4. Rotation Curves

Gas and young stars will move on nearly closed orbits, and if the underlying potential is axisymmetric these will be nearly circular. So if you measure the bulk velocity v (of gas or young stars, *not* old stars) at any place on a galactic disc, you've measured $R\,(\partial\Phi/\partial R)$; and if you measure $v(R)$—the 'rotation curve'—you have information on the mass distribution.

When people first starting measuring rotation curves (c. 1970), it quickly became clear that the mass in disc galaxies doesn't follow the visible disc. Disc galaxies generically have rotation curves that are fairly flat to as far out as they can be measured (several scale radii). The simplest interpretation of a flat rotation curve is that enclosed mass $M(r) \propto r$, or $\rho(r) \propto 1/r^2$, a 'dark halo'. The deep picture of M104 in part 1 of these notes suggests that dark halos are not entirely dark, but as yet nobody knows really knows how far they extend. And there is no good estimate of the total mass of any disc galaxy. This is what makes disc rotation curves very important.

However, one needs to be a little careful about interpreting flat rotation curves. The maximum contribution to the rotation curve from an e^{-R/R_0} disc is not (as we might naively expect) around R_0 but around $2.5R_0$. Adding the effect of a bulge can easily give a fairly flat rotation curve to $4R_0$ without a dark halo. To be confident about the dark halo, one needs to have the rotation curve for $\gtrsim 5R_0$. In practice, that means H I measurements; optical rotation curves don't go out far enough to say anything about dark halos.

The rest of part 4 is a more detailed working out of the previous paragraph. It follows an elegant derivation and explanation due to A.J. Kalnajs.

The potential from a disc with surface density $\Sigma(R)$ is

$$\Phi(R) = -G \int_0^\infty R'\Sigma(R')\,dR' \int_0^{2\pi} \frac{d\phi}{\sqrt{R^2 + R'^2 - 2RR'\cos\phi}}. \tag{4.1}$$

To make this tractable, let us first define[1]

$$2\pi L(u) \equiv \int_0^{2\pi} \frac{d\phi}{\sqrt{1+u^2-2u\cos\phi}} = 1 + \left(\tfrac{1}{2}u\right)^2 + \left(\frac{\frac{1}{2}\frac{3}{2}u^2}{2!}\right)^2 + \ldots \quad (u<1). \tag{4.2}$$

Then

$$\Phi(R) = -2\pi G \int_0^R \Sigma(R') \left(\frac{R'}{R}\right) L\left(\frac{R'}{R}\right) dR' - 2\pi G \int_R^\infty \Sigma(R') L\left(\frac{R}{R'}\right) dR', \tag{4.3}$$

and hence

$$\begin{aligned} v^2(R) = R\frac{\partial\Phi}{\partial R} = 2\pi G \int_0^R &\left[\left(\frac{R'}{R}\right) L\left(\frac{R'}{R}\right) + \left(\frac{R'}{R}\right)^2 L'\left(\frac{R'}{R}\right)\right] \Sigma(R')\,dR' \\ -2\pi G &\int_R^\infty \left(\frac{R}{R'}\right) L'\left(\frac{R}{R'}\right) \Sigma(R')\,dR'. \end{aligned} \tag{4.4}$$

[1] If you really want to know where that came from, look up any musty old celestial mechanics book under 'Laplace coefficients'.

(Here L' means a derivative!) The important thing to take away with you is not the algebraic mess but the form of the relation, which is

$$v^2(R) = 2\pi G \int_0^\infty K\left(\frac{R}{R'}\right) \Sigma(R')\, dR'. \tag{4.5}$$

Changing variables to

$$x = \ln R, \qquad y = \ln R',$$

we can write this as a convolution

$$v^2(R) = 2\pi G \int_{-\infty}^{\infty} K(e^{x-y})\, R'\Sigma(R')\, dy. \tag{4.6}$$

The kernel $K(R/R')$ is in Figure 4.1.

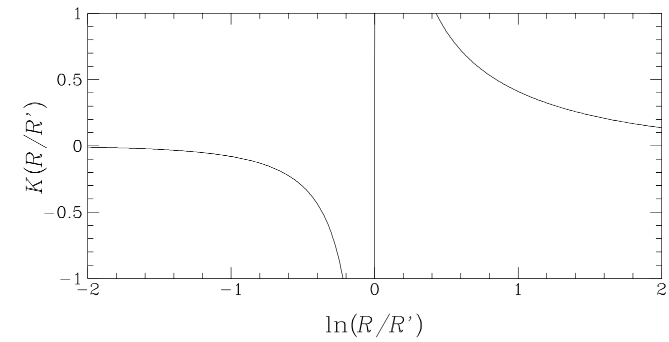

Figure 4.1: The kernel $K(R/R')$. Observe that the $R > R'$ part tends to have higher absolute value than the $R < R'$ part.

