Essential Tools for Astrophysicists

Based on notes contributed by:
Loren Anderson, Paul Baker, Kevin Bandura, Harsha Blumer,
Paul Brook, Sarah Burke-Spolaor, Zach Etienne, Michael Lam,
Duncan Lorimer, Maura McLauglin, Sean McWilliams,
Mayuresh Surnis & D.J. Pisano

Fall 2022 — Spring 2023

These notes serve as a summary of the material we will cover in the Astrophysics seminar classes this year. We have focused on what we consider "essential" knowledge for Ph.D. students. This will help prepare you for the advanced level graduate astrophysics classes. Since this is very much a living document, we welcome all feedback on the text and hope that it provides you with a solid foundation in your research. Please refer to the course syllabus for detailed information about class structure, expectations and schedule.

Contents

1 Approximation and Estimation

In your analysis toolkit, one of the most important things you *must have* is an ability to perform approximation. The basic idea is using existing knowledge to make educated guesses about the approximate values of parameters, and simple arithmetic to arrive at an estimated answer. While it may seem simple-minded, it takes remarkable amounts of practice to do well, and chances are you are presently quite bad at it.

These techniques will allow you to understand the relative importance of different factors influencing your data, whether you got anywhere near the result you should have expected (i.e. answering the simple question *did I do something wrong in my analysis?*), and many other useful applications. Often, approximation to an order of magnitude is enough to further a fast-paced brainstorming discussion or allow you to make a decision. To give you an idea of the relevance: order-of-magnitude calculations are, apparently, used to screen interviewees at companies like Google because almost all careers involving analysis require this type of thinking. Entire courses have been designed on the premises summarized here. In this Chapter, we will touch on order of magnitude (OOM) basics during lecture, and future classes will be organized so that you can acquire further practice.

These notes give a general background for understanding the conceptual thinking behind order of magnitude estimation. We also will review a common approximation technique: Taylor expansion.

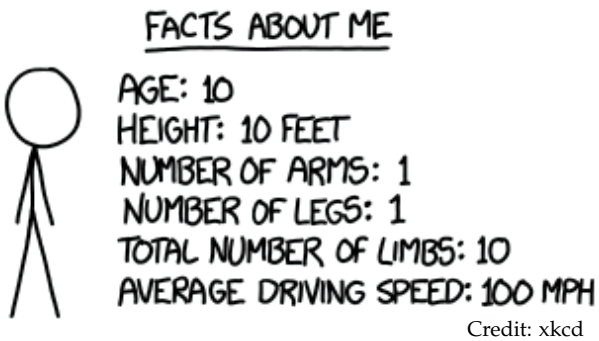
1.1 When to use approximation

When it's a good idea:

- You want a quick estimate for the sake of a discussion.
- You want to make a conservative estimate.
- You don't really have solid measurements.
- You have calculated a precise result but want a "sanity check" of your answer.

When it's a bad idea:

- You need to quote a formal error.
- The decision to be made is very sensitive to the numerical result.
- You're comparing two things that are close in value (around $\lesssim 100\times$ one another).



Credit: xkcd

1.2 How to think

Let's start with an example

Physicist Enrico Fermi was famous for coming up with relatively accurate estimates for complex questions based on little to no knowledge of relevance to the question. Here is a popular example from his repertoire. Before you look at the answer, consider how you would come up with a solution (And perhaps attempt one! How close do you get?):

Question:

How many piano tuners are there in Chicago?

Annotated Answer:

My intuition tells me that in a biggish city like Chicago, there should be something on the order of maybe tens to a few hundred, maybe up to a thousand, piano tuners in business in Chicago; however my intuition only gives me an estimation of the number within about 3 orders of magnitude (10-1000), and I'd like to see if it's anywhere near right.

I need to consider what the answer might rely on. I think: *The more households, the more pianos*; and from a general knowledge and partially a guess, I think *a piano needs to get tuned maybe every few years*. So let's start there and try to make an equation.

$$\text{Number of tuners} = \text{number of pianos in Chicago} \times \text{tunings per year per piano} \tag{1}$$

Will this give us the answer we want? Let's try dimensional analysis on it:

$$[\text{Tuners}] = [\text{Pianos}]\frac{[\text{Tunings}]}{[\text{year}][\text{Pianos}]} = \frac{[\text{Tunings}]}{[\text{year}]} \tag{2}$$

Dimensional analysis tells me this equation isn't quite right (number of tuners should not have units of tunings per year). So, let's modify the equation to get it in the right units. Because I want this number to come out in units of "Tuners," the above dimensional

analysis shows me that I should probably divide my equation by the number of tunings that can happen per tuner each year:

$$\text{Number of tuners} = \text{number of pianos in Chicago} \times \frac{\text{tunings per year per piano}}{\text{tunings per year per tuner}} \tag{3}$$

Which gives:

$$[\text{Tuners}] = [\text{Pianos}] \frac{[\text{Tunings}]}{[\text{year}][\text{Pianos}]} \Big/ \frac{[\text{Tunings}]}{[\text{year}][\text{Tuners}]} = [\text{Tuners}] \tag{4}$$

So now I have a nice equation, which has a few easier questions: 1) *How many pianos are there in Chicago?*; 2) *How often are pianos tuned?*; and 3) *How many pianos does the average tuner tune each week?* Let's answer these questions in turn using order of magnitude estimation.

1) How many pianos are in Chicago?
This sub-question breaks down into further numbers that we have to estimate: for instance, the number of households in Chicago with a piano. The question doesn't specify "greater Chicago" or "Chicago city limits", but as far as cities go I've seen quotes on the order of "the population of New York City is around 20 million people." I know Chicago's a much smaller city, so I'll say that it's probably on the order of around 2-5 million people.

The average dwelling houses between 1 and 5 people, so let's say there are on average 3 people per household in Chicago. To get the number of households in Chicago I can then say $\sim$3 million people/($\sim$3 per house)$\simeq$ 1 million households.

How many of those households have pianos? I'm not really sure about this number, but it seems like it should be much less than "all houses" and more than "1 in 100 houses." At the same time, 1 in 10 seems a bit high (of around 50 of my friends, I know maybe 2 with pianos). I'll pick 1 in 20, so around 5% of households might have a piano.

So I get number of pianos $= 10^6$ households $\times 0.05 = \sim 5 \times 10^4$ pianos in Chicago. Seems maybe like a big number, but I think my estimates were reasonable.

2) How often are pianos tuned?
Another number I'm not sure about, however to get something in the right range I'll take a guess. It seems like pianos would be tuned more than once every 10 years (otherwise a piano tuning might be a big deal, and when people have mentioned their piano getting tuned it hasn't seemed such). I'm guessing it's not more than once per year. Some number between 1 and 5 sounds reasonable, therefore, so I'll choose 1 per year to low-ball my answer. If it seems like this number was pulled roughly out of thin air, it was: but as you see I have a number that's probably correct within about 1 order of magnitude.

3) How many pianos can a tuner tune per week?
I know one (retired) piano tuner, who I've never seen tune anything but a guitar, so that doesn't really help me. However I'll consider that the tuners are in business as a full-time

job, and so they'll need to fill 30–40 hours per week with tunings. I think a tuning takes about 2-4 hours including travel time (remember it's Chicago: lots of big-city traffic), so they can probably get up to done 2 in a day. So, perhaps they can tune around 10 pianos per week, or around 500 per year.

Let's put this all together now and return to our Equation 3.

$$\text{Number of tuners } = 5 \times 10^4 \times 1 \,/\, (5 \times 10^2) \simeq 100 \text{ piano tuners.}$$

The internet tells me that the number of piano tuners in Chicago is actually around 290, so my estimate is not wildly off the mark, and is in the right order of magnitude! I can also see that if I had made slightly different decisions in the specific numbers in answering questions (1–3) above, I would still have come up with numbers in the ballpark of "300ish".

This is the essence of order-of-magnitude estimation.

Techniques of estimation

Note that while this example does not come from physics, it nicely demonstrates the basics of order-of-magnitude techniques when approaching what seems to be a complex question.[1] The basic idea was to divide the problem into simplified sub-domains, make educated assumptions for each, identify and ignore things that don't matter, and arrive at an answer. We will try this procedure a few times in class. Meanwhile here are a few tools of the trade (consider how each of these were used in the above example):

- *Define the question:* In real life, questions may not always be clear or precise in their wording. It might be up to you to decide exactly what question you are answering. Try starting such problems by rewriting a more precise question, and answering that.

- *Divide and conquer:* Split a complicated problem into manageable chunks, especially when you must deal with tiny or huge numbers, or when a formula naturally factors into parts (such as $V \sim l \times w \times h$). Ask yourself: what main factors will the answer rely on?

- *Guess:* Make a guess at the answer before solving a problem. The guess may suggest a method of attack. Guessing—and checking and modifying your guess—improves your intuition and guesses for future problems. Talk to your gut: When you make a guess, ask your gut how it feels. Is it too high? Too low? If the guess is both, then it's probably reliable.

[1] The below list is adapted from Goldreich, Mahajan, & Phinney's online textbook, "*Order-of-Magnitude Physics: Understanding the World with Dimensional Analysis, Educated Guesswork, and White Lies*"

- *Lie skillfully:* Simplify a complicated situation by assuming what you need to know to solve it. For example, when you do not know what shape an object has, assume that it is a sphere or a cube.

- *Cross-check:* Solve a problem in more than one way, to check whether your answers correspond.

- *Simplify:* If you're worried about whether some physical effect will matter, do not worry about it in your first attempt at a solution. The productive strategy is to start estimating, to explore the problem, and then to handle the exceptions once you understand the domain. The most common astronomy example: whether magnetic fields will influence a result. They often get ignored in the first pass of understanding a system, only to be returned to later once the basic system dynamics are understood.

- *Make assumptions:* You should always be aware of your assumptions and how (at least vaguely) they might affect your solution. However, as long as you remain aware of those assumptions, and the assumption allows a solution, make it, and return to worry about the damage afterward.

- *Lower your standards:* If you cannot solve the entire problem as asked, solve those parts of it that you can, because the sub-problem might still be interesting. Solving the subproblem also clarifies what you need to know to solve the original problem.

- *Use symbols:* Even if you do not know a certain value, define a symbol for it. It may cancel. If it doesn't, and the problem is still too complex, then lower your standards.

With a little knowledge and a repertoire of techniques, you can estimate just about any quantity and perform fairly advanced theoretical physics.

Rounding

One of the things that can at first be hard to get comfortable with in OOM estimation is to round off numbers, including constants, to the nearest order of magnitude while calculating. As an example, try out these numbers:

	(becomes)	
0.08	$\rightarrow$	0.1
π	$\rightarrow$	3 or 1
1234	$\rightarrow$	10^3
seconds per year ($31557600\,\mathrm{s/yr}$)	$\rightarrow$	$\pi \times 10^7 \sim 10^7$ seconds/year
$L_\odot = 3.828 \times 10^{26}\,\mathrm{W}$	$\rightarrow$	$10^{26}\,\mathrm{W}$

To gain a bit of precision in the result, you can use less crude approximation (keeping only one significant digit): for instance, $\pi = 3$, or perhaps $\mathrm{M}_\odot = 2 \times 10^{30}$ kg. Whether you

round a number upwards or downwards to the nearest OOM really depends on whether you'd like to low-ball or high-ball your answer.

1.3 A few examples

Try a few of these to practice your techniques. Some rely on just a bit of basic physics. I challenge you to come up with your own questions relevant to your life or field of study: how might you estimate them?

- How much does street lighting cost in the city of Morgantown?
- How old are the dark parts of the Earth-facing lunar surface (hint: consider that they are relatively craterless)?
- According to some recent highly accurate measurements made from satellites, the continent of North America is drifting at a rate of about 1 cm per year. Assuming a continent is about 50 km thick, estimate the kinetic energy the continental US has a a result of this motion.
- How much does it cost to get 1 kg of material out of Earth's orbit?

1.4 Taylor series: a review

Taylor series are important because they allow us to compute functions that perhaps cannot be computed directly, and they allow us to do so to arbitrary precision (for the case of estimation, they can simplify an expression immensely if you can tolerate a certain amount of imprecision). This is effective because when you form a Taylor polynomial expansion you can truncate it to your desired precision.

For a very practical example: Taylor series can be used to compute $\sin(x)$ without a calculator. The basic idea is say you have a function $f(x) = \sin(x)$, and you want to estimate f close to some arbitrary number, lets say at $x = a$. To do this, we can use a Taylor expansion to estimate values close to a. The Taylor expansion can be written as an approximation around that value a by expanding the function $f(x)$ as:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{5}$$

where $f^{(n)}(a)$ refers to the nth derivative of $f(x)$ evaluated at your reference number a. Note that when you evaluate the zeroth order ($n = 0$), the function $f^{(0)}(a)$ is simply $f(a)$, and $0!$ and $(x-a)^0$ are both 1.

Expanded out, you can write the above function as:

$f(x) \simeq \frac{f(a)}{0!}$ [$n = 0$; zeroth order approximation]

$f(x) \simeq \frac{f(a)}{0!} + \frac{f'(a)}{1!}(x-a)$ [$n = 1$; first order approximation]

$f(x) \simeq \frac{f(a)}{0!} + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2$ [$n = 2$; second order approximation]

(...)

Here is a visual example to help you wrap your head around this. Let's say we want to understand how to evaluate $f(x) = e^x$, a very popular function in physics. I know that $e^0 = 1$, so let's try using $a = 0$ (note: as you will see, this is a common thing to do, although a does not always have to be zero to use this series).

To get organized I can write out:

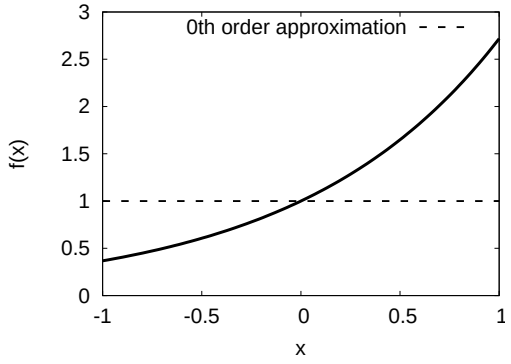
Order	**Derivative**	**Derivative at** a	**Factorial**
$n = 0$	$f(x) = e^x$	$f(0) = 1$	$0! = 1$
$n = 1$	$f'(x) = e^x$	$f'(0) = 1$	$1! = 1$
$n = 2$	$f''(x) = e^x$	$f''(0) = 1$	$2! = 2$
$n = 3$	$f'''(x) = e^x$	$f'''(0) = 1$	$3! = 6$

Obviously this function is a nice one in terms of book keeping because its derivatives are all the same (other functions will not be so kind).

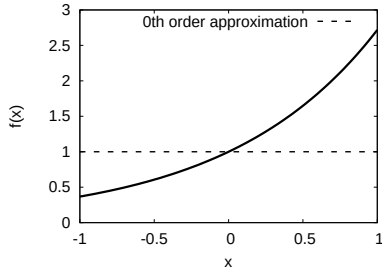
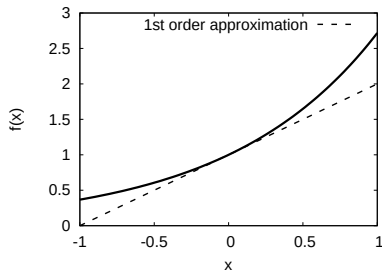
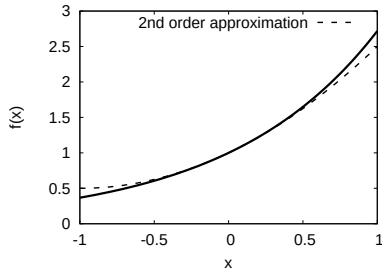
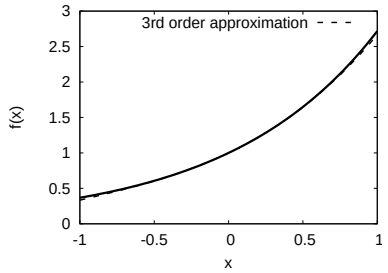
Now let's try the expansion. My zeroth-order approximation tells me that

$$f(x) \simeq \frac{f(a)}{0!} = 1 \tag{6}$$

Here is how that expression approximates the function e^x in the range (-1, 1):



Is that a good approximation to e^x? A world of NO, unless you're exactly at $x = 0$. Let's try adding in higher-order terms to first approximate this function as a line, then a quadratic, etc.

Order	$f(x) \simeq$	**What it looks like**
$n = 0$	1	
$n = 1$	$1 + x$	
$n = 2$	$1 + x + (1/2)x^2$	
$n = 3$	$1 + x + (1/2)x^2 + (1/6)x^3$	

For values of x very close to zero, the simple expression $1 + x$ will suffice as a replacement for e^x in an equation. If you need more precision or need to maintain precision further away from your reference value, you can add in the higher-order terms.

I hope you've noticed that what a Taylor Series amounts to is adding a bunch of polynomial expressions with various coefficients in front of them, to approximate a function. You may on occasion hear the term "polynomial coefficients"; in this context, this refers to the numerical coefficient in front of the polynomial for each order (for this example: $1, 1, 1/2, 1/6$).

Practical Application 1: Simplifying an Equation or Obtaining a Limit

Obtaining a limit:

Let's say I have a function

$$y(x) = \frac{\cos(x)}{2 - x^2} \tag{7}$$

And I want to determine the limit as $x \to 0$. Let's go straight to Mathematica, right? Think again. Let's do it analytically (as you would do for, say, a derivation in a publication), and try a Taylor series. We want the limit close to zero, so will set $a = 0$. I'll try to expand the most annoying part of this expression, defining $f(x) = \cos(x)$ and noting that:

Order	**Derivative**	**Derivative at a**	**Factorial**
$n = 0$	$f(x) = \cos(x)$	$f(0) = 1$	$0! = 1$
$n = 1$	$f'(x) = -\sin(x)$	$f'(0) = 0$	$1! = 1$
$n = 2$	$f''(x) = -\cos(x)$	$f''(0) = -1$	$2! = 2$
$n = 3$	$f'''(x) = \sin(x)$	$f'''(0) = 0$	$3! = 6$

So I can write:

$$f(x) = \cos(x) \simeq 1 + 0 - \frac{1}{2}x^2 + 0(...) \simeq 1 - \frac{1}{2}x^2 \tag{8}$$

...so long as the value of x is close to zero. Since we want lim $x \to 0$ this applies nicely and I can substitute in my Taylor series and find, simply, that:

$$\lim_{x \to 0} \frac{1 - \frac{1}{2}x^2}{2 - x^2} = \frac{1}{2} \tag{9}$$

Simplifying an equation:
The previous derivation works the same way if we had simply wanted to assume that x is always going to be a value close to zero, or to whatever value we want a to be (as in the example for e^x above). Let's say the expression in Eq. 7 is some function of wavelength λ and distance r where:

$$y = \frac{\cos(\lambda/r)}{2 - (\lambda/r)^2} \, . \tag{10}$$

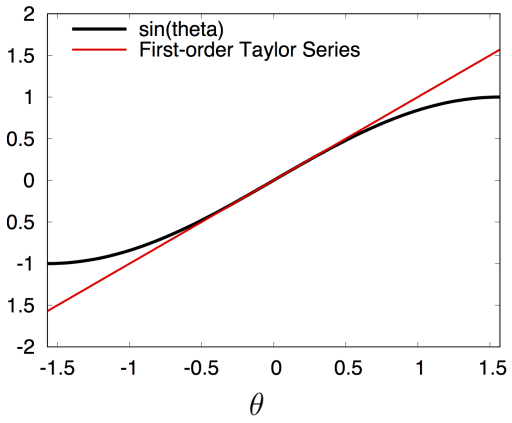
If this were the case, we could rightfully argue that in the case that $\lambda \ll r$, then we know that the signal will always have a value of $y \simeq 1/2$, which is much preferable to have in a derivation than the form of y in Eq. 10. Often in astronomy we will have situations where you know that $\lambda \ll r$ (or some equivalent) is a physically reasonable assumption. If you're using y as part of other expressions, replacing it with $1/2$ in your derivation will provide much simpler, and sufficiently accurate, results.

Practical Application 2: Small-Angle Approximations

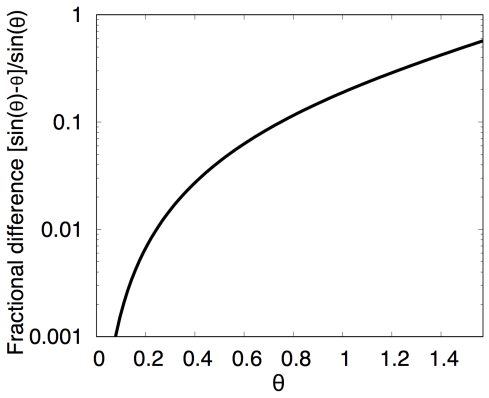
A very useful and common example of Taylor series is "small-angle approximation". In class we will try writing out the Taylor series for $\sin(\theta)$. If you did this for the basic trig functions you will find that:

$$\sin\theta \simeq \theta; \quad \tan\theta \simeq \theta; \quad \cos\theta \simeq 1 - \frac{\theta^2}{2} \quad \text{(as derived above)} \tag{11}$$

For reference, for sinθ here is how the approximated expression compares with the actual one, between $(-\pi/2, \pi/2)$:



And here is the fractional error as your θ value deviates from zero (shown for values in radians between 0 and $\pi/2$):



So you can see that using the small angle approximation is appropriate for values close to zero. Within θ of about 0.2 radians ($\sim$11 degrees), your error in assuming $\sin(\theta) = \theta$ will be about 1% deviant.

1.5 Problems

Create and work through an OOM question of your own design, on any topic that interests you (physics/astronomy topics are warmly welcomed but not required; make sure you come up with a question on your own, rather than using one you find on the web). Please arrive at the next class prepared to come up to the board and lead a discussion on your problem. Remember, the idea of these questions is not to arrive at *the* answer, but to understand the procedure through which to get a roughly correct "first-order" answer. There might be prizes for the most thought-provoking questions (and answers)!

2 Coordinate systems

These notes contain the essential information you'll need to understand where astronomical sources are in the sky. Further details can be found in *Textbook on Spherical Astronomy* (Smart 1977). Astronomers use many different coordinate systems, depending on what is most convenient, but all of these systems use spherical coordinates. It is important to understand these systems, why they exist, and how we can convert between them.

For spherical coordinates:

$$x = r \sin\theta \cos\phi \tag{12}$$

$$y = r \sin\theta \sin\phi \tag{13}$$

$$z = r \cos\theta \tag{14}$$

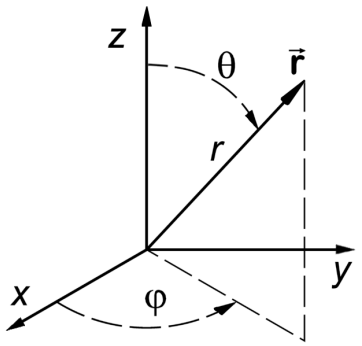


Figure 1: Spherical coordinates.

It would be much easier to deal with Cartesian coordinates, but everything is projected onto what we call the "celestial sphere." The celestial sphere is just the imaginary surface on which we observe celestial objects. It is helpful to think of this being a real physical sphere that rotates once per day. Due to this apparent projection, we don't usually know/care about coordinate r. We can therefore set $r = 1$ and express directions only using the two angles. This makes life considerably easier! One important consequence of this is that measurements on the sky are in $angles$, rather than physical distances.

To define a coordinate system, we therefore need just five things:
1) A longitude coordinate
2) A latitude coordinate
3) A starting point for longitude
4) A starting point for latitude (called the fundamental plane)
5) The origin
(The poles are necessarily defined by the above choices.)

We'll go through these five characteristics of each coordinate system.

2.1 Coordinate Systems

On the Earth

On the Earth we use:
1) Longitude to specify the E-W direction. Ranges from $180\,^\circ$ East to $180\,^\circ$ West, uses λ.
2) Latitude to specify the N-S direction, ϕ. Ranges from $90\,^\circ$ north ($+90\,^\circ$) to $90\,^\circ$ south ($-90\,^\circ$).
3) The longitude of Greenwich, England is at zero degrees longitude.
4) The equator is at zero degrees latitude.
5) The center of the Earth as the origin.

While you all probably know all this, we will soon see that all coordinate systems are similar to the Earth-based system that you are familiar with. There is, however, one significant difference between the Earth-based system and those used for astronomy: instead of considering coordinates as they appear looking down on the Earth, we imagine we are inside the sphere of the sky (the "celestial sphere") looking out. The effect of this change in perspective is that the direction of increasing longitude is flipped.

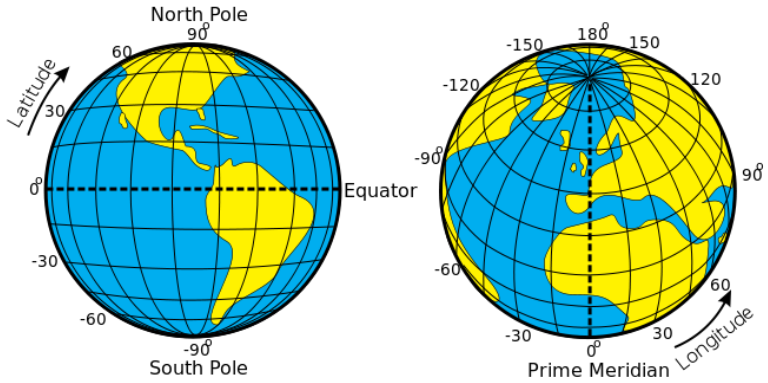


Figure 2: Coordinates used to define a location on Earth.

Local (Horizon) Coordinates

Perhaps the easiest celestial system to visualize is is the Local Coordinate system. This system is aligned with your local horizon, so it is tied to you the observer. It has:
1) Azimuth, Az, ranges from $0\,^\circ$ to $360\,^\circ$ and is measured from North toward East
2) Elevation or altitude measured from the horizon ($0\,^\circ$) to your zenith straight overhead ($+90\,^\circ$). Although you can't see it, at $-90\,^\circ$ is our nadir. Zenith angle, $\mathrm{ZA} = 90 - \mathrm{El}$ is an alternative angle often used for telescopes that only observe close to the Zenith.
3) North is Az= $0\,^\circ$.
4) Your horizon is at El = $0\,^\circ$.
5) The origin is your location as observer.

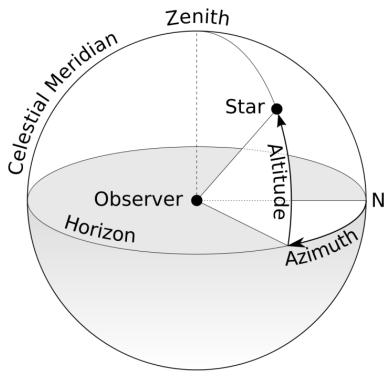


Figure 3: Local coordinates. Note the horizon, zenith, and celestial meridian. The star shown has an elevation measured from the horizon toward zenith, and a zenith angle measured from the zenith down toward the horizon. Notice also that increasing longitude (azimuth) goes in the opposite direction from that on earth.

While the horizon system is intuitive, because it is unique to each observer, it is not as generally useful as one would hope. We therefore have numerous other systems that are not tied to a specific observer. Horizon coordinates are useful when we are trying to determine when an object will rise and set.

Galactic Coordinates

Observations of objects in the Milky Way often make use of Galactic coordinates. This system is useful for specifying where objects are in relation to the rest of the Galaxy:
1) Galactic longitude, ℓ. Ranges from $0\,^\circ$ to $360\,^\circ$.
2) Galactic latitude, b. Ranges from $+90\,^\circ$ at the Galactic north pole to $-90\,^\circ$ at the Galactic south pole.
3) The Galactic center is at $\ell = 0\,^\circ$.
4) The Galactic mid-plane is at $b = 0\,^\circ$.
5) The Sun is the origin.

Stars in our Galaxy are found in higher densities toward the mid-plane and toward the Galactic center. Therefore, by knowing the Galactic coordinates of an object, we can possibly infer something about its environment. The caveat here is that we often do not know the distance to the object, which can make such inferences less accurate.

