Introduction to Gravitational Lensing and Cosmology

Peter Schneider
Institut f. Astrophysik, Universität Bonn, D-53121 Bonn, Germany

This is an introductory chapter to the Proceedings of the 33rd Saas-Fee Advanced Course:

Kochanek, C.S., Schneider, P., Wambsganss, J.:
Gravitational Lensing: Strong, Weak & Micro,

G. Meylan, P. Jetzer & P. North, eds. (Springer-Verlag: Berlin), in preparation.

1 Introduction

Light rays are deflected when they propagate through a gravitational field. Long suspected before General Relativity – the theory which we believe provides the correct description of gravity – it was only after Einstein's final formulation of this theory that the effect was described quantitatively. The rich phenomena which are caused by this gravitational light deflection has led to the development of the rather recent active research field of gravitational lensing, and the fact that the 2003 Saas-Fee course is entirely devoted to this subject is just but one of the indications of the prominence this topic has achieved. In fact, the activities in this area have become quite diverse and are reflected by the three main lectures of this course. The phenomena of light propagation in strong gravitational fields, as it occurs near the surface of neutron stars or black holes, are usually not incorporated into gravitational lensing – although the physics is the same, these strong-field effects require a rather different mathematical description than the weak deflection phenomena.

In this introductory chapter we shall provide an outline of the basics of gravitational lensing, covering aspects that are at the base of it and which will be used extensively in the three main lectures. We start in Sect. 1.1 with a brief historical account; the study of the influence of a gravitational field on the propagation of light started long before the proper theory of gravity – Einstein's General Relativity – was formulated. Illustrations of the most common phenomena of gravitational lensing will be given next, before we will introduce in Sect. 2 the basic equations of gravitational lensing theory. A few simple lens models will be considered in Sect. 3, in particular the point-mass lens and the singular isothermal sphere model. Since the sources and

deflectors in gravitational lensing are often located at distances comparable to the radius of the observable Universe, the large-scale geometry of spacetime needs to be accounted for. Thus, in Sect. 4 we give a brief introduction to the standard model of cosmology. We then proceed in Sect. ?? with some basic considerations about lensing statistics, i.e. the question of how probable it is that observations of a source at large distance are significantly affected by a lensing effect, and conclude with a description of the large-scale matter distribution in the Universe. The material covered in this introductory chapter will be used extensively in the later chapters of this book; those will be abbreviated as SL (Strong Lensing, Kochanek 2004), ML (MicroLensing, Wambsganss 2004) and WL (Weak Lensing, Schneider 2004).

Gravitational lensing as a whole, and several particular aspects of it, has been reviewed previously. Two extensive monographs (Schneider, Ehlers & Falco 1992, hereafter SEF; Petters, Levine & Wambsganss 2001, hereafter PLW) describe lensing in great detail, in particular providing a derivation of the gravitational lensing equations from General Relativity (see also Seitz, Schneider & Ehlers 1994). Blandford and Narayan (1992) review the cosmological applications of gravitational lensing, Refsdal & Surdej (1994) and Courbin et al. (2002) discuss quasar lensing by galaxies and provide an intuitive geometrical optics approach to lensing, Fort & Mellier (1994) describe the giant luminous arcs and arclets in clusters of galaxies, Paczyński (1996) reviews the effects of gravitational microlensing in the local group, the review by Narayan & Bartelmann (1999) provides a concise account of gravitational lensing theory and observations, and Mellier (1999), Bartelmann & Schneider (2001), Wittman (2002) and van Waerbeke & Mellier (2003) review the relatively young field of weak gravitational lensing.

1.1 History of gravitational light deflection

We start with a (very) brief account on the history of gravitational lensing; the reader is referred to SEF and PLW for a more detailed presentation.

The early years, before General Relativity. The Newtonian theory of gravitation predicts that the gravitational force $\mathbf{F}$ on a particle of mass m is proportional to m, so that the gravitational acceleration $\mathbf{a} = \mathbf{F}/m$ is independent of m. Therefore, the trajectory of a test particle in a gravitational field is independent of its mass but depends, for a given initial position and direction, only on the velocity of the test particle. About 200 years ago, several physicists and astronomers speculated that, if light could be treated like a particle, light rays may be influenced in a gravitational field as well. John Mitchell in 1784, in a letter to Henry Cavendish, and later Johann von Soldner in 1804, mentioned the possibility that light propagating in the field of a spherical mass M (like a star) would be deflected by an angle $\hat{\alpha}_{\rm N} = 2GM/(c^2\xi)$, where G and c are Newton's constant of gravity and the

velocity of light, respectively, and ξ is the impact parameter of the incoming light ray. At roughly the same time, Pierre-Simon Laplace in 1795 noted "that the gravitational force of a heavenly body could be so large, that light could not flow out of it" (Laplace 1795), i.e. that the escape velocity $v_{\rm e} = \sqrt{2GM/R}$ from the surface of a spherical mass M of radius R becomes the velocity of light, which happens if $R = R_{\rm s} \equiv 2GM/c^2$, nowadays called the *Schwarzschild radius* of a mass M.

Gravitational light deflection in GR. All these results were derived under the assumption that light somehow can be considered like a massive test particle; this was of course well before the concept of photons was introduced. Only after the formulation of General Relativity by Einstein in 1915 could the behavior of light in a gravitational field be studied on a firm physical ground. Before the final formulation of GR, Einstein published a paper in 1911 where he recalculated the results of Mitchell and Soldner (of whose work he was unaware) for the deflection angle. Only after the completion of GR did it become clear that the 'Newtonian' value of the deflection angle was too small by a factor of 2. In the general theory of relativity, the deflection is

$$\hat{\alpha} = \frac{4GM}{c^2\xi} = 1\farcs75 \left(\frac{M}{M_\odot}\right) \left(\frac{\xi}{R_\odot}\right)^{-1} . \qquad (1)$$

The deflection of light by the Sun can be measured during a Solar Eclipse when it is possible to observe stars projected near the Solar surface; light deflection then slightly changes their positions. A measurement of the deflection in 1919, with a sufficient accuracy to distinguish between the 'Newtonian' and the GR value, provided a tremendous success for Einstein's new theory of gravity.

Soon thereafter, Lodge (1919) used the term 'lens' in the context of gravitational light deflection, but noted that 'it has no focal length'. Chwolson (1924) considered a source perfectly coaligned with a foreground mass, concluding that the source should be imaged as a ring around the lens – in fact, only fairly recently did it become known that Einstein made some unpublished notes on this effect in 1912 (Renn et al. 1997) – hence, calling them 'Einstein rings' is indeed appropriate. If the alignment is not perfect, two images of the background source would be visible, one on either side of the foreground star. Einstein, in 1936, after being approached by the Czech engineer Rudi Mandl, wrote a paper where he considered this lensing effect by a star, including both the image positions, their separation, and their magnifications. He concluded that the angular separation between the two images would be far too small (of order milliarcseconds) to be resolvable, so that "there is no great chance of observing this phenomenon" (Einstein 1936).

Zwicky's visions. This pessimistic view was not shared by Fritz Zwicky, who in 1937 published two truly visionary papers. Instead of looking at lens-

ing by stars in our Galaxy, he considered "extragalactic nebulae" (nowadays called galaxies) as lenses – with his mass estimates of these nebulae, he estimated typical image separation of a background source to be of order $10''$ – about one order of magnitude too high – and such pairs of images *can* be separated with telescopes. Observing such an effect, he noted, would furnish an additional test of GR, allow one to see galaxies at larger distances (due to the magnification effect), and would determine the masses of these nebulae acting as lenses (Zwicky 1937a). He then went on to estimate the probability that a distant source would be lensed to produce multiple images, concluded that about 1 out of 400 distant sources should be affected by lensing (this is about the fractional area covered by the bright parts of nebulae on photographic plates), and hence predicted that "the probability that nebulae which act as gravitational lenses will be found becomes practically a certainty" (Zwicky 1937b). As we shall see in due course, basically all of Zwicky's predictions became true.[1]

The revival of lensing. Until the beginning of the 1960's the subject rested, but in 1963/4, three authors independently reopened the field: Yu.G. Klimov (1963), S. Liebes (1964) and S. Refsdal (1964a,b). Klimov considered lensing of galaxies by galaxies, whereas Liebes and Refsdal mainly studied lensing by point-mass lenses. Their papers have been milestones in lensing research; for example, Liebes considered the possibility that stars in the Milky Way can act as lenses for stars in M31 – we shall see in ML, this is a truly modern idea. Refsdal calculated the difference of the light travel times between the two images of a source – since light propagates along different paths from the source to the observer, there will in general be a time delay which can be observed provided the source is variable, such like a supernova. Refsdal pointed out that the time delay depends on the mass of the lens and the distances to the lens and the source, and concluded that, if the image separation and the time delay could be measured, the lens mass and the Hubble constant could be determined. We shall see in SL how these predictions have been realized in the meantime.

In 1963, the first quasars were detected: luminous, compact ('quasi-stellar') and very distant sources – hence, a source population had been discovered which lies behind Zwicky's nebulae, and finding lens systems amongst them should be a certainty. Nevertheless, it took another 15 years until the first lens system was observed and identified as such.

[1] Zwicky thought he had found a gravitational lens system and said so at a conference in the 1950s. Munch, one of his Caltech colleagues, said that if it were a lens, he'd "eat his hat". Sargent (from whom this story was communicated) found the photographic plate after Zwicky's death, hoping to improve Munch's diet, but concluded it was a plate defect.

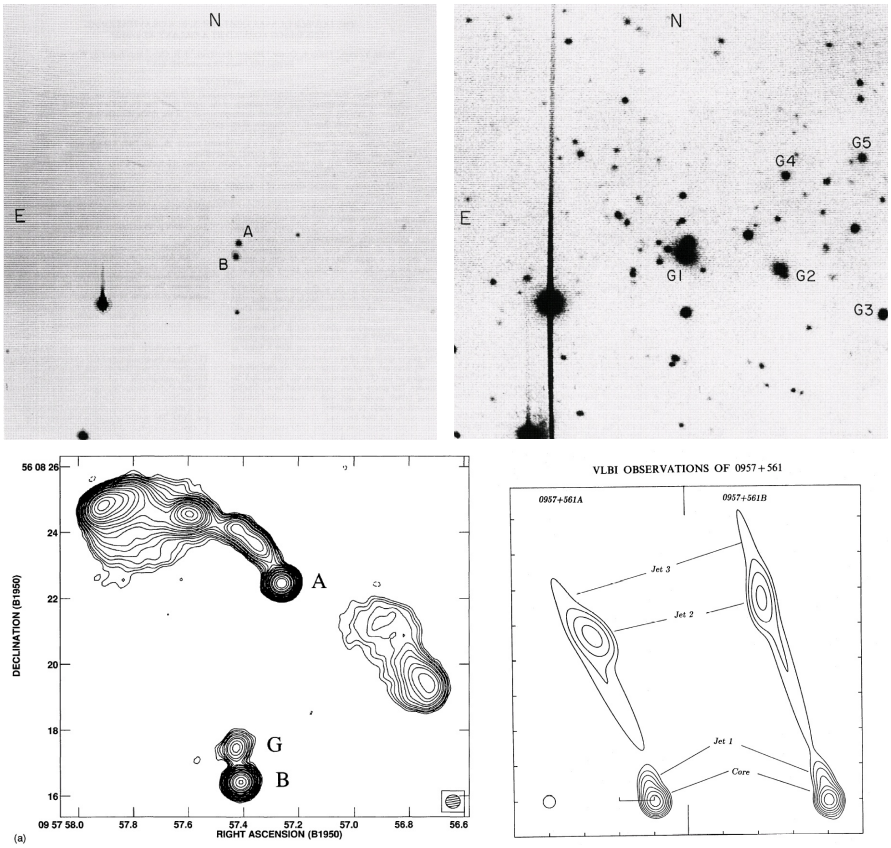


Fig. 1. The two upper panels show a short (left) and longer (right) optical exposure of the field of the double QSO 0957+561 (Young et al. 1981). In the short exposure, the two QSO images are clearly visible as a pair of point sources, separated by $\sim 6''$. The longer exposure reveals the presence of an extended source, the lens galaxy, between the two point sources, as well as a small cluster of galaxies of which the lens galaxy G1 is the brightest member. The lower left panel shows a 6 cm VLA map of the system (Harvanek et al. 1997), where besides the two QSO sources A and B, and the extended radio structure seen for image A, radio emission from the lens galaxy G is also visible. The milli-arcsecond structure of the two compact components A, B is shown in the lower right panel (Gorenstein et al. 1988a), where it is clearly seen that one VLBI jet is a linearly transformed version of the other, and they are mirror symmetric; this is predicted by any generic lens model which assigns opposite parity to the two images

1.2 Discoveries

First detections of multiple imaging (1979). In their program to optically identify radio sources, Walsh, Carswell & Weymann in 1979 discovered a pair of quasars separated by about 6 arcseconds, having identical colors, redshifts ($z_s = 1.41$) and spectra (see Walsh 1989 for the history of this discovery). The year 1979 also marked two important technical developments in astronomy: the first CCD detectors replaced photographic plates, thus providing much higher sensitivity, dynamic range and linearity, and the Very Large Array (VLA), a radio interferometer providing radio images of subarcsecond image quality, went into operation. With the VLA is was soon demonstrated that both quasar images are compact radio sources, with similar radio spectra. Soon thereafter, a galaxy situated between the two quasar images was detected (Stockton 1980; Young et al. 1980). The galaxy has a redshift of $z_d = 0.36$ and it is the brightest galaxy in a small cluster. We now know that the cluster contributes its share to the large image separation in this system. Furthermore, the first Very Long Baseline Interferometry (VLBI) data of this system, known as QSO 0957+561, showed that both components have a core-jet structure with the symmetry expected for lensed images of a common source (see Fig. 1). The great similarities of the two optical spectra (Fig. 2) is another proof of the lensing nature of this system.

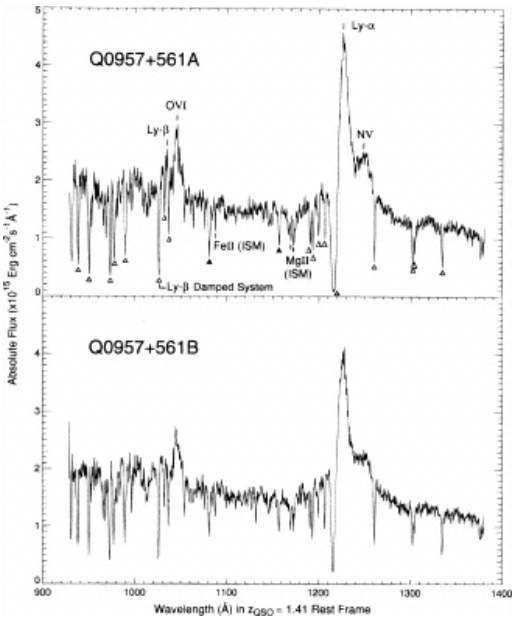


Fig. 2. Spectra of the two images of the lens system QSO 0957+561, obtained with the Faint Object Spectrograph onboard HST (Michalitsianos et al. 1997). The strong similarities of the spectra, in particular the same line ratios and the identical redshift, verifies this system as a definite gravitational lens system

One year later, the so-called triple quasar PG 1115+080 was discovered (Weymann et al. 1980). It apparently consisted of three images, one of which

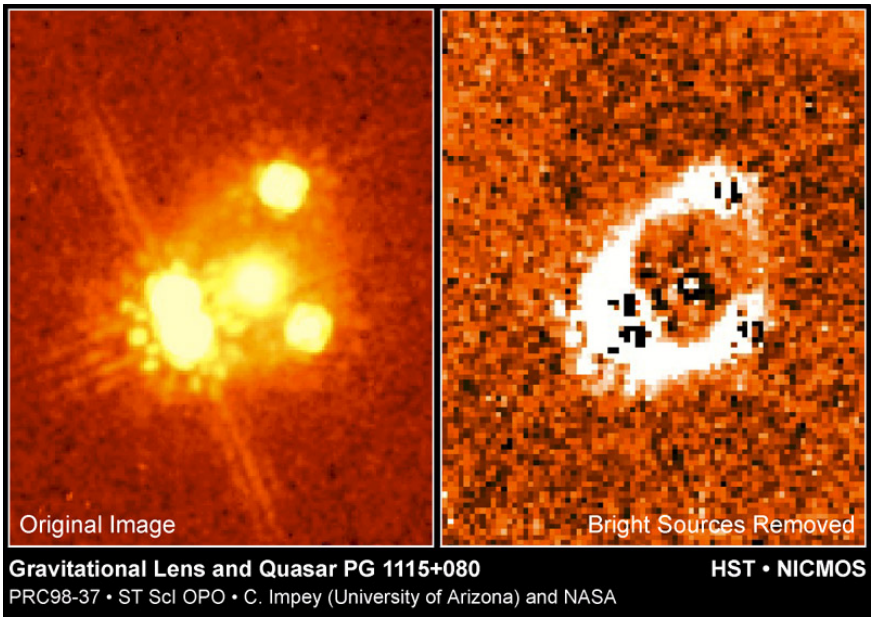


Fig. 3. In the left panel, a NIR image of the gravitational lens system PG 1115+080 is shown, taken with the NICMOS instrument onboard HST. The QSO has a redshift of $z_s = 1.72$. The double nature of the brightest component is clearly recognized, as well as the lens galaxy with redshift $z_d = 0.31$, situated in the 'middle' of the four QSO images. When the QSO images and the lens galaxy are subtracted from the picture, the remaining image of the system (right panel) shows a nearly complete ring, which is the lensed image of the host galaxy of the QSO, mapped onto a nearly complete Einstein ring. In near-IR observations of lens systems, such rings occur frequently (source: C. Impey and NASA, see Impey et al. 1998)

was much brighter than the other two (see Fig. 3). Soon thereafter it was shown that the bright image was in fact a blend of two images separated by $\sim 0\farcs5$, and thus very difficult to resolve with optical telescopes from the ground. The fact that the close pair is much brighter than the other two images is a generic prediction of lens theory, as will be shown below.

Until 1990, a few more lens systems or lens candidate systems have been discovered, some of them from a systematic search for lenses amongst radio sources (e.g., Burke et al. 1992), but most of them serendipitously (such as the one shown in Fig. 4). The 1990's then have witnessed several systematic searches for lens systems, including programs carried out with the Hubble Space Telescope (HST; Maoz et al. 1993), lens searches amongst 15000 radio sources (JVAS and CLASS; see King et al. 1999; Browne et al. 2003), and those amongst very bright high-redshift quasars – these surveys will be

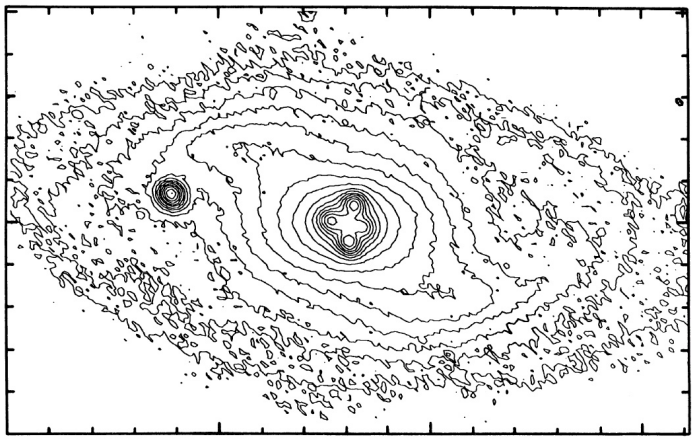


Fig. 4. Around the center of this nearby spiral galaxy ($z_{\rm d} = 0.04$), four point-like sources are seen is a fairly symmetric geometry (Yee 1988). Their spectra identify them as four images of a background QSO with $z_{\rm s} = 1.7$. This system, QSO 2237+0305, is the closest gravitational lens and one of the few systems where the lens is a spiral; it has been found in a spectroscopic redshift survey of nearby galaxies

detailed in SL. By now, more than 80 multiple-image lens systems with a galaxy acting as the (main) lens are known.



Fig. 5. The giant arc in the cluster of galaxies Cl 2244−02, taken with the ISAAC instrument at the VLT (source: ESO Press Photo 46d/98). The arc has a redshift of $z_{\rm s} = 2.24$, and was at the time of discovery the highest redshift normal galaxy. The high magnification caused by the gravitational lens renders this still (one of) the brightest galaxies with $z \geq 2$

Giant luminous arcs (1986). In 1986, two groups (Lynds & Petrosian 1986; Soucail et al. 1987) independently pointed out the existence of strongly elongated, curved features around two clusters of galaxies (see Figs. 5 and



Fig. 6. The cluster A2218 at $z = 0.175$ contains one of the most impressive systems of arcs, as can be seen in the multi-color images taken with the WFPC2 instrument onboard HST (source: NASA/STScI). This cluster contains several multiple image systems of background galaxies which, together with the morphology of arcs, allows the construction of very detailed mass models for this cluster. Also remarkable is the thinness of several of the arcs, so that they are not resolved in width even with the HST; this implies very large length-to-width ratios of these arcs and, correspondingly, very high magnifications

6). Their tangential extent relative to the cluster center was at least ten times their radial extent, although the exact value was difficult to determine as they were not well resolved in width from the ground (HST has shown that this ratio is substantially larger than 10:1 in many cases). These *giant luminous arcs* were seen displaced from the cluster center, and curving around it. Various hypotheses were put forward as to their nature, and all proven wrong, except for one (Paczyński 1987), when the redshift of the giant arc in A370 was measured (Soucail et al. 1988) and shown to be much larger than the redshift of the cluster. The arc was thus proven to be a highly distorted and magnified image of an otherwise normal, higher-redshift galaxy. By now, many clusters with giant arcs are known and have been investigated in detail. As with most optical studies of lenses, the high-resolution of the HST was essential to study the detailed brightness distribution of arcs and to identify multiple images by their morphology and colors. Less distorted images of background galaxies have been named *arclets* (Fort et al. 1988); they can be identified in many clusters, and they are generally stretched tangentially with respect to the cluster center. In addition, clusters can act as strong lenses also to produce multiple images of background galaxies. Some of these aspects will be covered in Sect. 4 of WL.

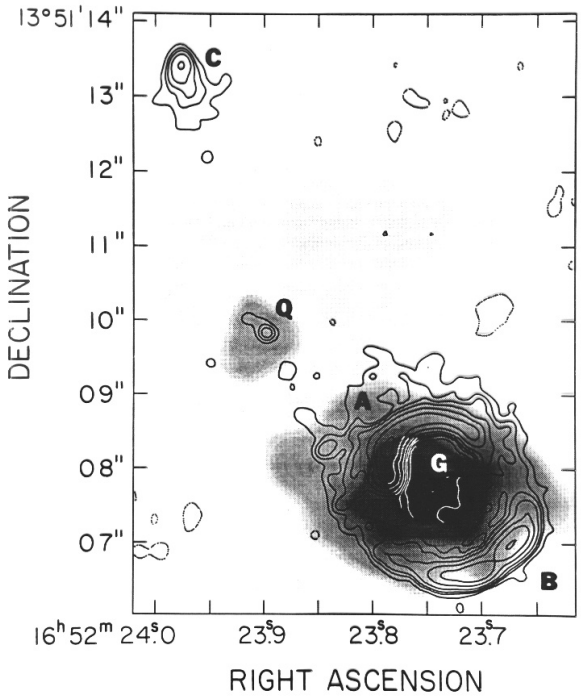


Fig. 7. The quasar MG 1654+13 at redshift $z_s = 1.72$ is shown, both as an optical image (grey scale) and in the radio (contours). The optical QSO is denoted as Q, and is the central component (or core) of a triple radio source. The Northern radio lobe is denoted by C, whereas the Southern radio lobe is mapped onto an Einstein ring. At the center of this ring, one sees a bright galaxy with spectroscopic redshift of $z_d = 0.25$. This galaxy lenses the second radio lobe into a complete Einstein ring. Within this ring, brightness peaks can be identified, and the components denoted A and B are similar to, but not multiple images of, the brightness peak in the Northern lobe C (source: G. Langston)

Rings, after all (1988). Whereas Einstein ring images were predicted in the case of a perfectly coaligned source with a spherically symmetric lens, the first multiple images lens systems have taught us that lenses are far from spherical – thus, the discovery of a radio ring in the source MG 1131+0456 (Hewitt et al. 1988) came as a big surprise. Unfortunately, owing to its faint optical counterpart, the lensing nature of this first system could not be proven easily, but the relative ease by which the radio source morphology, at several frequencies, could be modeled by a simple gravitational lens (Kochanek et al. 1989) made a very strong case for its lensing nature. The second radio ring discovered (Langston et al. 1989) made a much cleaner case: Of the two radio

lobes of a redshift 1.72 quasar, one of them is imaged into a ring (see Fig. 7). At the center of this ring lies a bright, redshift $z_{\rm d} = 0.25$ galaxy, responsible for the light deflection. High-resolution imaging with HST in optical and near-infrared filters revealed the presence of Einstein rings in many multiply imaged quasars (Fig. 8), where the host galaxy of the active nucleus is the corresponding (extended) source. We now know a lens need not be exactly spherical; it is a combination of the asymmetry (ellipticity) of the mass distribution and the source size that determines whether we will see an Einstein ring (see SL, Sect. 10).