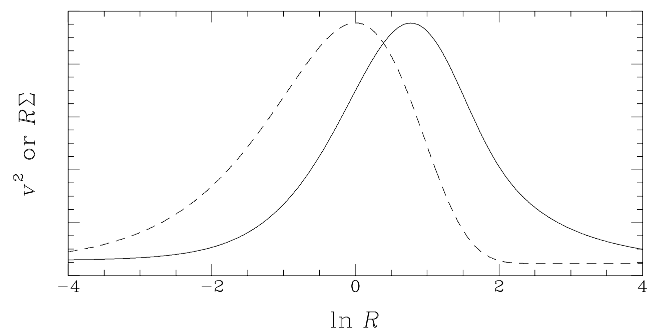

Figure 4.2: The dashed curve is $R\Sigma(R)$ for an exponential disc with $\Sigma \propto e^{-R}$ and the solid curve is $v^2(R)$. Note that R is measured in disc scale lengths, but the vertical scales are arbitrary.

Figure 4.2 shows $R\Sigma(R)$ and v^2 for an exponential disc, but the general shapes aren't very sensitive to whether $\Sigma(R)$ is precisely exponential. The important qualitative fact is that whatever $R\Sigma(R)$ does, v^2 does roughly the same, but expanded by a factor of $\simeq e$.

The distinctive shape of the $v^2(\ln R)$ curve for realistic discs makes it very easy to recognize *non*-disc mass. Figure 4.3, following Kalnajs, shows the rotation curves you get by adding either a bulge or a dark halo. (Actually this figure fakes the bulge/halo contribution by adding a smaller/larger disc; but if you properly add spherical mass distributions for disc/halo, the result is very similar.) Kalnajs' point is that a bulge+disc rotation curve has a similar shape to a disc+halo rotation curve—only the scale is different. So when examining a flat(-ish) rotation curve, you must ask what the disc scale radius is.

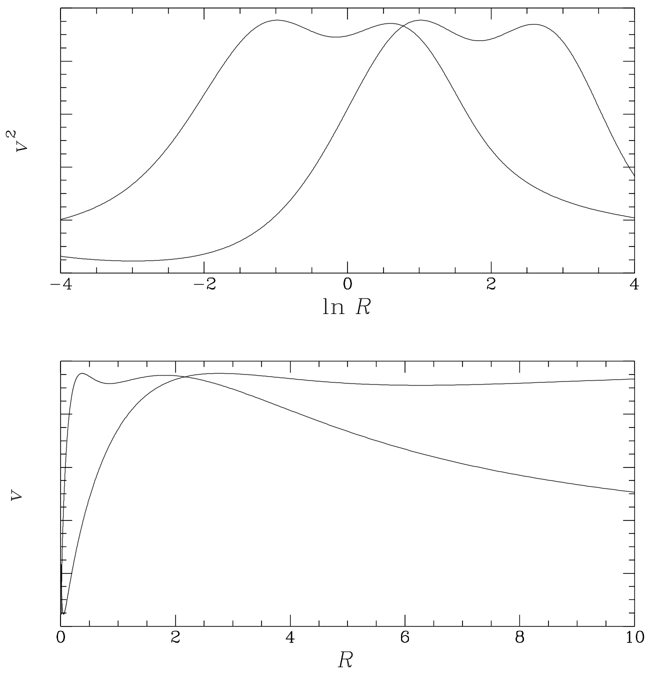

Figure 4.3: Plots of v^2 against $\ln R$ (upper panel) or v against R (lower panel) For one curve in each panel, a second exponential disc with mass and scale radius both scaled down by $e^2 \simeq 7.39$ has been added (to mimic a bulge); for the other curve a second exponential disc with mass and scale radius both scaled up by $e^2 \simeq 7.39$ has been added (to mimic a dark halo).

PROBLEM 4.1: Express the integral equation (4.3) relating $\Phi(R)$ and $\Sigma(R)$ as a convolution in $\ln R$. [10]

The convolution kernel differs from $K(R/R')$ of course, and in a particularly interesting way in the $R/R' \ll 1$ limit. Can you explain this difference using a physical argument? [10]

5. Gravitational Lensing

Gravitational lensing is about how the appearance of distant bright objects is altered by the gravity of foreground mass. Being a purely gravitational effect makes lensing astrophysically important as a probe of dark matter.

This part is more detailed than it needs to be. Only the section on microlensing in the Milky Way is really syllabus material. The rest you should consider as relevant background material plus general interest.