Equatorial system

Of all these systems, the Equatorial system is most often used because it corresponds most closely with that needed to perform observations. This system is aligned with the orientation of the Earth. As we will see later, this alignment causes one issue. The system

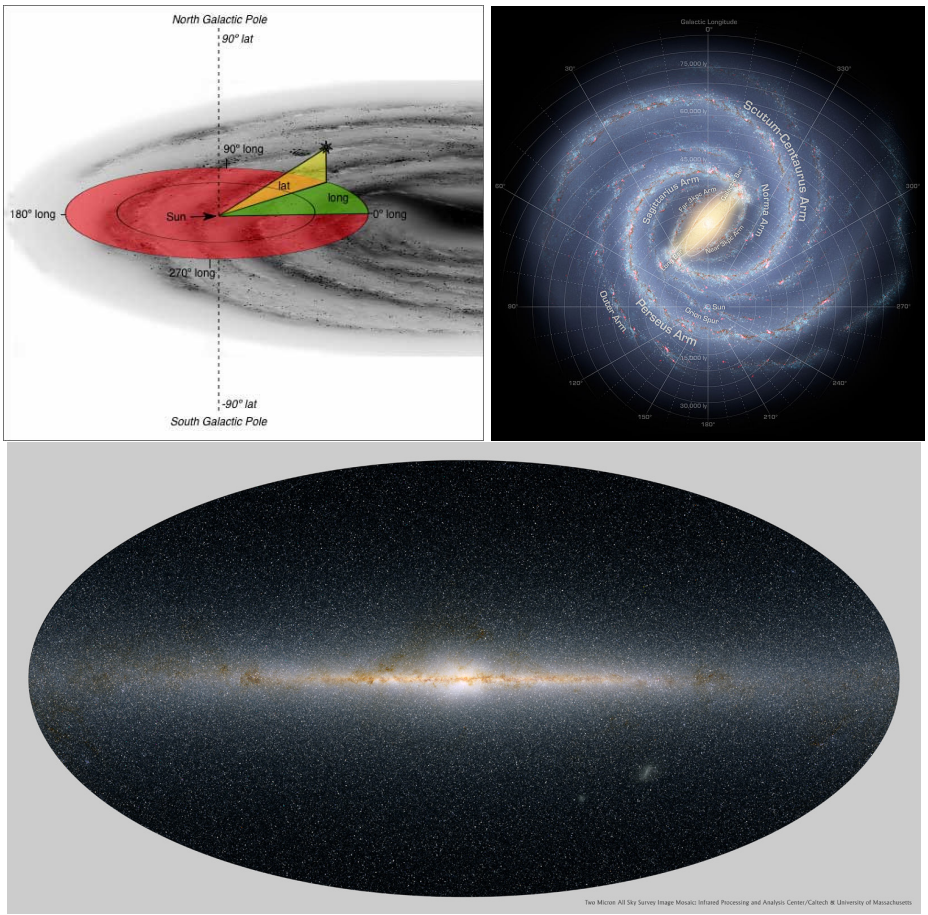


Figure 4: Galactic coordinates. The top-left panel shows how the angles for ℓ and b are measured. Note the direction of the north and south Galactic poles. The top right panel shows a face-on view of what we think our Galaxy looks like, with longitude directions shown. Longitude increases counter clockwise from the Galactic center. The bottom panel shows in infrared view of the Galaxy from our perspective on Earth. The center of the figure is the Galactic center $(\ell, b) = (0\,^\circ, 0\,^\circ)$. The bright band in the middle is the mid-plane ($b = 0\,^\circ$). Galactic longitude increases from the center toward the left hand side, with both the left and right edges at $\ell = 180\,^\circ$. Galactic latitude is $b = +90\,^\circ$ at the top and $b = -90\,^\circ$ at the bottom.

uses:
1) Right ascension, R.A., for longitude, which takes the symbol α. RA ranges from 0 to $360\,^\circ$ and there are no negative values.
2) Declination, Dec., which takes the symbol δ, is used for latitude. Dec. ranges from $-90\,^\circ$ to $+90\,^\circ$.
3) The origin of right ascension is the location of the Sun on the vernal equinox (March 21).
4) Zero degrees declination is the celestial equator, the projection of the Earth's equator onto the celestial sphere.
5) The center of the Earth is the origin.

Other tidbits:
1) NCP and SCP stand for the north and south celestial poles, i.e. $\delta = +90\,^\circ$ and $\delta = -90\,^\circ$
2) The altitude of the NCP (or SCP) is your latitude on Earth
3) The Ecliptic is the Sun's annual path across the celestial sphere. Since the apparent Solar motion is caused by the Earth's revolution around the Sun, and the Earths spin axis is tilted by $23.5\,^\circ$ to its orbital axis, the ecliptic is tilted by $23.5\,^\circ$ with respect to the celestial equator.
4) The Vernal equinox is where the Ecliptic crosses the celestial equator, so it's not such a crazy place to define $\alpha = 0\,^\circ$.

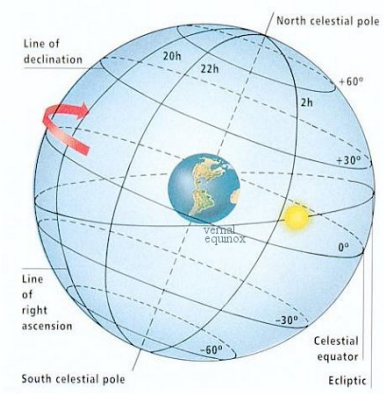


Figure 5: Equatorial coordinates. Note the north and south celestial poles, the celestial equator, and the ecliptic. The Vernal equinox is where the ecliptic and celestial equator intersect. *This system is the exact analog of what we use on the surface of the Earth, projected outward to the celestial sphere.*

Notation in the Equatorial System

R.A. can be expressed in degrees (or radians), but you'll usually see it written in hours, minutes and seconds of time, where $24\text{hours} = 360\,^\circ$ or $1\text{hour} = 15\,^\circ$. We will see why we add this complication a bit later. So, for example, $20\text{h}34\text{m}45\text{s} = (20 + 34/60 + 45/3600) \times 15 = 308.7\,^\circ$. Similarly, declination can be expressed in degrees or radians, but you'll usually see it written in degrees, minutes and seconds. So for example $09\,^\circ 45\text{m}34\text{s} = 9 + 45/60 + 34/3600 = 9.8\,^\circ$.

Although the 24 hours is strange, it is of course common to use minutes and seconds to distinguish fractions of an angle. In astronomy, we use "arcminutes" (symbol $'$) and "arcseconds (symbol $''$) to illustrate that these are angles on a curved surface. One important and strange caveat is that arcseconds of R.A. are not equal to arcseconds of Dec. (they differ by the factor of 15 above)! If the unit "arcsecond" is used without a coordinate attached, assume what is meant is that corresponding to the declination unit, i.e. $1/3600$th of a degree.

The Sun in Equatorial Coordinates

Throughout its yearly motion across the sky, the path of the Sun (the ecliptic) of course passes through a range of coordinates. It begins on March 21 at $\alpha = 0$h, and then advances approximately two hours per month over the next 12 months. Each week therefore adds about 30m to the R.A. of the Sun.

In Declination, the Sun is at $\delta = 0\,^\circ$ on March 21, then advances to $\delta = +23.5\,^\circ$ on the summer solstice (June 21), back to $\delta = 0\,^\circ$ on the Fall equinox (September 21), and to $\delta = -23.5\,^\circ$ on the winter solstice (December 21). This gets us to a fundamental result: in the Northern hemisphere, high declination sources closer to the NCP (i.e., the Sun in summer) are up in the sky longer than low declination sources (i.e., the Sun in winter). The same is true in the southern hemisphere for sources close to the SCP.

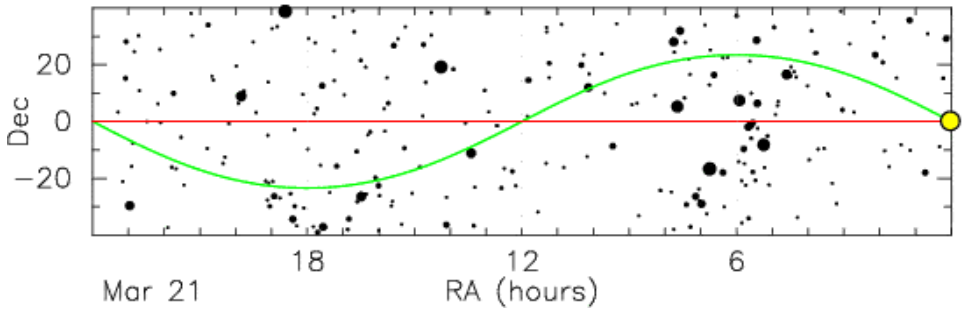


Figure 6: The motion of the Sun throughout the year (the ecliptic) in equatorial coordinates. On this diagram, the Sun follows the green line, moving right to left throughout the year. Animated version here: https://www.dropbox.com/s/6dx6z6z94unhmt9/solar_year.gif

Epochs

The Equatorial system is really useful, but has one strange quirk. Because the Earth's axis is not stable but rather precesses and nutates, the locations of the NCP, SCP, and Vernal Equinox shifts slightly. These changes are small, but important enough that we specify an "epoch" by convention every 50 years. The most common epoch currently is the Julian (J2000), and previously the Besselian (B1950) was used. As an example, a source with J2000 coordinates R.A. = 09 h 45 m 30 s and Dec. = -15 deg 31 m 20 s has equivalent B1950 coordinates of R.A. = 09 h 43 m 06 s and Dec. = -15 deg 17 m 28 s. While you can specify coordinates for observation in any epoch you like, it is very important to know which system you are using! Failure to keep track of this has led to untold wasted telescope hours :(.

A Few More Words about Declination

The declination measures angular distance from the NCP or SCP. The motion of stars across the sky is determined by your latitude on Earth, ϕ. For the northern hemisphere:
Sources with $\delta > 90\,^\circ - \phi$ are always visible. We call such stars "circumpolar". These same stars are never visible in south for same observing latitude.
Sources with $\delta < \phi - 90\,^\circ$ are never visible (circumpolar in the south for same latitude).
Low dec. sources do not rise far above the horizon.

For the southern hemisphere:
Sources with $\delta < \phi - 90\,^\circ$ are circumpolar.
Sources with $\delta > 90 - \phi$ are never visible.

In the northern hemisphere on Earth, all sources appear to circle around the NCP throughout the day/night. In the southern hemisphere, the same is true for the SCP. The full motion of every object takes 24 hours.

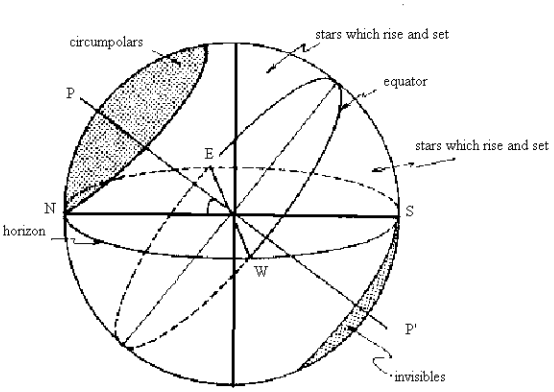


Figure 7: *Left:* A portion of the sky from a long-exposure photograph taken in the northern hemisphere. Shown are star tracks near the circumpolar region. These stars have high Dec. values. The NCP ($\delta = +90\,^\circ$) is at the center of the nested arcs. Each star track is roughly 1/3 of a circle, indicating that this exposure was roughly 8 hours long. *Right:* The celestial sphere showing the paths the stars take during one 24 hour period for one particular latitude on Earth. Note the circumpolar and "never rise" regions.

2.2 Time

Just as we need a coordinate system that is partially divorced from our observing location, we also need a time system not entirely based on the rotation of the Earth.

Hour Angle

Hour Angle (HA) is the angle between the celestial meridian (from the local coordinate system) to the objects RA. This is a useful additional parameter when observing.

If an object is crossing the prime meridian, which we call "transiting," the HA is zero. Objects that are rising (before transit) have negative HA. Objects that are setting (after transit) have positive HA. As time passes, the HA of an object increases.

For circumpolar sources, we say that it is in "upper culmination" when it crosses the prime meridian closest to zenith. There is a corresponding lower culmination 12 hours later when it crosses the prime meridian furthest from zenith.

Local Sidereal Time

As the Earth goes around the Sun, the Sun is of course found in different parts of the sky. Ideally, we do not want to use the Sun as a reference point for when a day begins and ends, but we would rather use something fixed to the locations of the stars. We therefore distinguish between "sidereal time," which is based on the apparent positions of the stars, from Solar time, which is derived from the apparent position of the Sun.

One sidereal day is equivalent to the true rotation period of the Earth with respect to background stars. As it revolves around the Sun in one year, the Earth completes one extra rotation on its axis compared with the number of days in a year. Thus the sidereal day is $365.25/366.25 = 99.727\%$ as long as the Solar day, or 23h56m. Universal time (UT), Greenwich mean time (GMT), Eastern daylight time (EDT) etc. are all based on Solar time. When talking about the positions of celestial sources, however, we must use sidereal time (ST).

Because it is tied to the local time in your time zone, ST is relative to the observer's longitude. We need an absolute sidereal time that is defined at the Greenwich meridian as Greenwich Sidereal Time (GST). The sidereal time we need for all calculations is the local sidereal time (LST) defined as LST = GST – West longitude.

There are many ways to think of LST. Here are the three easiest in my opinion:

- LST is defined as the HA of the vernal equinox at your location. When the vernal equinox crosses the local meridian, the LST is 0:00.
- The RA of any celestial body that is crossing the local meridian is the LST.
- So it follows that HA = LST - RA

This is where the idea of epochs comes into play. If the year is 2023 and we are using

J2000 coordinates, the HA is not exactly equal to the LST−RA. This is only exactly true at the current epoch, although we usually don't care too much about this small change in transit (and rise/set) time.

Julian Date

It is often convenient to have a running number for the day, so you may encounter a "Julian date." The Julian date begins with a value of "1" on January 1, 4713 BC at noon. Hours, minutes, and seconds are given in decimals, and increase in the normal way from that date and time.

2.3 Converting Between Coordinate Systems

With so many systems in use, it is essential to be able to convert between them. To do so accurately, one should use spherical trigonometry.

Local and Equatorial

It is often important to convert between local and equatorial systems, to determine for example when a source rises and sets. To do so, use:

$$\sin \mathrm{Az} = -\frac{\sin \mathrm{HA} \cos \delta}{\cos \mathrm{El}} \tag{15}$$

$$\sin \mathrm{El} = \sin \delta \sin \phi + \cos \delta \cos \phi \cos \mathrm{HA}\,, \tag{16}$$

where ϕ is your latitude on Earth.

Galactic and Equatorial

The Galactic plane is tilted by about $60\,^\circ$ from the CE and intersects the CE at $\ell \simeq 33\,^\circ$.

To convert between Galactic and equatorial coordinates:

$$\sin b = \cos \delta \cos \delta_g \cos(\alpha - \alpha_g) + \sin \delta \sin \delta_g \tag{17}$$

$$\tan(\ell - \ell_g) = \frac{\tan \delta \cos \delta_g - \cos(\alpha - \alpha_g) \sin \delta_g}{\sin(\alpha - \alpha_g)} \tag{18}$$

$$\sin \delta = \cos b \cos \delta_g \sin(\ell - \ell_g) + \sin b \sin \delta_g \tag{19}$$

$$\tan(\alpha - \alpha_g) = \frac{\cos(\ell - \ell_g)}{\tan b \cos \delta_g - \sin \delta_g \sin(\ell - \ell_g)}\,, \tag{20}$$

where $\alpha_g = 192.25\,^\circ$, $\delta_g = 27.4\,^\circ$, and $\ell_g = 33\,^\circ$ define the conversion from B1950 equatorial coordinates to the Galactic frame, which was is defined at the B1950 epoch.

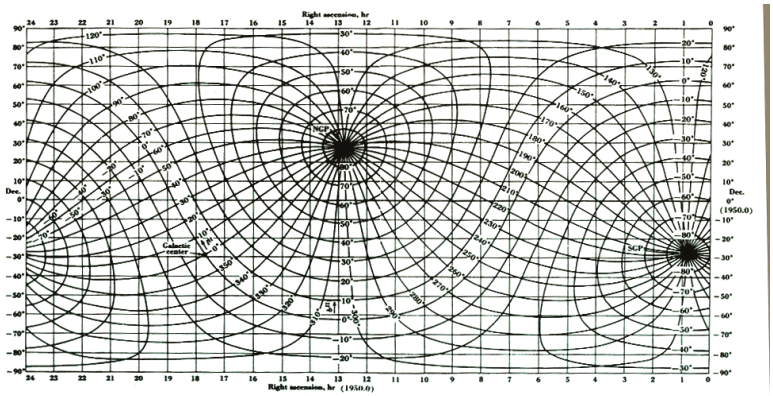


Figure 8: The conversion between equatorial and Galactic coordinates.

Honestly, it's better to use an online converter:
https://heasarc.gsfc.nasa.gov/cgi-bin/Tools/convcoord/convcoord.pl

2.4 Problems

The following problems are due next class. For these calculations, do not worry about the epoch of right ascension. The idea of these exercises is to provide you with rough estimates (accurate to a few minutes in time, and to about a degree or two in angle) necessary for preparing observing proposals. Treat all R.A. and Dec. values as the current epoch.

1) What is the approximate LST at noon on today's date?

2) If Altair (RA = 19h 50m, dec = +8deg 52m) is on your meridian, at a zenith angle of $15\,^\circ$
(a) What is your latitude (remember: the altitude of the NCP is your latitude)?
(b) What is the HA of Deneb (RA = 20h40m)?

3) How long is a source with $(\alpha, \delta) = (18\mathrm{h}, -10\,^\circ)$ above the horizon from Morgantown?

4) The Five Hundred Meter Aperture Spherical Telescope (FAST) in China uses a similar design to the now defunct Arecibo telescope to track sources as they move close to transit. Find out the observatory coordinates and maximum zenith angle of FAST and use this information to compute the LST ranges for each of the sources below:
(a) R.A. 20 h 34 m and Dec. +01 deg 45 m;
(b) R.A. 04 h 07 m and Dec. +16 deg 30 m.

5) Calculate the elevation at which the Galactic center transits as seen from:
(a) Green Bank;
(b) Parkes observatory (NSW, Australia).

3 Units and fundamental quantities in astronomy

The following notes are based on material given in Loren Anderson's Interstellar medium course. Further discussion of some of the material here can be found in "Essential Radio Astronomy" by Condon & Ransom which is freely available online and easy to find in a Google search. You should also read Chapter 3 of Carroll and Ostlie.

3.1 Intensity

The *specific intensity* of radiation the most basic observable quantity. It is essentially the surface brightness, and is appropriate for all resolved objects.

Brightness is independent of distance. Thus, the camera exposure time and aperture setting for an exposure of the Sun would be the same, regardless of whether the photograph was taken close to the Sun (from near Venus, for example) or far away from the Sun (from near Mars, for example), so long *as the Sun is resolved* in the photograph. This seems terribly wrong at first, but can easily be proven.

Brightness is the same at the source and at the detector. You can think of brightness in terms of energy flowing out of the source or as energy flowing into the detector.

In radio astronomy, we use units for specific intensity of temperature in Kelvin. It is related to the energy dE passing through an infinitesimally small area dA by:

$$dE = I_\nu \, dA \, \cos\theta \, d\Omega \, d\nu \, dt. \tag{21}$$

Here, "specific" refers to the fact that it is at a particular wavelength. We can of course rewrite this as:

$$I_\nu = \frac{dE}{dA \cos\theta d\Omega d\nu dt}. \tag{22}$$

In this expression, θ is measured normal to the surface dA and $d\Omega$ is the solid angle. The dimensions of I_ν are then erg cm^{-2} Hz^{-1} s^{-1} sr^{-1}.

Notice that we wrote the specific intensity in frequency units. I_ν has a dependence on $d\nu$, and $d\nu \neq d\lambda$. Instead,

$$d\nu = -(c/\lambda^2)d\lambda. \tag{23}$$

So combining with the above equations, we see that

$$\nu I_\nu = \lambda I_\lambda. \tag{24}$$

To get the *intensity* or *integrated intensity* we would of course integrate over frequency or wavelength so that

$$I = \int_0^\infty I_\nu d\nu = \int_0^\infty I_\lambda d\lambda. \tag{25}$$

3.2 Solid Angle

Specific intensity above included solid angles, which many students haven't yet heard of. A solid angle, measured in dimensionless steradians (sr), is simply a two-dimensional angle. Think of it as a cone spreading out from the center of a sphere to its edge. A solid angle is the area of a unit sphere such that there are 4π sr total on a sphere. The obvious application is the sky. Objects that appear larger on the sky have a larger solid angle.

The mathematical definition is

$$d\Omega = \sin\theta d\theta d\phi \tag{26}$$

or

$$\Omega = \int_S \int \sin\theta d\theta d\phi \,, \tag{27}$$

where θ and ϕ are angles in spherical coordinates and the integration is over surface S.

For a spherical solid angle, $\theta = \phi$ and when θ is small

$$\Omega \simeq \pi\theta^2 \,, \tag{28}$$

with θ in radians of course. Notice that this is just the area of a circle of radius θ. The true solid angle will be slightly smaller than this for a given value of θ, although this is almost always appropriate for astronomical measurements. The true formula is

$$\Omega = 2\pi(1 - \cos\theta) \tag{29}$$

3.3 Flux

While intensity is perfect for extended sources, we are frequently more interested in the quantity of *flux* integrated over solid angle:

$$F_\nu = \int I_\nu \cos\theta d\Omega = \int_0^{2\pi} \int_0^{\pi} I_\nu \cos\theta \sin\theta d\theta d\phi \,. \tag{30}$$

The units of flux are therefore erg cm^{-2} s^{-1} Hz^{-1}. In radio astronomy we frequently use the unit of Jansky (Jy), which is 10^{-23} erg cm^{-2} s^{-1} Hz^{-1}. Similar to the intensity, we can integrate over frequency or wavelength to get the *flux* or *integrated flux*.

In practice, when do we use spectral brightness and when do we use flux density to describe a source? If a source is unresolved, meaning that it is much smaller in angular size than the point-source response of the eye or telescope observing it, its flux density can be measured but its spectral brightness cannot. If a source is much larger than the point-source response, its spectral brightness at any position on the source can be measured directly, but its flux density must be calculated by integrating the observed spectral brightnesses over the source solid angle.

3.4 Luminosity

Intensity and flux are observable quantities, and not physical quantities. We are often more interested in luminosity, which is intrinsic to the source. The *observed* luminosity

$$L = 4\pi d^2 F\,, \tag{31}$$

where d is the distance to the source.

3.5 Distance

Distances to astronomical objects are fundamental. One such account can be found in "Measuring the Universe" by Webb (Springer/Praxis). The fundamentals that we need to know are that much of what we know about distances are based on the parallax technique applied to the nearby stars. Review section 3.1 of C&O.

3.6 Magnitudes

Magnitudes are the units of brightness, typically used in the optical and near-infrared regimes. They are always measured in a particular bandpass, for example the Johnson V-band. This allows us to compute "colors" by looking at magnitude differences. Colors are the crudest way of determining the shape of the spectral energy distribution.

Magnitudes are based on Hipparchus's classification of stars in the northern sky. Hipparchus classified stars with values of magnitudes from 1 to 6, 1 magnitude being the brightest. Because it was defined by eye, and the eye does not have a linear response, a first magnitude star is not twice as bright as a second magnitude star. Instead, astronomers later found that Hipparchus' system is roughly logarithmic, and 6th magnitude stars are roughly 100 times fainter than 1st magnitude stars. The magnitude system has two peculiarities: (1) it is defined backwards; (2) it is logarithmic.

Five equal steps in log-space (1st to 6th magnitude) result in factors of 2.512 in linear space ($100^{\Delta m/5} = 2.512^{\Delta m}$). Therefore, a 1st magnitude star is 2.512 times brighter than a second magnitude star, and a 4th magnitude star is $2.512^3 = 15.8$ times fainter than a 1st magnitude star. Another way of thinking about this is:

$$m_1 - m_0 = -2.5 \log_{10}(F_1/F_0) \tag{32}$$

or

$$F_1/F_0 = 10^{0.4(m_0 - m_1)}\,, \tag{33}$$

where F_1 and F_0 are the fluxes and m_0 and m_1 are the magnitudes at wavelengths or frequencies "0" or "1". How can we actually use this system? We need a reference star of

known flux and magnitude. Any star will do, but two commonly used ones are Vega and the Sun. Instead of arbitrary wavelengths, we usually use photometric filters. The most common filters used are the Johnson *U,V,B,R,I*, but there are now a large number of filters available. Magnitudes found using these filters are often denoted with the filter names themselves, e.g., B for m_B.

Dust attenuates star light, and this attenuation is often measured in magnitudes. Each kpc in the Galaxy produces about a magnitude of visual extinction. Star formation regions can have visual extinctions of 100, so a star would have $2.512^{100} = 10^{40}$ times less light than it would if extinction were not present. Extinction generally decreases with increasing wavelength, so it is less in the infrared and essentially absent in the radio.

3.7 Hertsprung Russell (HR) diagram

This is one of the most important diagrams in all of astronomy and a favorite question for examiners on student oral exams! If you don't know much about the HR diagram, now is the time to read about it. Look it up in C&O (Chapter 8), or find out about it online or in any good introductory astronomy text book. In class, you need to be able to draw it, describe it and connect it with some basic ideas about stars as black bodies. The next chapter will be devoted to discussing blackbody radiation in more detail. For now, you should be familiar with both the Stefan Boltzmann and Wien's laws.

3.8 Problems

Work through the following before the next class. We'll do them on the board in Russian roulette style, so please be ready to defend any one of them.

1. Draw a diagram that describes the parallax technique, and use it to derive a relationship between the distance to a star and its parallax in terms of the astronomical unit.

2. The parallax angle for Sirius is $0.379''$. Find the distance to Sirius in pc, light years, astronomical units and meters. Determine the distance modulus for Sirius.

3. What is the solid angle of a single $6''$ square pixel?

4. At what distance from a 100-W light bulb is the flux equal to the solar irradiance?

5. A star has a $B - V$ color of -1. Stars with $B - V = 0$ appear slightly blue. Does this star look more or less blue? What does that imply about its temperature?

6. Draw an HR diagram and clearly label the various types of stars on it. How are the parameters on the x and y axes measured in practice? Assuming that stars are perfect black bodies, draw lines of constant stellar radius on the diagram.

4 Interpreting Blackbody Emission

4.1 Blackbody Emission

The blackbody (Planck function) is:

$$B_\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/kT}-1}\,. \tag{34}$$

or

$$B_\lambda = \frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda kT}-1}\,, \tag{35}$$

where the function is evaluated at frequency ν or wavelength λ, and the object is at temperature T. These are shown in Figure 9.

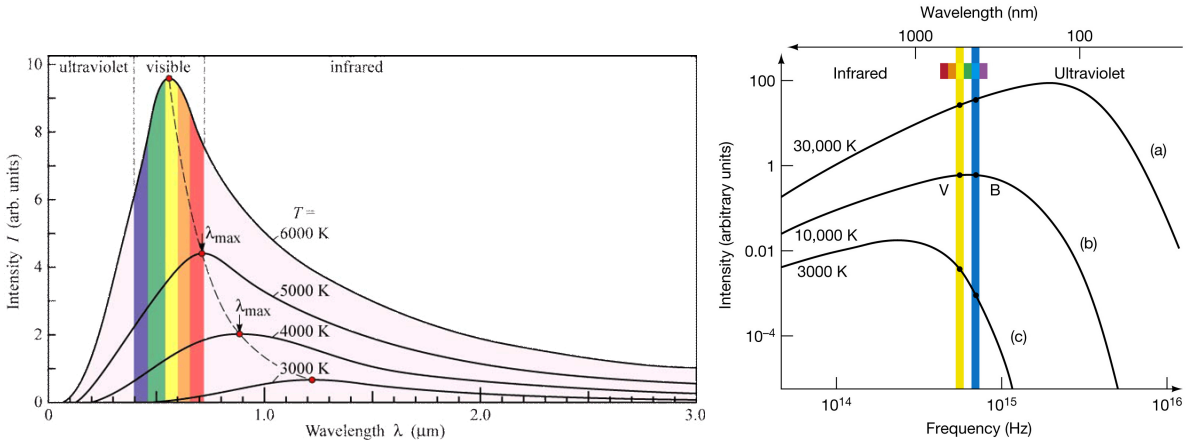


Figure 9: Blackbody curves in linear (left) and log (right)-space. Wien's Law can clearly be seen.

Students are often confused by the units: erg cm^{-2} s^{-1} Hz^{-1} sr^{-1} for B_ν or erg cm^{-2} s^{-1} cm^{-1} sr^{-1} for B_λ, where the additional "Hz" or "cm" term is the frequency or wavelength (often given in Angstroms). This also means that it is a surface brightness. For resolved objects, a surface brighness is most closely related to what your eyes see.

The fundamental observational quantity in astronomy is the specific intensity I_ν. But under what conditions is $I_\nu = B_\nu$? When the optical depth is high. Looking ahead a little to radiative transfer,

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} + B_\nu(T)(1-e^{-\tau})\,, \tag{36}$$

where $I_\nu(0)$ is the background radiation and τ_ν is the optical depth. So as $\tau \to \infty$, $I_\nu(\tau_\nu) \to 0 + B_\nu(T)(1-0) = B_\nu(T)$. To summarize, the Planck function has units of specific intensity or surface brightness, and in the limit of high optical depth, $I_\nu = B_\nu$.

There are a two important points about blackbody radiation. First, we can use Wien's

Law to determine the peak wavelength (or frequency):

$$\lambda_{\rm max} = \frac{0.2898}{T(\,{\rm K})}\ {\rm cm}\,, \tag{37}$$

or

$$\nu_{\rm max} = 5.879 \times 10^{10} T(\,{\rm K})\,. \tag{38}$$

We can derive these by setting the differential of B_λ or B_ν equal to zero. This tells us that hotter things peak at shorter wavelengths and higher frequencies. In the infrared, we have some handy rules of thumb: a 30 K cloud peaks at 100 μm, and a 100 K cloud will peak at 30 μm. Hot stars (30000 K) peak at 100 nm in the UV. The Sun (6000 K) peaks at 500 nm in the visible (green).

A second important point is that a hotter blackbody has a higher surface brightness intensity at *all* frequencies. This can be seen in Figure 9.

It is important to remember that more intensity at all frequencies does not necessarily mean more energy! Think about burners on a stove. A small hot burner will have very intense radiation. A large cooler burner will have less intense radiation. But the larger one may boil water faster because although its intensity (surface brightness) is lower, it emits more total energy. What matters is the product of the surface brightness and the emitting area.

Let's quantify this. To find the intensity (not the specific intensity), we integrate over all frequencies or wavelengths:

$$B(T) = \int_0^\infty B_\nu(T) d\nu\,. \tag{39}$$

After some math, this integral results in the expression

$$B(T) = \frac{\sigma T^4}{\pi}\,, \tag{40}$$

where σ is of course the Stephan-Boltzmann constant. In the case of an isotropic radiation field, which we can frequently assume, it can be shown that $F_\nu = \pi B_\nu$, so therefore $F = \sigma T^4$. This is of course the ***Stephan-Boltzmann Law***. We are often interested in the total luminosity of an object (in erg s^{-1} or W):

$$L = \int_S F dA\,, \tag{41}$$

the flux integrated over the emitting surface. For spherical objects, this leads to $L = 4\pi r^2 \sigma T^4$, where r is the object's radius. Thus, the total energy output is related to the surface area and the temperature.