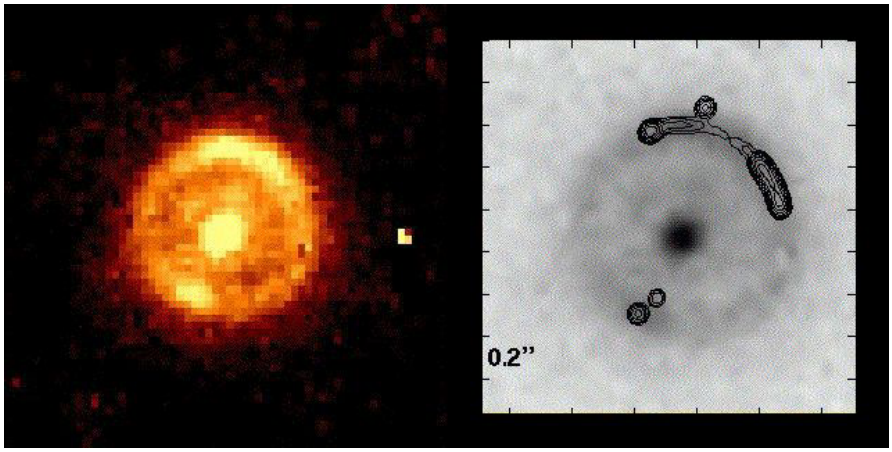


Fig. 8. The gravitational lens system B 1938+666. The left panel shows a NICMOS@HST image of the system, clearly showing a complete Einstein ring into which the Active Galaxy is mapped, together with the lens galaxy situated near the center of the ring. The right panel shows the NICMOS image as grey-scales, with the radio observations superposed as contours. The radio source is indeed a double, with one component being imaged twice (the two images just outside and just inside the Einstein ring), whereas the other source component has four images along the Einstein ring, with two them close together (source: L.J. King, see King et al. 1998)

Quasar microlensing (1989). The mass of galaxies is not distributed smoothly, since at least a fraction of it is in stars. These stars will split the (macro)images of a quasar into many microimages whose typical separations of few microarcseconds are unresolvable. However, these perturbations of the gravitational field change the magnification of the macroimages, provided the source is sufficiently compact. Since the source, the lens and the observer are not stationary, and the stars in the galaxies move, this magnification will also change in time; the characteristic time-scales are of order a decade or less, and in one case (QSO 2237+0305, see Fig. 4) where the lens is

very close to us ($z_{\rm d} = 0.0395$), even smaller. Hence, as predicted by Chang & Refsdal (1979, 1984), Paczyński (1986a) Kayser et al. (1986) and Schneider & Weiss (1987), this *microlensing* effect should yield flux variations of the images which are uncorrelated between the different images – an intrinsic variation of the source would affect the flux of all images in the same way, though with a time delay. In 1989, this microlensing effect was detected in the four image quasar lens QSO 2237+0305 as uncorrelated brightness variations in the four images (Irwin et al. 1989).

Weak lensing (1990). As mentioned before, arclets are images of background galaxies stretched by the lensing effect of a cluster. In order to identify an arclet as such, the image distortion must be significant; otherwise, owing to the intrinsic ellipticity distribution of galaxies, the stretching could not be distinguished from the intrinsic shape. However, if the distortion field varies slowly with position, then galaxy images lying close to each other should be distorted by a similar degree. Since we live in a Universe where the sky is densely covered with faint and small galaxies (e.g., Tyson 1988; Williams et al. 1996), an average over local ensembles of galaxies can be taken; the mean distortion of this ensemble is then a measure for the lens stretching. This *weak gravitational lensing effect* was first detected in two clusters in 1990 (Tyson, Valdes & Wenk 1990). The advances in optical imaging cameras, in particular the availability of large mosaic CCD cameras which enable the mapping of nearly degree-sized fields in a single pointing, and the development of specific image analysis tools, have permitted the detection and quantitative analysis of weak lensing in many clusters. Even weaker lensing effects, those by an ensemble of galaxies and of the large-scale matter distribution in the Universe were discovered in 1996 (Brainerd et al. 1996) and 2000 (Bacon et al. 2000; Kaiser et al. 2000; van Waerbeke et al. 2000; Wittman et al. 2000); we shall report on this in WL.

Time delays ($\geq$ 1992). Following Refsdal's idea to determine the Hubble constant from lensing by combining a good mass model for the lens with the time delay, the light curves of the first double QSO 095+561 were monitored by several groups in the optical and radio waveband (e.g., Vanderriest et al. 1989; Schild 1990; Lehár et al. 1992). From these lightcurves, estimates of the time delay were derived by a number of groups, and significantly different results were obtained. Difficulties include seasonal gaps in the optical light curves and the possibility of uncorrelated variability in the images due to microlensing by the lensing galaxy. To account for these effects, different methods were developed, yielding different results; broadly speaking, either delays of 410 days or 540 days were obtained. The issue was put to rest when a relatively sharp variation of the flux of the leading image was detected in December 1994 (Kundic et al. 1995; Fig. 9). Each of the two estimates for the time delay predicted a different epoch for the occurrence of the corresponding

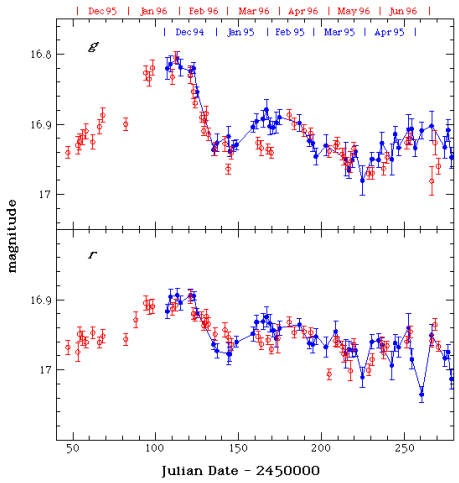


Fig. 9. Light curves of the two images of the QSO 0957+561A,B in two different filters. The two light curves have been shifted in time relative to each other by the measured time delay of 417 days, and in flux according to the flux ratio. The sharp drop measured in image A in Dec. 1994 and subsequently in image B in Feb. 1996 provides an accurate measurement of the time delay (data from Kundić et al. 1997)

feature in the other image. With the observation of the feature in the trailing image in February 1996 (Kundic et al. 1997), the controversy was resolved in favor of the short delay, yielding 417 ± 3 days. Time delays have now been measured in 10 lens systems, although the resulting estimates for the Hubble constant are still problematic – see SL.

Galactic microlensing (1993). Stars in our Galaxy can act as lenses for other stars or extragalactic sources; however, the probability for this to occur is extremely small, as already noted by Liebes (1964). However, if one considers a sufficient number of background sources, even very small probabilities can be beaten. Such a lensing effect would be noted as a magnification of the background star; owing to transverse motion of source, lens and observer, the magnification changes in time and leads to a characteristic light curve. Paczyński (1986b) proposed in 1986 to monitor the brightness of stars in dense stellar fields of the Large Magellanic Cloud to search for such characteristic variability. The main idea behind this suggestion was to test whether the dark matter in the halo of our Galaxy, necessary to explain the flat rotation curve of the Milky Way (and other spiral galaxies) is made up of compact objects – brown dwarfs, neutron stars, 'Jupiters', black holes. The 'only' problem was that about 1 out of 10^7 stars in the LMC is expected to be lensed at any given time – the number of stars needed to be monitored is indeed large. Nevertheless, two groups started this adventure in the early 1990s, and reported in 1993 the first microlensing events towards the LMC (Alcock et al. 1993; Aubourg et al. 1993). Shortly thereafter, a third group announced the discovery of microlensing events towards the Galactic bulge (Udalski et al. 1993). Since then, this fields has flourished, and will be covered in depth in ML. In addition to the discovery of a large number of microlens-

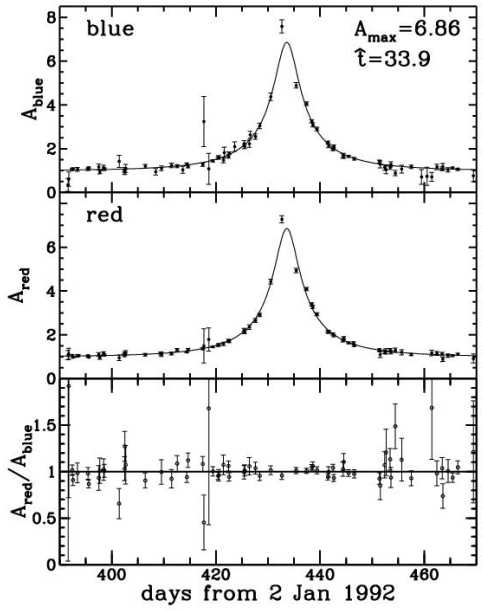


Fig. 10. Blue and red lightcurve of the first Galactic microlensing event MACHO-LMC-1 (Alcock et al. 1993). Data points with error bars show the measured brightness of a star in the LMC as a function of time, and the curve in both upper panels show the best fitting 'standard' microlensing lightcurve. Overall, the quality of the fit is impressive, and the lack of chromatic effects, demonstrated by the constancy of the flux ratio shown in the lowest panel, strongly argues for this being a microlensing event. However, some points (in particular one close to the maximum flux) deviate very significantly from the simple model lightcurve, indicating that this may be a binary microlens

ing events, these surveys provide unique data sets which are also useful for other branches of astronomy, most notably studies of stellar statistics and variability.

1.3 What is lensing good for?

Hopefully, by the end of these lectures we will have provided convincing answers to this question, but for the impatient, we shall summarize some of the highlights of lensing applications.

Measure mass and mass distributions. Gravitational light deflection is determined by the gravitational field through which light propagates. This in turn is related to the mass distribution via the Poisson equation (or its GR generalization). It is essential to realize that this simple fact implies that *gravitational light deflection is independent of the nature of the matter and of its state* – lensing is equally sensitive to dark and luminous matter, and to matter in equilibrium or far out of it. On the negative side, this implies that lensing alone cannot distinguish between these forms of matter, but on the positive side, it also cannot miss one of these matter forms. Hence, lensing is an ideal tool for measuring the total mass of astronomical bodies, dark and luminous.

From the Einstein deflection law (1), it is obvious that characteristic image separations scale with the lens mass like $M^{1/2}$; hence, the observation of multiple images and rings immediately allows an estimate of the mass of the lensing galaxy – or more precisely, the mass within a cylinder with a diameter of the image separation or the ring diameter, centered on the lens.[2] More detailed modeling, and additional observables, such as flux ratios, can yield very precise mass estimates. Indeed, as will be discussed in SL, accurate mass estimates within galaxies, with an uncertainty of a few percent, have been achieved – by far the most precise mass determinations in (extragalactic) astronomy. Similarly, from the locations of giant arcs in clusters, the masses of the central parts of clusters can be determined (Sect. 4 of WL). With the advent of HST imaging and the discovery of multiple image systems in some strong lensing clusters, detailed mass models have been obtained, which led to very precise mass estimates in those clusters (needless to say, they confirm the dominance of dark matter in clusters).

Weak lensing studies of clusters estimate the mass distribution to much larger radii than the strong lensing regime, and, like strong lensing effects, probe for asymmetries and substructures in the cluster mass. For example, already the strong lensing properties of the cluster A2218 (Fig. 6) reveals the bimodal nature of the mass distribution. In fact, substructure in the mass distribution of lens galaxies has been detected, thereby confirming one of the robust predictions of the Cold Dark Matter model for our Universe (SL, Sect. 8). In addition, the mass distribution of galaxies at large radii, where one runs out of local dynamical tracers, can be studied statistically using an effect called galaxy-galaxy lensing (WL, Sect. 8).

Constraining the number density of mass concentrations. The probability for a lensing event to occur (e.g., the fraction of high-redshift sources that are multiply imaged, or the fraction of stars undergoing microlensing) depends on the projected number density of potential lenses. Hence, by investigating statistically well-defined samples of sources and their lensed fraction, we can infer the number density of lenses. Examples of such studies are estimates of the number density of compact objects in the dark halo of our Galaxy, the redshift evolution of the number density of galaxies acting as strong lenses, and the number density of clusters producing strong and weak lensing signals. Upper limits on the number of lensing events can also be translated into upper bounds on the number density of putative lenses: e.g., the fact that nearly all multiply-imaged sources have a visible lens galaxy puts strong upper bounds on the number density of dark lenses (they can at most provide a few percent of the galaxy-mass objects), and the non-detection of lens systems with image separations of tens of milliarcseconds provides

[2] Whereas this 'cylinder' contains all the mass inhomogeneities of the cosmic matter distribution between the source and the observer, it is dominated by the mass of the lensing galaxy.

bounds on the number density of compact galaxies with masses $\sim 10^9 M_\odot$. In fact, by now lensing has put stringent constraints on the population of compact massive objects in the Universe over an extremely broad range of mass scales, from $\sim 10^{-3} M_\odot$ (from upper limits on the variability of distant quasars) to $\sim 10^{16} M_\odot$ (from the absence of very wide pairs of quasars), with only a few mass gaps within this range. Even lower-mass objects ($\sim 10^{-6} M_\odot$) can be ruled out as significant contributors to the dark matter in our Milky Way (see ML).

Provide estimates of cosmological parameters. Following Refsdal's idea, the Hubble constant can be obtained from the time delay in multiple image systems. This method has the advantage of being independent of the usual distance ladder used in determinations of H_0, and it also measures the Hubble constant on a truly cosmic scale, in contrast to the quite local measurements based on Cepheid distances. Despite the determination of time delays in a number of systems, values for H_0 by lensing are burdened with the uncertainties of the lens models; however, there is a trend towards slightly lower values of the Hubble constant than obtained from Cepheids (see SL, Sect. 5). Other cosmological parameters can also be obtained from lensing. For example, the fraction of lensed high-redshift quasars when combined with the distribution of image separations can be used to estimate the cosmological model (SL, Sect. 6). Weak lensing by the large-scale structure is sensitive to the matter density parameter and the normalization of the density fluctuations, and significant constraints on these parameters have been obtained (WL, Sect. 7). In particular in combination with results from the anisotropy of the cosmic microwave background, future cosmic shear studies will provide an invaluable probe of the equation of state of the dark energy. Weak lensing has also successfully been used to determine the bias parameter, which describes the relation between the statistical distribution of galaxies and the underlying dark matter, and for which only few alternative methods are available (WL, Sect. 8).

Lenses as natural telescopes. Since a lens can magnify background sources, these appear brighter than they would without a lens. This makes it easier to investigate these sources in detail, e.g. through spectroscopic observations. In some cases, this magnification is even essential to detect the sources in the first place, provided their lensed brightness just exceeds the detection threshold of a survey or of the current instrumental sensitivity. This magnification effect has in fact yielded spectacular results, such a very detailed spectra of very distant galaxies, the detection of some of the highest redshift galaxies behind cluster lenses, and the detection of very faint sub-millimeter sources in cluster fields.[3] In fact, the lens magnification can be very large in some

[3] A magnification by a factor of, say, 5 implies that a spectrum of the source can be taken in 1/25-th of the time it would take to get the same signal-to-noise

rare cases, but these rare cases truly stick out: some of the most extreme sources, with regards to their apparent luminosity, are strongly magnified – such as the spectacular IRAS galaxy F10214 (e.g., Broadhurst & Lehàr 1995), the by-far brightest redshift ~ 3 galaxy cB58 (Seitz et al. 1998), or the extremely luminous $z = 3.87$ quasar APM 08279+5255 (Irwin et al. 1998).[4] A good fraction of known galaxies with redshift larger than ~ 4 have been detected behind cluster lenses, including the redshift record holder ($z = 10.0$) at present (Pelló et al. 2004); see Fig. 11 for an example. During high-magnification Galactic microlensing events, detailed spectra of stars at large distances (e.g. the Galactic bulge) have been taken. As the high magnification region crosses a distant star, observations have mapped the surface brightness distribution of the stars to test stellar atmosphere models.

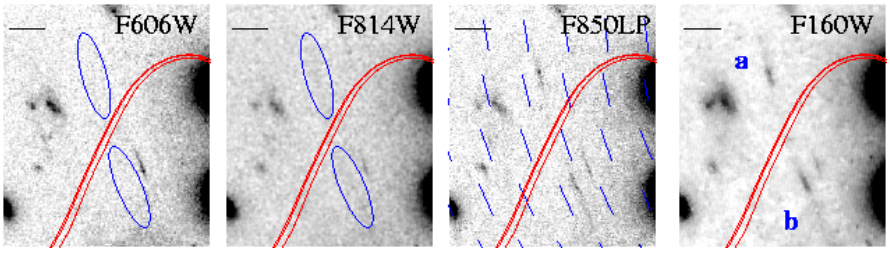


Fig. 11. Example for the use of a gravitational lens as a natural telescope. In a search for very high redshift objects, deep multi-band HST images are taken near the critical curves of clusters, where large magnifications are expected. Shown here are images of a field in the cluster A2218 (see Fig. 6) in four filters, ranging from $0.6\,\mu$m to the near-IR at $1.6\,\mu$m. In the two larger wavelength images, a double source is seen, which is absent at shorter wavelength. The two components are situated at opposite sides of the critical curve, which is drawn for three source redshifts of $z_s = 6$, 6.5 and 7; due to the large number of strong lensing constraints for this cluster, its mass distribution in the central part is very well determined. The sticks indicate the shear field of the cluster, and the elongation of the double images is parallel to this shear, as expected if they were gravitationally lensed images. From the location of the images with respect to the critical curve, and the drop-out of their flux at wavelengths shorter than $\sim 0.8\,\mu$m, the redshift of the source is estimated to be between $z_s = 6.5$ to 7 (from Kneib et al. 2004)

With the lenses as magnifiers, larger effective angular resolution of the sources is obtained. Galaxies acting as sources for giant arcs can therefore be

spectrum of the unlensed source. Needless to say that such a factor can make the difference between an observation being made and one that cannot be done.

[4] Such extremely bright quasars are of great importance for the study of the intergalactic absorption, e.g., the Ly-α forest; no surprise then that such objects, like the highly magnified $z = 3.62$ QSO 1422+231, are preferred targets for investigating absorption lines.

resolved in unprecedented detail, at least in one dimension. The host galaxy of quasars, which is difficult to study in unlensed objects owing to the large brightness contrast between the active nucleus and the surrounding host, can be studied much easier when lensing allows the spatial resolution of the host – in many cases, the host galaxy is in fact mapped into an Einstein ring (see Figs. 3 and 8).

Searches for planets. The light curves of Galactic microlensing events are affected by companions of the main lens. For example, light curves of binary stars are readily identified as such, provided their separation falls into a favorable range determined by the geometry of the lens system. Because of that, even planets will leave an observable trace in the microlensing lightcurves if they are situated at the right radius from the star and at the right orbital phase. Although these traces can be quite subtle, and last for a short time only, current observing campaigns aimed at the search for planets have the sensitivity for their detection, and several candidate events for the detection of planetary signals in microlensing light curves have been reported. Indeed, microlensing is considered to be the simplest (and cheapest) possibility to detect the presence of low-mass planets around distant stars (ML).

These few examples should suffice to illustrate the broad range of applications of gravitational lensing; the ever increased publication rate of articles investigating and applying gravitational lensing underlines the timeliness of the subject.

2 Gravitational lens theory

Assuming the validity of General Relativity, light propagates along the null geodesics of the space-time metric. However, most astrophysically relevant situations permit a much simpler approximate description of light rays, which is called gravitational lens theory. In this section, we summarize the basic equations for the description of light deflection in a gravitational field. The reader is referred to SEF and PLW for a more detailed account and further references.

2.1 The deflection angle

Consider first the deflection of a light ray by the exterior of a spherically symmetric mass M. Provided that the ray's impact parameter ξ is much larger than the Schwarzschild radius of the mass, $\xi \gg R_{\rm s} \equiv 2GM\,c^{-2}$, then General Relativity predicts that the deflection angle $\hat{\alpha}$ is

$$\hat{\alpha} = \frac{4GM}{c^2\,\xi}\,. \tag{2}$$

This is just twice the value obtained in Newtonian gravity (see Sect. 1.1). According to the condition $\xi \gg R_{\rm s}$, the deflection angle is small, $\hat{\alpha} \ll 1$. This condition also implies that the Newtonian gravitational field strength is small, $\phi_{\rm N}/c^2 \ll 1$.

The field equations of General Relativity can be linearized if the gravitational field is weak. The deflection angle of an ensemble of mass points is then the (vectorial) sum of the deflections due to the individual mass components. A three-dimensional mass distribution with volume density $\rho(\boldsymbol{r})$ can be divided into cells of size $\mathrm{d}V$ and mass $\mathrm{d}m = \rho(\boldsymbol{r})\,\mathrm{d}V$. Let a light ray pass this mass distribution, and describe its spatial trajectory by $(\xi_1(\lambda), \xi_2(\lambda), r_3(\lambda))$, where the coordinates are chosen such that the incoming light ray (i.e. far from the deflecting mass distribution) propagates along r_3. The actual light ray is deflected, but if the deflection angle is small, the ray can be approximated as a straight line in the neighborhood of the deflecting mass (note that this corresponds to the Born approximation in atomic and nuclear physics). A mass distribution for which this condition is satisfied is called a *geometrically-thin lens*. Then, $\boldsymbol{\xi}(\lambda) \approx \boldsymbol{\xi}$, independent of the affine parameter λ. Note that $\boldsymbol{\xi} = (\xi_1, \xi_2)$ is a two-dimensional vector. The impact vector of the light ray relative to the mass element $\mathrm{d}m$ at $\boldsymbol{r}' = (\xi_1', \xi_2', r_3')$ is then $\boldsymbol{\xi} - \boldsymbol{\xi}'$, independent of r_3', and the total deflection angle is

$$\begin{aligned}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) &= \frac{4G}{c^2} \sum \mathrm{d}m(\xi_1', \xi_2', r_3') \, \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2} \\ &= \frac{4G}{c^2} \int \mathrm{d}^2\xi' \int \mathrm{d}r_3' \, \rho(\xi_1', \xi_2', r_3') \, \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2} \, , \end{aligned} \tag{3}$$

which is also a two-dimensional vector. Since the last factor in (3) is independent of r_3', the r_3'-integration can be carried out by defining the *surface mass density*

$$\Sigma(\boldsymbol{\xi}) \equiv \int \mathrm{d}r_3 \, \rho(\xi_1, \xi_2, r_3) \, , \tag{4}$$

which is the mass density projected onto a plane perpendicular to the incoming light ray. Thus, the deflection angle produced by an arbitrary density distribution is

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \mathrm{d}^2\xi' \, \Sigma(\boldsymbol{\xi}') \, \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2} \, , \tag{5}$$

provided that the deviation of the actual light ray from a straight (undeflected) line within the mass distribution is small compared to the scale on which the mass distribution changes significantly. This condition is satisfied in virtually all astrophysically relevant situations (i.e. lensing by galaxies and clusters of galaxies), unless the deflecting mass extends all the way from the source to the observer (a case which will be dealt with in WL). It should also be noted that in a lensing situation such as that displayed in Fig. 12, the incoming light rays are not mutually parallel, but fall within a beam

with opening angle approximately equal to the angle which the mass distribution subtends on the sky. This angle, however, is typically *very* small (in the case of cluster lensing, the relevant angular scales are of order 1 arc min $\approx 3 \times 10^{-4}$ radians).

2.2 The lens equation

A typical situation considered in gravitational lensing is sketched in Fig. 12, where a mass concentration at redshift $z_{\rm d}$ (or distance $D_{\rm d}$) deflects the light rays from a source at redshift $z_{\rm s}$ (or distance $D_{\rm s}$). If there are no other deflectors close to the line-of-sight, and if the extent of the deflecting mass along the line-of-sight is very much smaller than both $D_{\rm d}$ and the distance $D_{\rm ds}$ from the deflector to the source,[5] the actual light rays which are smoothly curved in the neighborhood of the deflector can be replaced by two straight rays with a kink near the deflector. The magnitude and direction of this kink is described by the deflection angle $\hat{\boldsymbol{\alpha}}$, which depends on the mass distribution of the deflector and the impact vector of the light ray. The lens equation relates the true position of the source to its observed position on the sky. As sketched in Fig. 12, the source and lens planes are defined as planes perpendicular to a straight line (the optical axis) from the observer to the lens at the distance of the source and of the lens, respectively. The exact definition of the optical axis does not matter because of the smallness of the angles involved in a typical lens system. Let $\boldsymbol{\eta}$ denote the two-dimensional position of the source on the source plane, measured with respect to the intersection point of the optical axis with the source plane. From Fig. 12 we can read off the geometric condition that (again making use of the smallness of angles occurring, so that $\sin\hat{\alpha} \approx \hat{\alpha} \approx \tan\hat{\alpha}$)

$$\boldsymbol{\eta} = \frac{D_{\rm s}}{D_{\rm d}}\,\boldsymbol{\xi} - D_{\rm ds}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi})\ . \tag{6}$$

Introducing angular coordinates by

$$\boldsymbol{\eta} = D_{\rm s}\boldsymbol{\beta} \ \text{ and } \ \boldsymbol{\xi} = D_{\rm d}\boldsymbol{\theta}\ , \tag{7}$$

we can transform eq. (6) to

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_{\rm ds}}{D_{\rm s}}\,\hat{\boldsymbol{\alpha}}(D_{\rm d}\boldsymbol{\theta}) \equiv \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta})\ , \tag{8}$$

where we defined the *scaled deflection angle* $\boldsymbol{\alpha}(\boldsymbol{\theta})$ in the last step. The interpretation of the lens equation (8) is that a source with true position $\boldsymbol{\beta}$ can be seen by an observer to be located at angular positions $\boldsymbol{\theta}$ satisfying (8). If (8) has more than one solution for fixed $\boldsymbol{\beta}$, a source at $\boldsymbol{\beta}$ has images at

[5] This condition is very well satisfied in most astrophysical situations. A cluster of galaxies, for instance, has a typical size of a few Mpc, whereas the distances $D_{\rm d}$, $D_{\rm s}$, and $D_{\rm ds}$ are fair fractions of the Hubble length $cH_0^{-1} = 3\,h^{-1} \times 10^3\,{\rm Mpc}$.

