda\,\delta t \tag{A.6}$$

and no events with probability

$$P(0;\delta t) = 1 - \lambda\,\delta t\ . \tag{A.7}$$

What we want to find is the probability to find n events in t. We can start by finding the probability to find zero events in t, $P(0;t)$ and then generalize this result by induction.

Suppose we knew $P(0;t)$. We could then ask what is the probability to find no events in the time $t+\delta t$. Since the events are independent, the probability for no events in both intervals, first none in t and then none in δt, is given by the product of the two individual probabilities. That is,

$$P(0;t+\delta t) = P(0;t)(1-\lambda\,\delta t)\ . \tag{A.8}$$

This can be rewritten as

$$\frac{P(0;t+\delta t)-P(0;t)}{\delta t} = -\lambda P(0;t) \ , \tag{A.9}$$

which in the limit of small δt becomes a differential equation,

$$\frac{dP(0;t)}{dt} = -\lambda P(0;t) \ . \tag{A.10}$$

Integrating to find the solution gives

$$P(0;t) = e^{-\lambda t} + C \ . \tag{A.11}$$

For a length of time $t = 0$ we must have zero events, i.e., we require the boundary condition $P(0;0) = 1$. The constant C must therefore be zero and we obtain

$$P(0;t) = e^{-\lambda t} \ . \tag{A.12}$$

Now consider the case where the number of events n is not zero. The probability of finding n events in a time $t + \delta t$ is given by the sum of two terms:

$$P(n;t+\delta t) = P(n;t)(1-\lambda\,\delta t) + P(n-1;t)\lambda\,\delta t \ . \tag{A.13}$$

The first term gives the probability to have all n events in the first subinterval of time t and then no events in the final δt. The second term corresponds to having $n-1$ events in t followed by one event in the last δt. In the limit of small δt this gives a differential equation for $P(n;t)$:

$$\frac{dP(n;t)}{dt} + \lambda P(n;t) = \lambda P(n-1;t) \ . \tag{A.14}$$

We can solve equation (A.14) by finding an integrating factor $\mu(t)$, i.e., a function which when multiplied by the left-hand side of the equation results in a total derivative with respect to t. That is, we want a function $\mu(t)$ such that

$$\mu(t)\left[\frac{dP(n;t)}{dt} + \lambda P(n;t)\right] = \frac{d}{dt}\left[\mu(t)P(n;t)\right] \ . \tag{A.15}$$

We can easily show that the function

$$\mu(t) = e^{\lambda t} \tag{A.16}$$

has the desired property and therefore we find

$$\frac{d}{dt}\left[e^{\lambda t}P(n;t)\right] = e^{\lambda t}\lambda P(n-1;t) \ . \tag{A.17}$$

We can use this result, for example, with $n = 1$ to find

$$\frac{d}{dt}\left[e^{\lambda t}P(1;t)\right] = \lambda e^{\lambda t}P(0;t) = \lambda e^{\lambda t}e^{-\lambda t} = \lambda \,, \tag{A.18}$$

where we substituted our previous result (A.12) for $P(0;t)$. Integrating equation (A.18) gives

$$e^{\lambda t}P(1;t) = \int \lambda \, dt = \lambda t + C \,. \tag{A.19}$$

Now the probability to find one event in zero time must be zero, i.e., $P(1;0) = 0$ and therefore $C = 0$, so we find

$$P(1;t) = \lambda t e^{-\lambda t} \,. \tag{A.20}$$

We can generalize this result to arbitrary n by induction. We assert that the probability to find n events in a time t is

$$P(n;t) = \frac{(\lambda t)^n}{n!}e^{-\lambda t} \,. \tag{A.21}$$

We have already shown that this is true for $n = 0$ as well as for $n = 1$. Using the differential equation (A.17) with $n + 1$ on the left-hand side and substituting (A.21) on the right, we find

$$\frac{d}{dt}\left[e^{\lambda t}P(n+1;t)\right] = e^{\lambda t}\lambda P(n;t) = e^{\lambda t}\lambda\frac{(\lambda t)^n}{n!}e^{-\lambda t} = \lambda\frac{(\lambda t)^n}{n!} \,. \tag{A.22}$$

Integrating equation (A.22) gives

$$e^{\lambda t}P(n+1;t) = \int \lambda\frac{(\lambda t)^n}{n!}\, dt = \frac{(\lambda t)^{n+1}}{(n+1)!} + C \,. \tag{A.23}$$

Imposing the boundary condition $P(n+1;0) = 0$ implies $C = 0$ and therefore

$$P(n+1;t) = \frac{(\lambda t)^{n+1}}{(n+1)!}e^{-\lambda t} \,. \tag{A.24}$$

Thus the assertion (A.21) for n also holds for $n + 1$ and the result is proved by induction.