4.2 Applications in Radio Astronomy

If we are on the right-hand (long-wavelength) side of the peak, we can Taylor expand the exponential: $e^{hc/\lambda kT} - 1 \simeq 1 + hc/\lambda kT - 1 = hc/\lambda kT$. In frequency units, we find

$e^{h\nu/kT} - 1 \simeq 1 + h\nu/kT - 1 = h\nu/kT$. We can therefore write

$$B_\lambda \simeq \frac{2ckT}{\lambda^4} \tag{42}$$

or

$$B_\nu \simeq \frac{2\nu^2 kT}{c^2} \tag{43}$$

This is known as the *Rayleigh-Jeans limit* or *Rayleigh-Jeans approximation*. We almost always assume this limit in radio astronomy. For example,

$$S_\nu = \int_{beam} I_\nu d\Omega\,, \tag{44}$$

where S_ν is the flux density [radio astronomy doesn't use F_ν for some reason], and the integration is over the telescope beam (not necessarily over the entire source!). If we approximate the spectral shape of the source as that of a blackbody, we can refer to the temperature as the *brightness temperature*, T_B, and then

$$S_\nu = \int_{beam} B_\nu d\Omega = \int_{beam} \frac{2\nu^2}{c^2} kT_B d\Omega\,. \tag{45}$$

What is the brightness temperature? It is the value that is needed to give the measured flux S_ν. Wikipedia's definition: "Brightness temperature is the temperature a black body in thermal equilibrium with its surroundings would have to be to duplicate the observed intensity of a grey body object at a frequency ν." This is of course not necessarily the kinetic temperature. *If* the source has a constant surface brightness over the telescope beam,

$$S_\nu = \frac{2\nu^2}{c^2} kT_B \Omega\,. \tag{46}$$

4.3 Using Blackbodies

We can usually assume that stars emit similarly to blackbodies, in which case we know their approximate spectral shape for a given temperature. Therefore, observations of stars using astronomical filters can give you information about the temperatures of those stars. Since the temperature and mass are related, we can get a proxy for mass.

The flux (or magnitude) that we measure depends on the filter used. In the optical we may use the U, B, and V filters. We measure the convolution of the filter transmittance and the source spectrum.

Imagine two filters placed on a blackbody curve. The flux ratio of these filters will give you some information about how the spectrum is decreasing. For example, if the flux ratio is large (the longer-wavelength filter is reading much less), the decrease is steep and we must be on the long wavelength side of a high temperature peak. If the flux ratio is small, we must be on the short wavelength side of a low temperature peak. From our

discussion of magnitudes, we know that flux ratios are called colors. Colors therefore tell you about the spectral shape, and the temperature of the object.

That colors are useful relies on the fact that stellar spectra are similar to that of blackbodies. This is obvious from Figure 10 (Figure 3.11 in Carroll & Ostlie), where the U-V and B-V colors of stars are compared to those of blackbodies.

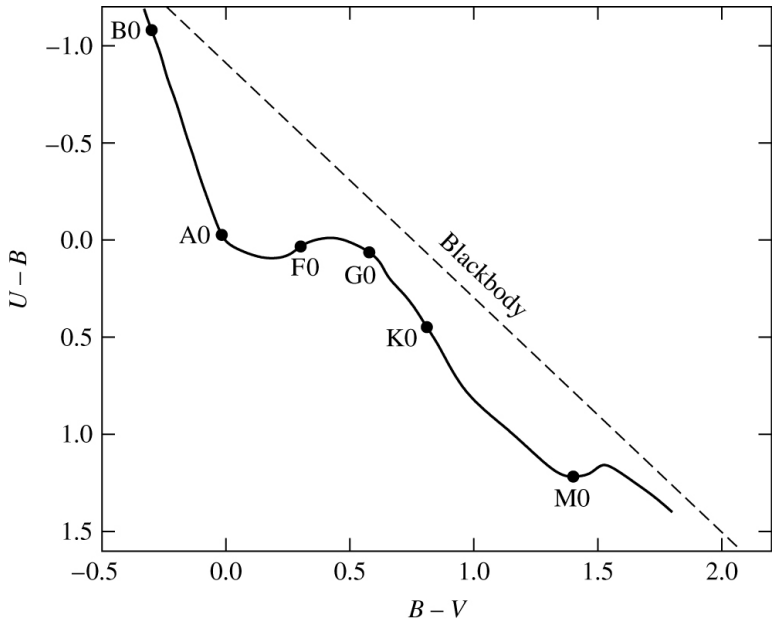


Figure 10: $B-V$ and $U-B$ colors for star of various spectral types. If you haven't encountered spectral types before, B0 is the largest and M0 are the smallest mass stars in the diagram.

Astronomers use colors as a proxy for temperatures, for example on the color-magnitude diagram, CMD. The CMD looks almost exactly like the H-R diagram because there is such a clean mapping between colors and temperatures. Why use the CMD? The quantities are entirely observable. In the H-R diagram, we often do not know the luminosity and temperature, but we can easily measure magnitudes for a bunch of stars.

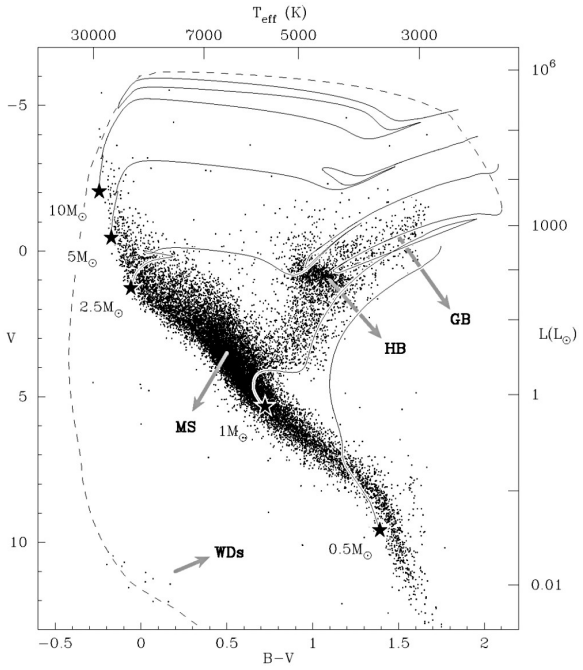


Figure 11: A Color-Magnitude Diagram (CMD). Each dot corresponds to one star. Shown are the main sequence (MS), location of white dwarfs (WD), the Horizontal Branch (HB), and the Giant Branch (GB). With time, stars evolve off the main sequence, go up into the giant branch, back down into the horizontal branch, and eventually become white dwarfs. The evolutionary tracks for stars of various masses are also shown.

4.4 Problems

1. Calculate the brightness of a blackbody radiator at a temperature of 600 K and a wavelength of 0.5 microns.

2. Verify that the maximum brightness per unit frequency of a blackbody occurs at 59 GHz $\times T$(K).

3. Making the reasonable assumption that the Sun is a blackbody emitter, calculate the number of photons with $\lambda < 912\mathring{A}$ emitted by the Sun each second.

4. Resistors are sometimes known as the "blackbody" of circuits in that they absorb any electrical power that is applied to them and convert that power into heat. In an analagous way to the generation of photons in a blackbody through the motions of charged particles in its walls, the motions of charged particles in a resistor generate electrical noise. Nyquist's formula gives the power per unit frequency interval

$$P_\nu = \frac{h\nu}{\exp(h\nu/kT) - 1},$$

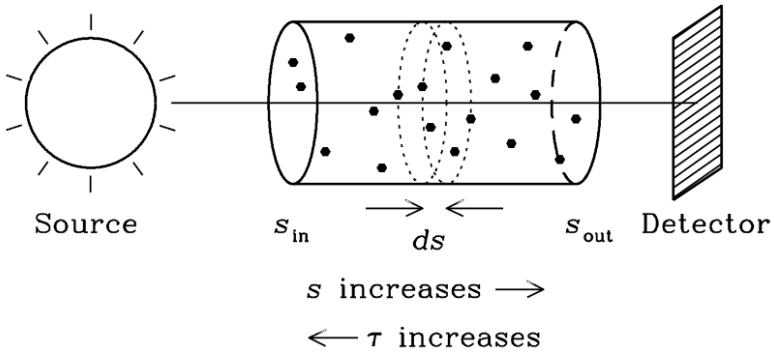
for a resistor of temperature T. Use this result to show that, in the Rayleigh-Jeans approximation, $P_\nu = kT$. This is sometimes called the Nyquist approximation.

5. Use the Nyquist approximation to find a relationship between the flux density S measured by a radio telescope of collecting area A and its so-called "antenna temperature", T. Hint: you can do this by equating the power received by the telescope to that radiated by an equivalent electrical circuit.

6. A single pulse of width 10 ms and flux density 1 Jy is detected from a pulsar located at a distance of 5 kpc from the Earth in an observation carried out at 1.4 GHz. Calculate the spectral brightness of the pulse and its corresponding brightness temperature. Hint: use a light-travel-time argument to estimate the size of the emission region and, hence, the solid angle subtended by this source in order to compute the spectral brightness.

5 Radiative transfer

Radiative transfer is the change in intensity dI_ν as radiation propagates from a source to the observer. Along the way, the emission will either be absorbed and scattered by intervening material, or it will encounter an emitting region. Consider an observation of a source shown schematically below (figure taken from ERA) where the specific intensity is modified by absorption (attenuation) and/or emission from an intervening medium.



For attenuation, we can define a "linear absorption coefficient" κ_ν with units of cm^{-1}. *This is misleading since it contains contributions from both absorption and scattering!* Note that this is not opacity or mass absorption coefficient, although both share the same notation and a similar definition! Sorry for the confusion. The amount of energy absorbed is proportional to the light intensity:

$$dI_\nu = -\kappa_\nu I_\nu ds, \tag{47}$$

where ds is the path. Absorption removes photons from the path, thus the negative sign. It is worth pointing out here that absorption excites atoms and molecules, and these atoms and molecules then re-emit. It this emission was beamed along ds there would be no change in intensity. Instead, the re-emitted light is more generally close to isotropic, so the emission is reduced.

For emission, we can define the emission coefficient j_ν as:

$$dI_\nu = j_\nu ds. \tag{48}$$

Notice that there is no dependence on I_ν, in contrast to absorption. The units of j_ν are erg cm^{-1} sr^{-1} s^{-1}.

The total change in intensity is therefore

$$dI_\nu = j_\nu ds - \kappa_\nu I_\nu ds, \tag{49}$$

or

$$\frac{dI_\nu}{ds} = j_\nu - \kappa_\nu I_\nu. \tag{50}$$

Equations 3 and 4 represent one form of the *Equation of Radiative Transfer*. This is one of the fundamental equations in astrophysics. All it is saying, however, is that the change in intensity along the path is just the emission (j_ν) minus the absorption ($\kappa_\nu I_\nu$).

Let's take the illustrative example of no emission. In this case

$$\frac{dI_\nu}{ds} = -\kappa_\nu I_\nu, \tag{51}$$

which has a solution

$$I_\nu(s) = I_{\nu,0} e^{-\kappa s}, \tag{52}$$

where $I_{\nu,0}$ is the unattenuated emission. The radiation intensity will decrease exponentially.

We can also define the dimensionless quantity of *optical depth* τ from

$$d\tau_\nu = -\kappa_\nu \, ds. \tag{53}$$

or

$$\tau_\nu = -\int \kappa_\nu \, ds, \tag{54}$$

where the integration is carried out over the path length. In most cases, we need only integrate over the source of interest. For example, if there is a gas cloud 20 kpc away that is 1 kpc thick, we may be able to only integrate over the 1 kpc of the cloud if the rest of the 20 kpc can be assumed to have no impact. For completeness,

$$I_\nu(\tau_\nu) = I_{\nu,0} e^{-\tau_\nu}, \tag{55}$$

The optical depth ranges from zero to infinity. Low values $\tau_\nu \ll 1$ are called "optically thin." These are things you can see through *at that particular frequency*. A good example is glass, which has a very low optical depth at optical frequencies, but actually has a high optical depth in the ultra-violet. High values $\tau_\nu \gg 1$ are called *optically thick*. A wall is optically thick at optical frequencies. A wall is optically thin at X-ray frequencies. Near $\tau \simeq 1$ we have to be careful - this is marginally optically thick.

If we rewrite things in terms of the optical depth, using $\frac{d\tau_\nu}{ds} = \kappa_\nu$,

$$\frac{dI_\nu}{d\tau_\nu} = \frac{j_\nu}{\kappa_\nu} - I_\nu. \tag{56}$$

We can further define the *Source function* S_ν

$$S_\nu = \frac{j_\nu}{\kappa_\nu} \tag{57}$$

Combining our expressions, we arrive a second form of the *Equation of Radiative Transfer*, this time using the optical depth and source function so that

$$\frac{dI_\nu}{d\tau_\nu} = S_\nu - I_\nu. \tag{58}$$

We will use this one from now on, because optical depth is a much better and more measurable parameter compared to actual linear depth.

In (full) thermodynamic equilibrium (TE) at temperature T, there is no change in intensity along the path and $\frac{dI_\nu}{d\tau} = 0$. In this case, $I_\nu = S_\nu = B_\nu(T)$, our old friend the Planck function. When is $I_\nu = B_\nu(T)$??? When $d\tau \to \infty$! Or in other words, when the optical depth is high, the intensity is that of a blackbody at temperature T. In this case, nothing else about the source matters, only its temperature.

This is a subtle, but extremely important point. For high optical depth sources, the only emission you can get out is that of a blackbody. You cannot for example get line emission. The source properties, aside from temperature, do not matter. The only thing you see is the *surface* emission. In fact, you only see down on average to the depth where the optical depth is unity. Think of a wall again, where you cannot determine how thick it is since you only see the paint layer (ok, so a wall actually is not a perfect blackbody since paint reflects light of different wavelengths....). Contrast this with glass. As glass get thicker, and thicker, we will notice more of a green hue. By determining *how* green it is, we can work out how thick it is. We will return to this point later.

5.1 Solutions to the Equation of Radiative Transfer

The deceptively simple equation of radiative transfer has had volumes written about its solutions. We can integrate the transfer function by multiplying by e^{τ_ν}. If we define $\tau_\nu = 0$ at $I_{\nu,0}$, we find

$$I_\nu(\tau_\nu) = I_{\nu,0}(\tau_\nu)e^{-\tau_\nu} + \int_0^{\tau_\nu} S_\nu(\tau')e^{-(\tau_\nu - \tau_\nu')}\, d\tau' \tag{59}$$

The intensity I_ν at optical depth τ_ν is the initial (background) intensity $I_{\nu,0}$ attenuated by a factor $e^{-\tau_\nu}$, plus the emission $S_\nu d\tau^{'}$ integrated over the path, itself attenuated by the factor $e^{\tau_\nu - \tau_\nu^{'}}$. This final exponent represents "self-absorption." The material itself will absorb its own radiation. "Self-absorption" refers to absorption by one species (HI, CO, etc) by that species. If background radiation from e.g. HI is absorbed by optically thick HI, this is called self-absorption. We will revisit this when we talk about HI and CO. This is known as the "formal solution to the equation of radiative transfer."

The difficulty in using Equation 59 is that in general we don't know how S varies with τ, because S depends on I, which is not known until S is known. It's a circular problem, which is why it is often solved computationally. It is worth examining this equation a bit more in limiting cases that allow us to simplify the integral:

$\tau = 0$

If the optical depth is zero, we get $I_\nu = I_{\nu,0}$, simply the background intensity back. If there is no optical depth, we get neither emission nor absorption (like a window!). This illustrates how emission and absorption are intimately related.

S constant

We can sometimes make the assumption that S_ν is a constant, so we can pull it out of the integral:

$$I_\nu(\tau_\nu) = I_{\nu,0}(\tau_\nu)e^{-\tau_\nu} + S_\nu \int_0^{\tau_\nu} e^{(-\tau-\tau')}\, d\tau' = I_{\nu,0}(\tau_\nu)e^{-\tau_\nu} + S_\nu(1 - e^{-\tau_\nu}) \tag{60}$$

The first term on the right hand side is attenuation along the line of sight. The second one is emission along the line of sight.

S constant, LTE

In Local Thermodynamic Equilibrium, LTE, $S_\nu = B_\nu(T)$, so

$$I_\nu(\tau_\nu) = I_{\nu,0}(\tau_\nu)e^{-\tau_\nu} + B_\nu(1 - e^{-\tau_\nu}). \tag{61}$$

We will discuss LTE later, but essentially it means that for a small volume we can assume a single temperature that is also reflected in the level populations of the atoms and molecules.

S constant, LTE, Radio Regime

In the radio, we use the brightness temperature instead of the intensity. They are related by $I_\nu = \frac{2\nu^2}{c^2}kT_B$. We can also use the Rayleigh-Jeans approximation $B_\nu(T) = \frac{2\nu^2}{c^2}kT$, with T here the kinetic temperature. Since these relationships both have the same constants, we can write

$$T_B = T_{B,0}e^{-\tau_\nu} + T(1 - e^{-\tau_\nu}). \tag{62}$$

Note that the use of the Rayleigh-Jeans approximation here does not imply that the material is optically thick. It just implies that the emission is still modified by the optical depth.

S constant, LTE, Optically Thin

If $\tau \ll 1$, we get emission from the background radiation, as well as from along the line of sight. We can make the Taylor expansion substitution $e^{-\tau_\nu} \simeq 1 - \tau_\nu$, so

$$I_\nu(\tau_\nu) = I_{\nu,0}(\tau_\nu)(1 - \tau_\nu) + B_\nu \tau_\nu \simeq I_{\nu,0}(\tau_\nu) + B_\nu \tau_\nu. \tag{63}$$

The first term again is the background radiation attenuated by the ISM. The second term is the Planck function modified by the optical depth of the ISM. Notice that we can still have a blackbody-like spectrum even if it is optically thin, although it is modified by the optical depth (which is less than 1). In the case that $\tau_\nu = 0$, we of course only see the background radiation. In the radio regime,

$$T_B = T_{B,0}(1 - \tau_\nu) + T\tau_\nu \simeq T_{B,0} + T\tau_\nu. \tag{64}$$

S constant, LTE, Optically Thick

If $\tau \gg 1$, $e^{-\tau_\nu} \to 0$, so

$$I_\nu = S_\nu. \tag{65}$$

If there is a blackbody in our line of sight, we don't see any emission from behind it. In radio astronomy, $T_B = T$ for optically thick emission, the kinetic temperature of the material (if in LTE).

5.2 Problems

1. Complete the derivation to justify Equation 59.

2. A signal passes through two cables with the same optical depth. The two cables have temperatures T_1 and T_2, where $T_1 > T_2$. Which cable should be connected first to obtain the lowest output power?

3. Suppose a radio astromomer observes sources at 1.3 cm wavelength at elevations between 8 and 11 degrees. If the zenith optical depth is 0.1, what are the optical depths due to the atmosphere at these elevations? The astronomer sees at most a 30% change in the optical depth over this range. Is this reasonable? Justify your answer.

4. Suppose that the emission from the Earth's atmosphere at 225 GHz is found to be 15 K at the zenith, 18 K at 60 degrees elevation, 30 K at 30 degrees elevation and 42 K at 20 degrees elevation. If the temperature of the atmosphere is 200 K, what is the zenith optical depth?

5. Show that the change in magnitude in some optical band caused by an optical depth τ is $-2.5\log_{10}(e^{-\tau})$. What optical depth would cause a magnitude change of one unit? Look up what is meant by optical extinction and relate it to this expression, show that the extinction is 1.086τ.

6. Consider observing an external galaxy which lies at some Galactic latitude b as seen from the earth. Adopting a simple slab model of the interstellar medium for our Galaxy, show that the optical depth cause by our Galaxy is proportional to cosec $|b|$. If $\tau = 1$ for an observation at the north Galactic pole. How much extinction would you expect for an observation of a galaxy at $b = 30^{\circ}$.

6 Stellar Structure and Evolution

Stars are objects that are bound by self-gravity and produce their own energy, typically through nucleosynthesis but, at some stages in life, through gravitational contraction. In this overview, we describe some very basic properties of stars and how they are observed and/or measured. We will then work through some of the key ingredients for understanding stellar stability and energy generation, and the processes through which stellar properties change as they age. Finally we will discuss stellar end-points.

6.1 Stellar Properties

The observables that allow us to model and understand stars include distance, fluxes, luminosity, temperature, color, mass, and size. Some of these can be measured directly and others much be inferred based on other quantities. For example, distances d are derived directly from parallax measurements, fluxes F with optical observations, and luminosities from fluxes and distances ($L = F/4\pi d^2$).

Temperature (or color) is measured from observations of the spectrum of a star, under the (very good!) assumption that it emits as a blackbody using Wien's Law $\lambda_{max} = 3 \times 10^6 \mathrm{nm}/T(\mathrm{K})$ where λ_{max} is the wavelength at which the blackbody spectrum peaks and T(K) is the temperature of the surface of a star in Kelvin. Given that the Sun's blackbody spectrum peaks at roughly 500 nm, we can derive a surface temperature of roughly 6000 K. If the blackbody spectrum peaks outside of the observed range, the temperature can still constrained by taking the ratio of the flux at two different wavelengths.

Size is tricky to measure directly, aside from in special systems such as eclipsing binaries. However, again assuming that stars are blackbodies we can use the Stephan-Boltzmann Law, $L \propto R^2T^4$ to calculate stellar radii. This very basic equation lies at the core of the HR Diagram, which classifies stars according to their luminosity and temperature (see Fig. 12). Oftentimes spectral classes, with O the hottest, blue-est class and M the coolest, red-est class, are used instead of temperatures. These spectral classifications are based off of the strength of hydrogen absorption lines, with A-type stars having the brightest lines and O-type stars the weakest. Fig. 13 summarizes the properties of different classes of stars.

A final and important property of stars is their composition. We define compositions in terms of the mass fractions of Hydrogen, Helium, and Metals, respectively, as X, Y, and Z (yes, we term all atoms heavier than Helium as metals). We will generally assume solar composition (roughly $X = 0.7$, Y = 0.28 , $Z = 0.02$ for all calculations. This composition will of course change throughout a star's life but aside from small (± 0.01) variations in metallicity for older (Population II) vs younger (Population I) stars, stars of all masses have comparable compositions.

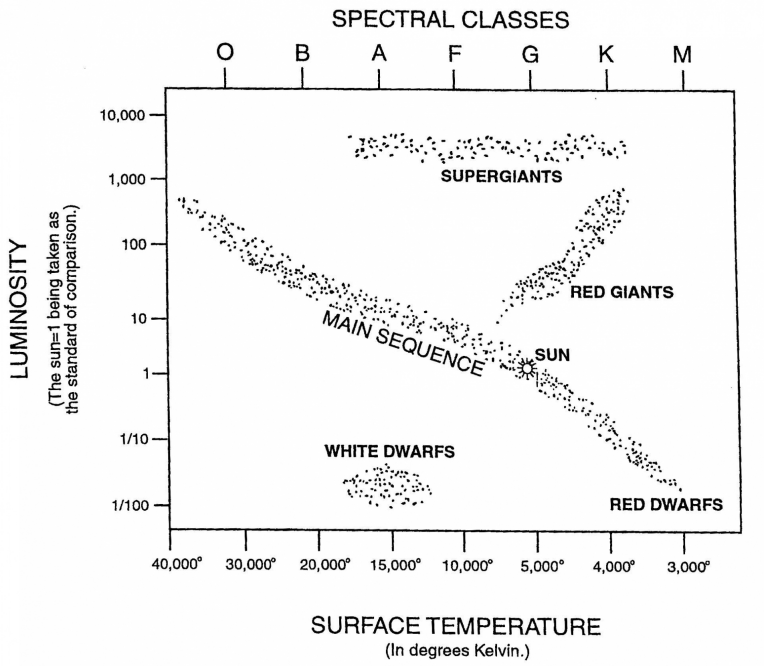


Figure 12: The HR Diagram, illustrating several different stellar populations.

There are many other stellar properties that one could discuss (e.g. shape, rotation, magnetic field, velocity, binarity). For the remainder of this writeup, however, we will assume that stars are non-rotating, non-magnetic, isolated, perfect spheres.

6.2 Energy Generation

Stars generate energy through nuclear fusion, or the combination of smaller atoms into larger atoms. This process releases energy for atomic numbers less than 56 because the masses of the constituent atoms are smaller than those of the final product. A simple calculation, given the masses of Hydrogen and Helium, shows that the conversion of four Hydrogen atoms into a single Helium atom releases $E = \Delta mc^2 = 4 \times 10^{-5}$ ergs. Given the Sun's luminosity of 4×10^{33} ergs s^{-1} we see that nuclear fusion could power the Sun for roughly 7×10^{10} years, roughly seven times greater than its estimated lifetime. Hence, nuclear fusion is an efficient and viable energy source.

However, the actual process of combining four Hydrogen atoms into a single Helium atom is not a simple task! In order to conserve charge, protons much decay into neutrons, resulting in the emission of neutrinos to ensure lepton conservation. This weak decay is the first, and slowest stage of the first Proton-Proton (PP) Chain, responsible for the majority of energy generation in the Sun. The other two PP Chains dominate in more massive (i.e. hotter) stars. Hydrogen is also converted to Helium through the CNO cycle. The inputs (four Hydrogen atoms) and outputs (a Helium atom, two neutrinos, two positrons, and photons) are the same, but in the CNO cycle Carbon, Nitrogen,

Stars: spectral types

Spectral Type	Colour	Temperature (K) Surface / core	Spectral characteristics
M	Red	3000	Molecular lines (e.g. TiO, vanadium oxide), very strong neutral metal lines
K	Orange	4000	Strong Ca lines, strong neutral metal lines, ± TiO, extremely weak hydrogen lines
G	Yellow	6000	Ca^+ lines strong, ionised metal lines weakening, neutral metal lines weakening, CH strong, hydrogen lines very weak
F	White	8000	Ionised (e.g. Fe^+, Mg^+, Si^+) and neutral metal lines, hydrogen lines weakening
A	White/blue	10 000	Hydrogen lines strong, ionised metal lines strong, weak neutral metal lines
B	Blue/UV	25 000	Strong He lines, strong hydrogen lines, Mg^+ and Si^+ lines
O	Blue/UV	50 000	Strong He^+ lines, weak He and hydrogen Balmer lines, Si^{3+}, O^{2+}, N^{2+} and C^{2+} lines

Spectral Class	Colour	Mass	Radius	Luminosity	Temperature (K)
M	red	0.1	0.1	0.001	3 000
K	orange	0.5	0.3	0.03	4 500
G	yellow	1	1	1	5 500
F	white	1.5	1.2	5.0	7 000
A	white	2.5	2	50	9 000
B	blue	10	5	10 000	17 000
O	blue	40+	20	500 000	40 000

Mass, Radius and Luminosity are given reltive to those of the Sun, which is a yellow G class star. (Mass of the Sun ≡ 1 solar mass ≡ $1M_\odot = 1.99 \times 10^{30}$ kg; radius of the Sun ≡ one solar radius ≡ $1R_\odot = 6.96 \times 108$ m; luminosity of the Sun ≡ one solar luminosity ≡ $1L_\odot = 3.83 \times 10^{26}$ W, where 1 Watt ≡ 1W ≡ 1J/s ≡ $1Js^{-1}$).

Figure 13: Properties of stars of different spectra classes.

and Oxygen act as catalysts for these reactions. The CNO cycle will dominate energy generation for stars roughly twice the mass of the Sun and higher. In Fig. 14 we illustrate these various processes. Note that comparison of the thermal energies of particles at the temperatures listed in Fig. 14 and the Coulomb barriers between the two positively charged particles of interest shows that nuclear reactions should be impossible! Quantum mechanical tunneling is necessary for any nuclear reactions to occur.

Stars spend nearly all of their lives on the 'main sequence', converting Hydrogen to Helium. Once they deplete the Hydrogen in their cores, their cores will contact and get hotter until (if a star is massive enough), fusion of heavier elements will be possible. The Sun will mostly likely end its life with a Carbon/Oxygen core, while the most massive

Process	Fuel	Major products	Approximate Temperature (K)	Approximate Minimum Mass (solar masses)
Hydrogen burning	Hydrogen	Helium	1-3 x 10^7	0.1
Helium burning	Helium	Carbon, Oxygen	2 x 10^8	1
Carbon burning	Carbon	Oxygen, Sodium, Magnesium	8 x 10^8	1.4
Neon burning	Neon	Oxygen, Magnesium	1.5 x 10^9	5
Oxygen burning	Oxygen	Magnesium to Sulfur	2 x 10^9	10
Silicon burning	Magnesium to Sulfur	Elements near Iron	3 x 10^9	20

Table taken from *Introductory Astronomy & Astrophysics* Third Edition, Zeilik, Gregory and Smith, Saunders College Publishing

Figure 14: Nuclear fusion processes and temperatures at which they dominate.

stars will convert all of the atoms in their cores to Iron.

Reaction rates for nuclear reactions can be calculated as $r_{ix} = n_i n_x \langle \sigma v \rangle$ s^{-1}cm^{-3}, where n_i and n_x are the number densities of the incident and target particles and $\langle \sigma v \rangle$ (often referred to as the "Gamow factor") is the average of the reaction cross-section and velocity. We assume that the particles are in local thermodynamic equilibrium. This then allows us to use the Maxwell-Boltzmann distribution for velocities. Calculating cross-sections is more complicated, but there are several critical terms. These include a E^{-1} dependence for the classical "size" of a particle (from the de Broglie wavelength), a "shape factor" (denoted by S) which accounts for "chemistry", and the probability of quantum mechanical tunneling. This probability goes like $\exp(-b/E^{1/2})$, where b is a constant that is larger for higher masses and charges.

In the end, the Gamow factor $\langle \sigma v \rangle$ will be an integral over all energies of the product of the tunneling probability and $\exp(-E/kT)$ from the Maxwell Boltzmann distribution (the E^{-1} term from the classical "size" of the atom will cancel the E dependence of the Maxwell Boltzmann integral). One of these exponentials (the tunneling probability) will increase with energy, while the other (the Maxwell Boltzmann distribution) will decrease with energy, leading to a roughly Gaussian shaped final product (see Fig. 15). Note that reactions happen in the extreme tail of the Maxwell-Boltzmann distribution.