Photons are affected by a gravitational field, but not in the same way as massive particles are. For the details we need general relativity, but fortunately, for astrophysical applications we only need to take over a few simple results. The most important is that if a light ray passes by a mass M with impact parameter R ($\gg GM/c^2$ and $\gg$ the size of the mass), it gets deflected by an angular amount

$$\alpha = \frac{4GM}{c^2 R}. \tag{5.1}$$

In contrast, a massive body at high speed v gets deflected by $\alpha = 2GM/(v^2 R)$.

The lensing equation

To make (5.1) useful we need two approximations, both very good in almost all astrophysical situations:

(i) The deflector is much smaller than the distances to the observer and the object being viewed (the 'source');

(ii) The deflections are always very small, so we can freely use $\sin\alpha = \alpha$, and also we can get the total deflection from a mass distribution by integrating (5.1).

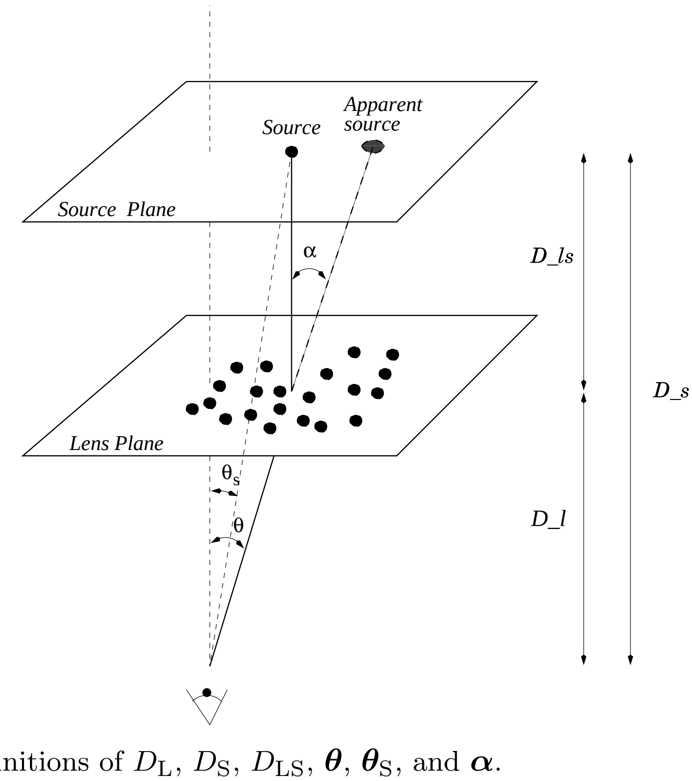

Figure 5.1: Definitions of $D_{\rm L}$, $D_{\rm S}$, $D_{\rm LS}$, $\boldsymbol{\theta}$, $\boldsymbol{\theta}_{\rm S}$, and $\boldsymbol{\alpha}$.

Accordingly, let us consider a situation as in Figure 5.1: observer is viewing a source at distance D_S, with a lens (a mass screen) intervening at distance D_L; D_{LS} is the distance from lens to the source.[1] We'll use angular coordinates for the transverse position.[2] Thus, $\boldsymbol{\theta}_S$ is the position of the source, $\boldsymbol{\theta}$ is its observed position after being deflected—note that these are two-dimensional angles. Let $\Sigma(\boldsymbol{\theta})$ be the lens's density surface mass density (as in solar masses per steradian). Let $\boldsymbol{\alpha}(\boldsymbol{\theta})$ be the deflection angle. Then, comparing vectors in the source plane, we get

$$D_S\boldsymbol{\theta} = D_S\boldsymbol{\theta}_S + D_{LS}\boldsymbol{\alpha}. \tag{5.2}$$

(By convention,[3] $\boldsymbol{\alpha}$ is directed outwards from the deflecting mass rather than towards it.) Using (5.1) to get $\boldsymbol{\alpha}$ in terms of Σ, we get

$$\boldsymbol{\theta} = \boldsymbol{\theta}_S + \frac{D_{LS}}{D_S}\boldsymbol{\alpha}(\boldsymbol{\theta}), \qquad \boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{4G}{c^2 D_L}\int \frac{\Sigma(\boldsymbol{\theta}')(\boldsymbol{\theta}-\boldsymbol{\theta}')\,d^2\boldsymbol{\theta}'}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}. \tag{5.3}$$

This is known as the lens equation. It gives $\boldsymbol{\theta}_S$ as an explicit function of $\boldsymbol{\theta}$, but $\boldsymbol{\theta}$ as an implicit function of $\boldsymbol{\theta}_S$. Moreover, $\boldsymbol{\theta}(\boldsymbol{\theta}_S)$ need not be single-valued, so sources can be multiply imaged.