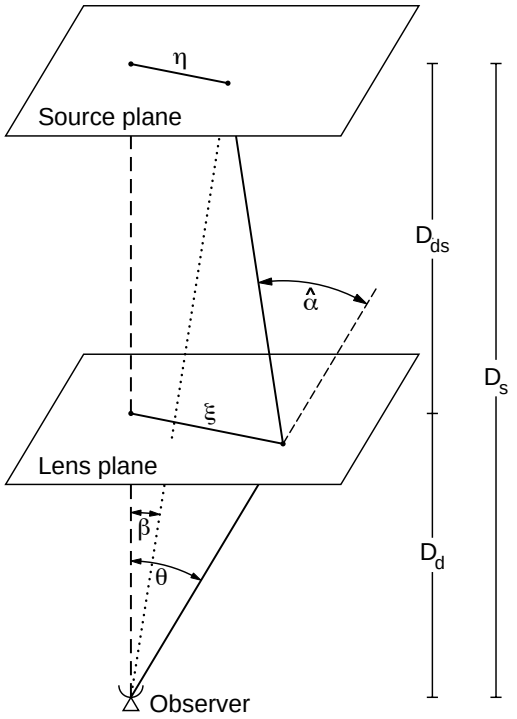


Fig. 12. Sketch of a typical gravitational lens system.

several positions on the sky, i.e. the lens produces multiple images. For this to happen, the lens must be 'strong'. We can express the scaled deflection angle in terms of the surface mass density as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \mathrm{d}^2\theta' \, \kappa(\boldsymbol{\theta}') \, \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} \,, \tag{9}$$

where we have defined the *dimensionless surface mass density* or *convergence*

$$\kappa(\boldsymbol{\theta}) := \frac{\Sigma(D_\mathrm{d}\boldsymbol{\theta})}{\Sigma_\mathrm{cr}} \quad \text{with} \quad \Sigma_\mathrm{cr} = \frac{c^2}{4\pi G} \frac{D_\mathrm{s}}{D_\mathrm{d}\, D_\mathrm{ds}} \,, \tag{10}$$

where the *critical surface mass density* Σ_cr depends on the distances to the source and the lens. As will be discussed later (Sect. 2.4), a mass distribution which has $\kappa \geq 1$ somewhere, i.e. $\Sigma \geq \Sigma_\mathrm{cr}$, produces multiple images for some source positions. Hence, Σ_cr is a characteristic value for the surface mass density which is the dividing line between 'weak' and 'strong' lenses.[6]

[6] In order to derive the foregoing equations, we have used Euclidean geometry to relate angles to length scales. We shall discuss in Sect. 4 that the equations still hold in an expanding universe, provided the distances D's are interpreted as angular diameter distances – hence, in the notation of Sect. 4, $D \equiv D^\mathrm{ang}$.

The lens equation (8) describes a mapping $\boldsymbol{\theta} \to \boldsymbol{\beta}$ from the lens plane to the source plane; for any mass distribution $\Sigma(\boldsymbol{\xi})$, this mapping can (in principle) be easily calculated. One problem of gravitational lens theory is the inversion of (8), i.e., to find all the image positions $\boldsymbol{\theta}$ for a given source position $\boldsymbol{\beta}$. Since the mapping $\boldsymbol{\theta} \to \boldsymbol{\beta}$ is non-linear, the inversion of the lens equation can be carried out analytically only for very simple mass models of the lens. As the number of images $\boldsymbol{\theta}$ for a given source $\boldsymbol{\beta}$ is not known a priori, a numerical inversion is non-trivial in general; however, we shall see below that there are methods to determine the image multiplicity as a function of the source position.

The identity $\nabla \ln|\boldsymbol{\theta}| = \boldsymbol{\theta}/|\boldsymbol{\theta}|^2$, valid for any two-dimensional vector $\boldsymbol{\theta}$, shows that the (scaled) deflection angle can be written as a gradient of the *deflection potential*,

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \mathrm{d}^2\theta' \, \kappa(\boldsymbol{\theta}') \, \ln|\boldsymbol{\theta} - \boldsymbol{\theta}'| \tag{11}$$

as

$$\boldsymbol{\alpha} = \nabla\psi \, , \tag{12}$$

so that the mapping $\boldsymbol{\theta} \to \boldsymbol{\beta}$ is a gradient mapping. Furthermore, using the identity $\nabla^2 \ln|\boldsymbol{\theta}| = 2\pi\delta_{\mathrm{D}}(\boldsymbol{\theta})$, where δ_{D} is the (two-dimensional) Dirac delta 'function', one obtains from (11) that

$$\nabla^2\psi = 2\kappa \, , \tag{13}$$

which is the *Poisson equation* in two dimensions. The similarity between these lensing relations and standard three-dimensional gravity (ψ corresponds to the gravitational potential ϕ_{N}, $\boldsymbol{\alpha}$ corresponds to the acceleration vector, κ corresponds to the volume mass density ρ) shall be noted.

For later purposes, we shall find it useful to define a further scalar function

$$\tau(\boldsymbol{\theta};\boldsymbol{\beta}) = \frac{1}{2}\,(\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta}) \, , \tag{14}$$

called the *Fermat potential*; this is a function of the lens plane coordinate $\boldsymbol{\theta}$, with the source position $\boldsymbol{\beta}$ entering as a parameter. It should be noted that

$$\nabla\tau(\boldsymbol{\theta};\boldsymbol{\beta}) = \mathbf{0} \tag{15}$$

is equivalent to the lens equation (8). As has been shown in Schneider (1985; see also SEF), the function $\tau(\boldsymbol{\theta};\boldsymbol{\beta})$ is, up to an affine transformation, the light travel time along a ray starting at position $\boldsymbol{\beta}$, traversing the lens plane at position $\boldsymbol{\theta}$ and arriving at the observer. Thus, (15) expresses the fact that physical light rays are those for which the light travel time is stationary – which thus expresses Fermat's principle in the context of lensing by a geometrically thin matter distribution. We shall see that the Fermat potential – or time-delay function – is very useful for a classification of the multiple images in a gravitational lens system. Displaying lens properties in terms of the Fermat potential (Blandford & Narayan 1986) provides useful insight in the behavior of the lens mapping.

2.3 Magnification and distortion

The solutions $\boldsymbol{\theta}$ of the lens equation yield the angular positions of the images of a source at $\boldsymbol{\beta}$. The shapes of the images will differ from the shape of the source because light bundles are deflected differentially, as we saw from the images of giant arcs in Fig. 6. In general, the shape of the images must be determined by solving the lens equation for all points within an extended source. Liouville's theorem and the absence of emission and absorption of photons in gravitational light deflection imply that lensing conserves surface brightness (or specific intensity). Hence, if $I^{(\mathrm{s})}(\boldsymbol{\beta})$ is the surface brightness distribution in the source plane, the observed surface brightness distribution in the lens plane is

$$I(\boldsymbol{\theta}) = I^{(\mathrm{s})}[\boldsymbol{\beta}(\boldsymbol{\theta})] \,. \tag{16}$$

If a source is much smaller than the angular scale on which the lens properties change, the lens mapping can be linearized locally. The distortion of images is then described by the Jacobian matrix

$$\mathcal{A}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \left(\delta_{ij} - \frac{\partial^2 \psi(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right) = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} \,, \tag{17}$$

where we have introduced the components of the *shear* $\gamma \equiv \gamma_1 + \mathrm{i}\gamma_2 = |\gamma| \mathrm{e}^{2\mathrm{i}\varphi}$,

$$\gamma_1 = \frac{1}{2}(\psi_{,11} - \psi_{,22}) \,, \quad \gamma_2 = \psi_{,12} \,, \tag{18}$$

and κ is related to ψ through Poisson's equation (13). Hence, if $\boldsymbol{\theta}_0$ is a point within an image, corresponding to the point $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\theta}_0)$ within the source, we find from (16), using the locally linearized lens equation,

$$I(\boldsymbol{\theta}) = I^{(\mathrm{s})}\left[\boldsymbol{\beta}_0 + \mathcal{A}(\boldsymbol{\theta}_0) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right] \,. \tag{19}$$

According to this equation, the images of a source with circular isophotes are ellipses. The ratios of the semi-axes of such an ellipse to the radius of the source are given by the inverse of the eigenvalues of $\mathcal{A}(\boldsymbol{\theta}_0)$, which are $1 - \kappa \pm |\gamma|$, and the ratio of the solid angles subtended by an image and the unlensed source is the inverse of the (absolute value of the) determinant of $\mathcal{A}$. The inverse of the Jacobian is called the magnification tensor,

$$M(\boldsymbol{\theta}) = \mathcal{A}^{-1} \,, \tag{20}$$

and yields the local mapping from the source to the image plane. The fluxes observed from the image and from the unlensed source are given as integrals over the brightness distributions $I(\boldsymbol{\theta})$ and $I^{(\mathrm{s})}(\boldsymbol{\beta})$, respectively, and their ratio is the *magnification* $|\mu(\boldsymbol{\theta}_0)|$. From (19), we find for the magnification of a 'small' source

$$\mu = \det M = \frac{1}{\det \mathcal{A}} = \frac{1}{(1-\kappa)^2 - |\gamma|^2} \,. \tag{21}$$

The images are thus distorted in shape and size. The shape distortion is due to the tidal gravitational field, described by the shear γ, whereas the magnification is caused by both isotropic focusing due to the local matter density κ and anisotropic focusing due to shear. The magnification as defined in (21) can have either sign; the sign of μ is called the *parity* of an image. Negative-parity images are mirror-symmetric images of the source. Of course, the observed fluxes of images are determined by the absolute value of μ. Since the intrinsic luminosity of sources is unknown, the magnification in a lens system is not an observable. However, the flux ratio of different images provides a direct measurement of the (absolute value of the) corresponding magnification ratio. In general, if two extended images of a source are observed, then their shapes depend on the shape of the source through $\mathcal{A}$. As the shape of the source is unknown, what can be determined from the shape of extended images is the relative magnification matrix $\mathcal{A}_{ij} = \mathcal{A}(\boldsymbol{\theta}_i)\mathcal{A}^{-1}(\boldsymbol{\theta}_j)$, which provides the linearized mapping of one image onto the other. Note that $\mathcal{A}_{ij}$ is in general not symmetric and thus has four independent components. For a pair of images with opposite parity, $\det \mathcal{A}_{ij} < 0$, and so these two images are mirror symmetric; an example of this can be seen in the VLBI images of QSO 0957+561 (see Fig. 1).

To consider the distortion of the shape of images in somewhat more detail, we shall rewrite the Jacobian in a slightly different form,

$$\mathcal{A}(\boldsymbol{\theta}) = (1-\kappa)\begin{pmatrix} 1-g_1 & -g_2 \\ -g_2 & 1+g_1 \end{pmatrix} , \tag{22}$$

where we have defined the *reduced shear*

$$g \equiv \frac{\gamma}{1-\kappa} = \frac{|\gamma|}{1-\kappa}\,\mathrm{e}^{2\mathrm{i}\varphi} . \tag{23}$$

As can be easily seen from (22), the factor $(1-\kappa)$ only yields an isotropic stretching of the image, but does not affect its shape. The reduced shear g – like γ – is considered to be a complex number, $g = g_1 + \mathrm{i}g_2$ and its components determine the change of shape between the source and the image. In particular, a circular source of unit radius is mapped onto an ellipse with axes $|(1-\kappa)(1+|g|)|^{-1}$ and $|(1-\kappa)(1-|g|)|^{-1}$, and the orientation of the ellipse is determined by the phase φ of g. As will be seen in WL, the reduced shear is the central quantity in weak gravitational lensing.

The images of a small source (what that means depends on the context; see below) are therefore magnified by $|\mu(\boldsymbol{\theta}_i)|$, and the total magnification of a small source at position $\boldsymbol{\beta}$ is given by the sum of the magnifications over all its images,

$$\mu_{\mathrm{p}}(\boldsymbol{\beta}) = \sum_i |\mu(\boldsymbol{\theta}_i)| , \tag{24}$$

where the index 'p' indicates that this equation applies to the point-source limit. The magnification of real sources with finite extent is given by the

weighted mean of $\mu_{\rm p}$ over the source's area,

$$\mu = \left[\int {\rm d}^2\beta\; I^{(\rm s)}(\boldsymbol{\beta})\right]^{-1} \int {\rm d}^2\beta\; I^{(\rm s)}(\boldsymbol{\beta})\, \mu_{\rm p}(\boldsymbol{\beta})\;, \tag{25}$$

where $I^{(\rm s)}(\boldsymbol{\beta})$ is the surface brightness profile of the source. Whereas gravitational lensing is achromatic, because the deflection of photons does not depend on their frequency, the finite resolution of observations can lead to color terms in practice, since the surface brightness distribution $I^{(\rm s)}(\boldsymbol{\beta})$ can be different at different frequencies. Then, if the magnification $\mu_{\rm p}(\boldsymbol{\beta})$ varies on scales comparable to the source size, the magnification of an extended but unresolved source *can* depend on the frequency.

Since the shear is defined by the trace-free part of the symmetric Jacobian matrix $\mathcal{A}$, it has two independent components. There exists a one-to-one mapping from symmetric, trace-free 2×2 matrices onto complex numbers, and we shall extensively use complex notation. Note that the shear (and the reduced shear) transforms as $\mathrm{e}^{2{\rm i}\varphi}$ under rotations of the coordinate frame, and is therefore *not* a vector (but a polar, i.e., it has the same transformation properties as the linear polarization of electromagnetic waves). Equations (11) and (18) imply that the complex shear can be written as

$$\begin{aligned} \gamma(\boldsymbol{\theta}) \;&= \frac{1}{\pi}\int_{\mathbb{R}^2} {\rm d}^2\theta'\, \mathcal{D}(\boldsymbol{\theta}-\boldsymbol{\theta}')\,\kappa(\boldsymbol{\theta}')\;, \quad \text{with} \\ \mathcal{D}(\boldsymbol{\theta}) \equiv \frac{\theta_2^2-\theta_1^2-2{\rm i}\theta_1\theta_2}{|\boldsymbol{\theta}|^4} &= \frac{-1}{(\theta_1-{\rm i}\theta_2)^2}\;. \end{aligned} \tag{26}$$

2.4 Critical curves and caustics, and general properties of lenses

In any lens there can be closed, smooth curves, known as *critical curves*, on which the Jacobian vanishes, $\det \mathcal{A}(\boldsymbol{\theta}) = 0$. The curves in the source plane which are obtained by mapping the critical curves with the lens equation are called *caustics*, which are not necessarily smooth, but can develop cusps. Critical curves and caustics are of great importance for a qualitative understanding of the lens mapping, owing to their following properties:

1. The magnification $\mu = 1/\det \mathcal{A}$ formally diverges for an image on a critical curve. Infinite magnifications are of course unphysical. All astronomical sources have a finite size that keeps their observed magnification (25) finite. For a hypothetical source of vanishing extent, the magnification would be finite because the geometrical optics approximation then breaks down and we must use wave optics. The resulting diffraction patterns predict finite, though potentially very high magnifications (see e.g. Ohanian 1983 or Chapter 7 of SEF). Nevertheless, a source located near a caustic can produce very highly magnified images close to the corresponding critical curve in the lens plane.

2. The number of images a source produces depends on its location relative to the caustic curves. Assuming a mass profile of a lens for which the deflection angle tends to zero for large $|\boldsymbol{\theta}|$ – as is true for all real lenses – and an upper bound to the deflection angle (i.e. excluding point-mass lenses for the moment), a source at large $|\boldsymbol{\beta}|$ will have only one image, at $\boldsymbol{\theta} \approx \boldsymbol{\beta}$, whereas it can have multiple images for small impact vectors. The lens mapping (8) is locally invertible at all locations for which $\det \mathcal{A} \neq 0$. This immediately implies that a change of the source position does not lead to the change of the number of images *unless* the source moves across a caustic – since caustics are obtained by mapping the critical curves (where the lens mapping in not invertible) onto the source plane. When a source position crosses a caustic, a pair of images near the corresponding critical curve is either created or destroyed, depending on the direction of crossing. The side of the caustic where the number of images is larger by two is often called the 'inner side'. A source close to, and on the inner side of a caustic possesses a pair of images with very high and nearly equal magnification on either side of the critical curve, in addition to any other images. The bright pair must have opposite parities because the magnification changes sign at the critical curve.

Whereas the critical curves are smooth, this does not need to be the case for caustics. To see that, let $\boldsymbol{\theta}(\lambda)$ be a parameterization of a critical curve; the caustic then is $\boldsymbol{\beta}(\boldsymbol{\theta}(\lambda))$. The tangent vector to the critical curve is the derivative $\dot{\boldsymbol{\theta}}(\lambda) \equiv \mathrm{d}\boldsymbol{\theta}(\lambda)/\mathrm{d}\lambda$, and the tangent vector to the caustic is

$$\frac{\mathrm{d}\boldsymbol{\beta}(\boldsymbol{\theta}(\lambda))}{\mathrm{d}\lambda} = \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}}\,\frac{\mathrm{d}\boldsymbol{\theta}}{\mathrm{d}\lambda} = \mathcal{A}(\boldsymbol{\theta}(\lambda))\,\dot{\boldsymbol{\theta}}(\lambda)\ .$$

This vector, however, can vanish if the tangent vector to the critical curve $\dot{\boldsymbol{\theta}}(\lambda)$ is parallel to the eigenvector of $\mathcal{A}$ whose eigenvalue is 0 (remember that we are analyzing a critical curve, along which one eigenvalue of $\mathcal{A}$ is always zero). Hence, if the direction of the tangent vector to the critical curve is the singular direction of $\mathcal{A}$, the caustic curve need not be smooth; in fact, it has a *cusp*. Apart from any cusps the caustic curves are smooth curves called *fold caustics*. These names are taken from singularity theory, a mathematical discipline that studies the critical points of general mappings. We shall see the occurrence of cusps later in several specific examples of lens mappings. A source close to and inside a cusp has three highly magnified images near the corresponding point on the critical curve; one can show (see e.g. Schneider & Weiss 1992; Mao 1992) that the sum of the absolute values of the magnification of the two outer images equals the absolute value of the magnification of the central of these three images. A source just outside the cusp has one highly magnified image near the corresponding critical curve.

We thus obtain a qualitative understanding of the geometry of a lens mapping from the critical curves and caustics. The critical curves divide the lens plane into regions of positive (i.e., $\mu > 0$) and negative ($\mu < 0$) parity. The

corresponding caustics divide the source plane into regions of different image multiplicity: whenever a source position changes across a caustic, the number of images changes by ± 2. Since for mass distributions without singularities (e.g., point masses) the number of images is 1 if the angular source position is sufficiently distant from the mass concentration, the number of images can be easily assigned to each of the regions in the source plane, once the caustics are known.

If an extended source is located on the caustic, either fold or cusp, the corresponding two or three images merge; only that part of the source which lies inside the inner region of the caustic is (locally) multiply imaged. Since $\det \mathcal{A} = 0$ implies that (at least) one of the two eigenvalues of $\mathcal{A}$ vanishes, the image(s) are highly distorted in the direction of the corresponding eigenvector; therefore, the image of a circular source can be very strongly elongated. This is the origin of the giant luminous arcs in clusters. From what has been said above, for a cusp the singular direction is tangent to the critical curve; hence, if an arc is produced by a source on a cusp, the direction of elongation displays approximately the local direction of the critical curve.

Types of images. The Fermat potential $\tau(\boldsymbol{\theta};\boldsymbol{\beta})$ introduced in Sect. 2.2 yields a convenient classification of images, according to whether an image $\boldsymbol{\theta}$ is located at a minimum, maximum, or saddle point of τ – remember that images of a source occur at points $\boldsymbol{\theta}$ for which τ is stationary. Since the Jacobian matrix is the Hessian of τ, $\mathcal{A}_{ij} = \partial^2\tau/(\partial\theta_i\partial\theta_j)$, these three types of images are distinguished by the signs of the two eigenvalues a_i of $\mathcal{A}$: At a minimum of τ, both are positive, implying that $\det \mathcal{A} > 0$, $\mathrm{tr}\, \mathcal{A} > 0$, whereas at a maximum, both are negative so that $\det \mathcal{A} > 0$, $\mathrm{tr}\, \mathcal{A} < 0$. At a saddle point, the signs of the a_i are different, so that $\det \mathcal{A} < 0$. Given that

$$\mathrm{tr}\, \mathcal{A} = 2(1 - \kappa)\,, \tag{27}$$

one sees that minima (maxima) occur at positions where $\kappa < 1$ ($\kappa > 1$), whereas nothing can be said about κ at saddles.

Odd number and magnification theorems. In a remarkable, one-page paper, Burke (1981) proved a theorem on the number of images a gravitational lens can produce: For a gravitational lens with a smooth surface mass density which decreases faster than $|\boldsymbol{\theta}|^{-1}$ as $|\boldsymbol{\theta}| \to \infty$, the number of images corresponding to extrema of τ, and thus to positive parity images, equals the number of saddle points plus 1, provided the source is not located on a caustic. Hence, the total number of images is odd. In addition, at least one of the images corresponds to a minimum of τ.

The proof of this theorem is obtained using the Poincaré–Hopf index theorem and can also be found in Sect. 5.4 of SEF. The fact that any source has at least one image corresponding to a minimum of τ is easily seen: the

Fermat potential $\tau(\boldsymbol{\theta};\boldsymbol{\beta})$ behaves like $|\boldsymbol{\theta}|^2/2$ for $|\boldsymbol{\theta}| \to \infty$, i.e. increases for large impact vectors, and since it is a smooth function, it must attain a minimum somewhere. In particular this implies that smooth lenses cannot make sources disappear. A simple way to see the validity of this theorem follows from what has been said above about the regions of different image multiplicity: A very misaligned source has one image, corresponding to a minimum of τ, and the number of images changes by ± 2 (one of either parity) whenever the source crosses a caustic, and thus is always odd.

As shown by Schneider (1984), a minimum image is magnified, provided $\kappa \geq 0$. This follows directly from the properties of minima,

$$0 < \det \mathcal{A} = (1-\kappa)^2 - |\gamma|^2 < 1 \; ,$$

where the final inequality follows from $\mathrm{tr}\, \mathcal{A} > 0$. Since each source is mapped onto at least one minimum image, the positive density constraint implies that the total magnification of all sources is larger than unity; in other words, the flux of a source behind a lens is larger than the unlensed source. What may sound as a contradiction on first sight – 'all sources are magnified' (which has triggered a rich and often confusing literature on the 'flux conservation' issue) is due to the assumed positivity of the surface density κ which is certainly the case near to strong lenses. However, most lines-of-sight in the inhomogeneous Universe pass through regions which are slightly underdense relative to the homogeneous Universe, resulting in negative κ – since κ is defined as the projected mass overdensity relative to the smooth Universe. The mean magnification over the sphere of sources at given redshift indeed is unity (Weinberg 1976) if the magnification is defined relative to the flux the same source would have in a homogeneous universe of the same mean density.

These two theorems can also be generalized to the case that the deflecting matter distribution is not a geometrically-thin lens, both using heuristic arguments (SEF) or a rigorous proof (Seitz & Schneider 1992).

Necessary and sufficient conditions for multiple imaging. A matter distribution described by its dimensionless surface mass density κ may or may not be sufficiently strong to cause multiple images of sources. Two general criteria for the occurrence of multiple images can be obtained:

1. An isolated transparent lens can produce multiple images if, and only if, there is a point $\boldsymbol{\theta}$ with $\det \mathcal{A}(\boldsymbol{\theta}) < 0$. This can be shown as follows: if $\det \mathcal{A}(\boldsymbol{\theta}) > 0$ for all $\boldsymbol{\theta}$, then the lens equation is globally invertible, and so no multiple images can occur. On the other hand, if $\det \mathcal{A}(\boldsymbol{\theta}_0) < 0$ at some point $\boldsymbol{\theta}_0$, then a source at $\boldsymbol{\beta}_0 \equiv \boldsymbol{\beta}(\boldsymbol{\theta}_0)$ has an image (at $\boldsymbol{\theta}_0$) which corresponds to a saddle point; according to the odd-number theorem, there must be at least two additional images corresponding to extrema of τ.

2. A sufficient (but not necessary) condition for possible multiple images is that there exists a point $\boldsymbol{\theta}$ such that $\kappa(\boldsymbol{\theta}) > 1$. The argument is similar to the one above: if $\kappa(\boldsymbol{\theta}_0) > 1$, then the source at $\boldsymbol{\beta}_0 \equiv \boldsymbol{\beta}(\boldsymbol{\theta}_0)$ has an image which cannot correspond to a minimum of τ, as for them $\kappa < 1$; hence, the source must have at least one additional image corresponding to a minimum.

The second criterion shows why lenses with $\kappa > 1$ are called 'strong': whereas $\kappa > 1$ is not a necessary condition for the possible occurrence of multiple images, the critical surface mass density $\Sigma_{\rm cr}$ is nevertheless the characteristic scale for the occurrence of strong lensing features likes arcs and multiple images. It should be noted that the critical surface mass density depends on the redshift (or distance) of the source; for a given physical surface mass density Σ, the lens strength increases with increasing source redshift since $\Sigma_{\rm cr}$ decreases. This also implies that the critical curves are different for sources at different redshifts; this effect is clearly seen in several clusters of galaxies where strong lensing phenomena occur at different separations from the cluster center for sources of different redshifts.

2.5 The mass-sheet degeneracy

Suppose you observe a multiply-imaged source for which the image positions, their fluxes and perhaps their shapes (in the case of resolved images) can be measured. One then wants to find a mass model for the lens which can reproduce the observational constraints in order to obtain information about the mass distribution in the lens. Whereas this topic will be treated in SL, Sect. 5, and in a somewhat different context in WL, we can already here consider the question of how unique such models can be, even if one assumes a great number of observational constraints. A partial answer to the question is provided by the existence of the *mass sheet degeneracy* (Falco et al. 1985; Gorenstein et al. 1988b; for the weak lensing case, see Schneider & Seitz 1995).

Let $\kappa(\boldsymbol{\theta})$ be a mass distribution which provides a good fit to the observables (i.e., image positions, flux ratios, relative image shapes in the case of extended images etc.); then the whole family of lens models with mass distribution

$$\kappa_\lambda(\boldsymbol{\theta}) = (1-\lambda) + \lambda\,\kappa(\boldsymbol{\theta}) \tag{28}$$

provides an equally good fit to the data. The first term corresponds to adding a homogeneous surface mass density $\kappa_{\rm c} = 1-\lambda$ to the mass distribution, whereas the second term describes a rescaling of the 'original' mass distribution $\kappa(\boldsymbol{\theta})$. We shall now prove the statement made above.