A.2 Mean and standard deviation of the Poisson distribution

First we can verify that the sum of the probabilities for all n is equal to unity. Using now $\nu = \lambda t$, we find

$$\begin{aligned}\sum_{n=0}^{\infty} P(n;\nu) &= \sum_{n=0}^{\infty} \frac{\nu^n}{n!} e^{-\nu} \\ &= e^{-\nu} \sum_{n=0}^{\infty} \frac{\nu^n}{n!} \\ &= e^{-\nu} e^{\nu} \\ &= 1 \,, \end{aligned} \tag{A.25}$$

where we have identified the final sum with the Taylor expansion of e^{ν}.

The *mean value* (or *expectation value*) of a discrete random variable n is defined as

$$\langle n \rangle = \sum_{n} n P(n) \,, \tag{A.26}$$

where $P(n)$ is the probability to observe n and the sum extends over all possible outcomes. In the case of the Poisson distribution this is

$$\langle n \rangle = \sum_{n=0}^{\infty} n P(n;\nu) = \sum_{n=0}^{\infty} n \frac{\nu^n}{n!} e^{-\nu} \,. \tag{A.27}$$

To carry out the sum note first that the $n = 0$ term is zero and therefore

$$\begin{aligned}\langle n \rangle &= e^{-\nu} \sum_{n=1}^{\infty} n \frac{\nu^n}{n!} \\ &= \nu e^{-\nu} \sum_{n=1}^{\infty} \frac{\nu^{n-1}}{(n-1)!} \\ &= \nu e^{-\nu} \sum_{m=0}^{\infty} \frac{\nu^m}{m!} \\ &= \nu e^{-\nu} e^{\nu} \\ &= \nu \,. \end{aligned} \tag{A.28}$$

Here in the third line we simply relabelled the index with the replacement $m = n - 1$ and then we again identified the Taylor expansion of e^{ν}.

To find the standard deviation σ of n we use the defining relation

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 \,. \tag{A.29}$$

We already have $\langle n \rangle$, and we can find $\langle n^2 \rangle$ using the following trick:

$$\langle n^2 \rangle = \langle n(n-1) \rangle + \langle n \rangle \; . \tag{A.30}$$

We can find $\langle n(n-1) \rangle$ in a manner similar that used to find $\langle n \rangle$, namely,

$$\begin{aligned} \langle n(n-1) \rangle &= \sum_{n=0}^{\infty} n(n-1) \frac{\nu^n}{n!} e^{-\nu} \\ &= \nu^2 e^{-\nu} \sum_{n=2}^{\infty} \frac{\nu^{n-2}}{(n-2)!} \\ &= \nu^2 e^{-\nu} \sum_{m=0}^{\infty} \frac{\nu^m}{m!} \\ &= \nu^2 e^{-\nu} e^{\nu} \\ &= \nu^2 \; , \end{aligned} \tag{A.31}$$

where we used the fact that the $n = 0$ and $n = 1$ terms are zero. In the third line we relabelled the index using $m = n - 2$ and identified the resulting series with e^{ν}. Putting this into equation (A.29) for σ^2 gives $\sigma^2 = \nu^2 + \nu - \nu^2 = \nu$ or

$$\sigma = \sqrt{\nu} \; . \tag{A.32}$$

This is the important result that the standard deviation of a Poisson distribution is equal to the square root of its mean.

Appendix B

The virial theorem

Here we will derive the virial theorem, which describes how the kinetic energy T and potential energy U are shared in a gravitationally bound system. It says that in equilibrium, the mean kinetic energy is related to the mean potential energy by $\langle T\rangle = -\frac{1}{2}\langle U\rangle$. To derive this result, consider a system of masses m_i having position vectors $\mathbf{r}_i$ with $i = 1, \ldots, n$. First we define the quantity A by

$$A = \sum_{i=1}^{n} m_i \dot{\mathbf{r}}_i \cdot \mathbf{r}_i \ , \tag{B.1}$$

where the dot denotes a derivative with respect to time. The time derivative of A is thus

$$\begin{aligned} \dot{A} &= \sum_{i=1}^{n} (m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i + m_i \ddot{\mathbf{r}}_i \cdot \mathbf{r}_i) \\ &= 2T + \sum_{i=1}^{n} \mathbf{F}_i \cdot \mathbf{r}_i \ , \end{aligned} \tag{B.2}$$

where to obtain the second line we identified the first term in parentheses as twice the kinetic energy, and in the second term we used Newton's law to replace $m_i \ddot{\mathbf{r}}_i$ by the force on the ith particle, $\mathbf{F}_i$.