The arrival time surface

It's possible to work entirely with the form (5.3), but there's a much more intuitive reformulation, which we'll now derive.

We start by noting that the lens equation (5.3) amounts to equating a gradient to zero:

$$\begin{gathered} \nabla T = 0, \quad T = \tfrac{1}{2}T_0(\boldsymbol{\theta}-\boldsymbol{\theta}_S)^2 - \Psi(\boldsymbol{\theta}), \\ \Psi(\boldsymbol{\theta}) = \frac{4G}{c^3}\int \Sigma(\boldsymbol{\theta}')\ln|\boldsymbol{\theta}-\boldsymbol{\theta}'|\,d^2\boldsymbol{\theta}', \quad T_0 = \frac{D_L D_S}{cD_{LS}}. \end{gathered} \tag{5.4}$$

The two terms in T express the change in light travel time for an arbitrary deflection:[4] the first term is what we would get from geometrical considerations alone; the second term is an extra time delay caused by the gravitational field.[5] The requirement that T be stationary is just Fermat's principle.

Next we consider a point mass M, which happens to be precisely between us and a point source. In other words $\boldsymbol{\theta}_S = 0$ and $\Sigma(\boldsymbol{\theta}) = M\delta(\boldsymbol{\theta})$. Then the lens equation is solved by $\theta = \theta_E$, with

$$\theta_E^2 = \frac{4GM}{c^2}\frac{D_{LS}}{D_L D_S}, \qquad R_E^2 = \frac{4GM}{c^2}\frac{D_L D_{LS}}{D_S}. \tag{5.5}$$

Here R_E is just the non-angular form of θ_E—it is called the Einstein radius. The image will consist of a ring of angular radius θ_E, called the Einstein ring.

---

[1] On galactic scales D_L, D_S, D_{LS} are ordinary distances, but on cosmological scales they must be understood as angular diameter distances, and $D_S \neq D_L + D_{LS}$. The reason for this complication is that the universe will have expanded substantially over the light travel time.

[2] Later on, we'll use $\theta_r, \theta_x, \theta_y$ as coordinates rather than r, x, y, to remind us that these are angles on the sky, not distances.

[3] The astrophysical convention being that you first think how a rational person would do it, and then you *change the sign.*

[4] In cosmology both terms need to be multiplied by $(1 + z_L)$.

[5] The gravitational time delay can be derived directly from general relativity, independently of (5.1), and is known as the Shapiro time delay. Radio astronomers can measure it directly.

PROBLEM 5.1: For very distant sources (i.e., $D_\mathrm{S} \gg D_\mathrm{L}$) we can write

$$\theta_\mathrm{E} = (\ldots) \times \sqrt{M/D_\mathrm{L}}.$$

Find $(\ldots)$ in arcsec, if M is measured in solar masses and D_L in parsecs. [4]

By a Gauss's-law type argument, for any circular mass distribution $\Sigma(\theta_r)$, $\Psi(\theta_r)$ and $\boldsymbol{\alpha}(\boldsymbol{\theta})$ will be influenced only by interior mass. So we'll get the same images for any circular distribution of the mass M, *provided it fits within an Einstein radius.* Bodies that fit within their own Einstein radius are said to be 'compact'. But the Einstein radius depends on where the source and observer are:

$$R_\mathrm{E} \sim (\text{Schwarzschild radius} \times D_\mathrm{L})^{\frac{1}{2}}.$$

This sort of means that the further away you look, the easier it gets to see examples of gravitational lensing. It's a surprising fact at first, but it's really just the gravitational analogue of a familiar fact about glass lenses—to get the maximum effect from a lens you have to be near the focal plane, if you're too near the lens doesn't have much effect.

For given $D_\mathrm{L}, D_\mathrm{S}$, to get a compact object you have to pack a mass (in projection) into a circle of radius θ_E; but the area of the circle is proportional to the mass. So clearly there has to be a critical density, say Σ_crit, such that if $\Sigma \geq \Sigma_\mathrm{crit}$ somewhere then there is a compact (sub)-object. Working out the algebra we easily get

$$\Sigma_\mathrm{crit} = \frac{D_\mathrm{L} D_\mathrm{S}}{D_\mathrm{LS}} \frac{c^2}{4\pi G}. \tag{5.6}$$