The lens equation corresponding to κ_λ reads

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}_\lambda(\boldsymbol{\theta}) \quad \text{with} \quad \boldsymbol{\alpha}_\lambda(\boldsymbol{\theta}) = (1-\lambda)\boldsymbol{\theta} + \lambda\boldsymbol{\alpha}(\boldsymbol{\theta})\;, \tag{29}$$

where quantities without index 'λ' correspond to the unscaled mass distribution $\kappa(\boldsymbol{\theta})$. Indeed,

$$\boldsymbol{\alpha}_\lambda(\boldsymbol{\theta}) = \nabla\psi_\lambda(\boldsymbol{\theta}) \quad \text{where} \quad \psi_\lambda(\boldsymbol{\theta}) = \frac{1-\lambda}{2}|\boldsymbol{\theta}|^2 + \lambda\psi(\boldsymbol{\theta}) \;, \tag{30}$$

so that the Poisson equation (13) is satisfied, $\nabla^2\psi_\lambda = 2\kappa_\lambda$. By combining the two equations (29), one finds

$$\frac{\boldsymbol{\beta}}{\lambda} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}) \;, \tag{31}$$

so that the lens equation for the transformed mass distribution κ_λ has the same form as for the untransformed mass distribution, except that the coordinates in the source plane is multiplied by $1/\lambda$. However, this rescaling is not directly observable. As a consequence, the Jacobi matrix and the magnification behave as

$$\mathcal{A}_\lambda = \lambda\mathcal{A} \quad ; \quad \mu_\lambda = \frac{\mu}{\lambda^2} \;; \tag{32}$$

the first of these relations then implies with (17) that $\gamma_\lambda(\boldsymbol{\theta}) = \lambda\gamma(\boldsymbol{\theta})$ and $(1-\kappa_\lambda) = \lambda(1-\kappa)$, in agreement with (28). However, the reduced shear g (23) is unchanged under the transformation, which means that the axis ratios of the elliptical images of a round source are unaffected by the transformation. In general, if nothing sets an absolute scale for the source (size or luminosity) or an absolute mass scale for the lens (e.g., from observations of its stellar dynamics), then one cannot distinguish the model described by κ from one described by κ_λ. In particular, the critical curves and the curves with $\kappa = 1$ are unaffected by the transformation (28). However, the Fermat potential transforms as

$$\tau_\lambda(\boldsymbol{\theta};\boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\beta})^2 - \psi_\lambda(\boldsymbol{\theta}) = \lambda\tau(\boldsymbol{\theta};\boldsymbol{\beta}/\lambda) + \text{const.} \;, \tag{33}$$

where the const. only depends on $\boldsymbol{\beta}$. As noted before, the Fermat potential is, up to an affine transformation, the light travel time from the source to the observer when passing through the lens plane at $\boldsymbol{\theta}$. Therefore, since the difference in the Fermat potential calculated at two image positions is proportional to the differences in light travel time, the mass-sheet degeneracy changes this observable time delay. If we know the value of H_0 from other cosmological observations, we can break the degeneracy and determine the absolute surface mass density of a lens. The implications of the mass sheet degeneracy for lens determinations of the Hubble constant will be described in SL. Furthermore, since the transformation (28) leaves the image shapes of extended sources unchanged, the weak lensing techniques to be described in WL are unable to break the mass sheet degeneracy, unless magnification information can be used – see (32). In addition, the mass-sheet degeneracy can be broken if sources with different distances D_s are lensed, since for a given physical mass density Σ, the resulting convergence κ will be different

for different source distances, owing to the dependence of $\Sigma_{\rm cr}$ on the source redshift.

Up to now we have not constrained the value of λ in (28); however, not all values are physically meaningful. For example, for some values of λ the resulting mass distribution κ_λ may attain negative values. Depending on κ, the non-negativity of the surface mass density will restrict the possible value of λ.

3 Simple lens models

For a general mass distribution, the deflection angle has to be obtained through numerical integration; however, for some relatively simple mass distributions, analytical expressions can be obtained. We shall introduce here a few simple mass models for lenses which turn out to be useful for understanding many of the lensing phenomena. The simplest lens models are obtained if the mass distribution is assumed to be spherically symmetric or, of relevance for lensing, if the projected mass distribution is axially symmetric, as then the lens equation reduces essentially to a one-dimensional equation. We shall consider the general properties of such lenses before specializing to two highly relevant cases, the point-mass lens, or more generally, the light deflection exterior to a spherically-symmetric mass distribution, and the isothermal sphere lens. The former one is of utmost relevance for Galactic microlensing, as will be demonstrated in ML, whereas the latter is often used as a simple prescription for the (dark) matter distribution of galaxies and clusters. Clusters and galaxies are not expected to have axisymmetric gravitational potentials; we shall consider the next simple lens models – those which have two axes of symmetry, like elliptical mass distributions – and their generic behavior next.

3.1 Axially symmetric lenses

An axisymmetric matter distribution is characterized by $\Sigma(\boldsymbol{\xi}) = \Sigma(|\boldsymbol{\xi}|)$, if the origin is chosen at the center of symmetry, implying $\kappa(\boldsymbol{\theta}) = \kappa(|\boldsymbol{\theta}|)$. The scaled deflection angle $\boldsymbol{\alpha}(\boldsymbol{\theta})$ is then collinear to $\boldsymbol{\theta}$, as follows from the symmetry of the situation; indeed, from (9) one obtains that

$$\begin{aligned}
\boldsymbol{\alpha}(\boldsymbol{\theta}) &= \frac{\boldsymbol{\theta}}{|\boldsymbol{\theta}|^2} 2 \int_0^{|\boldsymbol{\theta}|} \mathrm{d}\theta'\, \theta'\, \kappa(\theta') \quad \text{or} \\
\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) &= \frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|^2} \frac{4G}{c^2} 2\pi \int_0^{\xi} \mathrm{d}\xi'\, \xi'\, \Sigma(\xi') \equiv \frac{4GM(|\boldsymbol{\xi}|)}{c^2\,|\boldsymbol{\xi}|^2}\boldsymbol{\xi}\,, \qquad (34)
\end{aligned}$$

where $M(\xi)$ is the projected mass enclosed by the circle of radius $\xi = |\boldsymbol{\xi}|$. The deflection due to a geometrically-thin axisymmetric mass distribution at a point $\boldsymbol{\xi}$ is thus the point-mass deflection angle (2) for the mass $M(|\boldsymbol{\xi}|)$

enclosed by the circle with radius $|\boldsymbol{\xi}|$. This fact is analogous to Birkhoff's theorem in three-dimensional gravity which states that the gravitational force caused by a spherically-symmetric mass shell vanishes inside of it; here, the axisymmetric mass in rings causes no deflection at points within them.

Since $\boldsymbol{\alpha}$ is collinear with $\boldsymbol{\theta}$, so is $\boldsymbol{\beta}$, as seen from (8). Hence, if the source position is described by $\boldsymbol{\beta} = \beta\mathbf{e}$, where $\mathbf{e}$ is a unit vector, then $\boldsymbol{\theta} = \theta\mathbf{e}$ as well, and the lens equation becomes one-dimensional,

$$\beta = \theta - \alpha(\theta) \ , \tag{35}$$

where the deflection angle has the properties

$$\alpha(\theta) = -\alpha(-\theta) = \frac{m(\theta)}{\theta} = \bar{\kappa}(\theta)\theta \ , \tag{36}$$

where $m(\theta)$ is the dimensionless mass inside a circle of angular radius θ and $\bar{\kappa}(\theta)$ is the mean surface mass density inside of θ,

$$m(\theta) = 2\int_0^\theta \mathrm{d}\theta'\ \theta'\,\kappa(\theta') \ , \quad \bar{\kappa}(\theta) = \frac{m(\theta)}{\theta^2} \ . \tag{37}$$

For calculating the Jacobian matrix, it is useful to write the lens equation in the form

$$\boldsymbol{\beta} = [1 - \bar{\kappa}(|\boldsymbol{\theta}|)]\ \boldsymbol{\theta} \ ; \tag{38}$$

then, according to (17) one finds from differentiation that

$$\mathcal{A}(\boldsymbol{\theta}) = [1 - \bar{\kappa}(|\boldsymbol{\theta}|)]\,\mathcal{I} - \frac{\bar{\kappa}'}{|\boldsymbol{\theta}|}\begin{pmatrix} \theta_1^2 & \theta_1\theta_2 \\ \theta_1\theta_2 & \theta_2^2 \end{pmatrix} \ , \tag{39}$$

where $\mathcal{I}$ is the two-dimensional identity matrix, and $\bar{\kappa}'(\theta) \equiv \mathrm{d}\bar{\kappa}/\mathrm{d}\theta = 2\big[\kappa(\theta) - \bar{\kappa}(\theta)\big]/\theta$. Comparing (39) with the final form of (17), one sees that indeed $\mathrm{tr}\mathcal{A} = 2(1-\kappa)$, and the shear is

$$\gamma(\boldsymbol{\theta}) = \big[\kappa(\theta) - \bar{\kappa}(\theta)\big]\,\mathrm{e}^{2\mathrm{i}\varphi} \ , \tag{40}$$

where we set $\boldsymbol{\theta} = \theta(\cos\varphi, \sin\varphi)$; hence, the phase of the shear is the same as the polar angle of $\boldsymbol{\theta}$, as expected from symmetry. The determinant of the Jacobian can be calculated either from (21) as

$$\det\mathcal{A} = (1-\kappa)^2 - |\gamma|^2 = (1-\kappa)^2 - (\bar{\kappa}-\kappa)^2 = (1-\bar{\kappa})\,(1+\bar{\kappa}-2\kappa) \ , \tag{41}$$

or, using the original definition (17) of $\mathcal{A}$ as

$$\det\mathcal{A} = \frac{\beta}{\theta}\frac{\mathrm{d}\beta}{\mathrm{d}\theta} = (1-\bar{\kappa})\,(1-\bar{\kappa}-\theta\bar{\kappa}') \ , \tag{42}$$

which can be seen, by inserting the derivative of $\bar{\kappa}$, to yield the same expression.

The fact that $\det \mathcal{A}$ factorizes allows a very simple characterization of the critical curves of these axisymmetric lenses: Critical curves, which of course are circles in this case, occur either when $1 - \bar{\kappa}(\theta) = 0$, or when $1 + \bar{\kappa}(\theta) - 2\kappa(\theta) = 0$. The former ones are called *tangential critical curves*, the latter ones *radial critical curves*. The reason for naming them this way is found by considering the distortion of images close to these critical curves. Consider an image position on the θ_1-axis; according to (39), the Jacobian matrix is diagonal there, $\mathcal{A} = \mathrm{diag}(1+\bar{\kappa}-2\kappa, 1-\bar{\kappa})$. Near a tangential critical curve, the second eigenvalue becomes very small. If the image was a circle, the corresponding source in the source plane would be a highly flattened ellipse, with the minor axis in the β_2-direction being much smaller than the major axis. This implies that if the source is a circle, then the corresponding image near the tangential critical curve will be a highly elongated ellipse, with the highly stretched axis in the θ_2 direction, that is, tangent to the direction towards the center of the lens. Analogous reasoning shows that the image of a circular source near a radial critical curve will be strongly stretched in the radial direction. Recalling the shape and orientation of giant luminous arcs, this consideration suggests that arcs are images of (probably relatively round) sources occurring close to the tangential critical curves of the cluster lenses.

Tangential critical curves are thus characterized by the condition $\bar{\kappa} = 1$. The simplicity of this relation implies that from the location of the tangential critical curve, one can immediately determine the mass inside of it, using (37), namely $\pi\theta_{\mathrm{E}}^2 D_{\mathrm{d}}^2 \Sigma_{\mathrm{cr}}$, where θ_{E} is the angular radius of the critical curve. The relation between θ_{E} and the mass enclosed within θ_{E} is

$$\theta_{\mathrm{E}} = \left(\frac{4GM}{c^2}\,\frac{D_{\mathrm{ds}}}{D_{\mathrm{d}}D_{\mathrm{s}}}\right)^{1/2} \approx 0\farcs9 \left(\frac{M(\leq\theta_{\mathrm{E}})}{10^{12}M_\odot}\right)^{1/2} \left(\frac{D_{\mathrm{ds}}\,1\mathrm{Gpc}}{D_{\mathrm{d}}D_{\mathrm{s}}}\right)^{1/2}, \qquad (43)$$

where we used the definition (10) of the critical surface mass density. Thus, if giant arcs indeed trace the location of the tangential critical curve, their observation can be used to obtain a (at least approximate) mass estimate for the corresponding cluster mass inside of it (we shall come back to this issue in much mode detail in Sect. 4 of WL). The caustic corresponding to a tangential critical curve is a very special one: according to (38), the whole circle $\theta = \theta_{\mathrm{E}}$ is mapped onto the origin $\boldsymbol{\beta} = \mathbf{0}$ in the source plane: the caustic degenerates into a point. This degeneracy occurs solely due to the highly symmetric situation of the lens model; as we shall see later, any slight perturbation of the mass distribution will 'unfold' this caustic point into a curve of finite extent. This symmetric situation then leads to the following result: if a source is placed onto the caustic point, it will be imaged by the lens into a ring with radius θ_{E}, plus an additional image at the center of the lens with $\boldsymbol{\theta} = \mathbf{0}$. Such rings were predicted by Chwolson (1924), but already in 1911, Einstein has discussed their possible occurrence in his notebook, as shown in Renn et al. (1997). Whereas real lenses are not expected to be

perfectly axisymmetric, and therefore one would also not expect to find such Einstein rings, they have indeed been detected, as shown in Fig. 7; as will be explained in Sect. 10 of SL, the occurrence of rings depends on a combination of the mass asymmetry in a lens and the extent of the source. If the source in an axisymmetric lens is moved away from the caustic point, the ring will break up into two images, located near the Einstein radius, on opposite side of the lens center; their image separation will be $\Delta\theta \approx 2\theta_{\rm E}$.

Radial critical curves are circles where $1 + \bar{\kappa}(\theta) - 2\kappa(\theta) = 0$, or, equivalently, $\mathrm{d}\alpha/\mathrm{d}\theta = 1$. Their corresponding caustics are circles in the source plane. In clusters of galaxies, these radial critical curves give rise to radial arcs seen close to the cluster center, whereas they are not seen in galaxy lenses.

General properties of axisymmetric lenses. One can derive several general properties of axisymmetric lenses; again we shall assume that $\alpha(\theta) \to 0$ as $|\theta| \to \infty$, that the deflection angle is bounded, $|\alpha| \le \alpha_{\max}$, and that $\alpha(\theta)$ is a differentiable function. Then one can show that a source with sufficiently large β has only a single image at $\theta \approx \beta$ (the validity of this property is intuitively clear, but can be proven rigorously; see SEF).

Further, a lens can produce multiple images if, and only if, there exists at least one value of θ where $\mathrm{d}\beta/\mathrm{d}\theta = 1 + \bar{\kappa}(\theta) - 2\kappa(\theta) < 0$. The necessity is obvious, since if $\mathrm{d}\beta/\mathrm{d}\theta \ge 0$ throughout, $\beta(\theta)$ is a monotonic function, which can be globally inverted, and no multiple solutions can occur. Sufficiency is seen as follows: if $\mathrm{d}\beta/\mathrm{d}\theta < 0$ at one point, then there must exist a pair of points such that $\mathrm{d}\beta/\mathrm{d}\theta = 0$ (note that these points lie on radial critical curves), since asymptotically for large $|\theta|$, $\mathrm{d}\beta/\mathrm{d}\theta \to 1$. Hence, $\beta(\theta)$ then has a local maximum (say at θ_1) and a local minimum (at $\theta_2 > \theta_1$), and between these two values the function $\beta(\theta)$ decreases. A source located at β_0 with $\beta(\theta_2) \le \beta_0 \le \beta(\theta_1)$ then has at least three images, one with $\theta < \theta_1$, one with $\theta > \theta_2$, and one with $\theta_1 < \theta < \theta_2$. The points $\beta(\theta_1)$ and $\beta(\theta_2)$ lie on radial caustics (see Sect. 3 of SL for graphical illustrations of this point).

The conditions for the possible occurrence of multiple images can also be phrased in terms of the surface mass density: A necessary condition for the occurrence of multiple images is that $\kappa > 1/2$ at least at one point. This can be seen by noting that $\mathrm{d}\beta/\mathrm{d}\theta < 0$ implies $\kappa > (1 + \bar{\kappa})/2 > 1/2$. A sufficient condition for the possible occurrence of multiple images is $\kappa > 1$ at least at one point; this property has been shown already for a general mass distribution, and in this special situation can be seen as follows: if the maximum of κ occurs at $\theta_{\rm m}$, then $\kappa(\theta_{\rm m}) > 1$ and $\kappa(\theta_{\rm m}) \ge \bar{\kappa}(\theta_{\rm m})$, which implies $\mathrm{d}\beta/\mathrm{d}\theta < 0$ at $\theta_{\rm m}$, which according to the property shown before is a sufficient condition for possible multiple images.

The most useful statement on multiple imaging applies to centrally condensed lenses; those are mass distributions where $\kappa(\theta)$ does not increase with θ, or $\kappa'(\theta) \le 0$ for $\theta \ge 0$. These mass profiles are the only relevant ones in

astrophysics. Centrally condensed lenses are capable of producing multiple images if, and only if, $\kappa(0) > 1$. Sufficiency was shown already. Necessity follows from this: if $\kappa(0) \leq 1$, then $\bar{\kappa} \leq 1$ for all θ; then, one finds that $\mathrm{d}\beta/\mathrm{d}\theta = (1 - \bar{\kappa}) - \theta\bar{\kappa}' > 0$, since $\bar{\kappa}$ is also a non-increasing function of θ. Another way to phrase the multiple image condition for centrally condensed lenses is $\mathrm{d}\alpha/\mathrm{d}\theta > 1$ at the origin.

3.2 The point-mass lens

Consider a point mass M or, equivalently, the outside region of a spherical mass distribution of total mass M; let the mass be located at the origin of the lens plane. Then the surface mass density is $\Sigma(\boldsymbol{\xi}) = M\delta_{\mathrm{D}}(\boldsymbol{\xi})$, and from (5) one finds for the deflection angle

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4GM}{c^2}\,\frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|^2}\,, \tag{44}$$

hence the amplitude of the deflection angle agrees with (2), and its direction is the same as that of $\boldsymbol{\xi}$, as expected from symmetry. Specializing (8) to the current lens model yields

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{4GMD_{\mathrm{ds}}}{c^2\,D_{\mathrm{d}}D_{\mathrm{s}}}\frac{\boldsymbol{\theta}}{|\boldsymbol{\theta}|^2} = \boldsymbol{\theta} - \theta_{\mathrm{E}}^2\frac{\boldsymbol{\theta}}{|\boldsymbol{\theta}|^2}\,, \tag{45}$$

where in the second step we have used the definition (43) of the Einstein angle which depends on the lens mass M and the distances to lens and source. If we choose without loss of generality the source position $\boldsymbol{\beta}$ to be on the positive β_1-axis, then $\boldsymbol{\theta}$ will also be on the θ_1 axis, and the lens equation becomes one-dimensional. Scaling the angles in terms of the Einstein angle as $y := \beta/\theta_{\mathrm{E}}$, $x := \theta/\theta_{\mathrm{E}}$, (45) becomes $y = x - 1/x$, with the two solutions

$$x_{\pm} = \frac{1}{2}\left(y \pm \sqrt{y^2+4}\right)\,, \tag{46}$$

i.e., one image on each side of the lens. Note that $x_+ \geq |x_-|$, hence the image on the same side of the lens as the source is further away from the lens than the other image. In the language of the previous section, $m(\theta) = \theta_{\mathrm{E}}^2$, $\bar{\kappa}(\theta) = (\theta_{\mathrm{E}}/\theta)^2 = x^{-2}$, so that we find from (41) the image magnification to be

$$\mu = \frac{1}{\det\mathcal{A}} = \frac{1}{1-\bar{\kappa}^2} = \left(1 - \frac{1}{x^4}\right)^{-1}\,. \tag{47}$$

As seen from (46), $x_+ \geq 1$, and so $\mu(x_+) \equiv \mu_+ \geq 1$. On the other hand, the magnification of the second image can be rather small if x_- becomes small. The magnification of the two images and the total magnification of the source is

$$\mu_{\pm} = \pm\frac{1}{4}\left[\frac{y}{\sqrt{y^2+4}} + \frac{\sqrt{y^2+4}}{y} \pm 2\right]\,, \quad \mu_{\mathrm{p}} = \mu_+ + |\mu_-| = \frac{y^2+2}{y\sqrt{y^2+4}}\,; \tag{48}$$

hence, unless $y \lesssim 1$, the secondary image will be strongly demagnified. The image separation

$$\Delta\theta = 2\theta_{\rm E}\sqrt{1+y^2/4} \gtrsim 2\theta_{\rm E} \tag{49}$$

is therefore only slightly larger than $2\theta_{\rm E}$ in relevant cases, since for values of $y \gtrsim 1$ the secondary image will be demagnified below the detection threshold. The sum of the two magnifications is $\mu_{\rm p} \approx 1.34$ for $y = 1$.

The magnification formally diverges for $x = 1$, or $\theta = \theta_{\rm E}$, which justifies using the same name as for the tangential critical curve in Sect. 3.1. But what about the odd-number theorem (see Sect. **??**)? Remember, for its validity the smoothness of the mass distribution was assumed, but a point-mass lens is not smooth; in particular, the deflection potential ψ has a logarithmic spike at the origin. In fact, one can easily picture where the third image has been 'lost': Assume one would smear out the mass M over a small but finite region (say in the shape of a Gaussian), the central surface mass density κ_0 would be very high but finite; in that case, there would be a maximum of the Fermat potential close to the center (the exact position depending on the location of the source), hence the third image would appear there. Its magnification $\mu_3 \approx (\kappa_0 - 1)^{-2} \ll 1$ would then be very small.

3.3 The singular isothermal sphere

A simple lens model which applies, at least to first order, to the lensing properties of galaxies and clusters is the so-called *singular isothermal sphere* (SIS). This spherical mass distribution yields flat rotation curves, such as are observed for spiral galaxies. Their density distribution is described by

$$\rho(r) = \frac{\sigma_v^2}{2\pi G r^2}\,. \tag{50}$$

Physically this model corresponds to a distribution of self-gravitating particles where the velocity distribution at all radii is a Maxwellian with one-dimensional velocity dispersion σ_v (hence, the term 'isothermal'). The three-dimensional velocity dispersion is $\sqrt{3}\sigma_v$, and the Keplerian rotation velocity (i.e. the velocity of particles on a circular orbit) is $v_{\rm c} = \sqrt{2}\sigma_v$.

The mass distribution (50) has two pathological properties: the central density diverges as $\rho \propto r^{-2}$ (hence the name 'singular'), and the total mass of this distribution diverges as $r \to \infty$. The former feature can be cured by introducing a finite core radius, whereas the distribution for large r does not affect the lensing properties at smaller radii. In SL and WL we shall discuss the constraints lensing provides on the core radius of galaxy and cluster lenses.

The SIS lens model. For the reasons just mentioned, the singular isothermal sphere is often used as a mass model for gravitational lenses; its surface

mass density $\Sigma(\xi)$ follows from projection of (50) along the line-of-sight,

$$\Sigma(\xi) = \int_{-\infty}^{\infty} \mathrm{d}r_3 \; \rho\left(\sqrt{\xi^2 + r_3^2}\right) = \frac{\sigma_v^2}{2G}\,\xi^{-1} \; . \tag{51}$$

As will be shown immediately, the Einstein radius of this lens model is

$$\theta_{\mathrm{E}} = 4\pi \left(\frac{\sigma_v}{c}\right)^2 \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}} \; , \tag{52}$$

in terms of which one obtains

$$\kappa(\theta) = \frac{\theta_{\mathrm{E}}}{2|\theta|} \; ; \quad \bar{\kappa}(\theta) = \frac{\theta_{\mathrm{E}}}{|\theta|} \; ; \quad |\gamma|(\theta) = \frac{\theta_{\mathrm{E}}}{2|\theta|} \; ; \quad \alpha(\theta) = \theta_{\mathrm{E}}\,\frac{\theta}{|\theta|} \; ; \tag{53}$$

note that the magnitude of α is constant. Here we use the same notation as introduced before (34). The fact that $\bar{\kappa}(\theta_{\mathrm{E}}) = 1$ shows that θ_{E} is the tangential critical curve of the SIS. The lens equation then reads

$$\beta = \theta - \theta_{\mathrm{E}}\,\frac{\theta}{|\theta|} \; , \quad \text{or} \quad y = x - \frac{x}{|x|} \; , \tag{54}$$

where the second form employs the scaled angles $x = \theta/\theta_{\mathrm{E}}$, $y = \beta/\theta_{\mathrm{E}}$. As before, we set $y \geq 0$; then, for $y < 1$, there are two images, at $x_+ = y + 1$ and $x_- = y - 1$, i.e., on opposite sides of the lens center, with image separation $\Delta\theta = 2\theta_{\mathrm{E}}$. For $y > 1$, only one image occurs, at $x_+ = y + 1$. x_+ corresponds to a minimum of the Fermat potential, whereas x_- to a saddle point, so that the subscripts denote the parity of the two images. The magnification can be calculated from (41), noting that $\bar{\kappa} = 2\kappa$, so that

$$\mu = \frac{1}{\det \mathcal{A}} = \frac{1}{1 - \bar{\kappa}} = \frac{|x|}{|x| - 1} \; ; \tag{55}$$

hence, since $x_+ > 1$, $\mu_+ > 1$, whereas the secondary image, with $|x_-| < 1$, can be strongly demagnified as $x_- \to 0$, or $y \to 1$. From (40) we find that $|\gamma(x)| = \kappa(x) = 1/(2x)$; thus, images are stretched in the tangential direction by a factor $|\mu|$, whereas the distortion factor in the radial direction is unity. The total magnification of a point source is $\mu_{\mathrm{p}} = 2/y$ for $y \leq 1$, and $(1+y)/y$ for $y \geq 1$.