If the system is gravitationally bound, then all of the position vectors $\mathbf{r}_i$ remain finite, and therefore A must remain finite. Therefore the average of $\dot{A}$ over a sufficiently long time must go to zero. In this limit we therefore obtain from equation (B.2)

$$\langle 2T\rangle + \Big\langle \sum_{i=1}^{n} \mathbf{F}_i \cdot \mathbf{r}_i \Big\rangle = 0 \ . \tag{B.3}$$

This is the general form of the virial theorem.

We can now apply equation (B.3) to the case where our masses interact only by gravity. That is, the force on the mass m_i is given by

$$\mathbf{F}_i = -Gm_i \sum_{\substack{j=1 \\ (j\neq i)}}^{n} m_j \frac{(\mathbf{r}_i - \mathbf{r}_j)}{r_{ij}^3} \ , \tag{B.4}$$

where $r_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|$ and the sum over j in (B.4) extends over all of the masses except i, since we are computing the force on m_i. To evaluate equation (B.3) we require the quantity

$$\sum_{i=1}^{n} \mathbf{F}_i \cdot \mathbf{r}_i = -G \sum_{i=1}^{n} \sum_{\substack{j=1 \\ (j\neq i)}}^{n} \frac{m_i m_j (\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{r}_i}{r_{ij}^3} \ . \tag{B.5}$$

The double sum can be broken into those terms where $i < j$ and those where $j < i$:

$$\sum_{i=1}^{n} \mathbf{F}_i \cdot \mathbf{r}_i = -G \left[\sum_{\substack{i,j \\ (i<j)}} \frac{m_i m_j (\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{r}_i}{r_{ij}^3} + \sum_{\substack{i,j \\ (j<i)}} \frac{m_i m_j (\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{r}_i}{r_{ij}^3} \right] \ . \tag{B.6}$$

Terms with $i = j$ are of course absent from the sums. Now the second term in (B.6) can be rewritten simply by exchanging the names of the indices i and j to read

$$\sum_{\substack{j,i \\ (i<j)}} \frac{m_j m_i (\mathbf{r}_j - \mathbf{r}_i) \cdot \mathbf{r}_j}{r_{ji}^3} \ . \tag{B.7}$$

As we have $r_{ij} = r_{ji}$, equation (B.6) can therefore be written as

$$\begin{aligned} \sum_{i=1}^{n} \mathbf{F}_i \cdot \mathbf{r}_i &= -G \sum_{\substack{i,j \\ (i<j)}} \frac{m_i m_j (\mathbf{r}_i - \mathbf{r}_j) \cdot (\mathbf{r}_i - \mathbf{r}_j)}{r_{ij}^3} \\ &= -G \sum_{\substack{i,j \\ (i<j)}} \frac{m_i m_j}{r_{ij}} \\ &= U \ . \end{aligned} \tag{B.8}$$

where we identify the second line as the gravitational potential energy U of the system of masses. Notice that the sum with $i < j$ includes each pair of masses once. Combining this with equation (B.3) gives us our final form of the virial theorem,

$$\langle T \rangle = -\tfrac{1}{2} \langle U \rangle \ . \tag{B.9}$$

Although this result has been derived for averages over long periods of time, we will often apply it to an average over a large number of members of a single system considered at a single point in time. The statement that the two are equivalent is called the *ergodic hypothesis*.

Appendix C

The Maxwell-Boltzmann distribution

In this section we will derive the probability density function (pdf) that describes the distribution of speeds of the molecules in a gas of a certain temperature, the famous Maxwell-Boltzmann distribution. Suppose we look at a molecule with mass m in a gas at temperature T and consider first only the x-component of its velocity, v_x. The value of v_x taken on by a given molecule at a given time will be the end result of a tremendous number of collisions, each of which changes its v_x by some random value. According to the *Central Limit Theorem*, a random variable that is the sum of a very large number of terms will follow a Gaussian distribution. The conditions for this to hold are fairly unrestrictive and satisfied to a high degree of accuracy in our problem, so the pdf for v_x is well modeled as a Gaussian. If we work in the centre-of-momentum frame of the gas (i.e., there is no wind), then the mean value $\langle v_x \rangle$ is zero, so the pdf of v_x is

$$f_x(v_x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-v_x^2/2\sigma^2} \ . \tag{C.1}$$