It is impossible to calculate this double-exponential integral for $\langle \sigma v \rangle$ analytically, but one can use a Gaussian approximation to derive an analytical solution. These reaction rates r_{ix} are then typically converted into energy generation rates $\epsilon = r_{ix} Q/\rho$, with units of ergs g^{-1}s^{-1}, where Q is the amount of energy released per reaction. This is then approximated as a power-law, valid within a narrow temperature range, so that the final expressions we rely on have the form $\epsilon \propto \rho^{\lambda} T^{\nu}$. As the charges of the interacting atoms

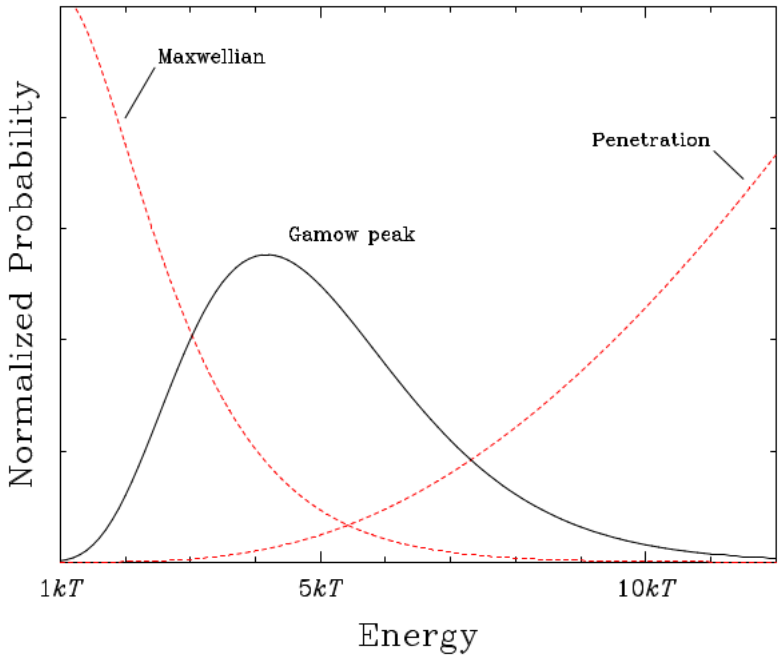


Figure 15: The Gamow factor for nuclear reactions.

increase, ν will also increase due to the higher Coulomb barrier, and consequent higher particle energies required for tunneling.

6.3 The Four Equations of Stellar Structure

Stellar modeling requires four basic equations. The first – the mass equation – simply describes M_r, the mass interior to a radius r, for a sphere with a particular density dependence $\rho(r)$.

$$M_r = \int 4\pi r^2 \rho(r) dr \tag{66}$$

The second equation of stellar structure – the equation of hydrostatic equilibrium – can be derived by equating an inwardly directed gravitational force and an outwardly directed force due to a source of internal pressure to dictate how pressure must vary with radius in a star in equilibrium.

$$\frac{dP}{dr} = -\frac{GM_r\rho(r)}{r^2} \tag{67}$$

The third equation of stellar structure – the luminosity equation – dictates how luminosity

produced from within a radius r varies with radius for an energy generation mechanism with a particular ϵ (or a particular density and temperature dependence).

$$\frac{dL_r}{dr} = 4\pi r^2 \rho(r) \epsilon dr \tag{68}$$

m

The fourth equation of stellar structure provides the relationship between luminosity and the temperature gradient in a star as

$$L_r = -\frac{4\pi r^2 c}{3\kappa\rho} \frac{daT^4}{dr} \tag{69}$$

where κ is the opacity. Opacity is defined such that the final intensity of light I_f is related to the initial intensity of light I_i through the relation, $I_f = I_i \exp(-\kappa\rho s)$, where s is the path length the light has traveled. From this expression, one can see that opacity has units of cm^{2}g^{-1}, or effective area per unit mass.

6.4 Energy Transport

Energy is transported in stars in three main ways: radiation, convection, and conduction. Because of its T^4 dependence, radiation will be important at high temperatures. As conduction is energy transport by particles, with energy proportional to ρT, it will be more important at high densities and lower (relative) temperatures. Convection will be important for energy transport whenever the temperature differential (dT/dr) is large, allowing convective bubbles to stay hotter than their surroundings as they rise. Note that the total opacity κ_{tot} is the harmonic sum of the radiative and conductive opacities, such that $1/\kappa_{tot} = 1/\kappa_{rad} + 1/\kappa_{cond}$. This means that even if the radiation opacity is very high, energy can still be transported through conduction if the conductive opacity is low (and vice versa). However, different sources of radiative opacities add linearly. For instance, opacities due to free-free, bound-bound, bound-free, and Thompson scattering would add like $\kappa_{rad} = \kappa_{ff} + \kappa_{bf} + \kappa_{bb} + \kappa_{TS}$ because if photons have been absorbed or scattered by one of these processes, the opacities of the other processes are not relevant.

Radiation and convection are the dominant processes of energy transport for main sequence stars, while conduction becomes important for degenerate stars (i.e. white dwarfs and neutron stars). In order to determine whether radiation or convection will dominate energy transport, one must calculate dT/dr throughout the star. We find that dT/dr is high in regions where Hydrogen and Helium are not fully ionized (or at temperatures of less than $\sim 10,000$ K). It is also high in regions where the CNO cycle dominates energy generation, because of its steep temperature dependence. Therefore, radiation is the dominant energy transport mechanism at the cores of low-mass stars while convection is

dominant in the cores of high-mass ($> 2\ M_\odot$) stars. Convection is also important in the outer "ionization zones" of both types of stars, with these zones extending deeper below the stellar surface for lower mass stars (see Fig. 16).

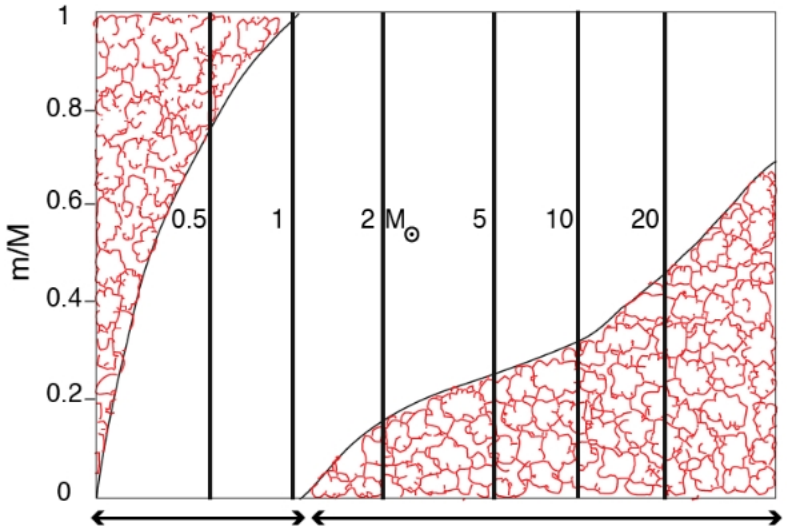


Figure 16: Low-mass stars have inner radiation and outer convective zones, while high-mass stars have inner convective and outer radiative zones.

6.5 The Virial Theorum and Equations of State

The virial theorum tells us that, for any system of particles under an external force in equilibrium, $2K = -\Omega$, where K is the total kinetic energy and Ω is the total potential energy. This theorum has many applications in stellar structure. For example, it can be used to calculate the average temperature of a star or system or particles. One useful expression of the virial theorum is

$$-\Omega = \int \frac{3P}{\rho} dM_r \tag{70}$$

Given a simple equation of state that relates density and internal energy, $P = (\gamma - 1)\rho E$, where E has units of ergs g^{-1}, we can then write the virial theorum as $-\Omega = 3(\gamma - 1)U$ where U is the total internal energy (the energy per unit mass integrated over mass). For an ideal gas, $\gamma = 5/3$ and therefore $K = U$. For gases dominated by radiation pressure, $\gamma = 4/3$. As we will show in Sec. 6.8, for degenerate gases, $\gamma = 5/3$ in the non-relativistic case and $4/3$ in the relativistic case.

The virial theorum can be used to calculate criteria for stability for stars. Importantly, in order for a star to be stable, the total energy $W = U + \Omega$ must be less than zero. One can then show that

$$W = \frac{3\gamma - 4}{3(\gamma - 1)}\Omega \tag{71}$$

This implies that a star in hydrostatic equilibrium must have $\gamma > 4/3$.

6.6 Homology Relationships

"Homology" relationships involve using simple dimensional analysis to determine how physical quantities in a system depend on each other. In our case, we aim to determine how the properties of a star (such as density, temperature, luminosity, etc) depend on stellar mass. One such relationship is $\rho \propto M/R^3$. Taking logarithm differentials of both sides, we find the first input to our problem

$$d\ln M = 3d\ln R + d\ln\rho \tag{72}$$

The second input to our problem is the luminosity equation, which tells us that $dL_r/dM_r = \epsilon$, with $\epsilon \propto \rho^\lambda T^\nu$. Taking logarithmic differentials of both sides, we can write this as

$$-d\ln M = -d\ln L + \lambda d\ln\rho + \nu d\ln T \tag{73}$$

We assume that pressure depends on density and temperature like $P \propto \rho^{\chi_\rho} T^{\chi_T}$. Setting this equal to $P \propto M^2/R^4$ from hydrostatic equilibrium, and taking logarithmic differentials, we find

$$2d\ln M = 4d\ln R + \chi_\rho d\ln\rho + \chi_T d\ln T \tag{74}$$

Finally, our fourth equation of stellar structure and an opacity that is expressed as $\kappa \propto \rho^n T^{-s}$ provide our last equation

$$d\ln M = 4d\ln R = nd\ln\rho - d\ln L + (4+s)d\ln T \tag{75}$$

With four equations and four unknowns, we can solve this system of equations in order to calculate α_R, α_T, α_L, and α_ρ, where $R \propto M^{\alpha_R}$, etc. In order to determine these quantities, we must make some assumptions. First, we assume that pressure is dominated by ideal gas pressure such that $\chi_\rho = \chi_T = 1$. Second, we assume that opacity is dominated by Thompson scattering of photons off of electrons such that $n = s = 0$. Third, we assume CNO cycle energy generation such that $\lambda = 1$ and $\nu = 15$. We can then calculate $\alpha_R = 0.78$, $\alpha_T = 0.22$, $\alpha_L = 3$, and $\alpha_\rho = -1.33$. These expressions are roughly correct for

the high mass $> 2\ M_\odot$ population. For lower mass stars, convection is more important for energy transport in the outer layers, and this treatment is not sufficient.

6.7 Distribution Functions, Energy Levels, and Ionization

The relation between the number density of some species in coordinate-momentum space can be expressed as

$$n(p) = \frac{1}{h^3} \sum_j \frac{g_j}{\exp([-\mu + E_j + E(p)]/kT) \pm 1}. \tag{76}$$

summed over j states and with "+" in the denominator for Fermions and "-" in the denominator for Bosons. Other terms in this equation are the chemical potential μ, the energy of the state E_j, kinetic energy $E(p)$, and degeneracy g_j.

For photons, $g_j = 2$, μ and E_j are zero, and $E(p) = pc$. Integrating over momentum space gives us the number density

$$n = \frac{8\pi}{h^3} \int_0^\infty \frac{p^2 dp}{\exp(pc/kT) - 1} \mathrm{cm}^{-3}. \tag{77}$$

From this one can calculate the radiation pressure $P_{rad} = aT^4/3$ and radiation energy density $E_{rad} = aT^4$, where $a = 8\pi^5 k^4/15c^3h^3$. Is is also possible to recover Wien's Law and the Stephan-Boltzmann law through calculation of the energy density per unit wavelength.

This distribution function is more complicated for Fermions, depending specifically on the species. One critical expression derived from the distribution function, however, relates the number densities of particles in two different energy states to the temperature

$$\frac{n_1}{n_2} = \frac{g_1}{g_2} e^{-(E_1 - E_2)/kT}. \tag{78}$$

The expression shows that levels become more sparsely populated as energy increases, but that higher energy states are more likely at higher temperatures.

One can also use this distribution function to derive the Saha equation for the ionization fraction y dependence on density and temperature

$$\frac{y^2}{1-y} = \frac{4.01 \times 10^{-9}}{\rho} T^{3/2} e^{-1.578 \times 10^5/T}. \tag{79}$$

This shows that material becomes more ionized at higher temperatures and lower densities.

6.8 Degeneracy and Compact Objects

Degeneracy is a natual outcome of the Pauli exclusion principle: as a material becomes denser, and the spacing between particles decreases, they are forced into higher and higher energy states. When all of the energy states up until the Fermi energy E_F are full, we say a gas is completely degenerate. This will never happen in principle, as it requires a temperature of zero. However, when $E_F > kT$, material will behave like a degenerate gas. As the exponential term in the distribution function is a step function (with all energy states up to a certain level occupied and none beyond), the number density in a degenerate gas is easy to derive as

$$n = \frac{8\pi}{h^3} \int_0^{p_F} p^2 dp \tag{80}$$

where p_F is the Fermi momentum. Defining $x_F = p_F/mc$, one can write the equation of state for a degenerate electron gas as

$$\frac{\rho}{\mu_e} = Bx^3 \tag{81}$$

where $B = (8\pi/3N_A)(h/m_e c)^{-3} = 9.7 \times 10^5$ g cm^{-3}, N_A is Avogadro's number, and μ_e is the ratio of the number of nucleons to electrons in the species. The transition from non-relativistic to relativistic mechanics will take place when $x \simeq 1$ or at densities of roughly 10^6 g cm^{-3}. This is the typical density of a white dwarf, and also close to the core density at which the "Helium flash" takes place in post-main sequence evolution. White dwarfs must therefore be treated relativistically.

For neutron star matter, with m_e replaced by m_n, $B = 6 \times 10^{15}$ g cm^{-3}, well above typical neutron star densities of 2×10^{14} g cm^{-3}, indicating that neutron stars can be treated non-relativistically. As with photons and atoms, one can calculate the pressure and energy of a degenerate gas. One finds that for non-relativistic neutrons or electrons, $P_e \propto E_e \propto \rho^{5/3}$ and for relativistic electrons, $P_e \propto E_e \propto \rho^{4/3}$. Hence, degenerate gases act like ideal gases in the non-relativistic limit and like photon gases in the relativistic limit.

From the virial theorum in the form $3(\gamma - 1)U = -\Omega$, assuming a uniform density sphere and $U = VE_e$, where V is the volume, one can derive a mass-radius relation of $M \propto R^{-3}$ for the non-relativistic regime. In the completely relativistic regime, the radial dependence will drop out of the expression, leading to a maximum mass limit. For white dwarfs, this limit is $1.456(2/\mu_e)^2\ M_\odot$, otherwise known as the Chandrasekhar mass. For

neutron stars, this simple treatment is not sufficient to arrive at a maximum mass estimate due to the unknown equation of state and the influence of the strong force.

6.9 Post Main Sequence Evolution

The lifetime of a star on the main sequence is a strong function of its mass, varying roughly as the third to fourth power of the mass (while more massive stars have larger supplies of Hydrogen, their luminosities are *much* higher than lower mass stars). Once a star exhausts the hydrogen in its core, the core will contract, heating up, and the outer layers will expand, cooling down. Eventually the core will become hot enough for fusion processes at higher atomic numbers. The triple-α process will convert Helium to Carbon and, if the core gets hot enough, heavier and heavier elements will be created. Stars follow complicated tracks on the HR diagram during this evolution, as shown in Fig. 17. More massive stars will complete this entire process faster and will not experience a Helium flash (as their cores are not degenerate).

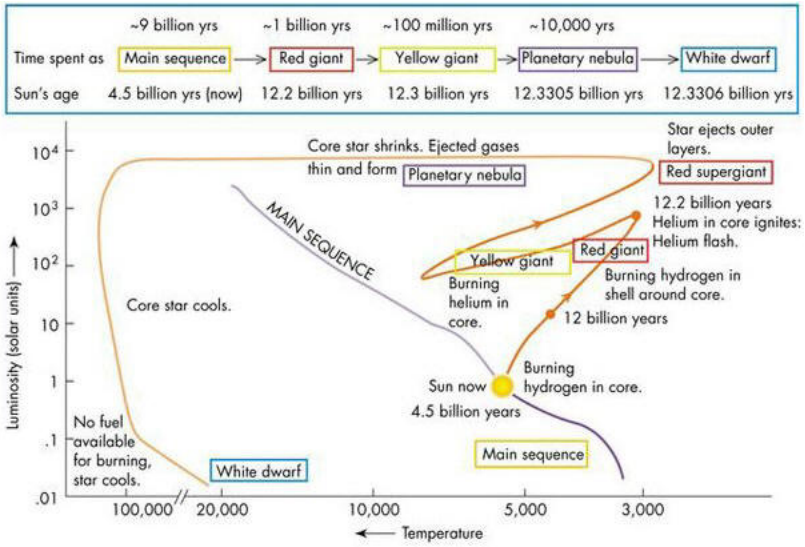


Figure 17: The path followed by a solar-mass star on the HR diagram.

The end-point of a star will be determined by its initial mass, with stars of $< 8\ M_\odot$ becoming white dwarfs, starts with masses between 8 and 20 $M_\odot$ becoming neutron stars, and more massive stars becoming black holes. Assuming any reasonable initial mass function, the vast majority of stars will end their lives as white dwarfs.

6.10 Nuclear Fusion questions

1. List all of the reactions in the PP I reaction chain. How much energy is released from each step? The masses of H, ${}^2_1$H, and ${}^3_2$He are 1.0074 amu, 2.0141 amu and 3.0160 amu,

respectively.

2. What fraction of the Sun's energy is produced from the PP I, PP II, and PP III reaction chains, respectively?

3. How does the PP chain energy generation rate depend (approximately) on density and temperature? If the Sun's density were increased by a factor of two, by how much would the PP reaction rate increase? If the Sun's temperature were increased by a factor of two, by how much would the PP reaction rate increase?

4. Assuming that the luminosity of the Sun is constant, calculate an upper limit (assuming all of the hydrogen in the Sun can be converted to helium) to the length of time that the Sun could power itself through fusion of hydrogen? Assume a hydrogen mass fraction $X = 0.73$.

5. List the two reactions of the triple-α process that will power the Sun once the hydrogen in the core is exhausted. How much energy is released from each step?

6. How does the triple-α reaction energy generation rate depend (approximately) on density and temperature? Explain and constrast to the dependence of the PP chain. If the Sun's density were increased by a factor of two, by how much would the triple-α reaction rate increase? If the Sun's temperature were increased by a factor of two, by how much would the triple-α reaction rate increase?

7. The energy generation rates for the PP chain and the CNO cycle are:

$$\epsilon_{\rm PP} \simeq \frac{2.4 \times 10^4 \rho X^2 e^{-3.380/T_9^{1/3}} {\rm erg\ g^{-1} s^{-1}}}{T_9^{2/3}} \tag{82}$$

$$\epsilon_{\rm CNO} \simeq \frac{4.4 \times 10^{25} \rho X Z e^{-15.228/T_9^{1/3}} {\rm erg\ g^{-1} s^{-1}}}{T_9^{2/3}} \tag{83}$$

In these expressions, T_9 is the temperature in units of 10^9 K and ρ is the density in units of g cm^{-3}. Assume that the Sun's central density and temperature are 100 g cm^{-3} and 1.5×10^7 K, respectively, and that $X = 0.73$, the helium mass fraction $Y = 0.25$, and the "metals" mass fraction is $Z = 0.02$. Assuming that the Sun gets 90% of its luminosity from the PP chain and 10% from the CNO cycle, what fraction of the Sun's total mass is participating in nuclear fusion?

8. Given the expressions above, show that the scaling relation you quoted in question 3 is roughly correct at Solar density and temperature.

9. At what temperature are the reaction rates for the PP chain and CNO cycle equal? Assume solar abundance ($X = 0.73$, $Y = 0.23$, $Z = 0.02$) and a central density of 100 g cm^{-3}.

Given that temperature scales with mass as $M^{0.22}$, what mass does this correspond to?

10. Sirius B, the first white dwarf discovered, has a mass a 1 $M_\odot$, density of 10^6 g cm^{-3}, temperature of 10^7 K, and luminosity of 0.03 $L_\odot$. Assume that the entire luminosity is produced by hydrogen fusion to calculate an upper limit on X. Use the equations for the energy generation rates of the P-P chain and CNO cycle. Assume $X_{\rm CNO} = 1$.

6.11 Stellar Structure questions

1. Estimate the ideal gas pressure and radiation pressure at the center of Sirius B. Compare these values with the ideal gas pressure and radiation pressure at the center of the Sun (you can assume a fully ionized hydrogen core for simplicity).

2. The equation of hydrostatic equilibrium is $dP/dr = -GM_r\rho/r^2$, where M_r is the mass interior to a radius r. Derive this equation by equating the inwards force due to gravity to the outwards force due to pressure.

3. Starting with the equation of hydrostatic equilibrium, estimate the pressure at the center of the Sun by evaluating at $r = R/2$, where R is the radius of the Sun, and assuming constant density. How does this estimate compare to the ideal gas and radiation pressure calculated in equation 1? What might be responsible for the discrepancy?

4. What is the virial theorem? Show that the virial theorem works for a point mass m in a circular orbit around a mass M.

5. Use the virial theorem to derive the Jean's mass, or the minimum mass needed for a cloud to collapse under the force of gravity, in terms of temperature and density. You may assume the cloud is composed purely of hydrogen.

6. Assume that a diffuse cloud composed entirely of hydrogen has a temperature of 10 K and number density of 10^3 cm^{-3}. What is the Jean's mass? Assuming uniform density, what is the Jean's radius (the maximum radius needed for collapse)?

6.12 Stellar Evolution Questions

1. Assume that the number of stars with different masses is determined by the Salpeter initial mass function where $dN/dM \propto M^{-2.35}$. Assuming that there are 300 billion stars in our galaxy, calculate the number of stars with masses less than 1 $M_\odot$, between 1 and 10 $M_\odot$, and greater than 10 $M_\odot$.

2. Describe the lifetimes, masses and end products for 1 $M_\odot$, 10 $M_\odot$, and 30 $M_\odot$ stars. Sketch the paths that they will take on the HR diagram over their lifetimes. What are the fractional lifetimes spent on the main sequence and after the main sequence?

7 Basic Statistics for Astronomy

These notes are designed to be an introduction to statistical analysis used in astronomy. It is by no means a rigorous introduction to basic statistics but is designed to be a practical one. That means that certain fundamental concepts have been ignored. Additionally, there are a number of things that I provide that aren't going to be covered, and I make note of those sections; they are there entirely for your use. As with any mathematics, the notation takes some getting used to.

Much of the statistics presented here can be found in some form in Gregory's *Bayesian Logical Data Analysis for the Physical Sciences*, which provides an excellent perspective of statistics as it relates to probability, shows both frequentist and Bayesian methodologies, and has excellent examples often related to astronomy. There are other books containing a wide range of basic mathematical support, such as Boas' *Mathematical Methods in the Physical Sciences* or Korn and Korn's *Mathematical Handbook for Scientists and Engineers* that I recommend as concise reference material.

7.1 Basic Tools and Jargon

Some useful phrases. Note that many of these will not show up here but I supply them anyway in case you have questions or want to look further for yourself.

- Random variables
- PDF, CDF, characteristic function
- Mean (arithmetic, geometric, harmonic), median, mode, sample vs population mean
- Moments, Central Limit Theorem
- Time series analysis: power spectrum, autocorrelation (or autocovariance) function, structure functions
- ARMA, Markov processes
- Frequentist, Bayesian, Likelihood
- Confidence Intervals
- Degrees of freedom, null hypothesis

7.2 Probability

Random variables: Random variables assign numbers to events:

$$\begin{aligned} a &\rightarrow X(a) \\ \text{event} &\rightarrow \text{number} \end{aligned} \tag{84}$$

(Cumulative) Distribution Function (CDF): The CDF provides the probability that given events a, $X(a) \leq x$, i.e.,

$$F_X(x) = P\{X \leq x\} \tag{85}$$

Properties:

1. $F_X(x)$ is a monotonically increasing function of x.
2. $F(-\infty) = 0,\ F(+\infty) = 1$
3. $P\{x_1 \leq X \leq x_2\} = F_X(x_2) - F_X(x_1)$

Probability Density Function (PDF): We see these most often. They describe the probability of a given value (sort of, see point 1).

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{86}$$

Properties:

1. $f_X(x)dx = P\{x \leq X \leq x + dx\}$
2. $\int_{-\infty}^{+\infty} f_X(x)dx = F_X(\infty) - F_x(-\infty) = 1$

Note that random variables need not be continuous. Consider a coin flipping, X can either be 1 for heads, -1 for tails. Then:

$$f_X(x) = \frac{1}{2}\left[\delta(x+1) + \delta(x-1)\right] \tag{87}$$

$$F_X(x) = \frac{1}{2}\left[\Theta(x+1) + \Theta(x-1)\right] \tag{88}$$

where δ is a Dirac delta function and Θ is the Heaviside step function.

Also consider a Gaussian distribution with mean μ and variance σ^2, often written as $X \sim \mathcal{N}(\mu, \sigma^2)$:

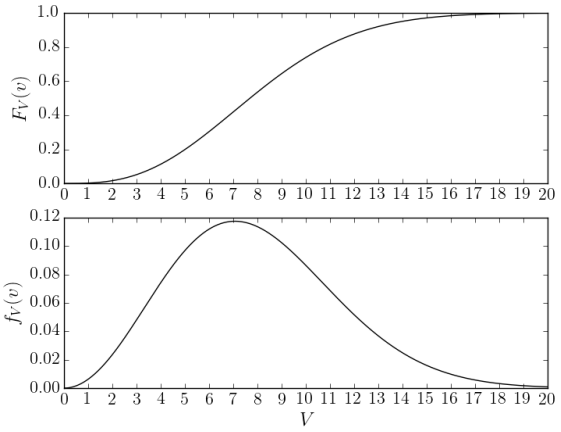
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{[x-\mu]^2}{\sigma^2}\right) \tag{89}$$

We'll get to the mean and variance shortly but you're probably at least somewhat familiar with the concepts already.

An example figure of the Maxwell-Boltzmann distribution, whose PDF is given by:

$$f_V(v) = \sqrt{\left(\frac{m}{2\pi kT}\right)^3} 4\pi v^2 e^{-\frac{mv^2}{2kT}} = \sqrt{\frac{2}{\pi}} \frac{v^2}{x^3} e^{-\frac{v^2}{2x^2}} \tag{90}$$

where $x^2 = kT/m$. Below is a plot for $x = 5$.



We can use this picture to illustrate two ways of quantifying the "average" value of the distribution. One is the median, which splits the distribution equally into a lower and upper 50% probability. You can use the CDF to easily read off this value (the x value where $y = 0.5$). The other is the "mode", which is the most likely value, which can be obtained from the peak of the PDF. In the case of the Maxwell-Boltzmann Distribution, the mode is $v_p = \sqrt{2}x$ and the median is $1.0877 v_p$.

Empirical Distribution function: This is worth mentioning briefly from a practical standpoint. If you sort your measurements X_i ($i \in 1, 2, \cdots, N$), you can estimate the CDF such that you start at zero from negative infinity, and for every one of N measurements, you increase the CDF by $1/N$ at each X_i. You'll end at positive infinity with a value of 1 and you'll get some kind of multi-step-function looking thing at the end. This approach is easier than estimating the PDF in that you typically need to define bins for your data, and although there are ways around it, it is problematic.

Moments: We use brackets to denote the expected value, the average of an *ensemble*. Therefore, the mean of X is simply $\langle X \rangle$.

For an arbitrary power n, the n-th (raw) moment is

$$E[X^n] \equiv \langle X^n \rangle = \int dx x^n f_X(x) \tag{91}$$

The expected value of X is the mean, i.e. $\bar{x} = \langle X \rangle$. The variance is $\sigma_x^2 = \left\langle X^2 \right\rangle - \langle X \rangle^2$. You may also see central moments defined as $\mu_n = \left\langle (X - \langle X \rangle)^n \right\rangle$, where we are simply subtracting the mean prior to finding the expected value.

Two higher-order moments you may see are the skewness and kurtosis. The skewness describes if there is a tail to the distribution (positive skewness implies a tail in the distribution at higher values, as with the Maxwell-Boltzmann distribution), whereas the kurtosis describes how "peaked" the distribution is. Moments are useful in comparing distributions without actually knowing the PDFs. They are also useful because often calculations of moments are vastly easier than calculations of full PDFs.

Functions of a random variable: We can map one random variable onto another random variable after some transformation $X \to g(X)$. That is, we will transform the variable X by some function $y = g(x)$. The property that probability is conserved will be useful:

$$\begin{aligned} |f_Y(y)dy| &= |f_X(x)dx| \\ \to f_Y(y) &= f_X(x)\left|\frac{dx}{dy}\right| \\ &= \frac{f_X(x)}{|dy/dx|} = \frac{f_X(x)}{|dg(x)/dx|} \\ &= \frac{f_X(g^{-1}(y))}{|dg(g^{-1}(y))/dx|} \end{aligned} \tag{92}$$

In the last step, we've written the PDF of Y purely as a function of y, which makes some sense.

Consider an example in which we want to scale and shift a random variable, i.e. $g(x) = ax + b$. Then:

$$\begin{aligned} Y &= g(X) = aX + b \\ \implies \frac{dg}{dx} &= a \\ \text{and } x &= g^{-1}(y) = \frac{y-b}{a} \\ \implies f_Y(y) &= \frac{f_X(x)}{|dg(x)/dx|} = \frac{f_X(g^{-1}(y))}{|dg(g^{-1}(y))/dx|} = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) \end{aligned} \tag{93}$$

This is how random distributions are generated on a computer, by transforming a uniform distribution of random numbers (what the computer can generate) into a specific distribution (normal, exponential, etc.).