Using this we can write (5.4) more concisely as

$$\begin{aligned} &\nabla T = 0, \quad T = T_0 \left[\tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_\mathrm{S})^2 - \psi(\boldsymbol{\theta})\right] \\ &\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int \kappa(\theta') \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| \, d^2\boldsymbol{\theta}', \end{aligned} \tag{5.7}$$

where κ is the projected mass density in units of the critical density. From the second line of 5.7 it should be evident that ψ satisfies a two-dimensional Poisson equation

$$\nabla^2 \psi = 2\kappa. \tag{5.8}$$

The fact there is a critical density, and that it depends on distances, has important astrophysical consequences. For example, a galaxy as a whole (a smooth distribution of $\sim 10^{12} M_\odot$ on a scale of $\sim 10^5\,\mathrm{pc}$) is not compact to lensing for $D_\mathrm{L} \lesssim 10^9\,\mathrm{pc}$—cosmological distances. But clumps within the galaxy may be compact at much smaller distances. In particular, a star is compact to lensing at distances of even $\lesssim 1\,\mathrm{pc}$.

The surface $T(\boldsymbol{\theta})$ is known as the time delay surface or the arrival time surface. Wherever the arrival time is stationary (i.e., the surface as a maximum, minimum, or saddle point) there'll be constructive interference, and an image. This is Fermat's principle. Furthermore, the less the curvature of the surface at the images, the more magnified the image will be. We'll formalize this in the next section.

Try to visualize the arrival time surface. The geometrical part is a parabola with a minimum at $\boldsymbol{\theta}_\mathrm{S}$. Having mass in the lens pushes up the surface variously. If $\kappa(\boldsymbol{\theta}) > 1$

anywhere, there will be a maximum somewhere near there, hence another image. There must be a third image too, because to have a minimum and a maximum in a surface you must have a saddle point somewhere. In fact

$$\text{maxima} + \text{minima} = \text{saddle points} + 1. \tag{5.9}$$

This is a really a statement about geometry that should be intuitively clear, though a formal proof is difficult.

A good way of gaining some intuition about the arrival time surface is to take a transparency with a blank piece of paper behind it and look at the reflections of a light bulb. Notice how images merge and split, and how you get grotesquely stretched images just as they do. Deep images of rich clusters of galaxies show just these effects!

Magnification

By magnification we mean: how much does the image move when we move the source? It should be clear that this magnification can't be a scalar, because an image doesn't in general move in the same direction as the source. In fact the magnification is a tensor. We'll denote it by M (A for 'amplification' is also used). Formalizing our definition, we have

$$M^{-1} = \frac{\partial \boldsymbol{\theta}_{\rm S}}{\partial \boldsymbol{\theta}} = \frac{\partial^2}{\partial \boldsymbol{\theta}^2} T(\boldsymbol{\theta}). \tag{5.10}$$

In cartesian coordinates

$$M^{-1} = \begin{pmatrix} 1 - \dfrac{\partial^2 \psi}{\partial \theta_x^2} & \dfrac{\partial^2 \psi}{\partial \theta_x \theta_y} \\ \dfrac{\partial^2 \psi}{\partial \theta_y \theta_x} & 1 - \dfrac{\partial^2 \psi}{\partial \theta_y^2} \end{pmatrix}. \tag{5.11}$$

Notice that M^{-1} is basically taking the curvature of the arrival time surface.

It is helpful to write M^{-1} in terms of its eigenvalues, and the usual form is like

$$M^{-1} = (1-\kappa) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix}. \tag{5.12}$$

The eigenvalues are of course $1 - \kappa \pm \gamma$. The first term in (5.12) is the trace part—and comparing equations (5.11) and (5.8) shows that it must be κ—while the second term is traceless. The κ thing produces an isotropic expansion or contraction, while the γ thing produces a stretching in the ϕ direction and a shrinking in the perpendicular direction; κ is known as 'convergence' and γ as 'shear'.

The determinant of M can be thought of as a scalar magnification.

$$|M| = [(1-\kappa)^2 + \gamma^2]^{-1}. \tag{5.13}$$

The places where one of the eigenvalues of M^{-1} becomes zero (and in consequence $|M|$ is infinite) are in general curves and are known as critical curves. When critical curves are mapped onto the source plane through the lens equation, they give caustics; a source lying on a caustic gets infinitely magnified.