By symmetry, the pdfs for the y and z components should have exactly the same form, i.e., we assume that there is no preferred direction. The parameter σ in (C.1) characterizes the width of the Gaussian pdf and we will show below that it is in fact equal to the standard deviation of the distribution. We do this by finding the mean value of v_x^2,

$$\langle v_x^2 \rangle = \int_{-\infty}^{\infty} v_x^2 f_x(v_x)\, dv_x = \int_{-\infty}^{\infty} \frac{v_x^2}{\sqrt{2\pi}\sigma} e^{-v_x^2/2\sigma^2}\, dv_x = \sigma^2 \ . \tag{C.2}$$

Thus the variance of the distribution, defined as $\langle v_x^2 \rangle - \langle v_x \rangle^2$ is equal to σ^2. The standard deviation, defined as the square root of the variance, is therefore equal to σ.

The speed v of a molecule is

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2} \tag{C.3}$$

and by symmetry we have

$$\langle v_x^2 \rangle = \langle v_y^2 \rangle = \langle v_z^2 \rangle \ , \tag{C.4}$$

and so therefore the mean value of v^2 is

$$\langle v^2 \rangle = 3\langle v_x^2 \rangle = 3\sigma^2 \ . \tag{C.5}$$

Now the Equipartition Theorem tells us that each quadratic term in the expression for the energy of a molecule contributes on average $kT/2$, where k is Boltzmann's constant and T is the temperature. Therefore we have

$$\tfrac{1}{2}m\left(\langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle\right) = 3\sigma^2 = \frac{3}{2}kT \tag{C.6}$$

or

$$\sigma = \sqrt{\frac{kT}{m}} \ . \tag{C.7}$$

It is reasonable to assume that the components of the velocity are *statistically independent*, i.e., a molecule's value of v_x has no influence on the probability to find a certain value for v_y, etc. If this is true, then the joint distribution for all three components of the velocity is simply the product of the individual pdfs. Using equation (C.1), this is found to be

$$f(v_x, v_y, v_z) = f_x(v_x) f_y(v_y) f_z(v_z) = \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-(v_x^2+v_y^2+v_z^2)/2\sigma^2} = \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-v^2/2\sigma^2} \ . \tag{C.8}$$

Using this we can find the probability to have the speed in an interval between v and $v + dv$ by integrating the joint probability density (C.8) over the infinitessimal volume in velocity space (i.e., axes v_x, v_y and v_z) where the speed is in the range $[v, v + dv]$. This volume is simply a spherical shell at "radius" v and with thickness dv. Furthermore, since the joint pdf (C.8) $f(v_x, v_y, v_z)$ in fact only depends on v, its integral over the shell is found by evaluating the integrand $f(v_x, v_y, v_z)$ at the speed v and multiplying by the volume of integration, which is the area of a sphere of radius v times the thickness of the shell dv:

$$f(v)\, dv = \frac{1}{(2\pi)^{3/2}\sigma^3} e^{-v^2/2\sigma^2}\, 4\pi v^2 dv = \sqrt{\frac{2}{\pi}} \frac{v^2}{\sigma^3} e^{-v^2/2\sigma^2}\, dv \ , \tag{C.9}$$

where as before $\sigma = \sqrt{kT/m}$ refers to the standard deviation of the velocity components. Equation (C.9) is the Maxwell-Boltzmann distribution. Its mode (most probable value) is at a speed $v_{\rm mode} = \sqrt{2kT/m}$, its mean is at $\langle v \rangle = \sqrt{8kT/\pi m}$, and its rms value is $v_{\rm rms} = \sqrt{3kT/m}$.

Bibliography

[1] Michael Zeilik and Stephen A. Gregory, *Introductory Astronomy and Astrophysics*, 4th edition, Brooks/Cole, 1998.

[2] H. Karttunen, P. Kröger, H. Oja, M. Poutanen and K.J. Donner (eds.), *Fundamental Astronomy*, 3rd edition, Springer, 2000.

[3] Bradley W. Carroll and Dale A. Ostlie, *An Introduction to Modern Astrophysics*, Addison-Wesley, 1996.