Let's consider a slightly different example. Suppose we want to generate exponentially-distributed numbers on our computer, i.e., we want $f_Y(y) = \exp(-y)$ (where $y \geq 0$), and we start with a uniform distribution:

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1; \\ 0, & \text{otherwise.} \end{cases} \tag{94}$$

Let's choose $y = g(x) = -\ln(x)$. Then:

$$\begin{aligned} \frac{dg}{dx} &= -\frac{1}{x} \\ \text{and } x &= g^{-1}(y) = \exp(-y) \\ \implies f_Y(y) &= \frac{f_X(g^{-1}(y))}{|dg(g^{-1}(y))/dx|} = \frac{1}{|-1/\exp(-y)|} = \exp(-y) \end{aligned} \tag{95}$$

Therefore, if we generate uniformly distributed values and take the negative natural logarithm of them, we will get exponentially-distributed numbers.

Joint Random Variables: I'm going to rush through this subsection but thought it would be good for you to at least have a minimal reference. For two random variables, we can write the joint distribution function as

$$F_{XY}(x, y) \equiv P\left\{X \leq x, Y \leq y\right\} \tag{96}$$

whose joint (2D) PDF is

$$f_{XY}(x, y) \equiv \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}. \tag{97}$$

If the variables are independent, then we have $f_{XY} = f_X(x) f_Y(y)$. Generally, we can obtain the 1D PDF of one of the variables by integrating the joint PDF over the other variable,

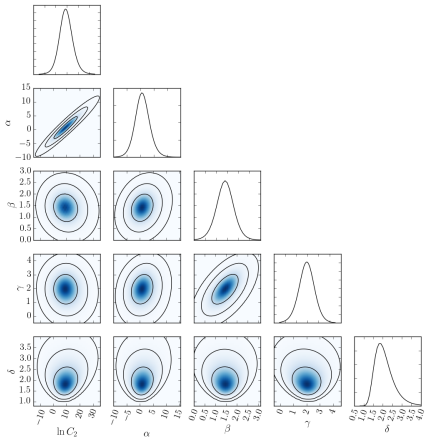
$$f_X(x) = \int dy f_{XY}(x, y). \tag{98}$$

We can write the joint probability in terms of a conditional probability, that is, a probability of one thing happening *given* another. Usually this is written as $f_X(x|c)$ where this reads "the probability of x given c. I write c to mean "constant" but there's no need for the value to be constant. Technically, we could write the Gaussian PDF as $f_X(x|\mu, \sigma)$! In terms of the joint PDF

$$f_{XY}(x, y) = f_{Y|X}(y|x) f_X(x) = f_{X|Y}(x|y) f_Y(y). \tag{99}$$

Here, $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$ are conditional probabilities (e.g. the first is the probability of Y given $X = x$), and $f_X(x)$ and $f_Y(y)$ are the marginal distributions. You often see the joint PDF and the marginal PDFs in multidimensional likelihood plots. Your model may contain many parameters and the peaks in probability are some complex surface in that space. It is often simpler to numerical integrate over each of the dimensions until your left with 1D PDFs for all of your parameters. Below is an example of a 5D parameter space.

Each of the contour panels show a 2D marginalized distribution, and the diagonals show the 1D marginalized distributions. You can imagine that if the probability has strange shapes over the parameter space, the marginalized 1D distributions might not be the best way to tell where your *maximum likelihood* truly is, since integrating over a dimension is reducing the information. In this case it's probably okay.



Algebra of Random Variables: How can we manipulate multiple random variables? First, let's consider the sum of two independent variables $Z = X + Y$. Without proof (you can work this out from the CDFs), the PDF of Z is the convolution of the PDFs of X and Y:

$$f_Z(z) = f_X(x) * f_Y(y) = \int dx f_X(x) f_Y(z-x) = \int dy f_X(z-y) f_Y(y) \tag{100}$$

We've substituted in for $y = z - x$ in the third form and $x = z - y$ in the fourth form. Both forms are valid and sometimes it is easier to do an integration one way than the other. I've written the integration variables explicitly different but remember that in reality they don't matter.

Now let's consider the product of two independent variables $Z = XY$. The product distribution is (again without proof)

$$f_Z(z) = \int dx f_X(x) f_Y(z/x) \frac{1}{|x|} = \int dy f_X(z/y) f_Y(y) \frac{1}{|y|} \tag{101}$$

Remember that if the random variables are not independent, the covariance between them must be accounted for.

7.3 Error Propagation

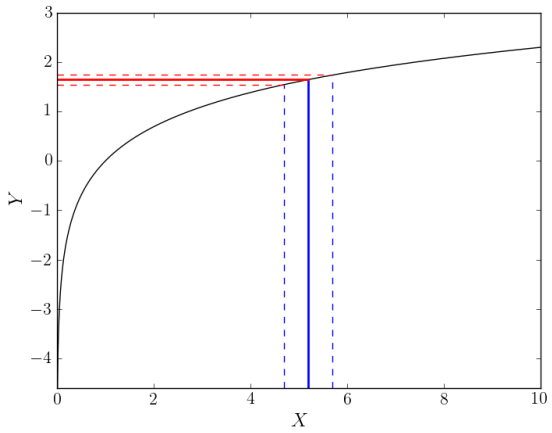
Propagation of uncertainty can simply be considered an extension of basic calculus. We have some quantity X with errors and we would like to know how the uncertainty transforms to some quantity $Y = f(X)$. We can look at the Taylor expansion:

$$
\begin{aligned}
Y = f(X) &= f(X_0) + \left.\frac{df}{dX}\right|_{X_0} (X - X_0) + \cdots \\
\rightarrow Y - Y_0 &\approx \left.\frac{df}{dX}\right|_{X_0} (X - X_0) \\
\rightarrow \Delta Y &= \left.\frac{df}{dX}\right|_{X_0} \Delta X \\
\rightarrow \sigma_Y &= \left.\frac{df}{dX}\right|_{X_0} \sigma_X
\end{aligned}
\tag{102}
$$

As an example, the figure on the next page shows $Y = \ln X$. Then, $dY/dX = 1/X$ which implies that $\sigma_Y = \sigma_X/X$. Therefore, if you have a measurement that is $X = 5.2 \pm 0.5$, we would expect that $Y = \ln X \pm \sigma_X/X = 1.6 \pm 0.1$. Because of the asymmetries in the $\ln$ function, one could consider asymmetric errors as well (as in the figure), which is generally true.

Useful aside: Significant figures are also very important here but often misunderstood. Rather than repeat all of the rules associated with significant figures, it's important to talk about what they really are: they are the figures (digits) that are significant. It's often easy to forget that. The significance comes from how well you "believe" you can measure a quantity and then propagate the uncertainties in those quantities through.

Graphically, you can see that an error in X will transform to an error in Y this way:



Let's say that you have a measurement which is a function of two variables, i.e., $Z = f(X, Y)$ and we have errors on X and Y. We'll have something similar to before except

using partial derivatives:

$$\begin{aligned} \Delta Z &= \frac{\partial Z}{\partial X}\Delta X + \frac{\partial Z}{\partial Y}\Delta Y \\ \rightarrow \sigma_Z^2 &= \left(\frac{\partial Z}{\partial X}\sigma_X\right)^2 + \left(\frac{\partial Z}{\partial Y}\sigma_Y\right)^2 \\ &= \left(\frac{\partial Z}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial Z}{\partial Y}\right)^2 \sigma_Y^2 \end{aligned} \tag{103}$$

Note that I've added in quadrature now (because variances always add). There are a few ways to think about this:

1. Graphically, we'd have some function $Z = f(X, Y)$ and let's for now say that X and Y are independent. The ΔZ that you travel is some distance away from your measurement, and that distance depends on the distances ΔX and ΔY. Therefore, the Pythagorean Theorem applies

2. You can show that when you add two Gaussians together, the variance of the final distribution will be the sum of the variances of the original two Gaussians. This mimics the process of adding two measurements with Gaussian errors.

3. More broadly, you can easily show that $< (X \pm Y)^2 >=< X^2 > + < Y^2 >$, if X and Y are independent.

Of course, we can generalize to multiple variables using this formalism.

I will provide some basic rules. There are a number of good examples on the wikipedia page for propagation of uncertainty. Assume that X and Y are independent. If they are not, you hve to account for the *covariance* between the parameters.

Basic Rules:

- $Z = aX \pm bY \rightarrow \sigma_Z^2 = a^2\sigma_X^2 \pm b^2\sigma_Y^2$
- $Z = XY$ or $X/Y \rightarrow \dfrac{\sigma_Z^2}{Z^2} = \dfrac{\sigma_X^2}{X^2} + \dfrac{\sigma_Y^2}{Y^2}$
- $Z = X^c \rightarrow \sigma_Z^2 = \left(cX^{c-1}\sigma_X\right)^2 = \left(c\dfrac{Z}{X}\sigma_X\right)^2 \implies \dfrac{\sigma_Z^2}{Z^2} = \left(\dfrac{c\sigma_X}{X}\right)^2$

Always remember: variances add!

7.4 Time Series Analysis

I thought it would be worthwhile to include some information on time series analysis. None of this will be discussed later and is included primarily as a reference, so if you'd

like to skip over it, that's perfectly fine. I would strongly recommend taking a look at it though at some point, and feel free to ask me any questions.

A time series can simply be thought of as multiple realizations of some stochastic process, i.e. $x(t) \sim X$ for all t. For example, a time series of white, Gaussian noise is just $x(t) \sim \mathcal{N}(0, \sigma_x^2)$ for all t. Typically, our time series will be finite in length and discrete, so really we measure some process x_i. We can *estimate* the mean of N measurements of x_i to be

$$\hat{x}_i = \frac{1}{N} \sum_i x_i, \tag{104}$$

which is simply the arithmetic average.

Autocorrelation Function: The ACF is a useful tool for understanding how correlated some function is in time. Therefore, it can tell you about characteristic timescales in your data, such as some characterstic decay or periodicity. It provides some measure of the amount of variance as a function of *time lag*. The base form of the ACF of $x(t)$ is

$$R_x(\tau) = \langle [x(t) - \mu_x][x(t+\tau) - \mu_x] \rangle \,, \tag{105}$$

where $\tau = t_2 - t_1$. In practical applications, subtracting the mean is important because interpreting the ACFs becomes difficult as you need to account for a large triangle function (from the convolution). Note that you may see the ACF defined as

$$R_x(\tau) = \frac{\langle [x(t) - \mu_x][x(t+\tau) - \mu_x] \rangle}{\sigma^2} \tag{106}$$

which is a normalization that places $R_x(\tau)$ in the range of -1 to 1, i.e. fully anti-correlated to fully correlated (0 = not correlated).

Cross-correlation Function: This is the same idea but applied to two different functions and time series. So if you're looking for some kind of correlation between two objects, you'll want to go to something like this. Here I show it in the normalized form

$$\rho_{xy}(\tau) = \frac{\langle [x(t) - \mu_x][y(t+\tau) - \mu_y] \rangle}{\sigma_x \sigma_y} \tag{107}$$

Note that one commonly used diagnostic to compare two data sets is the Pearson correlation coefficient. It is simply the CCF when $\tau = 0$ (i.e., no temporal considerations)!

Stationary Process: This is an important point that I thought I'd include for reference. A stationary process is one whose PDF doesn't change in time. Wide-sense stationarity is essentially stationarity up to second order, so the mean and the variance do not change with time. If the process is a function of time, then it is nonstationary.

Before, we considered $R_x(\tau) = R_X(t_2 - t_1)$. but there are processes where the ACF is not a function of only lag (e.g. random walks, red-noise processes) and so $R_x(\tau) \neq R_X(t_1, t_2)$, the more general form of the ACF.

On an additional practical note, if you are using the ACF to determine some correlation timescale in your data and your data span is smaller than that value, your ACF will not

Structure Functions: These are less commonly used in the literature but extremely useful.

$$D_x(\tau) = \left\langle [x(t) - x(t+\tau)]^2 \right\rangle \tag{108}$$

You can think of this as the mean squared *increment*, or the mean squared differences at some time lag τ. Both this and the ACF are second moment statistics. As such, they characterize the variance in particular ways.

The SF can be used when a process is not wide-sense stationary, such as in the case before in which the ACF is nonstationary. That's because the increments, $x(t) - x(t+\tau)$, might be statonary. It is also handy because no estimate of the sample mean is needed (it effectively removes the mean).

Note: One can also consider higher-order structure functions, which analyze higher-order increments. This is analogous to higher-order derivatives: rather than analyzing the slope (the increments/differences), one might want to characterize the concavity of a function, or more.

Fourier Transforms: I will not say much about Fourier Transforms even though they are one of the most important time series anlaysis tools, among other things. The reason they are so useful is that they allow you to characterize a time series in frequency space (put another way, you can characterize the frequency decomposition of a signal). If you've not seen them, then it's a pretty complicated topic to dive into.

Power spectra: These notes would be remiss without mentioning power spectral estimators. The most common one is a Fourier-based method although there are others (e.g., maximum entropy spectral estimators, which are related to ARMA processes). Via any of these methods, we can analyze a time series in the frequency domain, which has a huge number of applications, conveniences, and of course hurdles. As you might expect, ACFs and SFs are both related in various ways to Fourier-based power spectra.

7.5 Statistical Tests

Chi-squared Statistic: There are actually different kinds of chi-squared statistics you can come up with but all relate to the same idea. Let's say you have some measurements (data d) with errors (σ). The general idea is that you want to be able to say how closely the measurements match your model (m) to within the errors. So your test statistic for figuring out how good of a model you have is:

$$\chi^2 = \sum_i \frac{(d_i - m_i)^2}{\sigma_i^2} \tag{109}$$

which is just some measure of how far away your residual (data - model) are away from zero, where distance is in "σ" units. That is, a measurement that is far from the model but with large error bars contributes neglibly to this test statistic. A measurement that is far away but with small error bars contributes greatly. As you can tell, low χ^2 is typically better.

One concern is that if I throw more parameters into my model, I can always make the fit better. Thus, χ^2 will go down. How can I tell when I've added too many? The reduced-chi-squared statistic is better, in which you just divide by the number of *degrees of freedom*, usually just the number of data points minus by the number of fitted parameters (be careful, subtracting the mean counts as one degree of freedom!)

$$\chi^2_r = \frac{\chi^2}{N_{\rm dof}} = \frac{\chi^2}{N_{\rm points} - N_{\rm params} - 1} \tag{110}$$

The last form has the -1 to account for the sample mean separately. Generally, if $\chi^2_r > 1$, then the fit is not fully describing the data. If $\chi^2_r = 1$ then the fit is good. If $\chi^2_r < 1$, then your model is overfit.

Kolmogorov-Smirnov (K-S) test: Compares two CDFs, perhaps one that is empirically measured with one that you are testing, or two empicially measured CDFs to test for consistency.

T-test: Tests whether the means of two populations are statistically different from each other. This could be useful in comparing two data sets, for example.

F-test: Useful when comparing different models to see which fits best (model selection). For example, if two models have a different number of parameters (degrees of freedom), is the inclusion of extra parameters statistically significant? See also the Akaike Information Criterion.

7.6 Problems

Don't spend too long on these!

1. The transverse velocity of an object at distance d with total proper motion $\mu_{\rm T} = \sqrt{\mu_\alpha^2 + \mu_\delta^2}$ is

$$v_{\rm T} = 4.74 \text{ m s}^{-1} \left(\frac{\mu_{\rm T}}{\text{mas yr}^{-1}}\right)\left(\frac{d}{\text{pc}}\right). \tag{111}$$

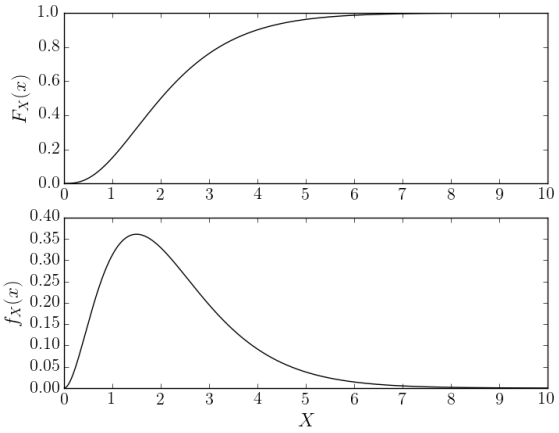
We measure the star Mintaka to be 0.21 ± 0.03 kpc away with $\mu_\alpha = -0.64 \pm 0.56$ mas yr^{-1} and $\mu_\delta = -0.69 \pm 0.27$ mas yr^{-1}. What's the transverse velocity and its error?

2. You observe Altair to have a magnitude of $m_V = 0.8 \pm 0.1$. The absolute magnitude is $M_V = 2.22$. How far away is Altair (with errors!). Hint: Don't forget your distance

modulus $m_V - M_V = 5 \log_{10}(d/10\ \mathrm{pc})$.

3. If c is a constant, what is the distribution of $c\mathcal{N}(0, 1)$?

4. An example of a Gamma distribution (shape parameter $k = 3.0$, scale parameter $\theta = 0.75$), both the CDF and the PDF.



Approximately what is the median? What is the mode? How do these relate to the mean (if you work through it, it's $k\theta$)?

5. You observe the blazar OJ 287 40 times over the span of 20 years looking for evidence of a periodic variation in the lightcurve and it looks like there might be one. You fit a sine wave and obtain a chi-squared of 37. Someone else takes your data and fits two exponential outbursts (each with three parameters, the amplitude, rise time, decay time) and obtains a chi-squared of 35. Which model is preferred?

6. Let the exponential distribution be $f_X(x) = \lambda e^{-\lambda x}$, where $x \geq 0$. If X and Y are drawn from the same exponential distribution, what is the PDF of $X + Y$?

8 Bayesian Statistics

Resources: I have found these three texts to be useful over the years. There are plenty of others out there.

- Allen B Downey, *Think Bayes: Bayesian Statistics Made Simple*, Green Tea Press (2012)
- Phil Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach*, Cambridge University Press (2010)
- William M. Bolstad, *Understanding Computational Bayesian Statistics*, Wiley (2009)

8.1 Introduction

Bayesian probability theory defines probability as a measure of the degree of belief in an outcome or proposition. This is in contrast to frequentist theory which defines probability as a ratio of instances of occurrence against the total number of trials. The frequentist definition restricts probability statements to propositions about random variables. The Bayesian definition is more robust, as it can apply to any logical proposition.

There are two fundamental tasks of science: model selection and parameter estimation (or measurement). For the first we wish to make statements about the probability that a hypothesis or model is correct. This is a natural statement in the Bayesian framework. In the frequentist framework this statement is much harder. Instead can must ask: what is the probability (*frequency*) we would observe this outcome, if our model were incorrect? If the resulting probability (sometimes called a 'p-value' in this context) is small, then the model is likely to be correct. There are other frequentist model selection techniques, but all require performing many identical observations to tease out the long term frequency of occurrences, which can be problematic. In Bayesian statistics we just directly calculate the probability that a model is correct.

For parameter estimation frequentist results are often presented in terms of confidence intervals. For example a star's mass may be estimated to be $(1.40 \pm 0.15) M_\odot$, where the uncertainty represents a 90% confidence interval (assuming a particular distribution of measurement uncertainty). This means that *if many new, independent measurements are made, 90% of new measurements would fall in this range*. This is not the same as *there is a 90% probability that the true mass is in this range*. In Bayesian statistics the we directly compute the probability distribution function for the parameter. There is no hidden assumption that the measurements follow a normal (or any other) distribution. And the Bayesian 90% **credible interval** really does mean there is a 90% probability that the true mass is in the given range.

8.2 Probability

The basic mathematical definitions of probability are as follows. A probability, p, must fall into the range $[0, 1]$, where 0 and 1 correspond to absolutely false and certainly true, respectively.

$$p \in [0, 1] \tag{112}$$

Probabilities describe exhaustive outcomes: if there are several, independent probable outcomes, the total probability of *any* outcome is 1. That is to say, one of the possible outcomes must occur. This can also be stated that sum of the probability that proposition A is true, $p(A)$, and the probability that A is false (or 'not A' is true) is unity.

$$\sum_i p_i = 1 \tag{113}$$
$$p(A) + p(\text{not } A) = 1$$

The probability of a proposition A can be conditional upon a second proposition B. The probability that A is true given B is true is $p(A \mid B)$. Because Bayesian probability applies to any logical proposition, we can apply the usual rules of deductive logic to them:

$$p(A \text{ and } B) = p(A) \cdot p(B \mid A) \tag{114}$$
$$p(A \text{ or } B) = p(A) + p(B) - p(A \text{ and } B) \tag{115}$$

The logical *or* and *and* operators are commutative, associative, and distributive.

By using equation (114) and the commutativity of the logical *and* we can derive **Bayes' Theorem**:

$$p(A \mid B) = \frac{p(A) \cdot p(B \mid A)}{p(B)} \tag{116}$$

This doesn't seem like a terribly exciting result, but it is incredibly useful.

Cookie Problem

There are two indistinguishable cookie jars.

- Jar 1 has 20 chocolate cookies and 20 vanilla cookies.
- Jar 2 has 10 chocolate cookies and 30 vanilla cookies.

You randomly select a jar, then randomly take a cookie from that jar. The cookie is vanilla. What is the probability you are holding Jar 2?[2]

[2] shamelessly adapted from *Think Bayes*, A. Downey

Solution: First, we write the question in terms of a conditional probability: the probability of holding Jar 2 given that we got a vanilla cookie. Now we use Bayes' Theorem:

$$p\left(J_2 \mid v\right) = \frac{p\left(J_2\right) \cdot p\left(v \mid J_2\right)}{p\left(v\right)}$$

All of the quantities on the right-hand-side are easily calculable. We have even odds of having selected Jar 2, so $p\left(J_2\right) = 1/2$. The probability of getting vanilla from Jar 2 is $p\left(v \mid J_2\right) = 3/4$. The probability of a getting vanilla under any Jar is $p\left(v\right) = 5/8$. Finally, $p\left(J_2 \mid v\right) = 3/5$.

Using Bayes' Theorem we rewrote the probability in terms of a **likelihood function** (which tells us the probability of observing our data, v, if our model, J_2 is correct) and the **prior probability** for J_2. The probability on the left-hand-side is called the **posterior probability**, the probability after the experiment is conducted.

A Medical Test's Accuracy

A rare condition, affects 1 in 100 000 people of your demographic group. There is a test for the condition that is 98% accurate. This means that 98% of test takers who have the condition test positive (the other 2% get false negative results), and 98% of those who do not have the condition test negative (the other 2% get false positive results). You decide to take the test and receive a positive result. What is the probability you have the condition?[3]

Solution: Again we use Bayes' Theorem:

$$p\left(c \mid +\right) = \frac{p\left(c\right) \cdot p\left(+ \mid c\right)}{p\left(+\right)}$$
$$p\left(+\right) = p\left(+ \mid c\right) \cdot p\left(c\right) + p\left(+ \mid h\right) \cdot p\left(h\right)$$

where c is the condition, h is healthy, and $+$ is a positive test result. We determine the probability of getting a positive result by adding the conditional probabilities for both health cases. Finally, we see $p\left(c \mid +\right) \approx 0.005$.

[3] adapted from the excellent chapter on probability on *Statistical and Thermal Physics*, H. Gould and J. Tobochnik. Problems of the "wild false positive" genre appear in almost every book on Bayesian statistics too.

8.3 Data Analysis

Bayes' theorem

The basic problem of data analysis is determining the degree of belief in a hypothesis $\mathcal{H}$ given some observed data d. Using Bayes' theorem this becomes:

$$p\left(\mathcal{H} \mid d\right) = \frac{p\left(\mathcal{H}\right) \cdot p\left(d \mid \mathcal{H}\right)}{p\left(d\right)}. \tag{117}$$

We will consider each of the terms on the right-hand-side separately below.

Likelihood: $\mathcal{L} = p\left(d \mid \mathcal{H}\right)$

The likelihood is the probability of observing your data, if your hypothesis were correct. This sounds backwards, but it is usually straightforward to determine (as in the cookie problem 8.2). A likelihood assumes that the data is drawn from a particular probability distribution function (PDF). If your data follows a known distribution like a powerlaw or a gamma distribution this is easy.

$$\mathcal{L}(x) = PDF(x)$$

When your data is described by a deterministic model (not a statistical one), things are a bit different. In these cases instrument noise will still be statistical, so the likelihood is based on the noise properties. Lets say our data, d, is built from a deterministic effect, $m^\star$, and instrument noise, n.

$$d = m^\star + n$$

We propose a model for the deterministic effect m. If our model is correct ($m = m^\star$), then the residual, r, is Gaussian noise.

$$r = d - m = n$$

And the likelihood that an individual datum is described by our deterministic model is the likelihood that the residual is described by the noise distribution!

The likelihood for the full data set is the product of the individual likelihoods

$$p\left(d \mid \mathcal{H}\right) = \prod_i p\left(d_i \mid \mathcal{H}\right) \tag{118}$$

For Gaussian noise with standard deviation, σ, the likelihood is:

$$p\left(n_i \mid \mathcal{H}\right) = \frac{1}{\sqrt{2\pi}\,\sigma}\, \exp\left(-\frac{{n_i}^2}{2\sigma^2}\right), \tag{119}$$

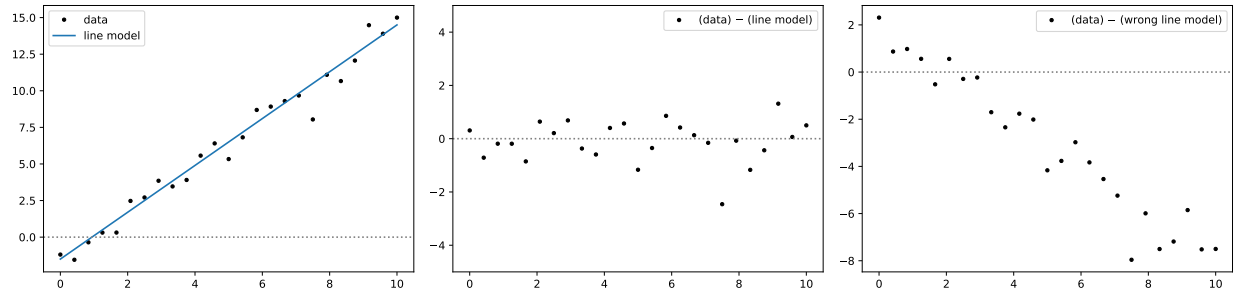


Figure 18: *left*: a deterministic line with Gaussian noise added. *middle*: subtracting the correct model from the data leaves Gaussian noise residuals. *right*: subtracting the wrong model leaves non-Gaussian residuals.

so

$$p\left(d \mid \mathcal{H}\right) = \prod_i \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(d_i - m_i)^2}{2\sigma^2}\right) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^N \exp\left(-\frac{\sum_i (d_i - m_i)^2}{2\sigma^2}\right) \tag{120}$$

More generally the noise may defined as a covariance matrix, $\mathbf{C}$. This is useful if the noise variance changes or is correlated between samples. In this case it is convenient to write the likelihood of equation (120) in the language of linear algebra. We define the residual as a vector $\vec{r} \to r_i = d_i - m_i$

$$p\left(d \mid \mathcal{H}\right) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{C}}} \exp\left(-\frac{1}{2}\,\vec{r} \cdot \mathbf{C}^{-1} \cdot \vec{r}\right) \tag{121}$$

Prior probability: $\mathcal{P} = p\left(\mathcal{H}\right)$

The prior probability is probability your hypothesis is true, before you have conducted your experiment. Many people feel a bit uneasy about priors. The choice of prior does in fact affect the outcome, so how is this okay? First, frequentist methods make a myriad of assumptions about what is known about the data. At least the Bayesian framework lays these assumptions bare, for all to see and scrutinize. Second, given a large number of iterative experiments the posterior converges to the correct. This point illustrates another of Bayes' theorem's strengths: it is straightforward to iteratively apply Bayes' theorem to successive experiments.

Lets say we observe some data d_1 and compute the posterior for our hypothesis, $p\left(\mathcal{H} \mid d_1\right)$. Now we collect some new data d_2. What is the posterior probability for $\mathcal{H}$ after this? The outcome of the first experiment is now part of our prior knowledge for the second. We can see this by considering the posterior after observing $(d_1 \text{ and } d_2)$, which we leave as an exercise for the reader in problem 8.5.

When using past posteriors as priors one must be careful not to *double count* data. Say you collect 5 years of observations and compute a posterior, then you collect 2 new years

of observations. If you want to use the 5 year posterior as a prior, the new analysis must use *only* the new 2 years of data. If you conduct a joint analysis of all 7 years, you must return to the original prior. The 7 year data set is not **independent** of the original 5 year data set.

There are a few common choices for priors, when starting from scratch. If you look up a parameter in the literature, and it's stated with uncertainty, that should be your prior on that parameter. For instance if $x = x_0 \pm \Delta x$ with uncertainty at the 1σ level, then

$$p\left(x\right) = \mathcal{N}(x_0, \Delta x^2).$$

An **ignorance prior** is usually a safe choice, where all outcomes are assumed to be equally likely. For a parameter this can be expressed as a uniform distribution. Uniform distributions are only properly defined for finite domains. This is perfect for bounded parameters like angles

$$p\left(\phi\right) = \mathcal{U}(0, 2\pi).$$

For unbounded parameters you can use a trick of the normal distribution

$$p\left(x\right) = \mathcal{U}(-\infty, \infty) = \lim_{\nu \to \infty} \mathcal{N}(0, \nu).$$

Practically, this is often achieved as

$$p\left(x\right) = \mathcal{N}(0, \texttt{MAX}),$$

where `MAX` is the maximum `int` or `float` allowed by the computer.

In some cases a log-uniform prior is the preferred ignorance prior

$$p\left(\log_b x\right) = \mathcal{U}(e_0, e_1).$$

This is useful when a parameter could span a large range of decades (for $b = 10$, 10^{e_0} to 10^{e_1}) and you have no information about the scale of the effect. This prior is used for the gravitational wave background amplitude in some NANOGrav analyses.