Again, what about the odd-number theorem? As was true for the point-mass lens, the mass distribution of the SIS is not smooth, so the theorem does not apply. Another 'strange' property of the SIS is that the number of images changes by ± 1 when the source position crosses the circle $y = 1$ – this is in apparent conflict to what we said in Sect. 2.4. Both of these effects are due to the singular mass distribution as $\theta \to 0$, which causes $|\boldsymbol{\alpha}|$ to be constant. If we smoothed out the central mass singularity, by introducing a small but finite core, then the deflection angle would be constant, except very close to the center where it would make a smooth transition from $-\theta_{\mathrm{E}}$ for $\theta < 0$ to

$+\theta_{\rm E}$ for $\theta > 0$. In this transition region, there will be two points (at $\theta = \pm\theta_{\rm r}$) where $\mathrm{d}\alpha/\mathrm{d}\theta = 1$, corresponding to a radial critical curve. The corresponding caustic circle will have radius $\beta_{\rm r} \lesssim \theta_{\rm E}$. A source with $|\beta| < \beta_{\rm r}$ will have three images, one at $x \approx y+1$, one with $x \approx y-1$ and one in the inner core region, whereas a source with $|\beta| > \beta_{\rm r}$ has just one. The lens equation maps the small circle with radius $\theta_{\rm r}$ onto the circle $\beta_{\rm r} \approx \theta_{\rm E}$. When we now let the core radius go to zero, $\beta_{\rm r} \to \theta_{\rm E}$, the magnification of the central image $\mu_3 \to 0$, and the central region of the lens that is mapped onto $\beta_{\rm r} \approx \theta_{\rm E}$ decreases to zero area. Hence, this limit process suggests that one can consider the third image to be present, located at $\theta = 0$, and having zero magnification.

3.4 Non-symmetric lenses

To describe the mass profile of real lenses, more complicated (and realistic) radial mass profiles can be used; even though the lens equation may no longer be analytically solvable, the fact that it is one-dimensional renders numerical investigations simple. The qualitative features of (centrally-condensed) axisymmetric lenses do not depend strongly on the details of the radial profile and can basically be read-off from the corresponding Young diagram (see Sect. 3 of SL).

Breaking the symmetry leads to qualitatively new properties of the lens. Most obvious of them, the central caustic point gets unfolded into a curve of finite size; a source situated inside this curve can then have five images. The fact that many of the observed lens systems have four images (i.e. five minus the one being invisible probably due to very strong demagnification at the center) shows that the axisymmetric models are definitely not sufficient to explain them.

The next more complicated gravitational potential is then one with two lines of symmetry, such as an ellipse has. Hence, one would be tempted to consider mass distributions where κ is constant on (confocal) ellipses. In fact, Bourassa et al. (1973), Bourassa & Kantowski (1975), and later Schramm (1990) have considered the lensing properties of such elliptical lenses – they turn out to be fairly complicated analytically in general; nevertheless, for some of the most relevant radial density profiles, explicit expressions for the deflection angle can be derived (e.g., Kormann et al. 1994; Keeton & Kochanek 1998; see SL), and such elliptical mass models are generally used for fitting observed lens systems. Here we consider a simpler class of lens models with similar symmetry, namely axisymmetric matter distributions with an external perturbation, henceforth called 'quadrupole lenses'.

Quadrupole lenses. Even if the mass distribution of a lens is axisymmetric (like that of a star), the corresponding gravitational potential is not expected to share this symmetry, because lenses are typically not isolated: a galaxy is often situated inside or near a group of galaxies, and the other member galaxies, and the dark-matter halo of the group, will perturb the symmetry of the

potential. In many cases of astrophysical interest, like the one just mentioned, the perturbing gravitational field changes very little over the relevant length scale of the main lens. As an example, consider a lens galaxy in a cluster of galaxies. The relevant length scale of the galaxy is about the region where the multiple images occur, i.e. a region with radius of the Einstein radius, or typically $1''$. In contrast, the relevant length scale of the cluster perturbation is either the separation of the galaxy from the cluster center, or the Einstein radius of the cluster, whatever is larger, and thus typically much larger than $1''$. It is thus natural to expand the deflection potential of the perturber about the center of the main deflector; the lowest-order, non-trivial term in the expansion is the quadratic term (tidal field). The analogous situation occurs for a star in a distant galaxy, where the symmetry of the point-mass lens is broken by the tidal field of its host galaxy; this is the situation considered by Chang & Refsdal (1984; see ML).

Here we study the lens action of an axisymmetric matter distribution which is perturbed by a larger-scale gravitational field, and we assume that the latter is locally (that is, over the region where we want to study the lensing properties of the main deflector) well described by its second-order Taylor approximation. Choosing the origin to be the center of the main lens, and the orientation of the coordinate system such that the Hessian of the deflection potential of the perturber (or tidal matrix) is diagonal at the origin, then the deflection caused by the perturber can be written as

$$\boldsymbol{\alpha}_{\rm p}(\boldsymbol{\theta}) = \boldsymbol{\alpha}_{\rm p}(\mathbf{0}) + \begin{pmatrix} \kappa_{\rm p}+\gamma_{\rm p} & 0 \\ 0 & \kappa_{\rm p}-\gamma_{\rm p} \end{pmatrix} \boldsymbol{\theta} \ , \tag{56}$$

where the surface mass density and shear of the perturber are labeled with subscript 'p'. Note that the strength of the perturbation is not assumed to be small. The lens equation then reads

$$\boldsymbol{\beta} = [1-\bar{\kappa}(|\boldsymbol{\theta}|)]\,\boldsymbol{\theta} - \begin{pmatrix} \kappa_{\rm p}+\gamma_{\rm p} & 0 \\ 0 & \kappa_{\rm p}-\gamma_{\rm p} \end{pmatrix} \boldsymbol{\theta} \ , \tag{57}$$

where we have translated the origin in the source plane by the vector $\boldsymbol{\alpha}_{\rm p}(\mathbf{0})$. The perturber thus adds a uniform sheet of matter plus an *external shear*. The uniform sheet can be transformed away, recalling our discussion of the mass-sheet degeneracy in Sect. 2.5; indeed, (57) can be rewritten as

$$\hat{\boldsymbol{\beta}} := \frac{\boldsymbol{\beta}}{1-\kappa_{\rm p}} = \begin{pmatrix} 1-g_{\rm p} & 0 \\ 0 & 1+g_{\rm p} \end{pmatrix} \boldsymbol{\theta} - \hat{\bar{\kappa}}(|\boldsymbol{\theta}|)\boldsymbol{\theta} \ , \tag{58}$$

where $g_{\rm p} = \gamma_{\rm p}/(1-\kappa_{\rm p})$ is the reduced shear of the perturber, $\hat{\boldsymbol{\beta}}$ the rescaled source coordinate, and $\hat{\bar{\kappa}}(|\boldsymbol{\theta}|) = \bar{\kappa}(|\boldsymbol{\theta}|)/(1-\kappa_{\rm p})$ the rescaled surface mass density. We shall in the following discard the hats on the variables in (58).

Although an axisymmetric lens with an external shear is too simple to represent real lenses, the resulting lens equation is sufficiently simple to allow

some analytical progress; for didactical purposes, we shall discuss this lens model in somewhat more detail. The lens equation now is two-dimensional, and therefore more complicated to invert (i.e., to find all image positions for a given source position) than in the axisymmetric case. However, the lens equation can be recast into a one-dimensional equation, by introducing polar coordinates $\boldsymbol{\theta} = \theta(\cos\varphi, \sin\varphi)$ in the lens plane; then, (58) can be written as

$$\cos\varphi = \frac{\beta_1}{\theta\left[1-\bar\kappa(\theta)-g_\mathrm{p}\right]} \quad , \quad \sin\varphi = \frac{\beta_2}{\theta\left[1-\bar\kappa(\theta)+g_\mathrm{p}\right]} \, , \tag{59}$$

and by adding the squares of these two equations,

$$\theta^2\left[(1-\bar\kappa)^2 - g_\mathrm{p}^2\right]^2 - \beta_1^2\left(1-\bar\kappa+g_\mathrm{p}\right)^2 - \beta_2^2\left(1-\bar\kappa-g_\mathrm{p}\right)^2 = 0 \, , \tag{60}$$

the polar angle φ has been eliminated: (60) is an equation for θ only and can be solved numerically. For each solution θ, the polar angle can be calculated from (59). Not all solutions will have $|\cos\varphi| \le 1$ and $|\sin\varphi| \le 1$; those solutions θ have been generated by a number of algebraic manipulations needed to arrive at (60) and thus shall be discarded then. We just saw a nice example of reducing the effective dimension of a problem to make it more tractable.

The Jacobian for the quadrupole lens can be obtained from its definition (17), and its determinant reads

$$\det\mathcal{A} = (1-\bar\kappa)^2 - g_\mathrm{p}^2 - \theta\bar\kappa'\left(1-\bar\kappa+g_\mathrm{p}\cos 2\varphi\right) \, , \tag{61}$$

so that the critical curves can be easily calculated: for each value of θ, the condition $\det\mathcal{A} = 0$ yields a value for $\cos 2\varphi$; if this lies between ± 1, one has found a pair (θ, φ) of coordinates on the critical curve; in fact, one has obtained four different critical points, one in each quadrant of the lens plane, due to the symmetry of our lens model with respect to both reflections $(\theta_1, \theta_2) \mapsto \pm(\pm\theta_1, \pm\theta_2)$. Hence, the structure of critical curves and caustics for quadrupole lenses can be easily investigated, at least numerically.

The non-singular isothermal sphere with external shear. We now consider a specific example of a quadrupole lens which has frequently been used in lens modeling: the perturbed non-singular isothermal sphere, for which

$$\kappa(\theta) = \frac{\theta_\mathrm{e}}{\theta_\mathrm{c}}\left(1+\frac{\theta^2}{2\theta_\mathrm{c}^2}\right)\left(1+\frac{\theta^2}{\theta_\mathrm{c}^2}\right)^{-3/2} \quad , \quad \bar\kappa(\theta) = \frac{\theta_\mathrm{e}}{\sqrt{\theta^2+\theta_\mathrm{c}^2}} \, , \tag{62}$$

so that $\kappa(0) = \theta_\mathrm{e}/\theta_\mathrm{c}$ is finite. The complex form of $\kappa(\theta)$ is chosen so that the deflection profile is simple. We note that for $\theta \gg \theta_\mathrm{c}$, the mass distribution approaches that of an SIS with Einstein angle θ_e, but for a finite θ_c, θ_e is *not*

the location of the critical curve, but in general, $\theta_{\rm E} = \left(\theta_{\rm e}^2 - \theta_{\rm c}^2\right)^{1/2}$ for $\theta_{\rm c} < \theta_{\rm e}$; otherwise, the lens is not critical. For this lens model, the one-dimensional form (60) of the lens equation can be even further simplified, by noting that (62) implies $\theta^2 = \theta_{\rm e}^2/\bar{\kappa}^2 - \theta_{\rm c}^2$. Inserting this expression into (60), one obtains after multiplying by $\bar{\kappa}^2$ an equation which is a sixth-order polynomial in $\bar{\kappa}$. Given that standard methods are known (e.g., Press et al. 1992) to find all solutions of polynomials, this latter form is much more useful; the roots of this polynomial are potential solution if they are real, and have $0 < \bar{\kappa} \leq \theta_{\rm e}/\theta_{\rm c}$; those solutions can then be inserted into the original lens equation to check whether they are actual solutions.

The critical curves are found from (61), where now the relation $\theta\bar{\kappa}' = -\theta^2\bar{\kappa}^3/\theta_{\rm e}^2$ can be used to replace $\bar{\kappa}'$ there. The equation $\det \mathcal{A} = 0$ can then be written in the form $\cos 2\varphi = f(\bar{\kappa})$, where f is a function of $\bar{\kappa}$ only. Thus for all radial coordinates $\theta > 0$, or equivalently, for all $0 < \bar{\kappa} \leq \theta_{\rm e}/\theta_{\rm c}$ one can determine $f(\bar{\kappa})$; a value with $|f(\bar{\kappa})| \leq 1$ yields four critical points at the radius corresponding to this mean surface mass density.

The critical curves of this special lens ("NIS plus external shear") can be studied analytically; because of its importance for understanding lens geometry, we shall provide a detailed description of the essential features in the following and illustrate the results in Fig. 13. First to note is that if the core radius is too large, or the central surface mass density too low, there is no critical curve. One finds that $\det \mathcal{A}(\mathbf{0}) = (1-\theta_{\rm e}/\theta_{\rm c})^2 - g_{\rm p}^2$ at the center, and can show that for $x_{\rm c} \equiv \theta_{\rm c}/\theta_{\rm e} > (1-g_{\rm p})^{-1}$ no critical curves exist. In other words, for $\kappa_0 = 1/x_{\rm c} < 1 - g_{\rm p}$ the lens is not critical. Compare this condition with the one for an unperturbed lens ($g_{\rm p} = 0$); there, in order for the lens to become critical, κ_0 must be larger than unity. If the core radius satisfies $(1+g_{\rm p})^{-1} < x_{\rm c} < (1-g_{\rm p})^{-1}$, there is a single closed critical curve (see upper left panel in Fig. 13), and the corresponding caustic has two cusps. Owing to the shape of the caustic curve, one often calls it a lips caustic. A source located inside the caustic has three images, whereas one outside has a single image.

At $x_{\rm c} = (1+g_{\rm p})^{-1}$, the Jacobian vanishes again at the origin, and for smaller values of the core radius, $x_{\rm c} < (1+g_{\rm p})^{-1}$, there are two critical curves and caustics, as seen in the upper right panel of Fig. 13. A second lips caustic is located inside the first one, oriented perpendicular to it. Sources inside both caustics now have five images, and those inside the outer one but outside the inner one have three. When the core radius is further decreased, the two critical curves approach each other at two points, and correspondingly, the cusps of the inner lips caustic approach the outer caustic. At $x_{\rm c} = (1-g_{\rm p})^{1/2}(1+g_{\rm p})^{-3/2}$, the critical curves and cusps merge, and for smaller values of $x_{\rm c}$, there are again two separate critical curves and two caustics, but now, as shown in the lower left panel of Fig. 13, one of the caus-

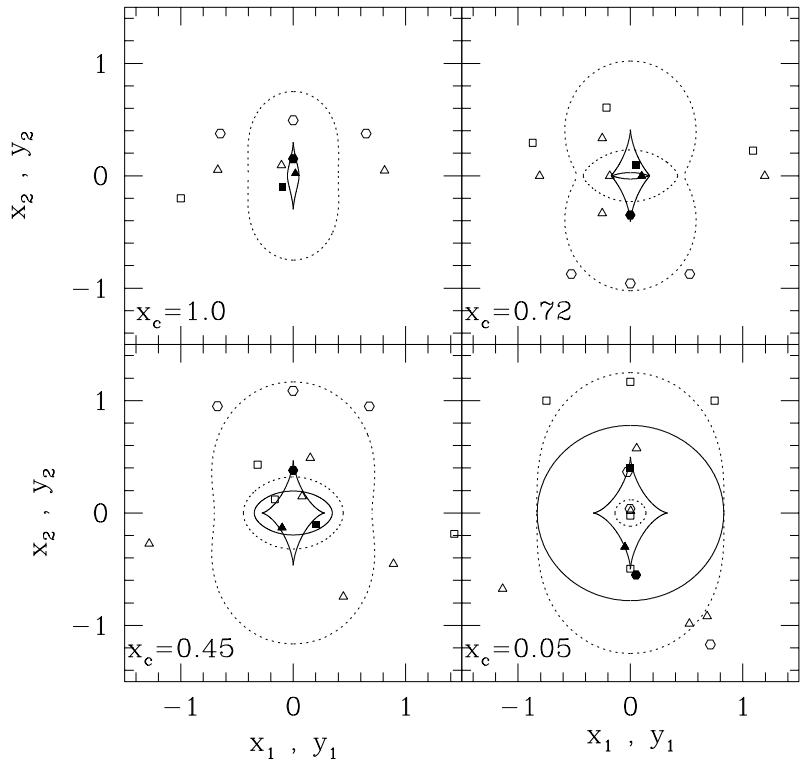


Fig. 13. Critical curves (dashed) and caustics (solid curves) for a non-singular isothermal sphere model with external shear. Angles in the source and lens plane have been scaled by $\theta_{\rm e}$, i.e., $\boldsymbol{\theta} = \theta_{\rm e}\mathbf{x}$, $\boldsymbol{\beta} = \theta_{\rm e}\mathbf{y}$. In all panels, the reduced shear is $g_{\rm p} = 0.2$. The four possible configurations are shown: for $x_{\rm c} > (1-g_{\rm p})^{-1}$, no critical curve exists; panel (a) shows a case with $(1+g_{\rm p})^{-1} < x_{\rm c} < (1-g_{\rm p})^{-1}$, for which a single critical curve exists, created from the previous case through a lips catastrophe. In panel (b), the case $(1-g_{\rm p})^{1/2}(1+g_{\rm p})^{-3/2} < x_{\rm c} < (1+g_{\rm p})^{-1}$ is shown, for which two critical curves exist, the second one created from the previous case by another lips catastrophe. The two corresponding caustics have two cusps each. In panel (c), $x_{\rm c} < (1-g_{\rm p})^{1/2}(1+g_{\rm p})^{-3/2}$; there, one caustic with four cusps, and one caustic without cusps occur. This case is obtained from the previous one through two hyperbolic umbilics where the two cusps of the inner caustic in (b) were transferred to the outer caustic; correspondingly, at that point the two critical curves intersect on the x_1-axis in this transition. Finally, panel (d) shows the same case as panel (c), except that now the two 'naked cusps' – the cusps outside the other caustic – are inside the other caustic. In addition, in all panels we have positioned three sources, indicated by the filled triangle, the filled square and the filled hexagon, together with their corresponding images, shown with the corresponding open symbols

tics has four cusps, the other has none.[7] Two of these cusps lie inside the other caustic, the other two fall outside of it; these are called 'naked cusps'.[8] The caustic with the cusps, also called the astroid or tangential caustic, corresponds to the outer critical curve, the one without cusp to the inner or radial critical curve. Decreasing x_c further, the inner critical curve decreases in size, whereas the corresponding caustic increases and finally completely encompasses the astroid caustic (Fig. 13, lower right panel).

In the figure, we have also illustrated the image locations (open symbols) for several source positions (filled symbols). As will be discussed in SL, many of these image configurations have actually been observed. For galaxy lensing, configurations of the type shown in the lower right panel are most relevant since galaxies seem to have a small core radius. In that case, one of the images is located very close to the center of the lens where κ is much larger than unity, and therefore the magnification is very small – which is the canonical explanation for the absence of an observed odd image. Furthermore we see that in this case, the characteristic maximum image separation is $\Delta\theta \sim 2\theta_e$, as expected. A source close to and inside a cusp produces three (highly magnified) images lying close together near the corresponding critical curve; in the case of a naked cusp, these are the only images of the source, whereas if the cusp lies inside the other caustic, two additional images are formed (one of which may be highly demagnified).

If the core radius is decreased to zero, the inner critical curve shrinks to zero size, the corresponding caustic becomes a circle with radius θ_e, and the number of images changes by ± 1 when a source crosses this curve – the behavior is identical to the one already encountered in the discussion of the SIS model. The other critical curve attains a simple parametric form,

$$\cos(2\varphi) = \frac{1-\left(1-g_p^2\right)x}{g_p} \quad \Longleftrightarrow \quad x = \frac{1-g_p\cos(2\varphi)}{1-g_p^2}\,, \tag{63}$$

describing a single closed curve around the origin, which is in fact an ellipse. By inserting this parameterized form into the lens equation and eliminating

[7] We have discussed folds and cusps before; one can show from singularity theory that these are the only two 'generic' singularities that occur in a lens mapping. However, if one considers a family of lens models, such as done here by varying x_c, higher-order singularities can occur. At the corresponding values of the lens model parameter – here x_c – the topology of critical curves can change. Examples are the creation of lips singularities, or the 'exchange of cusps' just mentioned, which technically speaking corresponds to a hyperbolic umbilic. Chap. 6 of SEF provides a general description of singularities and their metamorphoses in lens mappings.

[8] Lensing geometries where a source is located inside a naked cusp, producing three bright images, are probably seen in clusters, indicating a relatively large core radius for them.

the parameter φ results in the equation describing the astroid caustic,

$$\left[2y_1^2\left(\frac{1+g_\mathrm{p}}{g_\mathrm{p}}\right)^2\right]^{1/3}+\left[2y_2^2\left(\frac{1-g_\mathrm{p}}{g_\mathrm{p}}\right)^2\right]^{1/3}=2\,, \tag{64}$$

from which the locations of the cusps can be read off. In particular, (64) shows that the size of the astroid caustic increases with increasing g_p and, to first order, its linear size is $\propto g_\mathrm{p}$.

In the limit $g_\mathrm{p}\to 0$, the two critical curves become circles, with the outer (inner) one being the tangential (radial) critical curves. Because of that, one often uses the same names for the critical curves also in the perturbed case $g_\mathrm{p}\neq 0$. The radial caustic then separates the three-image region from the single-image region in the source plane, and the tangential caustic degenerates into a single point. The fact that this point unfolds in the presence of a perturbation is nicely illustrated by (64).

General discussion of 'elliptical' lenses. Mass distributions with elliptical isodensity contours are needed to realistically model gravitational lens systems. Although such models are considerably more difficult to handle analytically, their qualitative properties are similar to the NIS with external shear that was discussed above. In particular, the evolution of the critical curves and the caustics as a function of 'lens strength' or 'core size' for these models is the same as that shown in Fig. 13; the same is true for the properties regarding multiple imaging. Of course that does not mean that the choice of the lens model is arbitrary: for systems with sufficiently detailed observational constraints, a quantitative modeling technique can distinguish between the various classes of models; as we shall see in SL, the NIS with external shear is often too simple; many lens systems require an elliptical mass distribution plus some external shear in addition.

4 The cosmological standard model I: The homogeneous Universe

We assume that the reader is familiar with the basic concepts of standard cosmology, such as the hot Big Bang occurring some 13.7 billion years ago, after which the Universes expanded and cooled down. During this expansion, the simplest atomic nuclei, predominantly helium, were formed about a minute after the Big Bang, and some 370 000 years later, the Universe became neutral and released a thermal radiation that is still visible today, the Cosmic Microwave Background radiation with a temperature of 2.73 K. In addition, it is assumed that you are aware of the existence of dark matter, material that reveals itself only through gravity, like in governing the rotation curves of the Milky Way and other spiral galaxies or in providing the deep potential

wells of clusters of galaxies which can keep very hot X-ray emitting plasma and fastly moving galaxies gravitationally bound. Therefore, we shall only briefly summarize those relations which will be used later in this course. Excellent textbooks on cosmology are available, among them are Kolb & Turner (1990), Peacock (1999), Padmanabhan (1993), Peebles (1993) and Liddle & Lyth (2000).

4.1 The cosmic expansion

Metric and coordinates. Observational evidence suggests that the Universe around us, when averaged over large angles, is isotropic (the Cosmic Microwave Background, or CMB; the faint galaxy distribution, etc.). Furthermore, if we assume that our location in the Universe is not special, the same property also holds for other observers: also for them the Universe should appear isotropic when averaged over large scales. Together, this implies that the Universe is spatially homogeneous and isotropic around every point. Is has been shown by Robertson and Walker that for such a spacetime, the metric can be written in the form

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - a^2(t)\left[\mathrm{d}w^2 + f_K^2(w)\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)\right]\ , \tag{65}$$

where t is the cosmic time [which agrees with the time measured by comoving observers, i.e. those with constant (w,θ,φ)], $a(t)$ the cosmic scale factor, normalized that today, $a(t_0) = 1$, w the comoving radial coordinate, θ and φ are the angular coordinates on a unit sphere, and $f_K(w)$ the comoving angular diameter distance, which depends on the curvature parameter K in the following way:

$$\begin{aligned} f_K(w) &= |K|^{-1/2}\,\mathrm{sinn}\left(|K|^{1/2}w\right) \\ &\equiv \begin{cases} K^{-1/2}\sin(K^{1/2}w) & (K>0) \\ w & (K=0) \\ (-K)^{-1/2}\sinh[(-K)^{1/2}w] & (K<0) \end{cases} . \end{aligned} \tag{66}$$

Hence, (w,θ,φ) are spherical coordinates in a three-dimensional space of constant curvature K. Radiation from a comoving source emitted at time t_2 and received by a comoving observer at time $t_1 > t_2$ is *redshifted* by a factor $1 + z_{12} = a(t_1)/a(t_2)$.

Expansion equation. Inserting the metric (65) into Einstein's field equation of General Relativity shows that the matter contents must be that of a (homogeneous) perfect fluid with density $\rho(t)$ and pressure $p(t)$. The components of the field equation reduce to two independent dynamical equations for the scale factor $a(t)$,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2} + \frac{\Lambda}{3} \tag{67}$$

and

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda}{3}\,. \tag{68}$$

Equation (67) is called *Friedmann's equation* (Friedmann 1922). The two equations (67) and (68) can be combined to yield the *adiabatic equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3(t)\rho(t)c^2\right] + p(t)\frac{\mathrm{d}a^3(t)}{\mathrm{d}t} = 0\,, \tag{69}$$

which has the following intuitive interpretation: the first term $a^3\rho$ is proportional to the energy contained in a fixed comoving volume, and hence the equation states that the change in 'internal' energy equals the pressure times the change in proper volume. Hence, (69) expresses the first law of thermodynamics in the cosmological context. The parameter Λ in (67) and (68) is the cosmological constant; Einstein introduced it into his field equation (in 1916) because without it, no static model of the Universe would be predicted by General Relativity – the Hubble expansion of the Universe was discovered only a decade later, after which Einstein dismissed this term. In recent years, the cosmological constant has regained great popularity – because, as will be discussed later, there is strong evidence in favor of $\Lambda \neq 0$. On the other hand, the interpretation of Λ has also changed, as we shall see.