Example [Point mass and isothermal lenses] For a point mass, the lens equation is

$$\theta_{Sx} = \theta_x - \frac{\theta_x}{\theta_r^2}\theta_{\rm E}^2, \qquad \theta_{Sy} = \theta_y - \frac{\theta_y}{\theta_r^2}\theta_{\rm E}^2,$$

and this gives

$$M^{-1} = \begin{pmatrix} 1 - \left(\dfrac{1}{\theta_r^2} + 2\dfrac{\theta_x^2}{\theta_r^4}\right)\theta_{\rm E}^2 & 2\dfrac{\theta_x\theta_y}{\theta_r^4}\theta_{\rm E}^2 \\ 2\dfrac{\theta_x\theta_y}{\theta_r^4}\theta_{\rm E}^2 & 1 - \left(\dfrac{1}{\theta_r^2} + 2\dfrac{\theta_y^2}{\theta_r^4}\right)\theta_{\rm E}^2 \end{pmatrix}.$$

Taking the determinant and simplifying, we get

$$|M|^{-1} = 1 - \frac{\theta_{\rm E}^4}{\theta_r^4}. \tag{5.14}$$

For a circular mass distribution $\Sigma \propto \theta_r^{-1}$ (known as the 'isothermal lens', because it is just the $\rho \propto 1/r^2$ isothermal sphere in projection) the lens equation is

$$\theta_{Sx} = \theta_x - \frac{\theta_x}{\theta_r}\theta_{\rm E}^2, \qquad \theta_{Sy} = \theta_y - \frac{\theta_y}{\theta_r}\theta_{\rm E}^2,$$

and gives

$$M^{-1} = \begin{pmatrix} 1 - \left(\dfrac{1}{\theta_r} + \dfrac{\theta_x^2}{\theta_r^3}\right)\theta_{\rm E}^2 & \dfrac{\theta_x\theta_y}{\theta_r^3}\theta_{\rm E}^2 \\ \dfrac{\theta_x\theta_y}{\theta_r^3}\theta_{\rm E}^2 & 1 - \left(\dfrac{1}{\theta_r} + \dfrac{\theta_y^2}{\theta_r^3}\right)\theta_{\rm E}^2 \end{pmatrix}.$$

And from this we get

$$|M|^{-1} = 1 - \frac{\theta_{\rm E}}{\theta_r}. \tag{5.15}$$

It's shorter in polar coordinates, but tensor components in polar coordinates can get confusing. □

Magnification in lensing conserves surface brightness. We can prove this in a rather fun way. Let us consider the axial direction as a formal time variable t; then light rays can be thought of as trajectories. Now allow observers to be at arbitrary transverse position (say $\mathbf{w}$—two dimensional) and arbitrary t. Then $\boldsymbol{\theta}$ as observed at $(\mathbf{w}, t)$ is just the local $d\mathbf{w}/dt$ for the corresponding light ray, up to a constant factor. This means we can make a formal analogy with Hamiltonian formulation of stellar dynamics, with $\boldsymbol{\theta}$ (up to a constant) playing the role of the momentum, $\mathbf{w}$ playing the role of the coordinates, and $\psi(\mathbf{w}, t)$ replacing the Newtonian potential. The phase space density f is the density of photons in $(\mathbf{w}, \boldsymbol{\theta})$ space, or the number of photons per unit solid angle on the sky per unit telescope area, i.e., the surface brightness. The collisionless Boltzmann equation applies (as it does for any Hamiltonian system) and it tells us that surface brightness is conserved along trajectories! Surface brightness must be conserved by the act of placing the lens there too—think of surface brightness before and after going through the lens. QED. We must be careful, though, to understand 'along the trajectories' correctly. It means we must always be looking at photons from the same source, so if the image is moved in the sky by lensing we must follow it when we measure surface brightness.

In other words, lensing changes the apparent sizes (and shapes) of objects, but without altering their surface brightness.

PROBLEM 5.2: For unresolved sources, we don't observe the surface brightness but only the luminosity, say L. In a survey of objects with luminosity function $f(L)$ to a limit of $L_{\rm min}$, the number of objects detected will be

$$\int_{L_{\rm min}}^{\infty} f(L)\, dL \times \langle \text{area of survey} \rangle\,.$$

Now suppose there is a foreground lens in the survey area with uniform scalar magnification $|M|$. This will increase the effective luminosity limit of the survey to $L_{\rm min}/|M|$, and hence change the number of objects detected. The changed number of objects is not, however,

$$\int_{L_{\rm min}/|M|}^{\infty} f(L)\, dL \times \langle \text{area of survey} \rangle\,.$$

Correct this formula. [10]

This effect is known as 'amplification bias'.

That's more than enough theory, let's discuss real systems a little.

MULTIPLE-IMAGE QSOs

These happen when a foreground galaxy is within $\lesssim \theta_{\rm E}$ (in projection) of a QSO, and produces two or four images with arcsecond order separations. Two-image systems have a minimum and a saddle point, while four-image systems have two minima and two saddle points. In both cases there's a maximum too, at the bottom of the galaxy's potential well; but since that is also generally the densest part of the galaxy, κ is very high and $|M|$ nearly vanishes, so these central images are too faint to detect.