To see why this helps, we can consider a parameter x that could be anywhere in the range 0.1 to 100. Assuming a uniform prior ($p(x) = \text{const}$), we can compare the prior probability that x is in the smallest ($x \in [0.1, 1]$) and largest ($x \in [10, 100]$) decades.

$$\frac{\int_{10}^{100} p\left(x\right) \mathrm{d}x}{\int_{0.1}^{1} p\left(x\right) \mathrm{d}x} = \frac{90}{.9} = 100$$

The uniform prior heavily favors the largest decade. It states that the large values are favored over the small. In cases when we wish to detect a new phenomenon, we want the exact opposite. We expect a small value, or else we would have detected it already! It is these cases where the log-uniform prior is best.

Bayesian evidence: $\mathcal{E} = p(d)$

The Bayesian evidence is the probability of observing the data under any circumstances. It is best thought of as a normalization. It can be constructed by summing or integrating over the space of hypotheses. In most cases this is easier said than done. Explicitly this can look like:

$$p(d) = \sum_i p(d \mid \mathcal{H}_i) \cdot p(\mathcal{H}_i), \tag{122}$$

which is exactly what one must do to normalize the posterior. The Bayesian evidence is sometimes called the **marginal likelihood** as the normalization processes is equivalent to marginalizing the likelihood over **all** parameters.

The evidence is often challenging to compute because the tails of the distribution can be very important. Common numerical integration methods, like trapezoid and Simpson's sums, can become intractable for as few as three free parameters. Many computational methods, such as Markov chain Monte Carlo, avoid this problem by working with ratios of the posterior, so the evidence cancels out.

The evidence is important in problems of model selection, where multiple parameterized models must be compared. This occurs, for example, when trying to decide whether data are best fit by a straight line or a quadratic.

Marginalization

Sometimes we want to calculate a quantity like $p(d \mid \mathcal{H})$, but we only know how to calculate $p(d \mid \mathcal{H}, x)$. This can happen if our hypothesis is parameterized. For example, if we are trying to fit a line to data the slope of the line will appear in the likelihood. We can calculate the likelihood for any given x, but which x do we choose? In these cases x is said to be a **nuisance parameter**. To compute $p(d \mid \mathcal{H})$, we must first **marginalize** over x. That is, we average over all possible x making the likelihood a statement about any x instead of a single choice.

$$p(d \mid \mathcal{H}) = \int \mathrm{d}x \, p(x \mid \mathcal{H}) \, p(d \mid \mathcal{H}, x) \tag{123}$$

Note that the average is weighted by the prior on x.

Sometimes it is possible to compute these integrals analytically, but more often than not one must numerically integrate. As the number of nuisance parameters grows, the methods for computing these integrals must become more and more sophisticated.

For parameterized hypotheses notation like $\mathcal{H}(x, y)$ is common. In the case of two parameters we can marginalize over one only, if we choose:

$$p(d \mid \mathcal{H}(y)) = \int \mathrm{d}x \, p(x \mid \mathcal{H}) \, p(d \mid \mathcal{H}(x, y)). \tag{124}$$

Another common notation replaces the hypothesis with the parameters entirely. Using a parameter vector, we can write the marginal likelihood (or evidence) as

$$p\left(d\right) = \int p\left(\vec{x}\right) p\left(d \mid \vec{x}\right) \, \mathrm{d}^N x = \mathcal{E}, \tag{125}$$

where $\vec{x}$ is the vector of N parameters that define the model.

Parameter Estimation

In cases of parameter estimation, we want to determine the measured or best fit parameters from observed data. This may be the result of fitting a curve to data, where the fit parameters correspond to physical observables. If we calculate the posterior probability distribution (either analytically or numerically) for the parameters, then the inference problem is a matter of interpreting this result.

The **best fit** parameters are not rigorously defined. The true output of the analysis *is* the posterior. Any best fit or uncertainty is a summary of the full result. Many prefer the **mean** to report as the best fit parameter, calculated

$$\langle x \rangle = \int x \, p\left(x \mid d\right) \, \mathrm{d}x. \tag{126}$$

where $p\left(x \mid d\right)$ is the posterior probability of parameter, x, given the data, d.

When the posterior is computed numerically it can be challenging to compute the mean, especially if there are many fit parameters. In these cases it is more convenient to report the **median** with the added benefit that the cumulative sums used to compute the median will produce **credible intervals** too. The median, $x_{50\%}$, is found by inverting

$$0.50 = \int_{-\infty}^{x_{50\%}} p\left(x \mid d\right) \, \mathrm{d}x. \tag{127}$$

One can perform a similar inversion to find any percentile of the posterior.

To state the uncertainty on the parameters credible intervals are used. For a 90% credible interval we compute the 5% and 95% percentiles of the parameter. It is becoming more common to quote a median with pseudo-1σ uncertainty by computing the 18%, 50%, and 84% percentiles. This gives a 66% credible interval which is approximately the size of a 1σ frequentist confidence interval. One can compute upper or lower limits by similarly integrating the posterior.

Another commonly quoted best fit is the maximum probability parameters. These are sometimes abbreviated as MAP for *maxima a posteriori*. The MAP is easy to calculate, and it ensures the best fit parameters come from the same point in parameter space. Individually computing the median or mean for each of several parameters independently can lead to a worse combined result. This is rare and usually only happens when there are nontrivial correlations in the parameters. In any case, the MAP simultaneously maximizes all parameters, and can never have this problem.

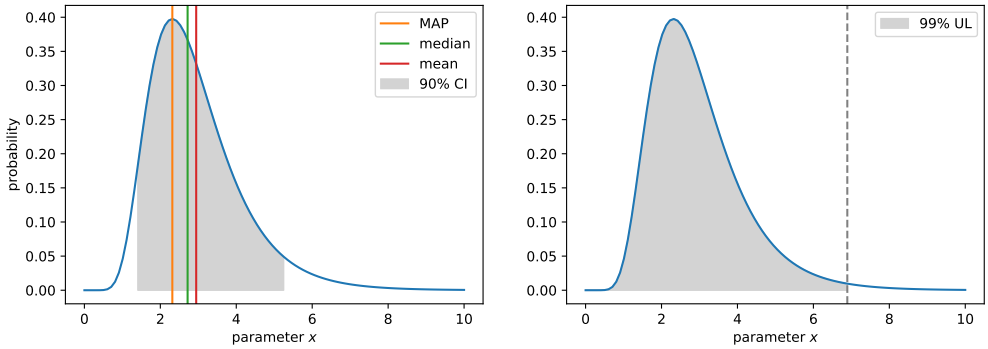


Figure 19: Example parameter estimation for a log-normal distribution. The distribution is skewed so the mean, median, and MAP do not agree. By integrating the probability distribution we can determine credible intervals (*left*) or an upper limit (*right*).

Model Selection

In many cases there is more than one hypothesis or model that fits the data to some extent. How do we decide which is preferred? A commonly used method is the maximum likelihood ratio. In this method one computes the likelihood of the best fit parameters for each model and compares. This method has an important shortcoming: it ignores the complexity of the models. A model with more free parameters should always achieve a better fit. So why not fit 10 data points with a 10th order polynomial? It fits perfectly every time. This complexity consideration is sometimes referred to as the **Occam factor**.

Bayesian analysis naturally takes care of the Occam factor by using the **marginal likelihood ratio** or evidence ratio (same thing). The marginalization integral takes into account the probability across the whole prior volume, penalizing models with many or widely ranging parameters.

The probability for a model $\mathcal{H}_i$ is calculated from Bayes theorem

$$p\left(\mathcal{H}_i \mid d\right) = \frac{p\left(\mathcal{H}_i\right) p\left(d \mid \mathcal{H}_i\right)}{p\left(d\right)} \tag{128}$$

If we want to compare two models, 0 and 1, we can compute the **odds ratio**

$$\mathcal{O}_{1,0} = \frac{p\left(\mathcal{H}_1 \mid d\right)}{p\left(\mathcal{H}_0 \mid d\right)} = \frac{p\left(\mathcal{H}_1\right)}{p\left(\mathcal{H}_0\right)} \frac{p\left(d \mid \mathcal{H}_1\right)}{p\left(d \mid \mathcal{H}_0\right)} = \frac{p\left(\mathcal{H}_1\right)}{p\left(\mathcal{H}_0\right)} \mathcal{B}_{1,0} \tag{129}$$

It is common to break the odds ratio into the prior odds ratio and the marginal likelihood ratio or **Bayes factor**, $\mathcal{B}_{1,0}$.

$$\mathcal{B}_{1,0} = \frac{p\left(d \mid \mathcal{H}_1\right)}{p\left(d \mid \mathcal{H}_0\right)} \tag{130}$$

It is typical to report Bayes factors rather than full odds ratios, as the priors on various models are usually seen as a personal choice. In the case that there is no *a priori* reason to favor either model, there prior odds ratio is 1 and the odds ratio is the Bayes factor.

Table 1: Comparing Bayes factors to frequentist 'σ'.

Bayes factor, $\mathcal{B}$	probability, p	$n\,\sigma$
3	0.75	$1.15\,\sigma$
10	0.91	$1.69\,\sigma$
100	0.99	$2.58\,\sigma$
370	0.997	$3\,\sigma$
1 744 277	0.999 999 4	$5\,\sigma$

After computing the odds, we can reverse engineer the probability for a particular model. Starting from the definition of probability (equation 113), and using the definition of $\mathcal{O}_{i,0} = p\left(\mathcal{H}_i \mid d\right) / p\left(\mathcal{H}_0 \mid d\right)$, we can show that

$$p\left(\mathcal{H}_i \mid d\right) = \frac{\mathcal{O}_{i,0}}{\sum_i \mathcal{O}_{i,0}}. \tag{131}$$

For two models $\mathcal{O}_{0,0} = 1$ so

$$p\left(\mathcal{H}_1 \mid d\right) = \frac{\mathcal{O}_{1,0}}{1 + \mathcal{O}_{1,0}} = \frac{1}{1 + \frac{1}{\mathcal{O}_{1,0}}}$$

Table 1 compares Bayes factors, their probability for two model comparison, and the frequentist σ used to summarize p-value style model selection. A Bayes factor less than 10 is pretty weak evidence.

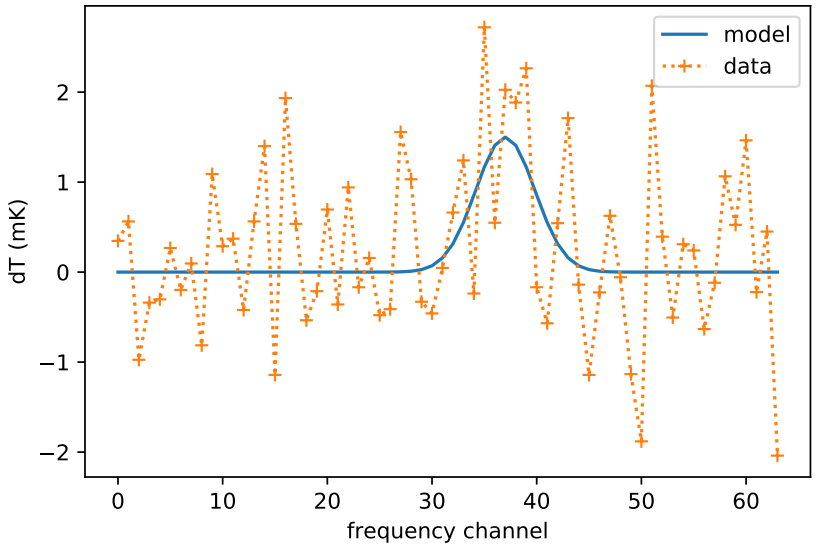


Figure 20: Measured and modeled spectrum. The amplitude of the modeled spectrum matches that of the simulation that was used to generate the data.

8.4 Examples

Model Selection: Is there a spectral line?

We have some noisy data containing a possible novel spectral line feature. The spectrum is measured as a temperature anomaly, dT, from the mean observed temperature in each frequency channel. Our instrument has 64 frequency channels and is known to have white, Gaussian radiometer noise at a level of $\sigma_N = 1$ mK.

Some theory predicts the line will have a gaussian shape:

$$dT = T_0 \exp\left(\frac{-(\nu - \nu_0)^2}{2{\sigma_L}^2}\right),$$

where T_0 is the amplitude of the feature that according to the theory can range from 0.01-100 mK. The frequency ν is measured in channel number for our instrument, and the theory says line should appear in channel $\nu_0 = 37$ with width $\sigma_L = 2$.

Of course, this theory could be wrong. In which case we should expect to see only the radiometer noise in our data. If we collect the data shown in Fig 20 with our instrument, what does it say about the theory?

We begin by stating our hypotheses. Let $\mathcal{H}_1$ be *there is a spectral line as described by the*

theory, and let $\mathcal{H}_0$ be *there is no spectral line*. We wish to calculate the odds ratio between these two hypotheses. Since we have no *a priori* reason to favor one or the other we set the prior odds between the hypotheses to 1, and the odds ratio reduces to the Bayes factor

$$\mathcal{O}_{1,0} = \mathcal{B}_{1,0} = \frac{p\left(d \mid \mathcal{H}_1\right)}{p\left(d \mid \mathcal{H}_0\right)}. \tag{132}$$

First, we calculate the likelihood for the noise only model, $\mathcal{H}_0$. We know the properties of the radiometer noise (white, Gaussian), so

$$p\left(d \mid \mathcal{H}_0\right) = \frac{1}{\sqrt{2\pi}\,\sigma_N} \exp\left(-\frac{\sum {d_i}^2}{2{\sigma_N}^2}\right). \tag{133}$$

Next, we calculate the likelihood for the line model, $\mathcal{H}_1$. Under $\mathcal{H}_1$ the data in the i-th frequency channel is given by $d_i = dT_i + n_i$, where dT_i is the modeled temperature. After subtracting the modeled line the residuals are noise, so the likelihood is

$$p\left(d \mid \mathcal{H}_1, T\right) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_N}\right)^2 \exp\left(-\frac{\sum (d_i - m_i)^2}{2{\sigma_N}^2}\right) \tag{134}$$

$$m_i = dT_i = T \exp\left(-\frac{{\nu_i - \nu_0}^2}{2{\sigma_L}^2}\right).$$

To compute $p\left(d \mid \mathcal{H}_1\right)$ we must first marginalize over the nuisance parameter, T.

$$p\left(d \mid \mathcal{H}_1\right) = \int \mathrm{d}T\, p\left(T \mid \mathcal{H}_1\right) p\left(d \mid \mathcal{H}_1, T\right) \tag{135}$$

Because T can fall into a huge range, $[10^{-2}, 10^2]$, we adopt a log-uniform prior

$$p\left(\log T \mid \mathcal{H}_1\right) = \frac{1}{\log T_{\max} - \log T_{\min}} \tag{136}$$

The companion Jupyter notebook[4] walks through the numerical calculation of these expressions for the data shown in Fig 20. The final result is an odds ratio of $\mathcal{O}_{1,0} \approx 124$. This means the line model is favored over the noise only model at $124:1$ or about 99.2%. This is good, but not definitive evidence.

Parameter Estimation: What is the best fit value of T?

After determining the odds for the spectral line being present, we can now calculate the best fit amplitude for the feature. We effectively calculated the posterior for T (assuming the spectral line model) during the model selection process. By Bayes' theorem it is proportional to the integrand of equation (135).

$$p\left(T \mid \mathcal{H}_1, d\right) \propto p\left(T \mid \mathcal{H}_1\right) p\left(d \mid \mathcal{H}_1, T\right)$$

[4] https://www.github.io/paulthebaker/bayes_talk

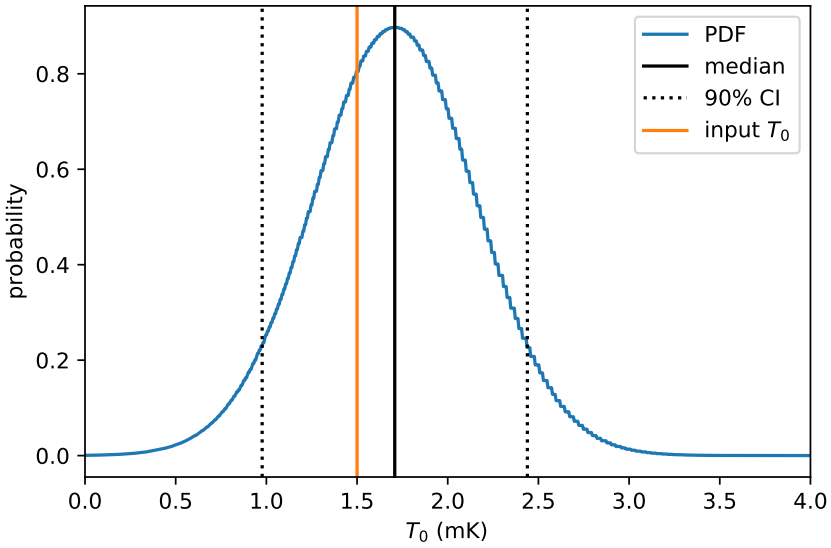


Figure 21: Posterior distribution function (PDF) for spectral line amplitude, T. The median and 90% credible interval are shown along with the input T_0 used to simulate the data.

This function is determined numerically for the data shown in Fig 20 in the companion Jupyter notebook. We integrate this function numerically to determine the median and 90% credible interval: $T = 1.7 \pm 0.7$.

8.5 Problems

Monty Hall

A favorite game from a classic a game show presented the contestant with three doors. Behind one was a fabulous prize, while the others held only junk. First, the contestant chooses a door. Next, the host opens one of the unchosen doors to reveal junk[5]. Finally, the contestant is given the opportunity to switch doors, taking home whatever prize is behind the chosen door.

1. What is the initial probability that the prize is in the door selected by the contestant?
2. After the host opens a door, what is the probability that the prize is in the remaining door?
3. What if there were 5 doors (still one prize), and the host opened three to reveal junk before the contestant's opportunity to switch?

Multiple observations

Show that

$$p\left(\mathcal{H} \mid d_1, d_2\right) = \frac{p\left(\mathcal{H} \mid d_1\right) \cdot p\left(d_2 \mid \mathcal{H}\right)}{p\left(d_2\right)},$$

where d_1 and d_2 are data from two **independent** observations and a comma is used to denote the logical *and*.

Cookie Jar Redux

Starting from the Cookie Problem of example 8.2, you draw a second cookie from the same jar, and it's vanilla too. What is the probability you are holding Jar 2 now?

Using Hubble's law to find distance

Hubble's law relates the distance to a galaxy, x, to its apparent recessional velocity, v, due to the expansion of the universe.

$$v = H_0\, x,$$

where H_0 is the Hubble constant.

[5]The host knows which door the prize is in and always reveals junk. If the contestant selected the prize door, he opens one of the others at random.

You observe a galaxy with a recessional velocity of $v = 100 \pm 5$ km/s, where the uncertainty represents 1σ Gaussian noise. Determine the posterior probability for the distance to this galaxy given the velocity observation. Write down the equation for $p(x \mid v)$. Set up any integrals, but don't evaluate them. In each case your answer should be a function of v only.

1. What is $p(x \mid v)$, assuming a point estimate of $H_0 = 70$ km/s/Mpc? Remember to account for the uncertainty in v.

2. What is $p(x \mid v)$, accounting for 1σ uncertainty in $H_0 = 72 \pm 3$ km/s/Mpc[6] by marginalizing? Start by determining $p(x \mid v, H_0)$.

3. What is $p(x \mid v)$, accounting for uniform uncertainty in $H_0 = 70 \pm 10$ km/s/Mpc?

Using Hubble's law to find distance, calculated

Warning: this problem requires a computer.

1. Evaluate and plot the posteriors from parts 1, 2, and 3 of the previous problem (8.5).

2. Determine the median and a 90% CI on the distance, x, for each case.

3. What if you use the Hubble constant from the Planck mission[7], $H_0 = 67.7 \pm 0.5$ km/s/Mpc?

Spectral line redux

Warning: this problem requires a computer and is **very** challenging.

1. Redo the calculation of the odds for the spectral line problem from section 8.4. This time let there be uncertainty in the predicted channel for the line $\nu_0 = 37 \pm 5$, where the uncertainty is given at the 1σ level.

2. Use the joint posterior on T and ν_0 to determine the best fit amplitude and channel for the line.

[6] V. Bonvin, *et al*. "H0LiCOW – V. New COSMOGRAIL time delays of HE 0435–1223: H_0 to 3.8% precision from strong lensing in a flat ΛCDM model". *MNRAS*, **465** (4): 4914–4930 (2016).

[7] Planck Collaboration, P.A.R. Ade, *et al*. "Planck 2015 results – XIII. Cosmological parameters". *A&A*, **594** A13 (2016).

9 Kepler's Laws and the Virial Theorem

Almost all of the material in these notes comes from Carroll & Ostlie, Chapter 2 as well as Chapter 8 of Taylor's Classical Mechanics. You can find more detail on the historical context of Kepler's laws in almost any introductory astronomy textbook. You can find more details on the derivations of Kepler's Laws and the virial theorem in most classical mechanics texts. Applications of Kepler's laws and the virial theorem are regularly found in texts on galaxies (e.g. Sparke & Gallagher) or stars (e.g. Kippenhan & Weigert).

9.1 Kepler's Laws

In 1609, Kepler wrote down the first two laws of planetary motion as regards the solar system, with the third law following ten years later:

1. A planet orbits the Sun in an ellipse, with the Sun at one focus of the ellipse.
2. A line connecting a planet to the Sun sweeps out equal areas in equal times.
3. The average orbital distance of a planet cubed is proportional to its sidereal period squared: $P[yr]^2 = a[AU]^3$

Once Newton developed his law of gravity, it was possible to write these laws in a more general form applicable to motion of any two bodies moving under the influence of gravity:

1. The path of the reduced mass about the center of mass under the influence of gravity is a conic section.
2. The rate of change of the area swept out by a line connecting a planet to the focus of an ellipse is constant under any central law force.
3. $P^2 = \frac{4\pi^2}{G(m_1+m_2)}a^3$

While the derivations of these laws are all in Carroll & Ostlie, I will give a brief synopsis of them here with a focus on aspects pertaining to their use.

Kepler's First Law

The derivation of Kepler's First Law depends on the conservation of angular momentum and uses the definition of gravitational acceleration as well as the concept of the reduced

mass. The reduced mass is defined as:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \tag{137}$$

where m_1 and m_2 are the masses of the two bodies in question, say the Earth and the Sun. In this framework, the reduced mass is orbiting around the center of mass of the system. To derive the orbit, $r(\phi)$, set the acceleration equal to the central force plus the centrifugal force:

$$\mu \ddot{r} = F(r) + \frac{l^2}{\mu r^3} \tag{138}$$

where l is the angular momentum of the system. We want to rewrite this equation in terms of ϕ. We start by substituting $u = 1/r$, and convert d/dt into $d/d\phi$ using the equation for the angular momentum. This gives us:

$$u''(\phi) = -u(\phi) - \frac{\mu}{l^2 u(\phi)^2} F \tag{139}$$

Now we can substitute the gravitational force into this equation and solve for u. After going back to r, we get the equation for our orbit:

$$r(\phi) = \frac{c}{1 + \epsilon cos\phi} \tag{140}$$

where $c = \frac{l^2}{G m_1 m_2 \mu}$, and ϵ is the eccentricity. If you convert this to Cartesian coordinates, for $\epsilon < 1$, you can show that this is an ellipse. The eccentricity is related to the total energy such that:

$$E = \frac{(G m_1 m_2)^2 \mu}{2 l^2} (\epsilon^2 - 1) \tag{141}$$

If the total energy is less than zero, then we have a bound system and ϵ has a value between 0 and 1 yielding an elliptical orbit. If the total energy is exactly zero, then $\epsilon = 1$ and we have an orbit that is a parabola. Finally, if the total energy of the system is greater than zero, then we have $\epsilon > 1$ and a hyperbolic orbit. We will discuss some properties of these orbits in class.

Kepler's Second Law

Kepler's second law holds for all central force laws, since they all conserve angular momentum. As a planet moves along its orbit, regardless of the exact orbital shape, it sweeps out an area equal to:

$$dA = \frac{1}{2} r^2 d\theta \tag{142}$$

Taking the time derivative of both sides yields:

$$\frac{dA}{dt} = \frac{1}{2} r^2 \frac{d\theta}{dt} = \frac{1}{2} r v_\theta \tag{143}$$

and since r and v_θ are perpendicular:

$$\frac{dA}{dt} = \frac{1}{2}\frac{l}{\mu} \tag{144}$$

Since l is constant for all central forces, the area swept out in a given time is also constant.

Kepler's Third Law

To start we will do a simplified derivation using Newton's laws and assuming a circular orbit (which is a good approximation for the solar system). To start a planet in orbit around the Sun must balance the gravitational force with the centripetal force:

$$\frac{GMm}{r^2} = \frac{mv^2}{r} \tag{145}$$

This simplifies to

$$\frac{GM}{r} = v^2 \tag{146}$$

and we can further simplify given that $v = \frac{2\pi a}{P}$ and given that the radius is constant for a circular orbit. Therefore:

$$P^2 = \frac{4\pi^2 a^3}{GM} \tag{147}$$

where M in this case is the total mass of the system, that is $M = m_1 + m_2$.

In general, we will start with Kepler's first two laws, remembering that the area of an ellipse is πab. Combined with Kepler's Second law, the orbital period is:

$$P = \frac{A}{dA/dt} = \frac{2\pi ab\mu}{l} \tag{148}$$

Now, we can square both sides and relate the semi-major and semi-major axes to each other using $b^2 = a^2(1-\epsilon^2)$:

$$P^2 = 4\pi^2 \frac{a^4(1-\epsilon^2)\mu^2}{l^2} \tag{149}$$

We can also relate the semi-major axis to c via $a = \frac{c}{1-\epsilon^2}$ and substitute in the expression for l yielding:

$$P^2 = 4\pi^2 \frac{a^3 c\mu^2}{l^2} = \frac{4\pi^2}{G\mu M} a^3 \tag{150}$$

and we now have the familiar equation for Kepler's third law. While in the solar system one body is much larger than the other (for planets orbiting the Sun or moons orbiting planets), Kepler's third law applies to any two body system so it applies to binary stars or binary galaxies as well.

9.2 The Virial Theorem

The virial theorem applies to systems of particles and describes the statistical properties of that system averaged over time. To derive this, consider the general equation of motion for a star in a cluster:

$$F_i = \dot{p}_i \tag{151}$$

and when we sum over all the stars:

$$\sum_i \frac{d}{dt}(m_i\vec{v_i}) = -\sum_{i,j} \frac{Gm_im_j}{|\vec{x_i} - \vec{x_j}|^3}(\vec{x_i} - \vec{x_j}) + \sum_i \vec{F_{ext,i}} \tag{152}$$

Note that the sum of the gravitational forces is only done for $i \neq j$. We now take the dot product of both sides with $\vec{x_i}$ and consider just the left hand side of the equation (before summation):

$$\frac{d}{dt}(m_i\vec{v_i}) \cdot \vec{x_i} = \frac{1}{2}(m\frac{d\vec{v_i}}{dt} \cdot \vec{x_i} + m\vec{v_i} \cdot \vec{v_i}) = \frac{1}{2}\frac{d^2}{dt^2}(m_i\vec{x_i} \cdot \vec{x_i}) \tag{153}$$

Rearranging Equation 153, and plugging it back into Equation 152, we get:

$$\frac{1}{2}\sum_i \frac{d^2}{dt^2}(m_i\vec{x_i} \cdot \vec{x_i}) - \sum_i m_i\vec{v_i} \cdot \vec{v_i} = -\sum_{i,j} \frac{Gm_im_j}{|\vec{x_i} - \vec{x_j}|^3}(\vec{x_i} - \vec{x_j}) \cdot \vec{x_i} + \sum_i \vec{F_{ext,i}} \cdot \vec{x_i} \tag{154}$$

Now, we can identify each of these terms:

$$\frac{1}{2}\frac{d^2I}{dt^2} - 2KE = PE + \sum_i \vec{F_{ext,i}} \cdot \vec{x_i} \tag{155}$$

where I is the moment of inertia of the system, KE is the total kinetic energy of the system, and PE is the total potential energy of the system. Since the virial theorem applies to the system when averaged over time, we are basically saying that the appearance of this system doesn't change over time. This implies that the moment of inertia of the system will remain constant over time. Therefore, we get:

$$2\langle KE\rangle + \langle PE\rangle + \sum_i \vec{F_{ext,i}} \cdot \vec{x_i} = 0 \tag{156}$$

This is the scalar virial theorem. The tensor virial theorem has the same structure, but says that the virial theorem holds in each dimension separately. There are many applications of the virial theorem, but most astronomical problems assume that there are no external forces that need to be considered:

$$2\langle KE\rangle + \langle PE\rangle = 0 \tag{157}$$

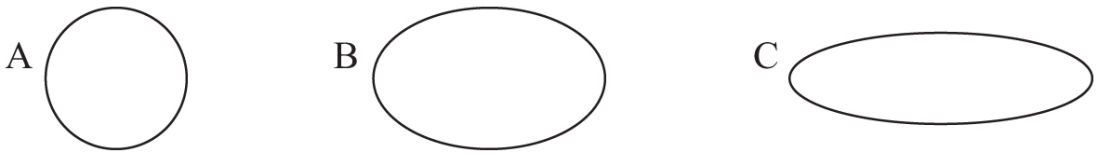
9.3 Problems

1. Given a planet with an eccentric orbit, show that the orbital velocity is highest at perihelion and lowest at aphelion.

2. Sketch out how the velocity of a planet with an eccentric orbit varies over the period of its orbit. How does this change with the eccentricity?

3. Sketch out how the velocities of the eight planets of the solar system compare as a function of their distance.

4. Consider the sketch below. Which orbit best represents the orbit of the Earth? Which orbit best represents that of a comet? How can you tell?



5. Show that Keplers Second Law holds for a particle of mass m with angular momentum l moving under a force: $F(r) = -\frac{k}{r^2} + \frac{\lambda}{r^3}$

6. Consider a spacecraft that is orbiting the Sun following the Earths orbital path. Given what you know about the energy of orbits, how would you get the spacecraft into orbit around Mercury? What about Neptune?

7. Using the scalar virial theorem, what happens to the temperature of a star cluster (not subject to external forces) as it loses stars? Explain why this is.