World models for which the metric is given by (65) and where the scale factor $a(t)$ obeys Friedmann's equation (67) and the adiabatic equation (69) are called Friedmann-Lemaître models. It should be noted that eq. (68) can also be derived from Newtonian gravity except for the pressure term and the cosmological constant. Unlike in Newtonian theory, pressure acts as a source of gravity in General Relativity.

Matter models. By themselves, these equations do not specify the expansion history $a(t)$; for this we have to add an equation of state (EOS). In general, matter components cannot be described by a simple equation of the form $p = p(\rho)$; however, for some limiting cases an equation of this form does exist. Fortunately, the matter contents in our Universe seems to be such that over most of its history it can be described by a few components, each of which having such a simple EOS.

If the constituents of matter have random (thermal) velocities much smaller than c, $p \ll \rho c^2$, then the pressure of this component can be neglected in the expansion equation; this kind of matter is approximated by $p = 0$ and called 'dust' (or simply 'matter'). For $p = 0$, (69) yields that

$$\rho_{\mathrm{m}} \propto a^{-3}\,, \tag{70}$$

a result that is intuitively clear: as the physical (or proper) volume of a fixed comoving volume behaves like $V \propto a^3$, and the number of matter particles is conserved, their number density, and thus mass density must decrease as $\rho \propto V^{-1} \propto a^{-3}$.

In the other limiting case, where the constituents of matter have a random velocity close to c (or even c, as must be the case for massless particles, like photons), one has $p = \rho c^2/3$. For obvious reasons, matter with this EOS is called 'radiation'. From (69) one then finds that the energy density of radiation evolves as

$$\rho_{\rm r} \propto a^{-4} \,, \tag{71}$$

a result that can also be easily understood. Whereas the number density of photons (assuming that they constitute the 'radiation') decreases as a^{-3}, again due to number conservation, their individual energy decreases as a^{-1}, owing to the redshift of their energy (or adiabatic decompression).

Finally, there may be a mass component which can be interpreted as the energy density of the vacuum, assumed to be a constant in time. If the density is independent of a, then (69) predicts that $p_{\rm v} = -\rho_{\rm v} c^2$ for this matter component.

The matter density and pressure of the Universe is then given by the sum of these three components,

$$\rho = \rho_{\rm m} + \rho_{\rm r} + \rho_{\rm v} = \frac{\rho_{\rm m0}}{a^3} + \frac{\rho_{\rm r0}}{a^4} + \rho_{\rm v} \,, \quad p = \frac{\rho_{\rm r} c^2}{3} - \rho_{\rm v} c^2 = \frac{\rho_{\rm r0} c^2}{3a^4} - \rho_{\rm v} c^2 \,, \tag{72}$$

where the additional index '0' indicates that these are the values at present time. Inserting these expressions into (67) and (68) (setting $\Lambda = 0$ in these equations) shows that a term of the same form as the Λ-term appears; the cosmological constant can therefore be interpreted as a vacuum energy density.

Cosmological parameters. The ratio

$$H(t) = \dot{a}\, a^{-1} \tag{73}$$

is the expansion rate of the Universe, and its current value H_0 is called *Hubble constant*. This is the ratio of recession velocity to the distance of objects in the nearby Universe, and has the value

$$H_0 \approx 3.2 \times 10^{-18}\, h\, {\rm s}^{-1} \approx 1.0 \times 10^{-10}\, h\, {\rm yr}^{-1} \,, \tag{74}$$

where h parameterizes our lack of knowledge on the exact value of H_0; the currently best estimates yield $h \approx 0.72$ (see Sect. 6.3). If $\Lambda = 0$ and the spatial curvature vanishes, $K = 0$, then the current density of the Universe is directly related to H_0, as seen from (67); this density is called *critical density*,

$$\rho_{\rm cr} := \frac{3H_0^2}{8\pi G} \approx 1.9 \times 10^{-29}\, h^2\, {\rm g\,cm}^{-3} \,. \tag{75}$$

This characteristic density is used to scale the matter densities by defining the *density parameters*

$$\Omega_{\rm m} := \frac{\rho_{\rm m0}}{\rho_{\rm cr}} \,; \quad \Omega_{\rm r} := \frac{\rho_{\rm r0}}{\rho_{\rm cr}} \,; \quad \Omega_\Lambda := \frac{\rho_{\rm v}}{\rho_{\rm cr}} = \frac{\Lambda}{3H_0^2} \,. \tag{76}$$

The radiation density in the Universe is fairly well known: it is dominated by the energy density of the cosmic microwave background (CMB) which has a Planck spectrum with temperature of $T_{\rm CMB} \approx 2.73\,{\rm K}$ and whose energy density can be calculated from the Stefan–Boltzmann law to be

$$\rho_{\rm CMB} = \frac{1}{c^2}\,\frac{\pi^2}{15}\,\frac{(kT_{\rm CMB})^4}{(\hbar c)^3} \approx 4.5\times 10^{-34}\,{\rm g\,cm^{-3}} \;\rightarrow\; \Omega_{\rm CMB} = 2.4\times 10^{-5}\,h^{-2}\;. \tag{77}$$

In addition, from the era just before the primordial nucleosynthesis took place, there is a relic background of neutrinos, at a temperature corresponding to $T_\nu = (4/11)^{1/3} T_{\rm CMB} \approx 1.95\,{\rm K}$; their number density today is $113\,{\rm cm^{-3}}$ per species. If neutrinos were massless, they would contribute to the radiation density, and with three neutrino families one would have $\Omega_{\rm r} \approx 1.68\Omega_{\rm CMB}$. If neutrinos have a small, but finite rest mass above $\sim 10^{-4}\,{\rm eV}$ they would be non-relativistic today, and contribute to the matter density instead. However, at earlier epochs the neutrinos were relativistic and therefore contributed to the radiation energy density. In any case, for the present epoch we can completely neglect the influence of the radiation on the expansion rate. This, of course, was not always the case; since the radiation density drops as a^{-4}, whereas the matter density only as a^{-3}, there was an epoch (or a scale factor) when both were equal, namely at

$$a_{\rm eq} = \frac{\Omega_{\rm r}}{\Omega_{\rm m}} = 3.2\times 10^{-5}\Omega_{\rm m}^{-1}\,h^{-2}\;, \tag{78}$$

and we used $\Omega_{\rm r} \approx 1.68\Omega_{\rm CMB}$ here, since at $a_{\rm eq}$, the neutrinos were relativistic. For scale factors $a \lesssim a_{\rm eq}$, radiation was the dominant component in the Universe.

Making use of (72) and the definitions (76) of the density parameters, the expansion equation (67) becomes

$$H^2 = H_0^2\left[\frac{\Omega_{\rm r}}{a^4} + \frac{\Omega_{\rm m}}{a^3} - \frac{Kc^2}{a^2\,H_0^2} + \Omega_\Lambda\right]\;. \tag{79}$$

Specializing this to the current epoch, $a = 1$, yields an expression for the curvature, $K = (\Omega_{\rm m} + \Omega_\Lambda - 1)H_0^2/c^2$ (where we used $\Omega_{\rm r} \ll \Omega_{\rm m}$), which can be inserted into (80) to yield

$$H^2 = H_0^2\left[\Omega_{\rm r}a^{-4} + \Omega_{\rm m}a^{-3} + (1-\Omega_0)a^{-2} + \Omega_\Lambda\right]\;, \tag{80}$$

where we defined

$$\Omega_0 = \Omega_{\rm m} + \Omega_\Lambda + \Omega_{\rm r} \tag{81}$$

as the total density parameter of the present-day Universe. One sees that the sign of $\Omega_0 - 1$ agrees with that of K, so that the total matter density determines the spatial curvature of the Universe. Note that (80) is a first-order differential equation for $a(t)$, which can be integrated (numerically, if

necessary) with the boundary condition $a(t_0) = 1$. The general discussion of the qualitative behavior of $a(t)$ (see, e.g., Peacock 1999, Sect. 3.2; also Fig. 19 below) yields the following results: The scale factor $a(t)$ is a monotonically increasing function for $a < 1$; hence, a decreases monotonically as we go backwards in time. Whereas in principle it is possible that a does not decrease below a finite positive value (so-called bouncing Universes), we happen not to live in one – such models predict a minimum a, and therefore a maximum redshift; the fact that we have discovered sources at redshift $z \gtrsim 6$, coupled with a matter density that certainly exceeds $\Omega_\mathrm{m} > 0.05$ excludes the possibility that our Universe is of that kind. Hence, formally $a \to 0$ as we go into the past, at a finite instant. This instant is called the Big Bang, an event when the Universe was extremely dense and hot. The behavior of $a(t)$ in the future depends on the values of the density parameters. If $\Omega_\Lambda = 0$, then $a(t)$ will continue to grow provided $\Omega_\mathrm{m} \leq 1$, otherwise it will reach a maximum value of a and then recollapse. If $\Omega_\Lambda > 0$, the threshold value of Ω_m for recollapse is slightly changed. Flat models, i.e. those with $\Omega_\mathrm{m} + \Omega_\Lambda = 1$ expand forever provided $\Omega_\mathrm{m} \leq 1$.

Defining $t = 0$ to be the instant of the Big Bang when $a = 0$, the cosmic time as a function of scale factor can be calculated from (80), since $\mathrm{d}t = \mathrm{d}a\,\dot{a}^{-1} = \mathrm{d}a(aH)^{-1}$; ignoring Ω_r (which is important only over a very brief period at the beginning of the expansion), one has

$$t(a) = \frac{1}{H_0} \int_0^a \mathrm{d}a' \, \left[a'^{-1}\Omega_\mathrm{m} + (1 - \Omega_\mathrm{m} - \Omega_\Lambda) + a'^2\Omega_\Lambda\right]^{-1/2} \; ; \qquad (82)$$

in particular, t_0 is obtained by setting $a = 1$. Apart from a numerical factor which depends on the density parameters, this yields $t_0 \sim H_0^{-1}$. Equation (82) can be inverted to yield $a(t)$.

Light propagates along null geodesics; in the coordinate system used to define the metric (65), it is easy to show from symmetry arguments that radial null curves (i.e. those with $\theta = \mathrm{const.}$, $\varphi = \mathrm{const.}$) are geodesics; for them $c\,\mathrm{d}t = -a\,\mathrm{d}w$, if we choose our location at $w = 0$. The minus sign occurs since photons propagating to us have $\mathrm{d}t > 0$ but $\mathrm{d}w < 0$. Light from a source that we observe today was emitted at a time obtained from integrating $c\,\mathrm{d}t = -a\,\mathrm{d}w$; every observation of the distance Universe is inevitably a look into the past.

We therefore have a number of variables which can be used to describe the location of a source: its comoving distance w, the time t at which the light was emitted which we observe today from that source, the scale factor a at this time or, equivalently, the redshift $z = a^{-1} - 1$, and the temperature of the Universe (which is defined as the temperature of the microwave background radiation – note that cosmic expansion evolves a blackbody into a blackbody, with temperature $T \propto a^{-1}$). These variables are related to each other by

$$\mathrm{d}t = \frac{\mathrm{d}a}{\dot{a}} = \frac{\mathrm{d}a}{a\,H} \; ; \quad -\mathrm{d}w = \frac{c\,\mathrm{d}t}{a} = \frac{c\,\mathrm{d}a}{a\,\dot{a}} = \frac{c\,\mathrm{d}a}{a^2 H} \; . \qquad (83)$$

4.2 Distances and volumes

The meaning of distance. Which of these descriptions of the location of a source is the 'correct distance'? Well, wrong question. This question is based on the Euclidean preconception that there is a uniquely defined correct distance, and that this is the outcome of all (correct) methods to measure the distance. However, in a general spacetime, two complications occur. The harmless one is that space may be curved. The more important one is that any observation measures distances not at a given instant of time, but along the backward light cone, and distances change in time as the Universe expands. There is not a unique meaning of distance. Nevertheless, one can construct methods on how to measure distance, and define distances according to these measurement procedures. The two most important definitions of distance are described next.

Distance measures. Suppose one knows the physical diameter $2R$ of a source at redshift z (or scale factor a) which is observed to have an angular diameter of δ. In Euclidean space one would then measure the distance to this source to be $D = 2R/\delta$; accordingly, one defines the angular diameter distance as exactly this ratio,

$$D_{\rm ang}(z) = 2R/\delta = a(z)\, f_K(w) \;, \tag{84}$$

where the final expression follows from the metric by setting $\delta = {\rm d}\theta$ and ${\rm d}s = 2R$. In the foregoing expression, w is to be understood as a function of redshift; the corresponding relation can be obtained from (83) to be

$$\begin{aligned} w(z_1, z_2) &= \frac{c}{H_0} \int_{a(z_2)}^{a(z_1)} \left[a\Omega_{\rm m} + a^2(1-\Omega_{\rm m}-\Omega_\Lambda) + a^4\Omega_\Lambda\right]^{-1/2} {\rm d}a \\ &= w(z_2) - w(z_1) \;, \end{aligned} \tag{85}$$

which is the *comoving distance* between two sources that we see to have redshifts $z_1 < z_2$, and we set $w(z) \equiv w(0,z)$. The comoving distance can be interpreted as the spatial distance between the intersections of the worldline of these two comoving sources with the spatial hypersurface $t = t_0$ – cf. the definition of comoving coordinates. Generalizing (84), we can define the angular-diameter distance $D_{\rm ang}(z_1, z_2)$ of a source at redshift z_2 seen by an observer at redshift $z_1 < z_2$ as

$$D_{\rm ang}(z_1, z_2) = a(z_2)\, f_K[w(z_1, z_2)] \;. \tag{86}$$

Note that in general, $D_{\rm ang}(z_1, z_2) \neq D_{\rm ang}(z_2) - D_{\rm ang}(z_1)$; on the other hand, such an additive relation *is* valid for the comoving angular diameter distances for a Universe with vanishing curvature $K = 0$, as seen from (85) and (66). Thus, it is often useful to employ the comoving angular diameter distance,

i.e. the ratio between the comoving diameter of an object and its angular diameter.

Another method to measure distances is to relate the observed flux S of a source to its luminosity L; if we know the luminosity, then the distance to the source can be determined by

$$D_{\rm lum}(z) \equiv \sqrt{\frac{L}{4\pi S}}\,, \tag{87}$$

which is called the luminosity distance. In Euclidean space, this measurement would yield the same result as that from comparing diameters and angular sizes; in curved spacetimes, this is no longer true. In fact, one can show (Etherington 1933) that in general,

$$D_{\rm lum}(z) = (1+z)^2\, D_{\rm ang}(z) = (1+z)\, f_K(w)\,. \tag{88}$$

In this equation, flux and luminosity have to be interpreted as bolometric quantities, i.e., integrated over all frequencies. The flux at a given frequency ν is related to the specific luminosity of the source at a different frequency $\nu_{\rm e} = (1+z)\nu$, owing to redshift. This frequency shift is taken into account by the so-called K-correction in the relation between specific flux and luminosity.

We still need another distance concept, the proper distance. Suppose we measure the redshifts z and $z + \Delta z$ of two comoving sources, being very similar, and which also have small angular separation $\Delta\theta$ on the sky. What is the separation between these two sources that an observer would measure who lives somewhere near them? This separation can be measured by this fiducial observer in the same way as we can measure the distance to Virgo-cluster galaxies, without caring about the values of the cosmological parameters – locally space can be approximated as being Euclidean where distances have a unique meaning. The proper separation transverse to the line-of-sight is $D_{\rm ang}(z)\,\Delta\theta$, and that along the line-of-sight is

$$\begin{aligned}\Delta r_{\rm prop} &= a(z)\,\Delta w = a(z)\frac{{\rm d}w}{{\rm d}a}\,\frac{{\rm d}a}{{\rm d}z}\Delta z = \frac{c\,a(z)}{H(z)}\Delta z \\ &= \frac{c}{H_0}\,\frac{\Delta z}{\sqrt{\Omega_{\rm m} a^{-1} + (1-\Omega_{\rm m}-\Omega_\Lambda) + \Omega_\Lambda a^2}}\,.\end{aligned} \tag{89}$$

Volume elements. We can now also calculate volume elements: suppose in a solid angle ω one measures ${\rm d}N$ sources with redshift between z and $z + {\rm d}z$, the proper number density of these sources is $n = {\rm d}N/{\rm d}V$, where the volume is given by the physical thickness of the redshift slice times the area transverse to the line of sight, which is $D^2_{\rm ang}(z)\,\omega$, so

$${\rm d}V_{\rm prop} = D^2_{\rm ang}(z)\,\omega\,\frac{{\rm d}r_{\rm prop}}{{\rm d}z}\,{\rm d}z\,, \tag{90}$$

where we indicated that this is the proper volume element. The corresponding comoving volume element is then

$$\mathrm{d}V_{\rm com} = a^{-3}\,\mathrm{d}V_{\rm prop} = f_K^2[w(z)]\,\omega\,\frac{\mathrm{d}w}{\mathrm{d}z}\,\mathrm{d}z\;. \tag{91}$$

Finite volumes can be obtained from the foregoing equations by integration.

Special cases. Whereas the foregoing expansion equations are easily evaluated through numerical integration, there are some cases where explicit expressions can be obtained. The simplest model is the *Einstein–de Sitter* (EdS) Universe, characterized by $\Omega_{\rm m}=1$, $\Omega_\Lambda=0$; this model has zero curvature. The expansion equation (neglecting radiation) reduces to $H=H_0a^{-3/2}$, yielding $t=2/(3H_0)a^{3/2}$; in particular, the current age of the Universe in this model is $t_0=2/(3H_0)\sim 6.7\times10^9h^{-1}$ years. The comoving and angular diameter distance for an EdS model are easily obtained as

$$\begin{aligned} D_{\rm ang}(z_1,z_2) &= \frac{2c}{H_0}\,\frac{1}{1+z_2}\left[(1+z_1)^{-1/2}-(1+z_2)^{-1/2}\right]\;,\\ D_{\rm com}(z_1,z_2) &= \frac{2c}{H_0}\left[(1+z_1)^{-1/2}-(1+z_2)^{-1/2}\right]\;. \end{aligned} \tag{92}$$

In particular, as for all flat models, the comoving angular diameter distance is the same as the comoving distance. Unfortunately, we seem to not be living in an EdS Universe (see Sect. 6.3).

For models without a cosmological constant, the angular-diameter distance can be written in closed from, using the famous Mattig (1958) relation,

$$\begin{aligned} D_{\rm ang}(z_1,z_2) &= \frac{2}{\Omega_{\rm m}^2(1+z_1)(1+z_2)^2}\\ &\times\left[(\Omega_{\rm m}z_2-\Omega_{\rm m}+2)\sqrt{1+\Omega_{\rm m}z_1}-(\Omega_{\rm m}z_1-\Omega_{\rm m}+2)\sqrt{1+\Omega_{\rm m}z_2}\right]\;. \end{aligned} \tag{93}$$

Next we consider the expansion equation (80) qualitatively. The different dependencies of the four terms in (80) on the scale factor shows that for very small a, the expansion was dominated by radiation, for $a\gtrsim a_{\rm eq}$ it was dominated by matter; the effects of curvature (if different from zero) and the cosmological constant play a role only at later stages of the cosmic expansion. For small $a\ll1$, (80) can be approximated as $H=H_0\Omega_{\rm m}^{1/2}a^{-3/2}\sqrt{1+a_{\rm eq}/a}$ which can be integrated to yield

$$t(a)=\frac{2}{3H_0}\,\Omega_{\rm m}^{-1/2}\left[a^{3/2}\left(1-2\frac{a_{\rm eq}}{a}\right)\left(1+\frac{a_{\rm eq}}{a}\right)^{1/2}+2\,a_{\rm eq}^{3/2}\right]\;, \tag{94}$$

and so $t=a^2\left(2H_0\sqrt{\Omega_{\rm m}a_{\rm eq}}\right)^{-1}$ for $a\ll a_{\rm eq}$, and $t=2a^{3/2}\left(3H_0\sqrt{\Omega_{\rm m}}\right)^{-1}$ for $a_{\rm eq}\ll a\ll1$. For EdS, (94) describes the expansion for all a through the radiation and matter dominated phases.

4.3 Gravitational lensing in cosmology

The meaning of distance in lensing. When we used distances to write the gravitational lens equations in Sect. 2 we have not discussed what 'distance' means there. Now we learned that the concept of distance in curved spacetimes, even if they are as simple as the Friedmann–Lemaître spacetimes, are more complicated than in the Euclidean case. Therefore, which of the many distances defined above is the one to be used in the gravitational lens equations?

The answer is quite obvious: recall that the basic lens equation (6) relates images and source positions by a geometrical consideration; for that one needs to relate angles with transverse distances. This is exactly the way the angular-diameter distance was defined; hence, all equations in Sect. 2 are also valid in a Friedmann–Lemaître spacetime if the distances D are taken to be the angular-diameter distances $D_{\rm ang}$.

In many cases, the equations of gravitational lensing become simpler if the comoving angular diameter distances $f_K(w)$ are used; one example is the expression (96) for the time delay. In particular this is true for flat cosmological models, for which $f_K(w) = w$. Furthermore, most equations of gravitational lensing contain distances only in form of the ratio $D_{\rm ds}/D_{\rm s}$, for which it is irrelevant whether $D = D_{\rm ang}$ or $D = f_K$ is used. *In order not to confuse the reader, we shall consistently use the following convention throughout the rest of this book (recalling that in all equations in Sects. 2 and 3, $D \equiv D_{\rm ang}$ is implied): the angular diameter distance is denoted by $D^{\rm ang}$, and the comoving distance is denoted by D or f_K.* For example, in this notation the critical surface mass density and the Einstein radius of a point mass read

$$\begin{aligned} \Sigma_{\rm cr} &= \frac{c^2}{4\pi G}\,\frac{D_{\rm s}^{\rm ang}}{D_{\rm d}^{\rm ang}\,D_{\rm ds}^{\rm ang}} = \frac{c^2(1+z_{\rm d})}{4\pi G}\,\frac{D_{\rm s}}{D_{\rm d}\,D_{\rm ds}} \\ \theta_{\rm E} &= \left(\frac{4GM(1+z_{\rm d})}{c^2}\,\frac{D_{\rm ds}}{D_{\rm d}D_{\rm s}}\right)^{1/2} . \end{aligned} \tag{95}$$

The time delay. We mentioned in Sect. 2.2 that the light travel times along the light rays that form the multiple images in a lens system are not the same, but have not given an expression for it. Now that we are armed with the necessary cosmological relations we can do so. There are two ways to derive an expression for the time delay, both of which shall briefly be described here. Cooke & Kantowski (1975) argued that the time delay must have two different components: first, a light ray that is bent is longer, and thus light needs more time to propagate along it, than for a straight ray. Since the individual light rays are bent by different angles, their geometrical lengths are different, giving rise to a 'geometrical time delay' between them. Second, light rays propagate through a gravitational potential which retards them; this is the well-known 'Shapiro effect', which has been amply tested by radar

echo delay experiments towards the inner satellites in the Solar System. This is the 'gravitational time delay'. The total time delay is then simply the sum of the two. For this derivation, it is important to note that the gravitational time delay 'occurs' at the redshift of the lens, and hence gets redshifted by a factor $1+z_{\rm d}$ owing to the cosmic expansion.

An alternative derivation of the time delay was given by Refsdal and his collaborators (see e.g. Kaiser & Refsdal 1983 and references therein); they considered the wavefronts emitted from a source. Wavefronts are surfaces of equal light travel time (image for example a set of photons all emitted in a single flash from a bursting source; their location at fixed instant form a wavefront), and Fermat's principle states that light rays are perpendicular to the wavefronts. Close to the source, the wavefronts are spheres; owing to perturbations in the gravitational potential, they get distorted. If propagating near a mass concentration, the wavefronts can become strongly distorted, and after passing it, they can actually intersect themselves. An observer located in that region would then be passed by the same folded wavefront more than once; since the different sheets of the wavefronts have different orientations, the observer will then see multiple images of the corresponding source, in the direction perpendicular to the individual wavefront sheets. The time delay between two images is then obtained as the time between the passing of the two corresponding wavefront sheets. From a purely geometrical consideration, the time delay can then be derived, yielding the same expression as adding together the geometrical and potential time delays of the first method.[9] The time delay can be written most conveniently in terms of the Fermat potential as (Schneider 1985)

$$\begin{aligned}\Delta t &= \frac{D_{\rm d}^{\rm ang} D_{\rm s}^{\rm ang}}{c\, D_{\rm ds}^{\rm ang}}(1+z_{\rm d})\left[\tau(\boldsymbol{\theta}^{(1)};\boldsymbol{\beta})-\tau(\boldsymbol{\theta}^{(2)};\boldsymbol{\beta})\right]\\ &= \frac{D_{\rm d} D_{\rm s}}{c\, D_{\rm ds}}\left[\tau(\boldsymbol{\theta}^{(1)};\boldsymbol{\beta})-\tau(\boldsymbol{\theta}^{(2)};\boldsymbol{\beta})\right]\ , \end{aligned} \tag{96}$$

where in the second expression the comoving angular diameter distances were used, and $\tau(\boldsymbol{\theta};\boldsymbol{\beta})$ is the Fermat potential defined in (14). This result then confirms the statement made in Sect. 2.2: $\tau(\boldsymbol{\theta};\boldsymbol{\beta})$ is, up to an affine transformation, the light travel time along a ray originating at $\boldsymbol{\beta}$ in the source plane, traversing the lens plane at $\boldsymbol{\theta}$ and then propagating to the observer. The additive constant of this affine transformation is irrelevant, as only differences are observable; the factor in the linear term is given in (96). The potential time delay is described by the deflection potential $\psi(\boldsymbol{\theta})$ in τ, the geometric time delay by the $|\boldsymbol{\beta}-\boldsymbol{\theta}|^2/2 = |\boldsymbol{\alpha}|^2/2$-term.