[4] Robert C. Smith, *Observational Astrophysics*, Cambridge University Press, 1995.

[5] A.E. Roy and D. Clarke, *Astronomy, Principles and Practice*, 4th ed., IoP, 2003.

[6] A calculator for local sidereal time is available through the web site of the Time Service Department, Department of the Navy, `tycho.usno.navy.mil/time.html`.

[7] More accurate conversions between sidereal and mean time are given in Jean Meeus, *Astronomical Algorithms*, Willmann-Bell, 1991.

[8] Florence Institute and Museum of the History of Science, `galileo.imss.firenze.it`

[9] Website of the Yerkes Observatoroy, `astro.uchicago.edu/yerkes`.

[10] Eugene Hecht, *Optics*, 4th edition, Addison-Wesley, 2002.

[11] P. Tipler and G. Mosca, *Physics for Scientists and Engineers*, 5th ed., 2005, Chapter 31.

[12] Photo archive of the National Museum of Science and Industry, `www.ingenious.org.uk`.

[13] Website of the Science Museum (London) `www.sciencemuseum.org.uk`.

[14] Michael Hoskin, The Cambridge Illustrated History of Astronomy, CUP, 1997.

[15] Richard Preston, *First Light: The Search for the Edge of the Universe*, Random House, 1996.

[16] Website of the Thueringer Landessternwarte Tautenburg, `www.tls-tautenburg.de`.

[17] Website of the Sptizer Space Telescope, `www.spitzer.caltech.edu`.

[18] George B. Arfken and Hans J. Weber, *Mathematical Methods for Physicists*, 6th edition, Elsevier, 2005.

[19] R. Gilliland and A.K. Dupree, Ap.J., **463** (1996) L29.

[20] Website of the Space Telescope Science Institute, `www.stsci.edu`.

[21] Website of the Carlsberg Meridian Telescope, `www.ast.cam.ac.uk/ dwe/SRF/camc.html`.

[22] Website of the Hipparcos Space Astrometry Mission, `www.rssd.esa.int/Hipparcos`

[23] Website of the GAIA mission `www.rssd.esa.int/gaia`.

[24] Website of the Lick Observatory, `mtham.ucolick.org`.

[25] Website of the Special Astrophysical Observatory, Russian Academy of Sciences, `www.sao.ru`.

[26] Website of the Carlsberg Merdian Telescope, `www.ast.cam.ac.uk/ dwe/SRF/camc.htm`.

[27] Erika Böhm-Vitense, *Introduction to stellar astrophysics, volume 2: stellar atmospheres*, Cambridge University Press, 1989.

[28] Website of the Stanta Barbara Instruments Group (SBIG) `www.sbig.com`.

[29] NASA Planetary Data Service, Planet Profiles, `pds.jpl.nasa.gov/planets/special/planets.htm`.

[30] The National Space Science Data Center, `nssdc.gsfc.nasa.gov`.

[31] NASA Astronomy Picture of the Day, `antwrp.gsfc.nasa.gov/apod/ap040619.html`, June 19, 2004.

[32] NASA Astronomy Picture of the Day, `antwrp.gsfc.nasa.gov/apod/ap010605.html`, 5 June, 2001.

[33] NASA Astronomy Picture of the Day, `antwrp.gsfc.nasa.gov/apod/ap010824.html`, 24 August, 2001.

[34] NASA Astronomy Picture of the Day, `antwrp.gsfc.nasa.gov/apod/ap970320.html`, March 20, 1997.

[35] NASA Astronomy Picture of the Day, `antwrp.gsfc.nasa.gov/apod/ap950821.html`, August 21, 1995.

[36] Science@NASA Headline News: 10th Planet Discovered, 29 July 2005, `science.nasa.gov/headlines/y2005/29jul_planetx.htm`

[37] Galileo Galilei, *Sidereus Nuncius*, Apud Thomam Baglionum, Venetiis, MDCX.

[38] NASA Astronomy Picture of the Day, `antwrp.gsfc.nasa.gov/apod/ap970929.html`, September 29, 1997.

[39] NASA Astronomy Picture of the Day, `antwrp.gsfc.nasa.gov/apod/ap990511.html`, 11 May 1999.

[40] NASA Astronomy Picture of the Day, `antwrp.gsfc.nasa.gov/apod/ap990502.html`, 2 May 1999.

[41] Claus Grupen, *Astroparticle Physics*, Springer, 2005.

[42] Barbara Ryden, *Introduction to Cosmology*, Addison-Wesley, 2003.