8. Use the virial theorem to derive the total mass of a cluster of stars.

9. Use the tensor virial theorem, explain how v_x, v_y, and v_z will vary for a flattened disk galaxy. How about for a galaxy shaped like a football? What would a star cluster look like if $v_x = 3v_y = 2v_x$? (Do not worry about the stability of such a cluster.)

10 Special Relativity

Derived from Chapter 4 in Ostlie and Carroll: "The Theory of Special Relativity"

10.1 Fundamental postulates

Special relativity is built upon two postulates, which were borne out of the finding that Maxwell's equations do not change under certain transformations (called **Lorentz transformations**). Einstein argued that we can make deeper sense of this finding if the following two postulates hold true:

1. The laws of physics are identical for all observers moving at a constant velocity. In other words, there is no experimental result that would be modified by the motion of an experimenter whose velocity is fixed.

2. The speed of light in vacuum is the same for all observers, regardless of their motion.

10.2 Time dilation

To get a better intuition about what these postulates imply, consider the following physical scenario:

Suppose you are on a train and you have a funny kind of clock (Fig. 22), in which one tick is the time it takes a laser beam to strike a mirror at the top of the train, and the next tick is when it returns to the laser pointer at the bottom of the train. Since light moves at a constant speed, each tick will be the same length of time. We can define this length of time as Δt, or one "tick" for the observer on the train.

Now let's imagine we are an observer on the ground, watching a this train passes by from left to right at speed v, as pictured in Fig. 23. Then since the light moves at a constant but not infinite speed, the mirror at the top of the train moves as the light comes up and bounces off of it. This means that the light moves upward at an angle. We will define the length of time it takes for the light to travel the height of the train as $\Delta t'$.

As illustrated in Fig. 24, the distance that light travels for the person on the train as $c\Delta t$, the distance light travels for the person on the ground as $c\Delta t'$, and finally, we know that the train moves $v\Delta t'$ during a tick of the clock as measured by the observer on the ground. We can then use the Pythagorean theorem to relate the different ticks.

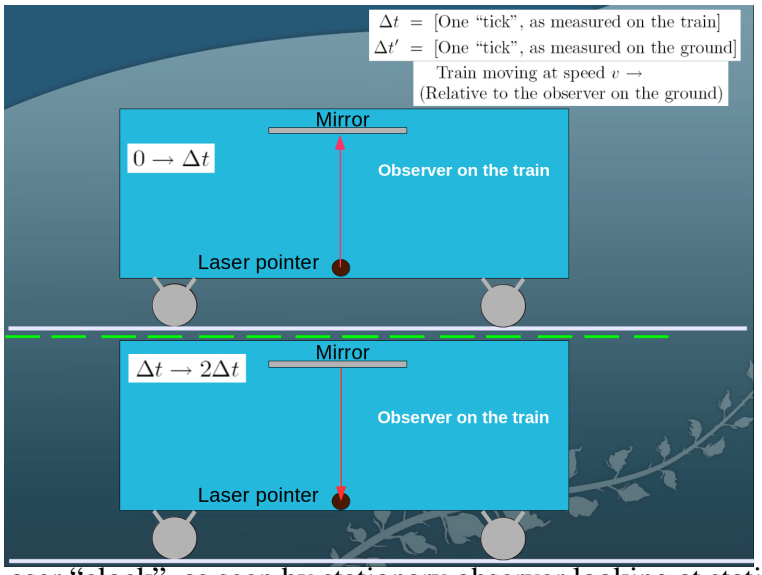


Figure 22: Laser "clock", as seen by stationary observer looking at stationary train.

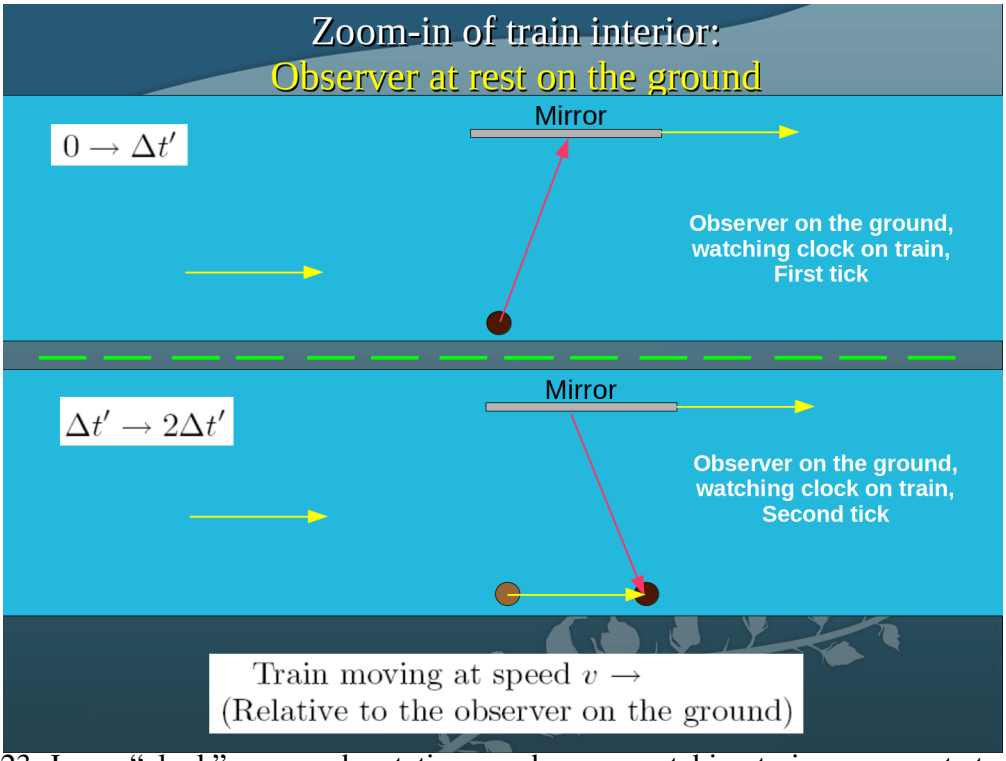


Figure 23: Laser "clock", as seen by stationary observer watching train move past at speed v.

Then after a bit of math, we can quickly derive an expression for $\Delta t'$ in terms of Δt:

$$\Delta t' = \frac{\Delta t}{\sqrt{1-(v/c)^2}} \tag{158}$$

As described in Fig. 25, this yields an alarming conclusion: $\Delta t' > \Delta t$ for $v > 0$! Thus **moving clocks tick more slowly**. The factor $\frac{1}{\sqrt{1-(v/c)^2}}$ is a common one in relativity, and is

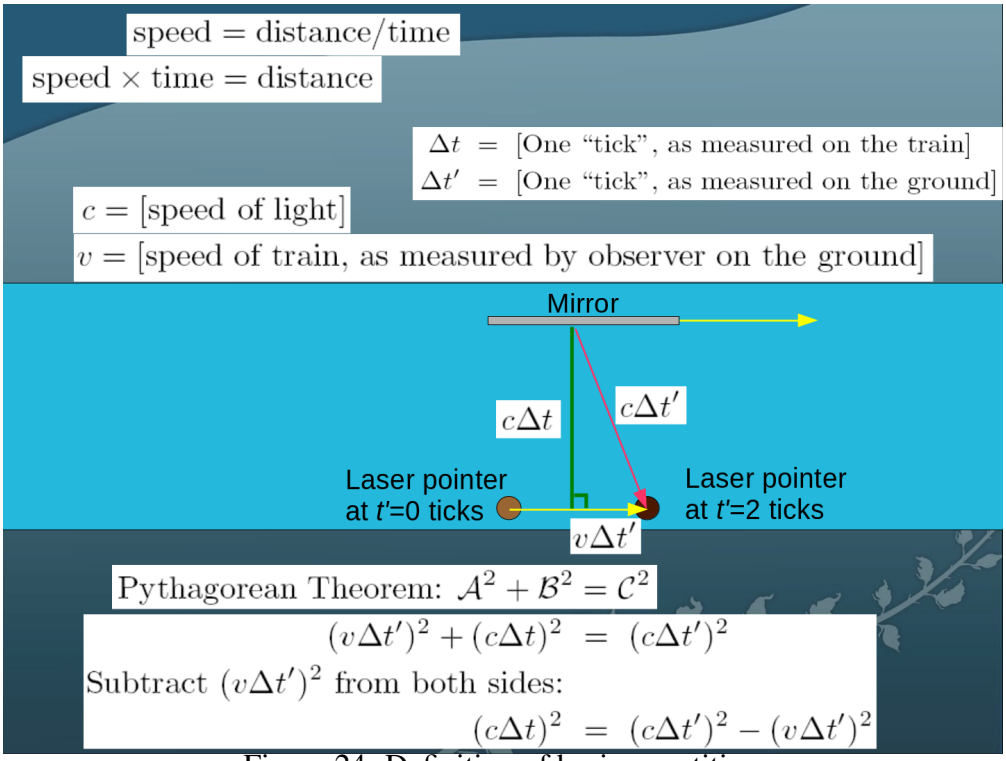


Figure 24: Definition of basic quantities.

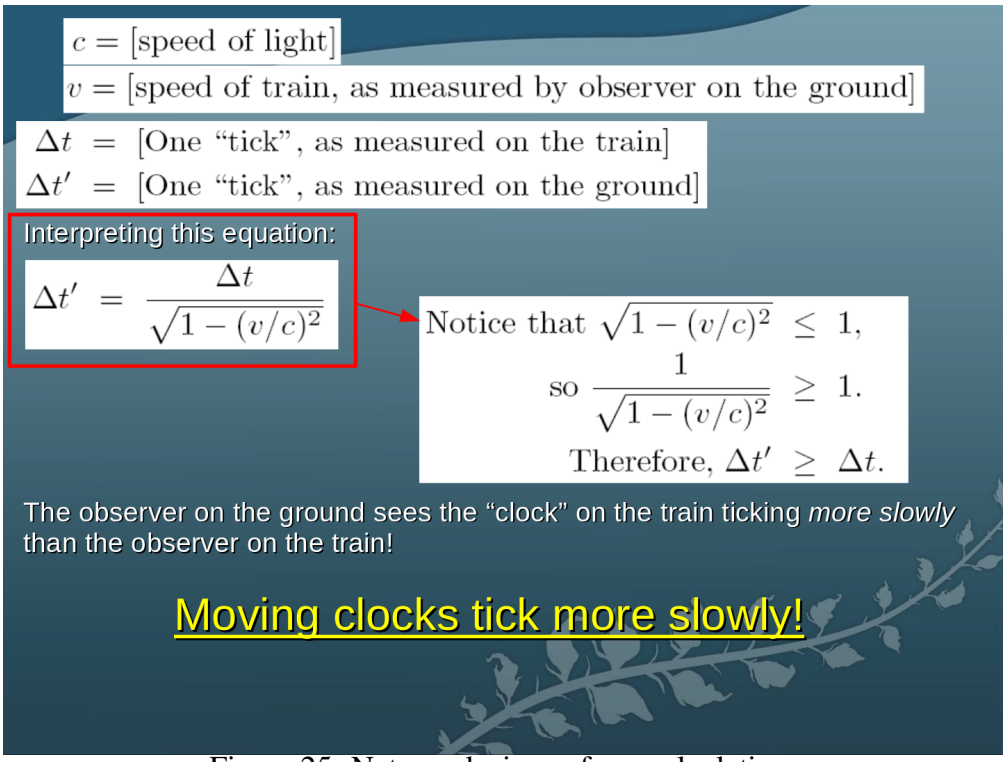


Figure 25: Net conclusions of our calculation.

sometimes known as the **relativistic gamma factor**.

So moving clocks ticking more slowly is pretty strange, but could we have predicted this? I think we could have. We know the speed of light is not infinite, and the same no matter how fast you move. **THAT is weird.** Think about it: could you imagine a car that doesn't accelerate but always travels 10 miles per hour, no matter if you run toward it or away from it? That is really really weird, and in order for it to stay at 10 miles per hour, something has to give about space—miles and time—hours. It turns out that moving clocks must tick more slowly, and moving objects will be squished in the direction they move.

This intuitive understanding under our belts, please continue by reading Sec. 4.2 through the end of Chapter 4 in Ostlie & Carroll.

10.3 Problems

1. Fill in the missing steps in the derivation between the bottom of Fig. 24 and Eq. 158.

2. Derive Eq 4.33 from Eq. 4.27 (same as Eq. 158) Fill in all missing steps.

3. Defining the force $\boldsymbol{F} = \frac{d\boldsymbol{p}}{dt}$, Taylor expand Eq. 4.44 to derive Newton's second law and the first nonzero relativistic correction. Using between one and two sentences, describe the implication of the sign of this correction.

11 General Relativity

In addition to these notes, you should also read Chapter 17 of Carroll and Ostlie.

11.1 Einstein's field equations

The field equations of General Relativity relate mass and energy, represented by the energy-momentum tensor $T_{\mu\nu}$, to spacetime curvature, represented by the Einstein tensor $G_{\mu\nu}$. The Einstein tensor is constructed from the Ricci curvature tensor $R_{\mu\nu}$ and the curvature scalar R through the relationship

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \,, \tag{159}$$

where $g_{\mu\nu}$ is the metric tensor. The Ricci tensor and curvature scalar are both formed by using the metric tensor to contract the indices of the Riemann curvature tensor, $R_{\alpha\beta\gamma\delta}$, which in turn is constructed from the metric tensor and its derivatives, so that the spacetime curvature is fully determined by the metric tensor. The metric tensor is simply a collection of coefficients that describe how a unit of length will vary in each spacetime direction, as measured from each location within that spacetime. The Einstein tensor is not the only tensor that can be constructed from curvature-related quantities, but it is the simplest choice that obeys local Lorentz invariance, meaning that physical phenomena won't depend on the choice of inertial frame.

The energy-momentum tensor has several contributions. The T_{tt} component gives the energy density, the T_{it} components give the momentum density in the i direction, and the T_{ti} components give the energy flux in the i direction. The T_{ij} components describe the $i^{\rm th}$ component of the force per unit across a surface with a unit normal in the j direction, so that the diagonal spatial components T_{ii} are simply the pressure in each spatial direction.

The field equations can be compactly stated as

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} : 8\pi T_{\mu\nu} \,, \tag{160}$$

where in the last expression, we have adopted "geometrized" units where $G = c \equiv 1$. This convention allows spatial and temporal quantities to be described in the same units, which are also the units of mass. We can relate times and lengths in geometrized and conventional units using

$$t[\mathrm{s}] = 4.92 \times 10^{-6} \left(\frac{M}{M_\odot} \right) t[M] \,, \tag{161}$$

$$r[\mathrm{km}] = 1.5 \left(\frac{M}{M_\odot} \right) r[M] \,. \tag{162}$$

11.2 Principle of equivalence

One of the foundational principles of relativity is the assumed equivalence between inertial mass (the "m" in $F = ma$ for relating any force on an object to the acceleration it induces) and gravitational mass (the "m" in $F = GMm/r^2 \equiv mg$). This principle asserts that gravity accelerates all objects equally (i.e. the two m's cancel and $a = g$), and that the effects of gravity are absent in free fall. If we define a freely falling frame x' and the Earth frame x such that

$$x' = x - \frac{1}{2}gt^2 \,, \tag{163}$$

then the acceleration in the freely falling frame is given by

$$a' \equiv \ddot{x}' = \ddot{x} - \frac{d^2}{dt^2}\left(\frac{1}{2}gt^2\right) = a - g = 0 \,, \tag{164}$$

so that Newton's laws without gravity apply in free fall. Since freely falling objects appear to move along curved paths due to gravity (i.e. we see parabolic trajectories in the Earth frame), gravity must be a curving of "straight" paths, as would be defined from a purely spatial perspective, rather than from Einstein's spacetime perspective.

11.3 Spacetime interval and general covariance

Curvature is most easily expressed as a change in the separation, ds, between nearby points in spacetime. This change, which we refer to as the spacetime interval or, more simply, the metric, is described by the metric tensor $g_{\mu\nu}$ and is given by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \,, \tag{165}$$

where dx^μ and dx^ν are simply coordinate intervals, so that the metric tensor tells us how physical separations vary as functions of the coordinates we have chosen to describe the spacetime. The metric tensor therefore depends on the coordinate system that we use, although physically observable quantities (e.g. energy fluxes) will not depend on the choice of coordinates. This independence of results on the choice of coordinates is referred to as general covariance, and was another guiding principle for Einstein in his search for his field equations.

To further illustrate this concept, we can consider a straight path in two dimensions. In Cartesian coordinates, the metric is simply $ds^2 = dx^2 + dy^2$, so that the metric tensor is just $g_{\mu\nu} = \mathrm{diag}(1, 1)$. However, in polar coordinates, $ds^2 = dr^2 + r^2 d\theta^2$ and $g_{\mu\nu} = \mathrm{diag}(1, r^2)$, so that the metric tensor appears to depend on the coordinates, which would suggest that the spacetime is curved. However, we know that this metric tensor just describes flat space in polar coordinates, so its coordinate dependence is just an artifact of this choice of coordinates. On the other hand, if we consider a "straight" path within the curved two-dimensional surface of a sphere, then $ds^2 = R^2 d\theta^2 + R^2 \sin^2\theta d\phi^2$

and $g_{\mu\nu} = \mathrm{diag}(R^2,\, R^2 \sin^2\theta)$. For this case, there is no coordinate transformation that can yield a metric tensor which does not depend on the coordinates, so this surface is truly curved. Generally, one must calculate all of the components of the Riemann curvature tensor and show that they all vanish in order to determine whether an apparently coordinate-dependent metric tensor is really describing flat spacetime in some non-Cartesian coordinate system.

We also note that the spacetime interval can take on positive or negative values. Events with a vanishing spacetime interval are said to be null- or lightlike-separated, because photons travel along paths that balance their space and time intervals to always yield a vanishing spacetime interval. Events with positive spacetime intervals are said to be spacelike separated, meaning they are separated moreso in space than they are in time. This means that even photons could not travel between these events, so they are not in causal contact with each other. Finally, events with negative spacetime intervals are said to be timelike separated, meaning they are more separated in time than they are in space, and photons, as well as massive particles, could travel between these events.

11.4 Effects of curved spacetime around a point mass

For a point mass, the most natural coordinate system will be spherical-like. Flat spacetime in standard spherical polar coordinates has a metric given by

$$ds^2 = -dt^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 = -dt^2 + dr^2 + r^2 d\Omega^2 \,. \tag{166}$$

To describe a spherically symmetric curved spacetime, we would introduce two unknown coefficients such that

$$ds^2 = -A(r)dt^2 + B(r)dr^2 + r^2 d\Omega^2 \,, \tag{167}$$

where spherical symmetry requires that the angular content of the metric be unmodified from the case of flat spacetime, and that the coefficients A and B can only depend on r. Since a point mass has no spatial extent (i.e. all mass is contained in a singularity), the vacuum form of the field equations, $G_{\mu\nu} = 0$, must be satisfied. By using the metric to calculate the Einstein tensor, and requiring that the resulting equations agree with Newton's law far from the point source, we can determine the functional form for A and B. The result is the Schwarzschild metric, given by

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \,. \tag{168}$$

If we are interested in the behavior of objects moving within this curved spacetime, then we need to consider objects moving along timelike intervals (i.e. with negative spacetime intervals). In this case, it is convenient to introduce a "proper" time such that $d\tau^2 = -ds^2$, and

$$g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = -1 \,. \tag{169}$$

For an observer at a fixed (r, θ, ϕ) outside a spherical object, the coordinate time and proper time are related by

$$\frac{dt}{d\tau} = \sqrt{-\frac{1}{g_{tt}}} = \left(1 - \frac{2M}{r}\right)^{-1/2} , \tag{170}$$

where dt can be considered the time interval measured by a distant observer, and $d\tau$ can be considered the time interval measured locally at r.

Similarly, the frequency of an emitted photon as measured by distant observers will differ from that measured locally, a phenomenon known as gravitational redshifting. The two measurements are related by

$$f_\infty = f_r \sqrt{1 - \frac{2M}{r}} . \tag{171}$$

Through a similar (but longer) calculation, if we were to consider an object moving periodically in r and θ, we would expect from Keplerian dynamics that the object would also have to complete a periodic cycle in ϕ over that same period (i.e. we would expect closed orbits). However, orbits in General Relativity do not close in this way, so objects actually advance in ϕ by more than 2π over the course of a periodic cycle in r. This extra accumulated phase $d\phi_{\mathrm{GR}}$ is referred to as periastron precession or periastron advance, and is given by

$$d\phi_{\mathrm{GR}} = 6\pi \frac{M}{a(1-e^2)} \tag{172}$$

for an orbit with semi-major axis a and eccentricity e. This precession (referred to as perihelion advance when occurring around the Sun) had been observed in Mercury's motion long before Einstein developed his theory, and in fact he checked eat iteration of his developing theory to see if they correctly postdicted the observed 43″/century in phase accumulated by Mercury's orbit around the Sun.

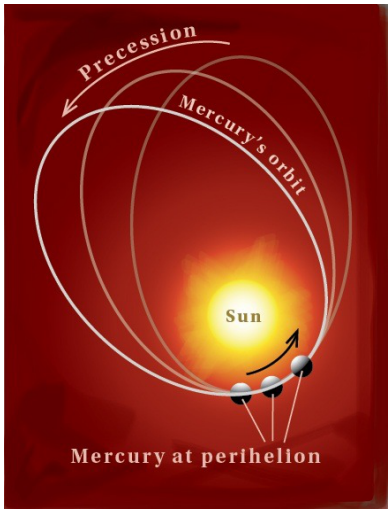


Figure 26: Perihelion advance of Mercury (not to scale).

Another closely related phenomenon is the bending of light rays as they pass a massive object. This was the first prediction of General Relativity that was spectacularly confirmed in 1919 by Sir Arthur Eddington and his team when they observed a slight shift in the apparent positions of background stars near the edge of the Sun during a solar eclipse. The shift depends on the mass of the gravitating object and the impact parameter, b, of the light ray, and is given by

$$d\phi = \frac{4M}{b} = 1.75'' \left(\frac{M}{M_\odot}\right)\left(\frac{R_\odot}{b}\right) . \tag{173}$$

When light comes close to a gravitational field and is bent, it is also delayed due to the gravitational time dilation/redshifting that occurs. This effect is known as the Shapiro time delay, and is given by

$$dt = 10\mu\text{s}\left(12 - \ln\left[\left(\frac{b}{R_\odot}\right)^2\left(\frac{1\,\text{AU}}{r}\right)\right]\right) , \tag{174}$$

where in addition to the impact parameter, the distance r between the gravitating object and the observer's location is also required, since the effect is a result of integrating along the path of the light ray as it climbs out of the gravitational potential.

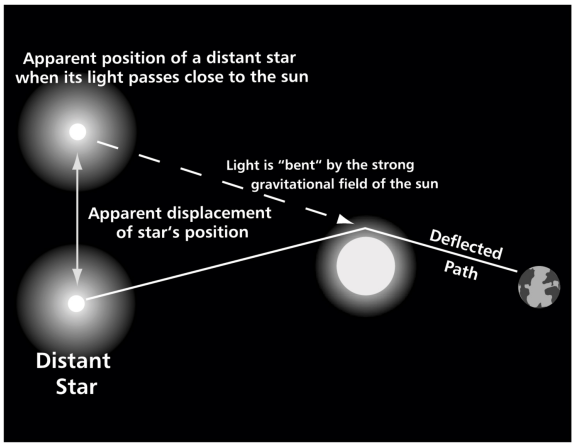


Figure 27: Bending of light and the shift in the apparent position of background stars (not to scale).

Both light bending and Shapiro delay occur for gravitational lenses. Gravitational lenses occur when an intervening massive object acts as a lens, focusing light from a distant source and enlarging the angle it subtends on the sky. The angular scale of the lensed image, referred to as the Einstein radius, is given by

$$\theta_E = \sqrt{\frac{4MD_{LS}}{D_S D_L}} = 1'' \sqrt{\left(\frac{M}{10^{11}\ M_\odot}\right)\left(\frac{D_{LS}}{D_L}\right)\left(\frac{\text{Gpc}}{D_S}\right)} , \tag{175}$$

where D_{LS} is the distance between the source and the lens, D_S is the distance between the observer and the source, and D_L is the distance between the lens and the observer.

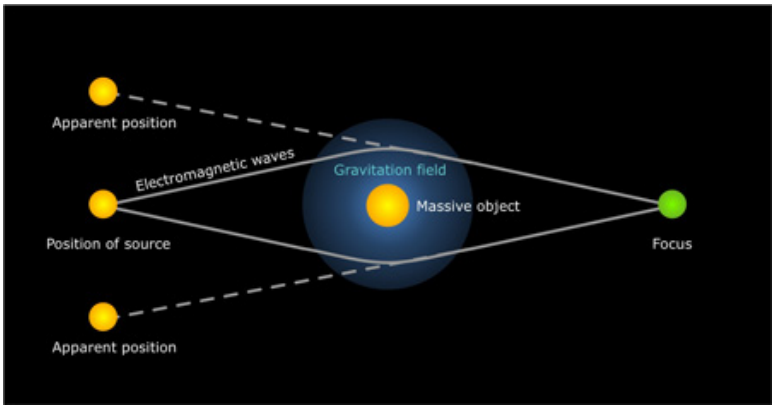


Figure 28: Gravitational lensing.

11.5 Effects of spin

If a spinning object is orbiting a point mass, it will have its spin axis reoriented. This effect, akin to periastron precession, is called geodetic precession, and is given by

$$d\phi_{\text{geodetic}} = 2\pi \left(1 - \sqrt{1 - \frac{3M}{r}}\right) , \tag{176}$$

where r is the radius of the orbit, and $d\phi_{\text{geodetic}}$ describes the rotation of the spin axis around the direction of the orbital angular momentum.

If the gravitating object itself is spinning, then the spacetime is no longer described by the Schwarzschild metric, but is instead described by the Kerr metric. The most direct observational consequence of this rotation is the phenomenon of Lense-Thirring precession, whereby an orbiting object accumulates an additional phase shift due to the dragging of the object by the spacetime itself as it swirls around the spinning object. The amount of phase shift is given by

$$d\phi_{LT} = \frac{2JT}{r^3} = \frac{2I\Omega T}{r^3} , \tag{177}$$

where J is the spin angular momentum of the gravitating object, I is its moment of inertia, and Ω is its angular frequency.

Both geodetic precession of an object's spin and Lense-Thirring precession due to Earth's rotation were observed by Gravity Probe B, which was essentially a well isolated and accurately measured gyroscope in orbit around the Earth. The observing team was able to measure a value of $d\phi_{\text{geodetic}} = 6.6''/\text{yr}$ and $d\phi_{LT} = 0.042''/\text{yr}$, although the latter measurement required a Herculean data analysis effort to overcome a larger than expected amount of measurement noise due to electrical charge buildup on the gyroscope.

11.6 Gravitational Waves

When spacetime is disturbed, it can return to equilibrium by emitting gravitational waves. These waves manifest themselves as an oscillating change in the separation between nearby points in spacetime. This change is typically characterized as a strain akin to a mechanical strain describing the fractional change in length of an object under force. If the distance between two points is initially given by x, and this distance changes by an amount Δx under the influence of a passing gravitational wave, then the strain due to that gravitational wave is simply $h = \frac{\Delta x}{x}$.

If we define the quadrupole moment, or deviation from spherical symmetry, of a distribution of mass to be

$$I_{ij} = \frac{M}{3}\int (3r_i r_j - r^2\delta_{ij}) d^3r \,, \tag{178}$$

where δ_{ij} is the Kronecker delta, then the strain is $h_{ij} \sim \frac{\ddot{I}_{ij}}{r}$. If $I \sim MR^2$ is the typical magnitude of I_{ij}'s components for a distribution of mass M with a scale of R, then $\ddot{I} \sim 2M\dot{R}^2 + 2MR\ddot{R}$. If $\ddot{R} = 0$, then $h \sim \frac{\ddot{I}}{r} \sim \frac{Mv_{\rm orb}^2}{r}$, where $v_{\rm orb}$ is the orbital velocity of the mass distribution and r is the distance between the mass and the observer.

If we consider a binary on a circular orbit, we can describe the binary either as two objects orbiting around the center of mass, or as an effective single body orbiting around a fixed object. In the effective case, the single body has a mass given by the reduced mass of the system, $\mu = m_1 m_2/M$, where $M = m_1 + m_2$ is the total mass of the binary. The single object orbits at a distance $a = r_1 + r_2$ from the fixed object, with $r_1 = am_2/M$ and $r_2 = am_1/M$ being the distances of each object in the binary from the center of mass.

We can easily calculate the components of I_{ij}, which are given by

$$\begin{aligned}
I_{xx} &= m_1(r_1\cos\phi)^2 + m_2(r_2\cos(\phi+\pi))^2 = \mu a^2\cos^2\phi = \frac{1}{2}\mu a^2(1+\cos 2\phi)\,, && (179)\\
I_{yy} &= m_1(r_1\sin\phi)^2 + m_2(r_2\sin(\phi+\pi))^2 = \mu a^2\sin^2\phi = \frac{1}{2}\mu a^2(1-\cos 2\phi)\,, && (180)\\
I_{xy} &= I_{yx} = m_1 r_1^2\sin\phi\cos\phi + m_2 r_2^2\sin(\phi+\pi)\cos(\phi+\pi)) && (181)\\
&= \mu a^2\sin\phi\cos\phi = \frac{1}{2}\mu a^2\sin 2\phi\,, && (182)
\end{aligned}$$

where $\phi = \omega t$ is the orbital phase, so that $\ddot{I}_{ij}$ is given by

$$\begin{aligned}
\ddot{I}_{xx} &= -2\mu a^2\omega^2\cos 2\phi\,, && (183)\\
\ddot{I}_{yy} &= 2\mu a^2\omega^2\cos 2\phi\,, && (184)\\
\ddot{I}_{xy} &= \ddot{I}_{yx} = -2\mu a^2\omega^2\sin 2\phi\,, && (185)
\end{aligned}$$

and the strain is given by

$$h_{ij} = 2\frac{G\ddot{I}_{ij}}{c^4 r} = -\frac{4G\mu a^2\omega^2}{c^4 r}\begin{bmatrix}\cos 2\phi & \sin 2\phi & 0\\ \sin 2\phi & -\cos 2\phi & 0\\ 0 & 0 & 0\end{bmatrix} \equiv \begin{bmatrix}h_+ & h_\times & 0\\ h_\times & -h_+ & 0\\ 0 & 0 & 0\end{bmatrix} \tag{186}$$

where h_+ and $h_\times$ are the two gravitational wave polarizations, and we note that the gravitational-wave frequency is twice the orbital frequency. For a binary with an orbital separation R, we can use Kepler's law to eliminate the frequency and find

$$h \sim |h_+| \sim |h_\times| \sim 10^{-21} \left(\frac{M}{10\ M_\odot}\right)^2 \left(\frac{20M}{R}\right) \left(\frac{100\,\mathrm{Mpc}}{r}\right) . \tag{187}$$

There are three different experimental efforts to observe gravitational waves from different sources. Ground-based interferometers like LIGO can observe strains of 10^{-23} and are sensitive to stellar mass binaries, future space-based interferometers like LISA can observe strains of 10^{-21} and are sensitive to massive black-hole binaries, and pulsar timing arrays like NANOGrav can observe strains of 10^{-15} and are sensitive to supermassive black-hole binaries.