[9] As Sjur Refsdal reports, the first referee of his paper on the wavefront method rejected it wholeheartedly, claiming that the resulting expression can contain only the geometrical contribution to the time delay. It remains unknown how this referee imagined the geometry of distorted and overlapping wavefronts without the effect of retardation provided by the gravitational field of the deflector.

5 Basics of lensing statistics

One is frequently interested in the probability that a specific gravitational lensing event occurs. For example, Zwicky (1937b) estimated the probability that a distant source is multiply imaged by "extragalactic nebulae" using the surface density of these objects on the sky. This basic problem of statistical gravitational lensing has since been studied in considerable detail, as will be discussed in SL. The results of such an investigation depend on the assumed distribution of lens masses and their individual density profiles. A comparison of these results with a statistically well defined sample of observed lens cases can in principle allow one to constrain the lens contents and/or the geometry of the Universe. The probability for microlensing events to occur depends on the density of compact objects along the sightline to the population of sources which are monitored, as will be detailed in ML.

Another typical problem of statistical lensing is the so-called magnification bias. Let us consider a sample which should include all sources of a certain kind in a region of the sky brighter than a given threshold (i.e., a flux-limited sample). From the observed fluxes of the sources and their distances (e.g., determined from their redshifts) we can derive the intrinsic luminosities of the sources. If a source is magnified by a gravitational lens, its derived luminosity will not be the true one, but will be higher in general. Moreover, there may be sources in the sample which do not belong there because they are intrinsically too faint to be included, but have been magnified above the threshold of the sample. Since flux-limited samples of extragalactic sources are used to derive information about the evolution of the sources and about the structure of the universe, the magnification can mislead astronomers. Statistical lensing investigations are used to estimate the importance of this effect and its consequences. In this section we shall provide the basics of lensing probability investigations, with details left to later chapters when specific applications are discussed.

5.1 Cross-sections

The lensing probabilities depend on the number density of lenses, as well as on their mass profile. The latter is used to define lensing cross-sections. We shall start with two specific examples which should motivate this concept.

Cross-sections for a point-mass lens. First, consider a (point) source at distance D_s, and a point mass at distance D_d from Earth. The separation of the two images and their magnifications depend on the relative alignment of source, observer, and lens. There is a one-to-one relationship between the source position $\beta = y\theta_\mathrm{E}$ and the corresponding total magnification μ_p, see (48), where total magnification means the summed magnifications of the individual images. Thus, for any $\mu_\mathrm{p} > 1$ there is a value of y such that, if the

distance β of the source is less than $y\theta_{\rm E}$ from the optical axis, the latter is magnified by more than $\mu_{\rm p}$:

$$y^2 = 2\left(\frac{\mu_{\rm p}}{\sqrt{\mu_{\rm p}^2 - 1}} - 1\right) \quad . \tag{97}$$

Hence, we can define the cross-section

$$\sigma(\mu_{\rm p}) = \pi\theta_{\rm E}^2\, y^2(\mu_{\rm p}) \tag{98}$$

for a point source magnification larger than $\mu_{\rm p}$. In other words, centered on the caustic point there is a solid angle $\sigma(\mu_{\rm p})$ within which the source must lie in order to be magnified by more than $\mu_{\rm p}$. One could also consider σ to be the solid angle in which a lens must be located such that a fixed source is magnified by more than $\mu_{\rm p}$; these two points of view are basically equivalent, though there are some subtleties involved (Ehlers & Schneider 1986) which shall not be discussed here.

As a second example, we consider the ratio $r = \mu_+/|\mu_-|$ of the absolute values of the magnifications (brightness ratio) for the two images produced by a Schwarzschild lens – see (48). In order for this ratio to be less than r, the impact parameter y needs to be less than $r^{1/4} - r^{-1/4}$, and so the cross-section for magnification ratio less than $r > 1$ is

$$\sigma(r) = \pi\theta_{\rm E}^2\left(r^{1/2} + r^{-1/2} - 2\right) \quad . \tag{99}$$

General definition of a lensing cross-section. After these two examples we now discuss the general definition of a cross-section. Consider a source and a lens, both at fixed distances from Earth. The lens may be described by a set of parameters, and the source is characterized, say, by its size and its brightness profile (if the source extent is relevant). If one is interested in a certain property Q of this gravitational lens system, one can ask where the source must be located such that the images have the property Q. Two examples for Q were given above, namely, that the total magnification is larger than μ and that the brightness ratio of the images is smaller than r. More complicated examples of Q will be considered in due course. The question can be answered through an analysis of the gravitational lens model, as demonstrated above for the point mass lens. One usually finds that the source must be in a certain region of the source plane. The solid angle of that region is then the Q–cross-section σ_Q for this lens–source system.

Lensing cross-section for a singular isothermal sphere. To illustrate the concept further, we shall consider the lensing cross-sections of an SIS. From Sect. 3.3 we know that this lens produces two images if $\beta < \theta_{\rm E}$, and

that the image separation is $2\theta_{\rm E}$. Hence, the cross-section of an SIS to produce two images with separation larger than $\Delta\theta$ is

$$\sigma(\Delta\theta) = \pi\theta_{\rm E}^2 {\rm H}(2\theta_{\rm E} - \Delta\theta) \ , \tag{100}$$

where ${\rm H}(x)$ is the Heaviside step function. Next the (total) magnification and the flux ratio r can be included; both are functions of $y = \beta/\theta_{\rm E}$, $\mu = 2/y$ and $r = (1+y)/(1-y)$. Therefore, the cross-section for an SIS to produce two images with separation larger than $\Delta\theta$, total magnification larger than μ and flux ratio of the images smaller than r is

$$\sigma(\Delta\theta, r, \mu) = \pi\theta_{\rm E}^2 \left[\min\left(\frac{r-1}{r+1}, \frac{2}{\mu}\right)\right]^2 {\rm H}(2\theta_{\rm E} - \Delta\theta)\, {\rm H}(\mu - 2) \ . \tag{101}$$

The mass-sheet degeneracy and scaling of cross-sections. In Sect. 2.5 we have seen that the imaging properties, such as angular separation and magnification ratios – and thus flux ratios, are unchanged if the surface mass density is transformed according to (28). As can be seen from (31), this transformation merely leads to a scaling of the lens plane, thereby affecting the magnifications. The scaling of the lens plane implies that the cross-section for an extended source of size ρ will, after the mass-sheet transformation, be related to the cross-section for a source of size $\lambda\rho$. Hence, the cross-section for flux ratio smaller than r, image separation larger than $\Delta\theta$, magnification larger than μ for a source of size ρ transforms like

$$\sigma_\lambda(r, \Delta\theta, \mu, \rho) = \lambda^2 \sigma\left(r, \Delta\theta, \lambda^2\mu, \frac{\rho}{\lambda}\right) \ . \tag{102}$$

5.2 Lensing probabilities; optical depth

The probability that a lensing event with specified properties Q occurs is given by the product of the number density of lenses and their cross-sections. Consider a solid angle ω towards sources at distance $D_{\rm s}$. To the distance interval ${\rm d}x$ around x (note that we use a different notation for distance along the line-of-sight here, for reasons which soon will become clear) corresponds a proper volume element ${\rm d}V = D_{\rm ang}^2(x)\ ({\rm d}r_{\rm prop}/{\rm d}x)\ \omega$ within this solid angle. We consider lenses which are described by a set of parameters, summarized as χ; such parameters could be lens mass, core radius, ellipticity, etc. If $n(x,\chi)\,{\rm d}\chi$ is the (proper) number density of lenses at distance x with properties within ${\rm d}\chi$ of χ, the total cross-section of all lenses within the tube of solid angle ω is then

$$\sigma_{\rm tot}(Q) = \int_0^{x_{\rm s}} {\rm d}x\, \omega\, D_{\rm ang}^2(x)\, \frac{{\rm d}r_{\rm prop}}{{\rm d}x} \int {\rm d}\chi\, n(x,\chi)\, \sigma(Q; x, x_{\rm s}, \chi) \ , \tag{103}$$

where the Q-cross-section of a lens depends on the lens parameters χ, the distance parameter x along the line-of-sight, and the source distance $x_{\rm s}$. The

picture underlying this equation is that the cross-sections of the individual lenses simply add up linearly. This picture is justified as long as the projected separation between lenses is much larger than the linear size of the cross-sections, or in other words, the cross-sections of individual lenses do not overlap. The probability for a lensing event with property Q – also frequently called *optical depth* for lensing – is then given by the ratio of the cross-section $\sigma_{\rm tot}(Q)$ and the solid angle ω, i.e., the fraction of solid angle covered by the cross-sections,

$$P(Q) = \int_0^{x_{\rm s}} {\rm d}x\; D_{\rm ang}^2(x)\, \frac{{\rm d}r_{\rm prop}}{{\rm d}x} \int {\rm d}\chi\; n(x,\chi)\, \sigma(Q;x,x_{\rm s},\chi)\;. \tag{104}$$

We shall now consider the case where the distance to the source is small, so that cosmological distances play no role. In this case, we can take $x = D_{\rm d}$, and (104) becomes

$$P(Q) = \int_0^{D_{\rm s}} {\rm d}D_{\rm d}\; D_{\rm d}^2 \int {\rm d}\chi\; n(D_{\rm d},\chi)\, \sigma(Q;D_{\rm d},D_{\rm s},\chi)\;. \tag{105}$$

In the other case, where the sources of interest are at cosmological distances, redshift is a convenient distance variable, and (104) reads

$$P(Q) = \int_0^{z_{\rm s}} {\rm d}z_{\rm d}\; D_{\rm ang}^2(z_{\rm d})\, \frac{{\rm d}r_{\rm prop}}{{\rm d}z} \int {\rm d}\chi\; n(z_{\rm d},\chi)\, \sigma(Q;z_{\rm d},z_{\rm s},\chi)\;, \tag{106}$$

where the proper distance interval along the line-of-sight is given in (89). In this cosmological context, it is often useful to specify the comoving number density $n_{\rm com}$ of lenses, instead of the proper density n; both are related by $n(z) = (1+z)^3\, n_{\rm com}$; furthermore, we can use the comoving distance w as integration variable, and work in terms of the comoving angular diameter distance $f_K(w)$; then, (106) becomes

$$P(Q) = \int_0^{w_{\rm s}} {\rm d}w\; f_K^2(w) \int {\rm d}\chi\; n_{\rm com}(w,\chi)\, \sigma(Q;w,w_{\rm s},\chi)\;, \tag{107}$$

which is particularly convenient in the case of flat models ($f_K(w) = w$).

5.3 Magnification bias

Besides the optical depth for a source to be multiply imaged (with image separation larger than an angular resolution limit of a survey), the magnification probability distribution has received great attention in the literature. Questions that have been studied include:

- Can all bright quasars be merely highly magnified images of much less luminous Seyfert galaxies (Barnothy 1965)? No, the lensing probabilities are far too small (e.g., Tyson 1981; Peacock 1982), even if the dark matter in the Universe consists of compact objects (Canizares 1982; Schneider 1987).

- Can magnification by (compact objects in the) halos of galaxies explain the apparent angular correlation (e.g., Arp 1987, and references therein) between nearby bright galaxies and high-redshift quasars? Again no, magnification probabilities are far too small (e.g., Canizares 1981; Vietri & Ostriker 1983; Schneider 1992).
- Does magnification affect the expected number of multiply-imaged QSOs in a gravitational lens survey? Yes (e.g., Wallington & Narayan 1993): for a lens survey with a bright flux threshold, magnification boosts the fraction of lensed sources by a large factor (see SL).

Magnification can cause sources to be included in a flux-limited sample which without lensing would be too faint to be included. Furthermore, sources with a very large magnification factor can attain apparent luminosities which exceed the maximum luminosities of the corresponding class of sources. In fact, several of the most luminosity-extreme sources are strongly magnified: The apparently most luminous IRAS galaxy F10214+4724 is magnified by a galaxy by about a factor of $\mu \sim 50$ (Broadhurst & Lehar 1995), the most luminous 'normal' Lyman-break galaxy cB58 (Yee et al. 1996) at redshift $z = 2.72$ is magnified by the cluster MS1512+36 with redshift $z \sim 0.3$ by a factor $\mu \sim 30$ (Seitz et al. 1998), the very bright $z = 3.87$ QSO APM 08279+5255 (Irwin et al. 1998) is gravitationally lensed and highly magnified by a foreground galaxy (Ibata et al. 1999), and several of the highest redshift galaxies have been found behind lensing clusters (e.g., Hu et al. 2002; Kneib et al. 2004; Pelló et al. 2004).

Consider a class of sources is a narrow redshift interval, and denote by $p(\mu)\,\mathrm{d}\mu$ the probability that one of these sources is magnified by a factor within $\mathrm{d}\mu$ of μ. Let $N_0(>S)$ be the number of these sources per unit solid angle that without lensing would be observed to have flux greater than S. If these sources get magnified by a factor μ, two things happen: first, a source with unlensed flux S will attain an observed flux μS. Second, since magnification enlarges the solid angle, sources that without lensing would be contained in a solid angle ω on the sky, will now be spread over the solid angle $\mu\omega$, i.e, the number density of sources decreases by a factor $1/\mu$. Together, if the magnification would be (locally) a constant μ, the observed source counts are

$$N(>S) = \frac{1}{\mu}\, N_0\left(>\frac{S}{\mu}\right) \ . \tag{108}$$

Considering a probability distribution $p(\mu)$ in magnifications, this result generalizes to

$$N(>S) = \frac{1}{\langle\mu\rangle}\int \mathrm{d}\mu\; p(\mu)\, N_0\left(>\frac{S}{\mu}\right) \ , \tag{109}$$

where $\langle\mu\rangle$ is the mean magnification within the region of the sky considered. If source counts are taken over random regions on the sky, then $\langle\mu\rangle = 1$, but if the magnification bias is considered around foreground galaxies, then in

these regions, $\langle\mu\rangle > 1$. The probability $p(\mu)$ satisfies

$$\int \mathrm{d}\mu\, p(\mu) = 1\ ; \quad \int \mathrm{d}\mu\, \mu\, p(\mu) = \langle\mu\rangle\ ; \tag{110}$$

the first relation expressing normalization, the second the definition of the mean magnification. Of course, $p(\mu)$ depends on the source redshift and the density of lenses, as mentioned in the previous section.

At this point, we need to enter briefly in the discussion on 'flux conservation'. As you recall, we have shown that any lens produces at least one image which is not demagnified, $\mu \geq 1$, provided $\kappa(\boldsymbol{\theta})$ is non-negative. If we consider source counts averaged over the whole sky, then of course $\langle\mu\rangle = 1$, and then (110) implies that if there are lines-of-sight where $\mu > 1$, then there must also exist those with $\mu < 1$. Hence, we get an apparent contradiction, which has led to much confusion in the literature. The resolution of this contradiction is seen when we consider the full matter distribution between us and the sphere of sources at a given redshift. The unmagnified flux of a source is defined as that flux which would be observed in a homogeneous Universe, since it is this model which underlies the definition of the luminosity distance. Now, the true mass distribution is inhomogeneous, consisting of large overdensities like galaxies and clusters, and underdensities like voids. A light bundle propagating through an underdense region of the Universe is less focused than one propagating through the mean density of the Universe, so that the effective $\kappa < 0$ for the former, and $\mu < 1$. Conversely, light bundles propagating through overdense regions get more focused, resulting in $\mu > 1$. We shall discuss these relations in more detail in WL.

Note that there is a minimum magnification for each source redshift. Since the cosmic density $\rho \geq 0$, a light bundle cannot be more defocused than propagating through empty space; hence, μ is bound from below, and this bound depends on the source redshift and the density parameters Ω_m and Ω_Λ.

We shall now consider the simple example of source counts which behave like a power law, $N_0(>S) = A\, S^{-\beta}$. Inserting this into (109) yields

$$N(>S) = \frac{1}{\langle\mu\rangle} \int \mathrm{d}\mu\, p(\mu)\, A \left(\frac{\mu}{S}\right)^{\beta} = N_0(>S)\, \frac{1}{\langle\mu\rangle} \int \mathrm{d}\mu\, p(\mu)\, \mu^{\beta}\ . \tag{111}$$

Thus, if the unlensed source counts behave like a power law, so do the lensed ones, with the same slope. The ratio between lensed and unlensed counts depends on the magnification probability distribution $p(\mu)$, as well as on the slope β of the counts. The first remarkable result is that, if $\beta = 1$, then (111) together with the second of (110) imply that $N(>S) = N_0(>S)$, i.e. the counts are unchanged in this case, independent of $p(\mu)$. Hence, in this case the enlargement of the solid angle over which sources are distributed just compensates the brightening of the sources. For $\beta < 1$, the number counts are depleted, whereas they are increased for $\beta > 1$. The larger the slope β,

the larger is the ratio $N(>S)/N_0(>S)$, i.e. the stronger is the magnification bias.

If one considers point sources, or more generally, sources whose angular sizes are much smaller than the characteristic angular scale of the lenses, then one can show (Blandford & Narayan 1986) that for very high magnification, $p(\mu) \propto \mu^{-3}$, up to an upper limit for μ at which the finite size of the source limits the magnification. This functional dependence is due to the universal behavior of the lens equation near fold caustics, as was discussed in Sect. 2.4. This functional behavior implies that the integral in (111) formally diverges as the slope β approaches 2. Hence, for a population of sources with steep number counts, the magnification bias can become very large. In fact, the formal divergence is due only to the assumption of a pure power law for $N_0(>S)$; whereas such a functional form is a good description, e.g., for the QSO counts over a limited range of fluxes (or luminosities), it cannot continue with a steep slope for arbitrarily faint sources, in order for the source population not to produce infinite total flux. Nevertheless, if the counts are steep, and one considers a value of S much larger than a break flux (where the steep counts turn into flatter ones towards lower fluxes), the ratio $N(>S)/N_0(>S)$ can be very high indeed. This is the reason why we see extreme QSOs like the one mentioned above, APM 08279+5255. Furthermore, if the source population is better described by a Schechter luminosity function, which implies an exponential decrease in the counts for high luminosities, the bias can be even larger: the probability with a Schechter function to find a single source far out in the exponential tail is very small, and if such an apparently luminous source is observed, it is most likely a lensed one, as is the case for F10214+4724 and cB58.

6 The cosmological standard model II: The inhomogeneous Universe

Whereas the Universe appears to be nearly homogeneous on large scales, it certainly is strongly inhomogeneous on smaller scales. Small fluctuations are imprinted onto the CMB, leading to tiny but measurable anisotropies in its temperature; in fact, these anisotropy measurements provide the strongest constraints on cosmological parameters currently available. Furthermore, the distribution of brighter (thus nearer) galaxies in the sky is highly anisotropic; galaxies tend to be strongly correlated, they tend to appear in groups or clusters of galaxies. Thus, on small scales the approximation of a homogeneous Universe must break down.

6.1 Structure formation

Whereas the CMB fluctuations indicate very small inhomogeneities at the time of recombination (corresponding to a redshift $z \sim 1100$) the inhomogeneities observed today in our neighborhood are much larger. A cluster of

galaxies, for example, is a massive perturbation with a mean density more than hundred times larger than the mean density in the Universe. It is believed that the density inhomogeneities that we see today have evolved from much smaller fluctuations in the very early Universe. This evolution happens naturally through *gravitational instability*. A slightly overdense region has a somewhat higher self-gravity than the average region of the Universe, so its expansion rate will be slightly smaller than that of the Universe as a whole. As a result of slower expansion, the density contrast of this region increases further, retarding expansion more, and so on. If the initial density contrast is sufficiently large, this instability can actually bring the expansion to a halt locally, after which the region recollapses under its own gravity to form galaxies and clusters.

In this picture of gravitational instability, one can study the evolution of structure in great detail. Since the major mass component in the Universe is dark matter, which by definition only interacts gravitationally, the dominant process is gravity. However, the laws of gravity are well understood. Furthermore, additional simplifications arise in certain regimes; e.g., at early stages in the evolution, density fluctuations are very small. One can therefore linearize the equations of gravity around the homogeneous Universe. If the length scale of the perturbations are much smaller than the characteristic scale of the Universe [a size given by $c/H(a)$], gravity can be approximated by the Newtonian equations. In addition, numerical simulations can follow the evolution of the density field under the influence of gravity, and great progress has been made in the level of detail these studies have achieved (e.g., Frenk et al. 1999; Springel et al. 2001).

We shall outline here a number of results which will be used in later Chapters; again, the reader is referred to the excellent textbooks mentioned at the beginning of this section for a much more detailed treatment.

Horizons. No signal propagates at speeds larger than c; at a given cosmic time t, this implies that the region within which matter has been in causal contact is finite, essentially given by $ct \sim c/H(t)$, where we used that $tH(t) \sim 1$. The size of this region is called the horizon size at time t. The comoving horizon size is

$$d_{\rm H} = \frac{c}{a\,H(a)} = \frac{c}{H_0}\,\Omega_{\rm m}^{-1/2}\,a^{1/2}\,\left(1+\frac{a_{\rm eq}}{a}\right)^{-1/2}\,, \tag{112}$$

where in the second step we used the approximation for $a \ll 1$ when curvature and vacuum energy play hardly any role; cf. (94). As we shall see, the comoving horizon size at the epoch of matter and radiation equality is of particular importance and is

$$d_{\rm H}(a_{\rm eq}) = 2^{-1/2} c H_0^{-1}\,\Omega_{\rm m}^{-1/2} a_{\rm eq}^{1/2} \approx 12(\Omega_{\rm m}h^2)^{-1}\,{\rm Mpc}\,. \tag{113}$$

Linear density evolution. If the density fluctuations are small, linear perturbation theory can be used to describe them. Specifically, one defines the *density contrast*

$$\delta(\mathbf{x},t) := \frac{\rho(\mathbf{x},t)-\bar{\rho}(t)}{\bar{\rho}(t)}\ , \tag{114}$$

where $\bar{\rho}(t)$ is the mean cosmic matter density at time t, and $\mathbf{x}$ is the comoving spatial coordinate. The matter equations are linearized about the homogeneous model, and only terms of order δ are considered. If several matter components are relevant (e.g., non-relativistic matter and radiation, in early phases of the evolution), one defines a density contrast for each of them, and considers the (coupled) set of linear evolution equations. As soon as $|\delta| \sim 1$, this perturbation theory breaks down, and the full set of non-linear evolution equations needs to be treated (numerically).

If one considers fluctuations on a scale much smaller than the horizon scale, one can use Newtonian gravity. The relevant equations in this case are the Vlasov (or collisionless Boltzmann) equation and the Poisson equation; the former is usually approximated by the fluid equations, e.g. the continuity equation and the Euler equation. A homogeneous, expanding Universe is a solution to these equations, and the expansion factor follows the Friedmann equation (67). Setting $\rho(\mathbf{x},t) = \bar{\rho}(t)[1+\delta(\mathbf{x},t)]$, these equations are then transformed into comoving coordinates, e.g. for the Poisson equation one finds

$$\nabla_x^2 \Phi = 4\pi G a^2 \bar{\rho}\delta = \frac{3H_0^2}{2a}\Omega_{\rm m}\delta\ . \tag{115}$$

Then, writing the velocity field as a sum of the Hubble expansion plus a small perturbation to it, one finds that the terms linear in δ lead to a single linear second-order differential equation in time whose coefficients do not depend on the spatial coordinates. Hence, there are two linearly independent solutions of the form $\delta(\mathbf{x},t) = D_\pm(t)\,\Delta_\pm(\mathbf{x})$. One of the two functions, $D_-(t)$ say, decreases quickly with time and is therefore unimportant for structure growth; the other one grows with time, so that

$$\delta(\mathbf{x},t) = D_+(t)\,\delta_0(\mathbf{x}) \tag{116}$$

is the relevant mode for structure growth. The function $D_+(t)$ is called the *linear growth factor*, which can be obtained from solving the aforementioned differential equation,

$$D_+ \propto H(a)\int_0^a \frac{{\rm d}a'}{[a'\,H(a')]^3}\ , \tag{117}$$

with the constant of proportionality chosen such that $D_+(t_0) = 1$. Because of this choice, one has $\delta(\mathbf{x},t_0) = \delta_0(\mathbf{x})$; hence, $\delta_0(\mathbf{x})$ is the current density contrast provided the evolution of δ follows linear perturbation theory. Even if it does not, defining the field δ_0 is meaningful, since (116) still describes

the evolution of the density contrast for epochs where δ was much smaller than unity. For obvious reasons, δ_0 is called the linearly-extrapolated density contrast. For an EdS Universe, the growth factor is $D_+(t) = a(t)$, for lower-density models, $D_+(t) \geq a(t)$ (see Fig. 14); Carroll et al. (1992) provide a fitting formula for D_+ for general cosmologies. Note that (116) predicts that the shape of fluctuations are time-independent in comoving coordinates, with only their amplitude being a function of time.

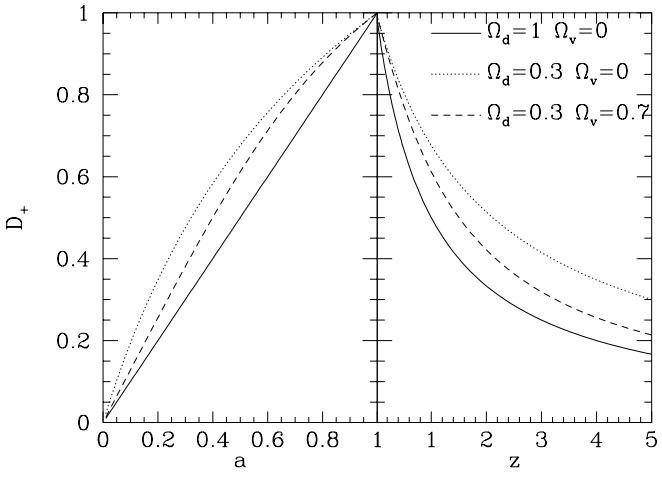


Fig. 14. The growth factor D_+ as a function of the scale factor a (left) and as a function of redshift (right), for three cosmological model: an EdS model ($\Omega_{\rm m} = 1$, $\Omega_\Lambda = 0$), a low density open model ($\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0$), and a low-density flat Universe ($\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$)

Random fields, correlation functions and power spectra. Cosmology will never be able to describe the specific density field of our Universe, since in order to do so, we would need to know the density fluctuation field $\delta(\mathbf{x}, t_{\rm i})$ at some initial time $t_{\rm i}$. Instead, what a theory of structure formation should explain are the statistical properties of the density field as a function of time: how many clusters of galaxies per unit volume form as a function of redshift, how does matter cluster together, etc. This is analogous to statistical physics, where the behavior of a physical system is described by its macroscopic statistical properties, not by the trajectories of all molecules.