Multiple-image QSOs are of great astrophysical interest, and two things make them so.

The first is that since QSOs are often very time-variable and the different images have different arrival times, the images will show the same time-variability, but with offsets. These offsets are simply the differences in $T(\boldsymbol{\theta})$ between different images. (So far they have been explicitly measured for two lenses.) Provided we know (or can model) $\kappa(\boldsymbol{\theta})$, the measured time offsets tell us T_0, and hence H_0. Basically it's this: normally we can only measure dimensionless things (image separations, relative magnifications) in lenses systems; but if we succeed in measuring a quantity that has a scale (the time delays) that tells us the scale of the universe (H_0). In practice, there is considerable uncertainty about the distribution of mass in the lensing galaxies, and this translates into an uncertainty in the inferred H_0 that is much larger than errors in the time delays. Maybe this problem can be overcome, maybe not...

The second thing has to do with the extremely small size of QSOs in optical continuum. Now the $\kappa(\boldsymbol{\theta})$ of a galaxy isn't perfectly smooth, it becomes granular on the scale of individual stars. This produces a very complicated network of critical lines (in the lens plane), and a corresponding complicated network of caustics in the source plane (like the pattern at the bottom of a swimming pool). The optical continuum emitting regions of QSOs are small enough to fit between the caustics, but the line emitting regions straddle several caustics. As proper motions move the caustic network, the continuum region will sometimes cross a caustic, and show a sudden change in brightness; the time taken for the brightness to change is the time it take to cross the caustic. This is the phenomenon of QSO microlensing: continuum shows it but lines don't. (It's just the gravitational version of stars twinkling and planets not twinkling.) This has been observed, and modelling the caustic network and putting in plausible values for the proper motion leads to an estimate of the intrinsic size of the continuum regions of QSOs. It's very small $\sim 100\,\mathrm{AU}$.

Galaxy clusters

Galaxy clusters are generally not in dynamical equilibrium (there haven't been enough crossing times since they formed). Their mass distributions and ψ potentials are thus warped in more complicated ways than for single galaxies. They are also much bigger on the sky and thus have many more background objects (faint blue galaxies) to lens.

The transparency with a paper behind it and several lightbulbs overhead is a good simulacrum of lensing by a cluster. Rich clusters show many highly stretched images of background galaxies, and these are known as arcs. A deep HST image of Abell 2218 shows over a hundred arcs, including seven multiple image systems.

An arc is close to a zero eigenvalue of M^{-1}, and is stretched along the corresponding eigenvalue. Thus each arc provides some sort of constraint on the ψ of the cluster.

Clusters also show weak lensing. That's when the eigenvalues $1 - \kappa \pm \gamma$ are too close to unity to show up as arcs, but if many galaxies in the same region are examined then statistically a stretching is measurable. The statistical stretching measures the ratio of the two eigenvalues, and thus $\gamma/(1-\kappa)$.

Several groups have been reconstructing cluster mass profiles from information provided by multiple-images, arcs, and weak lensing.

Microlensing in the Milky Way

One possibility for the dark matter in the Milky Way halo is that it consists of brown dwarfs, compact objects below the hydrogen burning threshold of $0.08M_\odot$. Such objects would act as point lenses. A point lens has two images, at

$$\theta = \tfrac{1}{2}\left(\boldsymbol{\theta}_\mathrm{S} \pm \sqrt{\boldsymbol{\theta}_\mathrm{S}^2 + 4\theta_\mathrm{E}^2}\right). \tag{5.16}$$

(There is formally a third image at $\theta = 0$, i.e., at the lens itself, but for a point mass this image has zero magnification.) The image separation for a $\sim M_0$ lens at distances of $\sim 10,\mathrm{kpc}$ is $< 1\,\mathrm{mas}$, far too small to resolve. What will be observed is a brightening equal to the combined magnification of both images. Using the result 5.14 for $|M|$ for a point lens, and adding the absolute values of $|M|$ at the two image positions, we get

$$M_\mathrm{tot} = \frac{u^2+2}{u(u^2+4)^{\frac{1}{2}}}, \qquad u = \frac{\theta_\mathrm{S}}{\theta_\mathrm{E}}. \tag{5.17}$$

Now because of stellar motions, θ_S will change by an amount θ_E over times of order a month, so microlensing in the Milky Way can be observed by monitoring light curves. If the background source star has impact parameter b and velocity v (projected onto the lens place) with respect to the lens, then

$$u = \frac{(b^2 + v^2t^2)^{\frac{1}{2}}}{D_\mathrm{L}\theta_\mathrm{E}}. \tag{5.18}$$