11.7 Problems

1. A muon has a half-life $\tau = 2.19 \times 10^{-6}$ s. If it is located at $r = 2.001M$, just outside the event horizon of a Schwarzschild black hole, how long would it be expected to live according to a distant observer?

2. What is the escape velocity for a particle orbiting a Schwarzschild black hole at $r = 6M$?

3. If observers can detect a perihelion advance of 1"/century, which planets would have detectable precession?

4. If an object is tidally disrupted at some critical differential tidal force $\left(\frac{dF}{dr}\right)_{\mathrm{crit}}$, relate this to the Newtonian tidal force for an orbit at $r = 6M$, which is the innermost stable circular orbit. From this, determine how a black hole's ability to disrupt an object depends on the black hole's mass.

5. If a gravitational wave $h \propto \exp(-(t-t_\mathrm{o})^2/\sigma^2) \sin \omega t$ passes two test particles separated by Δx, what is their separation after the gravitational wave passes?

12 Compact Objects

The following notes are based on the following textbooks: Carroll & Ostlie's Modern Astrophysics, Arnab Choudhuri's Astrophysics for Physicists, and Shapiro & Teukolsky's Black Holes, White Dwarfs and Neutron stars. Interested readers should refer these books for a detailed explanation.

12.1 White Dwarfs, Neutron stars and Black Holes

Compact objects are the end products of stellar evolution. The primary factor determining whether a star ends up as a white dwarf, neutron star, or a black hole is believed to be due to the star's mass. All these compact objects differ from normal stars in two fundamental ways. Firstly, they cannot support themselves against gravitational collapse by generating thermal pressure (since they do not burn nuclear fuel). Therefore, to balance the gravitational self-attraction, a nonthermal source of pressure is required (because the star will eventually cool). This pressure arises quantum mechanically from the Pauli principle, which makes a gas of cold fermions resist compression (known as the degeneracy pressure). In a white dwarf, the gravity is balanced by the electron degeneracy pressure. However, once the density of matter approaches nuclear density, the degeneracy pressure of neutrons becomes important (at such high density, inverse beta decay converts protons into neutrons). A neutron star is thus supported by the degeneracy pressure of neutrons. Black holes, on the other hand, are completely collapsed stars, i.e., stars that could not find any means to hold back the inward pull of gravity and therefore collapsed to singularities. Secondly, the compact objects have much smaller radii compared to normal stars and hence, much stronger surface gravitational fields. Some general properties of neutron stars and black holes are discussed below:

- **Electron degeneracy pressure:** Let us derive the degeneracy pressure for non-relativistic, degenerate electron gas. This can be done by combining Pauli's exclusion principle and Heisenberg's uncertainty principle. Let us assume that the gas is contained in a cubic box with dimensions $\triangle x$, $\triangle y$, $\triangle z$ and corresponding momenta of p_x, p_y, p_z. The uncertainty principle implies that

$$\triangle x p_x \approx \triangle y p_y \approx \triangle z p_z \approx \hbar \tag{188}$$

  Making a (slightly unreasonable) assumption that all electrons have the same momentum, we get the pressure as

$$P = \frac{1}{3} n_e p v \tag{189}$$

where n_e is the electron density. For a completely degenerate gas,

$$\triangle x \approx n_e^{-1/3} \implies p_x \approx \hbar n_e^{1/3} \tag{190}$$

Given that the electron gas is uniformly distributed, each direction is equally likely and thus,

$$p_x = p_y = p_z \implies p^2 = 3p_x^2 \implies p = \sqrt{3}p_x \approx \sqrt{3}\hbar n_e^{1/3} \tag{191}$$

putting the electron density in terms of atomic number (Z), atomic mass number (A), mass density (ρ) and the mass of Hydrogen atom (m_H), we get

$$n_e = \left[\left(\frac{Z}{A}\right)\frac{\rho}{m_H}\right] \tag{192}$$

Using the above expression of n_e and the fact that the velocity for non-relativistic electrons is given by $v = p/m_e$ in the expression for pressure, we get

$$P \approx \frac{\hbar^2}{m_e}\left[\left(\frac{Z}{A}\right)\frac{\rho}{m_H}\right]^{5/3} \tag{193}$$

For relativistic, degenerate gas, we have $v = c$ and the degeneracy pressure is given by

$$P \approx \frac{\hbar^2}{m_e}\left[\left(\frac{Z}{A}\right)\frac{\rho}{m_H}\right]^{4/3} \tag{194}$$

- **Maximum mass**: Chandrasekhar worked out that there is a maximum mass for a white dwarf, now generally referred to as the Chandrasekhar limit and is estimated to be 1.44 $M_\odot$ (for derivations, refer to the above textbooks). Neutron stars are also subject to a maximum mass, just like the white dwarfs. The mass radius relation is $R \propto M^{-1/3}$ for all degeneracy pressure-supported stars; so as the mass increases, the radius shrinks and the density rises. How well can we measure the neutron star masses and radii? Masses can be straightforward if the star is in a binary. The masses of some neutron stars have been measured with the highest precision of any objects outside the Solar System. These are the ones for which post-Keplerian parameters such as pericenter precession, Shapiro delay, and orbital decay due to gravitational radiation can be measured. The maximum mass could be much higher than 2 $M_\odot$, and the radius consistent with this range of masses could run from 10 –15 km for a canonical 1.4 $M_\odot$ star.

- **Neutron star structure:** A rough estimate of the characteristic radius and density of a neutron star can be made by considering that it is held up by neutron degeneracy pressure. The neutrons in the star are (marginally) non-relativistic, so the pressure is given by the general formula derived for a non-relativistic degenerate gas as

$$P = \left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{20m} n^{5/3} \tag{195}$$

where m is the mass per particle and n is the number density of particles. If we consider that the star is composed of pure neutrons, then $m = m_n = 1.67 \times 10^{24}$ g and $n = \rho/m_n$. Plugging this in the above equation, we get

$$P = \left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{20m_n^{8/3}} \rho^{5/3}. \tag{196}$$

This pressure must be sufficient to hold up the star. To see what this implies, we approximate the structure of the star as a polytrope[8]. This is a reasonably good approximation since the pressure in most of the star is dominated by non-relativistic degeneracy pressure, corresponding to an $n = 3/2$ polytrope. Using the relationship between central pressure and density appropriate to polytropes,

$$P_c = (4\pi)^{1/3} B_n G M^{2/3} \rho_c^{4/3}, \tag{197}$$

where B_n is a constant that depends (weakly) on the polytropic index n. Combining this with the pressure-density relation for a degenerate neutron gas, we get

$$\left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{20m_n^{8/3}} \rho^{5/3} = (4\pi)^{1/3} B_n G M^{2/3} \rho_c^{4/3} \tag{198}$$

$$\rho_c = \frac{4}{9} (20\pi B_n)^3 \frac{G^3 M^2 m_n^8}{h^6} \tag{199}$$

We can also make use of the relationship between central density and mean density for polytropes.

$$D_n = \frac{\rho_c}{\rho} = \rho_c \frac{4\pi R^3}{3M} \tag{200}$$

[8]Polytropes are self-gravitating gaseous spheres that were, and still are, very useful as crude approximation to more realistic stellar models. An equation of state of the form $P = \kappa\rho^\gamma$, where γ and κ are constants, is called polytropic. It is customary to define the corresponding polytropic index, denoted by n, as $\gamma = 1 + \frac{1}{n}$. Note that n does not have to be an integer. In the case that $\gamma = 5/3$ we get $n = 3/2$.

where $D_n = -[(3/\xi_1)(d\Theta/d\xi)_{\xi_1}]^{-1}$ is another constant that depends on the polytropic index (for details, cross ref. polytropic relation and Lane-Emden equation). Substituting this in for ρ_c, we get

$$\frac{3M}{4\pi R^3} D_n = \frac{4}{9}(20\pi B_n)^3 \frac{G^3 M^2 m_n^8}{h^6} \tag{201}$$

$$R = \frac{3 D_n^{1/3}}{20(2\pi)^{4/3} B_n} \left(\frac{h^2}{G m_n^{8/3}}\right) \frac{1}{M^{1/3}} \tag{202}$$

Plugging in the values appropriate for an n = 3/2 polytrope (D_n = 5.99 and B_n = 0.206) gives

$$R = 14 \left(\frac{M}{1.4 M\odot}\right)^{-1/3} km. \tag{203}$$

This R is only slightly higher than what we get using more sophisticated models ($\sim$10 km) for neutron stars that have had a chance to cool off from their initial formation and become fully degenerate. The slight discrepancy is due to several reasons: (i) the fluid is not purely neutrons; there are some protons too, which do not contribute to the neutron degeneracy pressure, (ii) the neutrons are not too far from being relativistic and this reduces their pressure compared to the fully non-relativistic pressure we have used above, and (iii) the neutron matter also has a considerably more complex structure than a simple degenerate electron gas due to the nuclear forces between the neutrons.

- **Magnetic fields**: Another important property of neutron stars is that they are superconductors, i.e., they have nearly infinite electrical conductivity. Therefore, electric currents flow with essentially no resistance and magnetic fields diffuse very little in superconductors; fields do not diffuse in or out of them. Therefore, the magnetic field within them is said to be "frozen into the fluid" meaning that any field line that passes through a given fluid element is trapped in that fluid element and moves and deforms with it. A magnetic field that is deformed (stretched or compressed) responds by applying a restoring "Lorentz force" on the fluid. The magnetic flux through a small region of surface area A on the initial core was AB_i. For the final neutron star, this surface area has shrunk in proportion to the radius squared. Therefore,

$$R_i^2 B_i = R_f^2 B_f \implies B_f = \left(\frac{R_i}{R_f}\right)^2 B_i \tag{204}$$

Thus when the core collapses to make a neutron star, the magnetic field that is trapped in the core is enhanced by a factor of $(R_i/R_f)^2$.

- **Pulsars and rotation**: The strong magnetic field is particularly important when coupled with another aspect of neutron stars – rapid rotation. Neutron stars are rapid rotators for exactly the same reason they are strongly magnetized: conservation during collapse, in this case conservation of angular momentum. Consider a massive star core rotating with an initial angular velocity ω_i and an initial moment of inertia $I_i = C_i M R_i^2$, where C_i is a constant or order unity that depends on the core's density structure. Its angular momentum is then given by

$$L = I_i \omega_i = C_i M R_i^2 \omega_i \tag{205}$$

As it collapses it must conserve angular momentum, so its angular momentum after collapse is

$$L = I_f \omega_f = C_f M R_f^2 \omega_f = C_f M R_i^2 \omega_i \implies \omega_f \approx \omega_i \left(\frac{R_i}{R_f} \right)^2 \tag{206}$$

Thus the angular velocity of the core is enhanced by the same factor of $\sim 10^6$ as the magnetic field. The period $P = 2\pi/\omega$ decreases by the same factor. The combination of a strong magnetic field and rapid rotation gives rise to pulsation. Refer to the textbooks (Carroll & Ostlie or Lorimer & Kramer) for pulsar properties.

- **Schwarzschild Radius**: In some cases, the stellar core of a star that explodes as a supernova can be pushed above the maximum neutron star mass. This can happen if not all of the stellar envelope is ejected and some of it falls back onto the proto-neutron star. In such a case, the neutron star may accrete the material and exceed its maximum mass. Another possibility is that a very massive star may encounter the pair instability region in the ($\log \rho$, $\log T$) plane before it gets to the iron photo-disintegration instability region. In this case, the collapsing core may be more massive than 2–3 solar masses and the result will again be a core that exceeds the maximum possible neutron star mass. If such a core is created, nothing can stop it from collapsing indefinitely.

  As the star collapses, the escape velocity from its surface rises:

$$v_{\rm esc} = \sqrt{\frac{2GM}{R}} \tag{207}$$

  Once the radius is small enough, this velocity exceeds the speed of light. The critical velocity at which this happens is called the Schwarzschild radius:

$$R_{\rm sch} = \frac{2GM}{c^2} \approx 3\frac{M}{M_\odot}\ {\rm km} \tag{208}$$

The Schwarzschild radius is the effective size of the black hole. Nothing that approaches within that distance of the mass can escape, since nothing can move faster than light. Because nothing that happens inside the Schwarzschild radius can ever influence events outside it, the Schwarzschild radius is called an event horizon.

- **Accretion power**: The luminosity of a neutron star is given by

$$L = 4\pi R^2 \sigma T^4. \tag{209}$$

Neutron stars are born very hot, T $> 10^{10}$ K, but after $\sim$1 Myr, the star cools and the temperature drops to $\sim 10^6$ K. For R = 10 km and temperature $\sim 10^6$ K in the luminosity equation above, L = 0.2 $L_\odot$. This is dim enough to make it quite hard to detect any but the nearest neutron stars by thermal emission. However, we can detect neutron stars and black holes when they emit nonthermally (e.g. pulsars) or if they are powered by accretion of material from another body. Consider a neutron star of radius R that accretes an amount of mass dM in a time dt. The material falls from rest at infinity, so it has zero energy initially. Just before it arrives at the surface, its potential and kinetic energies must add up to zero, so

$$\frac{1}{2}v^2 dM - \frac{GMdM}{R} = 0 \implies \frac{1}{2}v^2 = \frac{GM}{r} \tag{210}$$

When the material hits the surface and stops, its kinetic energy is converted into heat, and then it is radiated away. In steady state, all the extra energy must be radiated and hence, the amount of energy released is

$$dE = \frac{1}{2}v^2 dM = \frac{GMdM}{R} \tag{211}$$

The resulting luminosity is just the energy per unit time emitted via this process:

$$L_{\rm acc} = \frac{dE}{dt} = \frac{GM}{R}\dot{M} \tag{212}$$

Accretion luminosity increases as the radius of the star decreases, which means that it can be a much more potent energy source for compact things like neutron stars. For example, suppose a star accretes at a rate of 10^{-10} $M_\odot$ yr^{-1} such that it gains roughly 1 $M_\odot$ of mass over the age of the universe. For a white dwarf of R = 0.01 $R_\odot$, $L \approx$ 0.1 $L_\odot$ – high enough to be brighter than just an isolated white dwarf normally is. For a neutron star of R = 10 km, it would be 100 $L_\odot$ and for a black hole

of R $\approx$ 3 km, it approaches 1000 $L_\odot$. This process cannot produce arbitrarily high luminosities for the same reason that stars cannot have arbitrarily high luminosities: the Eddington limit. For any object in the depths of space, there is a maximum luminosity beyond which radiation pressure will overcome gravity and material outside the object will be forced away from it rather than falling inwards – Eddington luminosity. The Eddington limit is

$$L_{\rm Edd} = \frac{4\pi cGM}{\kappa} \tag{213}$$

where κ is the opacity. In high-energy accretion scenarios, we make a useful approximation on the basis that the accreting material is mostly ionized hydrogen and the opacity is provided by Thomson scattering. The cross-section will then come almost exclusively from radiation pressure on the electrons, but the mass lies almost exclusively in the protons. The Eddington luminosity depends only on the mass of the radiating object. The approximation is to set $\kappa = \sigma_T/m_p$ and we get the following approximation for the Eddington luminosity:

$$L_{\rm Edd} = \frac{4\pi cGMm_p}{\sigma_T} \tag{214}$$

If we require that $L_{\rm acc} < L_{\rm Edd}$, then we have

$$\frac{GM}{R}\dot{M} < \frac{4\pi cGM}{\kappa} \implies \dot{M} < \frac{4\pi cR}{\kappa}. \tag{215}$$

Thus, there is a maximum accretion rate onto compact objects.

12.2 Problems

The following questions are from Chapter 16 of Carroll and Ostlie.

1. Estimate the neutron degeneracy pressure at the center of a 1.4 $M_\odot$ neutron star (take the density to be 1.5×10^{15} g cm^{-3}) and compare this with the estimated pressure at the center of Sirius B?

2. (a) At what speed do relativistic effects become important at a level of 10%. In other words, for what value of v does the lorentz factor γ become equal to 1.1?
(b) Estimate the density of the white dwarf for which the speed of a degenerate electron

is equal to the value found in part (a).

3. Consider a pulsar that has a period P_0 and period derivative $\dot{P}_0$ at t = 0. Assume that the product $P\dot{P}$ remains constant for the pulsar (c.f. Eq. 16.29 in CO).
(a) Integrate to obtain an expression for the pulsar's period P at time t.
(b) Imagine that you have constructed a clock that would keep time by counting the radio pulsars received from this pulsar. Suppose you also have a perfect clock ($\dot{P}$ = 0) that is initially synchronized with the pulsar clock when they both read zero. Show that when the perfect clock displays the characteristic lifetime $P_0/\dot{P}_0$, the time displayed by the pulsar clock is $(\sqrt{3}-1)\ P_0/\dot{P}_0$.

4. During a glitch, the period of the Crab pulsar decreased by $|\Delta P| \approx 10^{-8}\ P$. If the increased rotation was due to an overall contraction of the neutron star, find the change in the star's radius. Assume that the pulsar is a rotating sphere of uniform density with an initial radius of 10 km.

5. (a) Determine the minimum rotation period for a 1.4 solar mass neutron star (the fastest it can spin without flying apart). Assume that the star remains spherical with a radius of 10 km.

(b) Newton studied the equatorial bulge of a homogeneous fluid body of mass M that is slowly rotating with angular velocity (Ω). He proved that the difference between its equatorial radius (E) and its polar radius (P) is related to its average radius (R) by

$$\frac{E-P}{R} = \frac{5\Omega^2 R^3}{4GM}. \tag{216}$$

Use this to estimate the equatorial and polar radii for a 1.4 solar mass neutron star rotating with twice the minimum rotation period found in part (a).

6. Compare the maximum angular momentum of a 1.4 $M_\odot$ black hole with angular momentum of the fastest known pulsar, which rotates with a period of 0.00139 s. Assume that the pulsar is a 1.4 $M_\odot$ uniform sphere of radius 10 km.

13 Cosmology

13.1 Background Material

Caroll and Ostlie Chapter 29

13.2 Other References: Books

- Barbara Ryden - Introduction to Modern Cosmology (Great first source)
- Andrew Liddle - An Introduction to Modern Cosmology (Another great first source, bit more brief)
- Steven Weinberg - Cosmology (Indepth overview)
- Scott Dodelson - Modern Cosmology (Focuses more on CMB)

13.3 Introduction

Cosmology is the study of the Universe on the largest scales, where even galaxies are considered "small". The current concordance model, the ΛCDM model, still fits all observations to very high precision. This model leads to a Universe which:

- Started with a hot big bang
- Initially went through a period of exponential expansion (Inflation)
- Continued to expand adiabatically and cool
- Cooled enough to form atoms, and created the Cosmic Microwave Background
- Began to form stars, and reionize the Universe
- Formed Galaxies, Planets, all other structures
- Became dominated by a "Dark Energy" and is again undergoing exponential expansion

Figure 29 has a nice graphic representation of our current understanding of the Universe and how it has evolved.

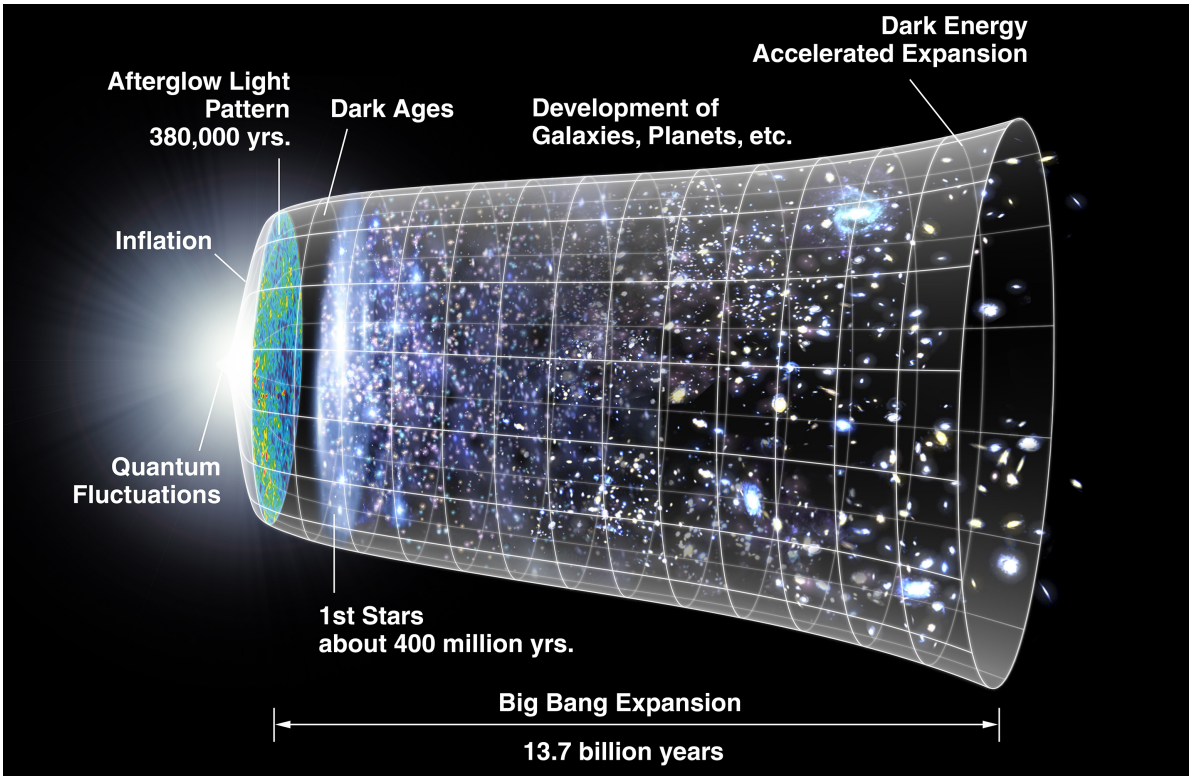


Figure 29: Current view of the history of the Universe

13.4 Expansion of the Universe: Friedmann Equations

Assuming that space is homogeneous (translationally invarient, no 'origin') and isotropic (rotationally invarient, no prefered direction), leads to the Robinson-Walker Metric

$$ds^2 = -dt^2 + a(t)^2[dR_o^2 + f_K(R_o)^2 R_o^2 d\omega^2] \tag{217}$$

where $d\omega^2 = d\theta^2 + sin^2\theta d\phi^2$, f_K is a curvature function, and $R = R_o a(t)$ is the "true" distance with R_o the co-moving distance and $a(t)$ the scale factor. (Carol and Ostle use a slightly different nomenclature).

Using the RW metric along with Einstein's Equation:

$$G^{\mu\nu} = \frac{8\pi G}{c^2} T^{\mu\nu} - \Lambda g^{\mu\nu} \tag{218}$$

leads to the Friedmann Equation:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{Kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{219}$$

and the Fluid Equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) = 0 \tag{220}$$

where ρ is the density, p is the pressure, Λ is a cosmological constant. These equations can be used to describe how the Universe evolves based on the content within.

13.5 Problems

1. A Friedmann Equation can be derived from conservation of energy of expanding non-relativistic shell. Use conservation of energy (Assume a constant energy to start with of $-\frac{1}{2}mKc^2R_o^2$) and an expanding shell to derive a Friedmann equation. What is missing?

2. Derive the Fluid equation from the first law of thermodynamics and non-relativistic matter. Again use a spherical shell of matter expanding with time.

3. Assuming Flat space, use the fluid and Friedmann equations, along with the equation of state for individual components to derive the density and scale factor as a function of time for:

   (a) A matter dominated Universe ($p = 0$)

   (b) A photon dominated Universe ($p = \frac{1}{3}\rho c^2$)

   (c) A Cosmological Constant dominated Universe ($p = -\rho c^2$)

4. Write the pseudo code or procedure to solve for the evolution of the scale factor with time with arbitrary density of photons, matter, and a cosmological constant.

5. In a Flat matter dominated universe, with Ho = 70 km/s/Mpc, how long do you need to wait to see a galaxy change by one part in a million (first can show)

$$\frac{dz}{dt_o} = H_o(1+z) - H_o(1+z)^{3(1+w)/2}$$

   and use $w = 0$ for matter.

6. CMB: What is the current energy density of the CMB? What is the current number density? What is the current energy density of baryons? Number density? What is the current number ratio of baryons to photons? What is Olber's Paradox? If we could 'see' at mm wavelengths, what would the sky look like? (compare this to optical sky dynamic range and scales). Would Olber's Paradox have been posited if we saw in mm? Why/Why not?

14 The Structure of the Universe

In addition to these notes, you should also read Chapter 27 of Carroll and Ostlie.

14.1 Global properties

On the largest cosmic scales, the Universe is *homogeneous* and *isotropic*. **Homogeneity** means that there is no preferred location, so all vantage points within the Universe will look the same. **Isotropy** means that there is no preferred direction, so regardless of which direction you look, the Universe will look the same. Referring back to our General Relativity lecture, the metric on the largest scales is essentially that of flat space:

$$ds^2 = c^2 dt^2 - d\ell^2 \,, \tag{221}$$

$$d\ell^2 = a(t)^2 (dx^2 + dy^2 + dz^2) \,, \tag{222}$$

where $a(t)$ is the scale factor for the Universe, and represents the expansion of spacetime that we observe.

On the largest scales, this expansion is accelerating due to the affects of dark energy, with dark matter overcoming the affects of dark energy on the scale of gravitationally bound structures. The behavior of this scale factor under the influence of dark energy and dark matter is the central concern of cosmology. For now, we simply note that 95% of the Universe is dark, and most of that is dark energy. There is 5x as much dark matter as ordinary matter.

14.2 Expansion of the Universe

Dark matter and dark energy have competing effects on expansion, as characterized by Friedmann's equation:

$$\frac{1}{a}\frac{d^2 a}{dt^2} = -\frac{4\pi G}{3}\rho + \frac{c^2}{3}\Lambda \,, \tag{223}$$

where ρ is the total mass density (baryons + dark) and Λ is the cosmological constant responsible (in the ΛCDM model of cosmology) for the observed accelerated expansion of the Universe.

On scales much smaller than the Universe as a whole, but larger than the scale of bound structure, the scale factor manifests itself through Hubbles law, whereby distant galaxies are observed to recede from us, with a velocity that is proportional to their distance:

$$v = \frac{dR}{dt} = HR \tag{224}$$

$$H \equiv \frac{\dot{a}}{a} \,, \tag{225}$$

where H is the Hubble parameter. The recession picture doesnt work on the largest scales, but it does give the right picture for establishing the concept of redshift z and its impact on observed wavelengths:

$$z = \frac{v}{c} = \frac{\Delta\lambda}{\lambda}, \tag{226}$$

$$1 + z = \frac{1}{a}. \tag{227}$$

14.3 Substructure of the Universe

The large-scale structure of the Universe is made up of voids and filaments, that can be further broken down into superclusters, clusters, galaxy groups, galaxies, and so on. Filaments are threadlike structures made up of isolated galaxies, groups, clusters and superclusters. Voids are vast empty bubbles of empty space. The origin of filaments and voids is the very small anisotropy present in the early Universe and visible and temperature fluctuations in the CMB. On the whole, the density of matter is too low to overcome the repulsion of dark energy, but overdensities above a certain density threshold can overcome the accelerated expansion. In this context, the expansion is often referred to as the Hubble flow, and structures dense enough to slow their expansion and eventually collapse have disconnected from the Hubble flow. When the overdensity collapses to a density above a critical threshold, the structure in the overdensity virializes. It is then gravitationally stable, and ceases to collapse, ultimately former groups and clusters.

Although some galaxies fly off on their own, most are bundled into groups and clusters. Groups are smaller, usually made up of less than 50 galaxies with diameters 1–2 Mpc. The Local Group, in which our Milky Way is a member, is made up of $\sim$40 galaxies. Clusters are bunches of 50–1000 galaxies with diameters of 2–10 Mpc. One peculiar property of clusters is that the velocities of their galaxies appear too large for gravity alone to keep them bunched together. This is the scale over which dark matter affects the structure of the Universe. A great number of groups, clusters and individual galaxies can come together to form superclusters.

14.4 Observational Evidence for ΛCDM

The principle evidence for dark energy is the agreement of the observed accelerated expansion with the exponential growth predicted by theory. The principle evidence for dark matter is the missing matter needed to explain clustering, galaxy rotation curves, etc. Alternative explanations for these exist, chief among them modifying gravity itself on large scales where it is very weak, but these theories dont agree well with other observations, in particular gravitational lensing.

14.5 Problems

1. If the Universe is isotropic, why does everything appear to be receding from us in every direction?

2. If the Universe is expanding, what is it expanding into?

3. How does the scale factor evolve with time if the Universe is matter dominated, radiation dominated, or dark energy dominated?

4. Prove that a circular orbit is virialized, and discuss the behavior of a noncircular binary in the context of virialization and time averaging.