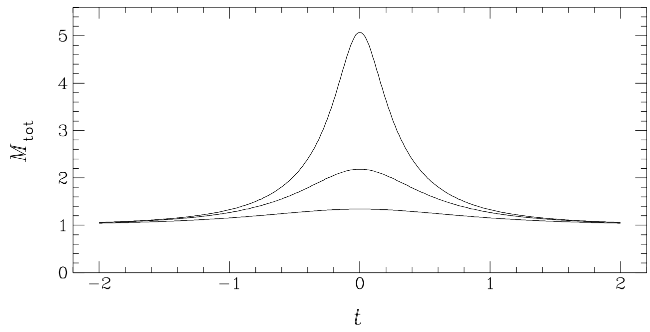

Figure 5.2: Light curves for impact parameters of R_E (lowest), $0.5R_E$ and $0.2R_E$. The unit of time is how long it takes the source to move a distance R_E.

Inserting (5.18) into (5.17) gives us $M_{tot}(0)$, i.e., the light curve, plotted for three different b in Figure 5.2. The height of a measured light curve immediately gives R_E/b, and the width gives R_E/v.

Though trying to resolve the images images in microlensing seems hopeless with foreseeable technology, there are some prospects for tracking the moving double image indirectly. By combining the positions and magnifications of the two images, we have for the centroid

$$\theta_{cen} = \frac{u(3+u^2)}{2+u^2}\theta_E. \tag{5.19}$$

Such microlensing events are rare, because θ_S has to be $\lesssim \theta_E$ for significant magnification. People speak of an optical depth τ to microlensing in a field. This is the probability of a star being (in projection) within θ_E of a foreground lens, at any given time. From equation (5.18) it amounts to the probability of $M_{tot} \geq 2/\sqrt{5} = 1.34$. It's just the covering factor of discs of radius θ_E (Einstein rings) from all lenses between us and the stars in the field.[6] The source stars might be bright stars in the Large Magellanic Cloud (LMC) and the lenses very faint stars or brown dwarfs in the Milky Way halo.

Using equation (5.5) for R_E and considering the total area covered by the Einstein rings of lenses at distances between D_L and $D_L + dD_L$ in a patch of sky, and then integrating over D_L, we have

$$\tau = \frac{4\pi G}{c^2 D_S}\int_0^{D_S} D_L D_{LS}\,\rho(D_L)\,dD_L. \tag{5.20}$$

Problem 5.3: Derive the formula (5.20) for the microlensing optical depth. [10]

Imagine an observer at radius $r = 1$ in an isothermal sphere made of machos, looking outwards (i.e., towards the anti-centre) at sources at radius $r = a$, and monitoring for microlensing. Show that τ for this observer will be

$$\tau = 2\frac{\sigma^2}{c^2}\left[\frac{a+1}{a-1}\ln a - 2\right]. \qquad [6]$$

[$\int_1^a x^{-2}(x-1)(a-x)\,dx = (1+a)\ln a - 2(a-1)$.]

[6] So optical depth is a bit of a misnomer.

The really nice thing about the formula (5.20) is that it doesn't depend on the mass distribution of the lenses, as long as each mass fits within its own Einstein radius (diffuse gas clouds don't count, nor does any kind of diffuse dark matter). So τ estimated from light curve monitoring could be used to make inferences about ρ.

How large is τ through the Galactic halo? To estimate that, we need an estimate for ρ. Now the Milky Way rotation curve suggests an isothermal halo, $\rho = \sigma^2/(2\pi G r^2)$, with $\sigma \sim 200\,\mathrm{km/sec}$. If we then say that r will be of order the D factors in (5.20), we get

$$\tau \sim \frac{\sigma^2}{c^2}, \quad \text{or} \quad \tau \sim 10^{-7} \text{ to } 10^{-6}. \tag{5.21}$$

So to have any hope of detecting such microlensing events, it is necessary to monitor the light curves of millions of stars. Four such surveys have been started up in the last two years, observing fields in the LMC and/or the Milky Way bulge. (The bulge surveys don't go through the halo of course, but through part of the Milky Way disc.)[7] The current estimates for τ are $\sim 10^{-7}$ towards the LMC and $\simeq 3 \times 10^{-6}$ towards the bulge. How much of the lensing mass is in brown dwarfs as distinct from faint stars, and whether the lensing mass alone can account for rotation curve data are not yet clear. Meanwhile, the huge number of variable stars discovered by these surveys are revolutionizing that field of study.

[7] An estimate of τ from a survey will include a correction for the detection efficiency. Surveys have to be very wary of spurious detections; hence any light curve possibly contaminated by stellar variability has to be discarded for microlensing purposes. Detection efficiencies are of order 30%.