The density fluctuations $\delta(\boldsymbol{x})$ at some fixed time are considered to be a random field. A *random field* is characterized by the probability that a specific realization $\delta(\boldsymbol{x})$ of the density fluctuations occurs. This probability therefore

is a mapping from the set of functions $\delta : \mathbb{R}^3 \to \mathbb{R}$ to $\mathbb{R}_0^+$. A conceptually simpler way to look at this is to assume that all probable realizations of the density field are 'smooth', so that $\delta(\boldsymbol{x})$ can be described, with sufficient accuracy, by its values on a regular grid in $\boldsymbol{x}$. Let $\boldsymbol{x}_i$ be a set of appropriately numbered gridpoints, and let $\delta_i = \delta(\boldsymbol{x}_i)$ be the density contrast at $\boldsymbol{x}_i$. The realization of the random field is then described by the (possibly infinite) set of the δ_i, and the random field is characterized by the joint probability distribution $p(\delta_1, \delta_2, \ldots)\mathrm{d}\delta_1\, \mathrm{d}\delta_2 \ldots$ that $\delta(\boldsymbol{x}_i)$ lies within $\mathrm{d}\delta_i$ of δ_i. Hence, we have reduced the description of the random field to a joint probability distribution of (possibly infinite) discrete random variables. Since the Universe is assumed to be statistically homogeneous and isotropic, the density field should share these properties; this is formulated by the requirement that if all grid points are translated and rotated the same way, $\boldsymbol{x}_i \to \mathcal{R}(\boldsymbol{x}_i + \boldsymbol{y})$, where $\mathcal{R}$ is a rotation matrix and $\boldsymbol{y}$ a translation vector, the probability density p must remain unchanged.

More generally, let $g(\boldsymbol{x})$ be a real or complex homogeneous and isotropic random field in n dimensions. It is characterized by the probability distribution that a particular realization $g(\boldsymbol{x})$ can occur – note that we do not distinguish notationally between the random field and a particular realization, though these are two very different concepts. Hence, let $p(g(\boldsymbol{x}))\,\mathrm{d}g$ be the probability for the occurrence of the realization g within $\mathrm{d}g$, where $\mathrm{d}g$ is the volume element in function space. Let $\langle X \rangle$ denote the ensemble average of a quantity X, that is, we imagine to have many realizations of this random field with the same statistical properties, and we average X over all these realizations. Formally,

$$\langle X \rangle = \int \mathrm{d}g\; p(g(\boldsymbol{x}))\, X \;. \tag{118}$$

We shall assume that

$$\langle g(\boldsymbol{x}) \rangle = 0 \;, \tag{119}$$

so that the expectation value of g at every position $\boldsymbol{x}$ vanishes. Consider the (two-point) *correlation function*

$$\langle g(\boldsymbol{x}) g^*(\boldsymbol{y}) \rangle = C_{gg}(|\boldsymbol{x} - \boldsymbol{y}|) \;. \tag{120}$$

The correlation function can only depend on the separation $\boldsymbol{x} - \boldsymbol{y}$ of the two points because the homogeneity of the field g means that the correlator cannot depend on $\boldsymbol{x}$ and $\boldsymbol{y}$ individually. Furthermore, it depends only on $|\boldsymbol{x} - \boldsymbol{y}|$ because g is an isotropic random field. Note that C_{gg} is a real function, even if g is complex, as can be seen by taking the complex conjugate of (120), which is equivalent to interchanging $\boldsymbol{x}$ and $\boldsymbol{y}$, thus leaving the right-hand side unaffected. We define the Fourier transform of $g(\boldsymbol{x})$ as

$$\hat{g}(\boldsymbol{k}) = \int_{\mathbb{R}^n} \mathrm{d}^n x\; g(\boldsymbol{x})\, \mathrm{e}^{\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{k}} \quad ; \quad g(\boldsymbol{x}) = \int_{\mathbb{R}^n} \frac{\mathrm{d}^n k}{(2\pi)^n}\; \hat{g}(\boldsymbol{k})\, \mathrm{e}^{-\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{k}} \;. \tag{121}$$

We shall now calculate the ensemble average of $\langle\hat{g}(\boldsymbol{k})\,\hat{g}^*(\boldsymbol{k}')\rangle$, by inserting the Fourier representation,

$$\langle\hat{g}(\boldsymbol{k})\,\hat{g}^*(\boldsymbol{k}')\rangle = \int_{\mathbb{R}^n} \mathrm{d}^n x\; \mathrm{e}^{\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{k}} \int_{\mathbb{R}^n} \mathrm{d}^n x'\; \mathrm{e}^{-\mathrm{i}\boldsymbol{x}'\cdot\boldsymbol{k}'}\,\langle g(\boldsymbol{x})g^*(\boldsymbol{x}')\rangle\ . \tag{122}$$

Using (120) and substituting $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{y}$, this becomes

$$\begin{aligned}\langle\hat{g}(\boldsymbol{k})\,\hat{g}^*(\boldsymbol{k}')\rangle &= \int_{\mathbb{R}^n} \mathrm{d}^n x\; \mathrm{e}^{\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{k}} \int_{\mathbb{R}^n} \mathrm{d}^n y\; \mathrm{e}^{-\mathrm{i}(\boldsymbol{x}+\boldsymbol{y})\cdot\boldsymbol{k}'}\, C_{gg}(|\boldsymbol{y}|) \\ &= (2\pi)^n \delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}') \int_{\mathbb{R}^n} \mathrm{d}^n y\; \mathrm{e}^{\mathrm{i}\boldsymbol{y}\cdot\boldsymbol{k}}\, C_{gg}(|\boldsymbol{y}|) \\ &=: (2\pi)^n \delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}') P_g(|\boldsymbol{k}|)\ ,\end{aligned} \tag{123}$$

where in the second step the x-integration was performed, and the final equality defines the *power spectrum of the quantity* g which obviously depends only on the modulus of $\boldsymbol{k}$. Hence, the power spectrum and the correlation function are Fourier transform pairs,

$$P_g(|\boldsymbol{k}|) = \int_{\mathbb{R}^n} \mathrm{d}^n y\; \mathrm{e}^{\mathrm{i}\boldsymbol{y}\cdot\boldsymbol{k}}\, C_{gg}(|\boldsymbol{y}|)\ . \tag{124}$$

Since the Fourier transform $\hat{g}(\boldsymbol{k})$ describes the same random field as $g(\boldsymbol{x})$, one can characterize the properties of the random field by the probability for the occurrence of a realization with Fourier transform $\hat{g}(\boldsymbol{k})$. As was done above for the real-space distribution, one can also discretize $\hat{g}$ on a grid in $\boldsymbol{k}$-space, denoted by $g_{\boldsymbol{k}}$.

A *Gaussian random field* is characterized by the properties that (1) the Fourier components $g_{\boldsymbol{k}}$ are mutually statistically independent, and that (2) the probability density for $g_{\boldsymbol{k}}$ is described by a Gaussian. The second property follows in many cases from the first, due to the central limit theorem. The first property implies that the phases of different Fourier components are mutually independent. A Gaussian random field is fully described by its power spectrum; a particular realization of such a random field can be obtained by drawing Gaussian deviates with dispersion $\sigma(\boldsymbol{k}) = \sqrt{P(|\boldsymbol{k}|)}$ and Fourier transforming the resulting $\hat{g}(\boldsymbol{k})$.

Gaussian random fields are almost universally used to describe the properties of the density perturbations in the early Universe. This is partly due to the argument given above, that the central limit theorem suggests that if the primordial perturbations were generated in a stochastic way (such as predicted from the inflationary theories), the resulting density field should be Gaussian. Another reason is that Gaussian random fields have very simple and convenient properties, which can be derived from the preceding results: The probability distribution of any linear combination of the random variable $g(\boldsymbol{x}_i)$ is a Gaussian, and more general, the joint probability distribution of a number M of linear combinations of the random variables $g(\boldsymbol{x}_i)$ is a multivariate Gaussian. In fact, this property can be used to define a Gaussian random field.

The power spectrum. Defining $\hat{\delta}(\boldsymbol{k},t)$ to be the Fourier transform of the density fluctuation field, then (116) immediately yields that $\hat{\delta}(\boldsymbol{k},t) = D_+(t)\hat{\delta}(\boldsymbol{k},t_0)$, which implies for the corresponding power spectrum

$$P_\delta(k,t) = D_+^2(t)P_0(k)\;, \tag{125}$$

where $P_0(k)$ is the linearly extrapolated power spectrum which would be the true present-day power spectrum if the fluctuations follows the linear evolution characterized by (116). Furthermore, the factorization of $\hat{\delta}(\boldsymbol{k},t)$ implies that each Fourier mode evolves independently in time.

There are several situations where (116) is invalid. The obvious one is when the density contrast approaches unity, where the linearization of the evolution equations breaks down. We shall discuss this case further below. Second, since the comoving horizon scale grows in time, the characteristic comoving length scale $\lambda = 2\pi/k$ of each Fourier mode was larger than the horizon size some times in the past. For such superhorizon fluctuations, Newtonian theory of gravity necessarily breaks down, and one needs to use linear perturbation theory of the Einstein equations. Third, for $a \lesssim a_{\rm eq}$ radiation dominated the matter contents of the Universe, which affects the growth of structure. Fourth, particle populations with an appreciable intrinsic velocity dispersion will not simply fall into the potential wells, but can stream away from them; this certainly applies to all relativistic species. One distinguishes between Cold Dark Matter (CDM), where the characteristic particle velocities are highly non-relativistic, $\sigma_v \ll c$, at the time when $a = a_{\rm eq}$, and Hot Dark Matter (HDM), when the matter particles are relativistic at this epoch. If the Universe is dominated by HDM, small-scale fluctuations would be smeared out due to free-streaming, and the first objects to form would be clusters or superclusters of galaxies. Since the large-scale matter distribution obtained for such models are very different in several respects from the observed one, it is concluded that HDM (such as massive neutrinos) can only contribute a very small fraction to $\Omega_{\rm m}$. The favored model is one which is dominated by CDM.

The effects of radiation domination in the early Universe and the initial superhorizon scale of density modes can be summarized as follows: Suppose at some very early time $t_{\rm i}$ when all Fourier modes of interest had scales much larger than the horizon scale then, the power spectrum of the density fluctuations was $P_{\rm i}(k)$. The power spectrum at some later time when all scales of interest are much smaller than the horizon is then

$$P_\delta(k,t) = T^2(k)\,\frac{D_+^2(t)}{D_+^2(t_{\rm i})}P_{\rm i}(k)\;, \tag{126}$$

where the *transfer function* $T(k)$ accounts for the aforementioned effects. It can be calculated, and accurate approximation formulae for it are available (e.g., Bardeen et al. 1986; Eisenstein & Hu 1998). For large scales, i.e., small k, $T \approx 1$. For large k, $T(k) \propto k^{-2}$ in a CDM Universe [$T(k)$ decreases

exponentially with k for a HDM model]. The transition between the two regimes depends on the scale of the comoving horizon at the time of matter-radiation equality, i.e. on $d_{\rm H}(a_{\rm eq})$. As cosmological length scales are measured in $h^{-1}\,$Mpc, this length is $\propto (\Omega_{\rm m}h)^{-1}$. Thus, the shape of the transfer function is determined by the *shape parameter*

$$\Gamma_{\rm spect} = h\,\Omega_{\rm m}\,\exp\left[-\Omega_{\rm b}\left(1+\sqrt{2h}\Omega_{\rm m}^{-1}\right)\right]\,, \tag{127}$$

and the final factor yields a small correction which accounts for the baryonic contribution (with density parameter $\Omega_{\rm b}$) to the density of the Universe. As the shape of the power spectrum, which is measurable, has imprinted on it the comoving horizon scale $d_{\rm H}(a_{\rm eq})$, this scale is actually observable.

Knowing the transfer function, one needs a prescription for the power spectrum $P_{\rm i}(k)$ at some very early times to predict P_δ for later stages of the evolution. Since at $t_{\rm i}$ all modes of relevance are much larger than the horizon scale, there is no characteristic length scale available; therefore, one assumes that the primordial power spectrum was a power law, $P_{\rm i}(k)\propto k^n$. Furthermore, if it is assumed that the total power of the fluctuations at the time when their scale equals the horizon size is independent of k, and for that matter, independent of time, then $n=1$. Such a primordial power spectrum is called Harrison–Zeldovich power spectrum, and is also the favored value in theories which explain the origin of primordial fluctuations as initial quantum fluctuations blown up in a period of exponential expansion, the inflation period.

The linear power spectrum is thus determined in terms of n and the shape parameter, except for the overall normalization. Several methods exist to fix this normalization, three of which are mentioned here.

1. Normalization by density fluctuations in a sphere. The relative fluctuations of the galaxy number density in the local Universe, $\delta n/n$, is of order unity if one considers spheres of radius $R=8h^{-1}\,$Mpc. If one assumes that the galaxies accurately trace the underlying dark matter field, this observation would imply that the fluctuation field $\delta(\boldsymbol{x},t_0)$, averaged over a scale of $R=8h^{-1}\,$Mpc, has a dispersion of 1. However, there is no guarantee that the galaxy number density field closely follows the dark matter distribution (see Fig. 15). Nevertheless, one might suspect that the galaxy density is large at those locations where the dark matter density is also large. In particular, galaxies might find it easier to form in peaks of the dark matter distribution, and therefore galaxies can be clustered more than the dark matter. One often summarizes our ignorance about the relative distribution of galaxies and dark matter into a *linear bias factor* b, defined such that the fluctuations of the galaxy number density are a factor b larger than the fluctuations of the underlying dark matter distribution (for a detailed discussion on biasing, see, e.g., Bardeen et al. 1986; Kauffmann, Nusser & Steinmetz 1997). We define the density field

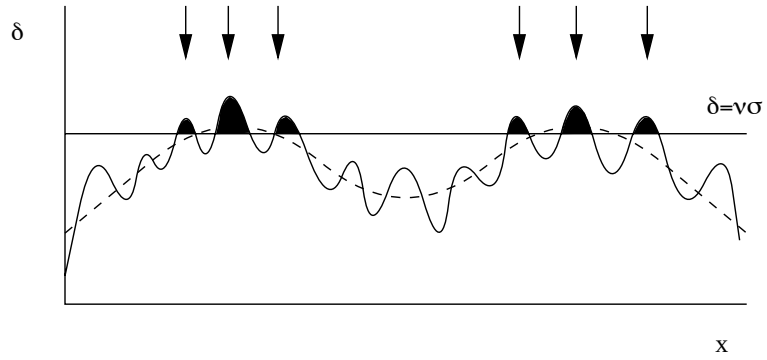


Fig. 15. This figure depicts a one-dimensional cut through a density field which contains power on small and large scales, the latter being shown by the dashed curve. If it is assumed that galaxies form only at locations where the density exceeds a critical threshold, here indicated by the horizontal line, then it is easily seen that they are more strongly clustered than the matter field itself, as this threshold is exceeded predominantly at the peaks of the long wavelength perturbations; therefore, this would lead to a bias of the galaxy distribution relative to that of the underlying matter (from Peacock 2003)

smoothed on a scale R by

$$\delta_R(\boldsymbol{x}) = \int_{\mathbb{R}^3} \mathrm{d}^3 y \; \delta_0(\boldsymbol{y})\, W_R(|\boldsymbol{x}-\boldsymbol{y}|)\;, \tag{128}$$

where $W_R(x)$ is a filter function, normalized such that $\int \mathrm{d}^3 x W_R(x) = 1$. From the convolution theorem for Fourier transforms, one finds $\hat{\delta}_R(\boldsymbol{k}) = \hat{\delta}_0(\boldsymbol{k})\, \hat{W}_R(\boldsymbol{k})$, and the power spectrum of the smoothed field is $P_R(k) = |\hat{W}_R(k)|^2\, P_0(k)$. The dispersion of the smoothed density field is then

$$\sigma^2(R) = \left\langle \delta_R^2(\boldsymbol{x}) \right\rangle = \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\, P_R(k) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3}\, \left|\hat{W}_R(k)\right|^2 P_0(k)\;, \tag{129}$$

where we made use of (124). Note that $\sigma(R)$ describes the dispersion of the smoothed version of the linearly extrapolated density field today. If we take a so-called top-hat filter, which is constant for $x \leq R$ and zero otherwise, one has

$$W_R(x) = \frac{3}{4\pi R^3} \mathrm{H}(R-x) \quad \longrightarrow \quad \hat{W}_R(k) = 3\frac{\sin kR - kR\cos kR}{(kR)^3}\;. \tag{130}$$

Coming back to the normalization, the dispersion in the galaxy number counts then implies that

$$\sigma(8h^{-1}\mathrm{Mpc}) \equiv \sigma_8 \approx \frac{1}{b}\;. \tag{131}$$

2. Normalization through the CMB. The DMR experiment on the COBE satellite mission (see, e.g., Bond 1996; Smoot 1997 for reviews) has detected the anisotropy of the microwave background on scales above $\sim$ 5 degrees. The degree of anisotropy is proportional to the fluctuation spectrum, and thus can be used directly to normalize the spectrum. The normalization of the power spectrum is hampered by the uncertainty whether the CMB anisotropy on large angular scales is caused solely by scalar perturbations (i.e., density and associated adiabatic temperature fluctuations), or whether tensor perturbations (gravitational waves) have been present at the recombination epoch. On smaller scales, the expected contribution from tensor modes becomes very small. The first year data of the WMAP satellite (Bennett et al. 2003) have provided a much more accurate determination of the power spectrum normalization, yielding $\sigma_8 = 0.9 \pm 0.1$ (Spergel et al. 2003). Since the CMB anisotropies measured by WMAP probe inhomogeneities on considerably larger scales than $8h^{-1}\,\mathrm{Mpc}$, translating their amplitudes into a value of σ_8 depends on the shape of the power spectrum. In Sect. 6.3, we shall give the best current estimates of the whole set of cosmological parameters.
3. Normalization by the local abundance of clusters. The number density of clusters as a function of their mass can be estimated analytically in terms of the power spectrum $P_\delta(k)$, as will be shown later. By comparing the observed number density of clusters with these prediction, the normalization of the power spectrum is determined (e.g., Eke et al. 1996). This comparison yields normalizations which, when expressed in terms of σ_8, are of the form

$$\sigma_8 \approx 0.52\Omega_\mathrm{m}^{-0.52+0.13\Omega_\mathrm{m}} \quad \text{for} \quad \Omega_\mathrm{m} + \Omega_\Lambda = 1\,. \tag{132}$$

   These estimates are relatively insensitive to the shape of the power spectrum (i.e., of Γ_spect), because the mass contained in a massive cluster is about the mass contained in a comoving sphere of radius $8h^{-1}\,\mathrm{Mpc}$, so that the cluster abundance directly measures σ_8. However, there has been some recent claims that (132) may overestimate σ_8 (e.g., Viana et al. 2002, and references therein). The main problem of the cluster normalization is to obtain a well-selected sample of clusters (e.g., from X-ray surveys) and to determine their masses reliably.

As we shall discuss in WL, lensing by the large-scale structure provides a powerful tool to determine the normalization of the power spectrum. As for clusters, this method yields, to lowest order, a degeneracy between σ_8 and the density parameter Ω_m.

To summarize, a CDM-dominated Universe has a (linear) power spectrum given by (126), which is determined by a few parameters. Together with the assumption that the primordial density field was Gaussian, then the evolved field will also be Gaussian as long as it stays in the linear regime. Thus, in a

statistical sense, the density field is fully specified, so that this cosmological model is predictive.

Non-linear evolution. When $|\delta|$ is no longer much smaller than unity, the linear perturbation theory breaks down. The first idea, to consider higher-order perturbations, does not really solve the problem: the perturbation series is not converging, and only slightly larger perturbations of δ can be described satisfactorily. In addition, the fluid equations cease to be valid, since due to the converging velocity field, streams of matter start to intersect, and thus the Vlasov equations needs to be employed. A different approach, Lagrangian perturbation theory (Zeldovich 1970; Buchert & Ehlers 1993) is substantially more successful. However, with the advent of high-speed computers with large memory, the need for (semi)analytic approximations decreases, as the evolution of the density field can be obtained from N-body simulations tracing the dark matter particles (e.g. White et al. 1987; Pearce et al. 2001; Navarro et al. 2004). Each such simulation yields an evolved realization from an initially Gaussian density field with power spectrum according to (126), starting at a high redshift. Such simulations can be either used directly to study the properties of the matter distribution, or can be used to derive fit formulae for various quantities of interest, some of which will be discussed below.

Using a scaling argument, Hamilton et al. (1991) derived an approximate equation relating the linearly evolved power spectrum to the fully non-linear power spectrum P_δ; this equation contains a single function, whose parameters can be fitted to the results of N-body simulations. This approximation, later generalized and refined by Jain, Mo & White (1995) and Peacock & Dodds (1996), is truly remarkable as it yields an accurate description of the fully non-linear power spectrum for all values of k and t; example power spectra are displayed in Fig. 16. More recently, an even more accurate expression has been obtained by Smith et al. (2003).

6.2 Halo abundance and profile

Gravitationally bound objects like galaxies and clusters are of course highly non-linear structures in the Universe – their average density contrast is much larger that unity. Nevertheless, there are analytical approaches to determine their number density as a function of mass and redshift. The best known of these is the Press–Schechter approach, and more refined ones are deviates of it.

The mass function of halos. The Press & Schechter (1974) approach is based on two considerations: the time evolution of a spherical overdensity and its collapse, and the statistical (Gaussian) property of the initial density field.

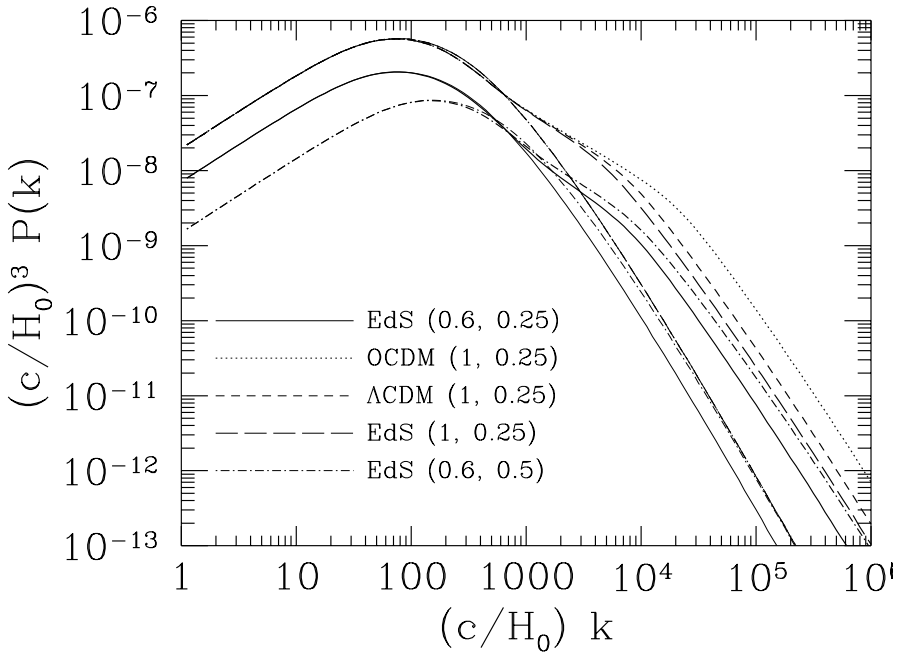


Fig. 16. The power spectrum of the cosmic density fluctuations at current epoch, for various cosmological models. $(c/H_0)^3 P(k)$ is plotted as a function of $(c/H_0)k$. The cosmological models are distinguished by line types [with values in parenthesis denoting $(\sigma_8, \Gamma_{\rm spect})$], with thin lines displaying the linearly extrapolated power spectrum, and thick lines the fully non-linear power spectrum following the prescription of Peacock & Dodds (1996). The solid lines correspond to an Einstein-de Sitter Universe, the dotted lines to an open Universe with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0$, the short-dashed lines to a flat low-density Universe with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$; those three models are approximately normalized by the present-day cluster abundance, as discussed at the end of Sect. 6.1, and have the shape parameter $\Gamma_{\rm spect} = 0.25$. The remaining two models are EdS cosmologies, with different shape parameter or different normalization. The linear power spectrum depends only on σ_8 and $\Gamma_{\rm spect}$ (once the primordial slope $n = 1$ is fixed), so that those three models are degenerate. This degeneracy is broken in the non-linear spectrum. The non-linear spectrum deviates from the linear prediction at $(c/H_0)k \geq 1000$, corresponding to length scales of $L = 2\pi/k \leq 20h^{-1}$ Mpc

The spherical collapse model considers an overdensity with spherically symmetric density distribution. According to Birkhoff's theorem, the evolution of a mass shell M is independent of the radial density profile at larger radii, as long as shells of matter do not cross each other. The radius of the mass shell as a function of time then follows an equation of motion. At early time, when the average density contrast δ inside the mass shell is small, the

expansion closely follows the Hubble expansion, but being slightly slower, the density contrast grows. This leads to an increased deviation from the Hubble expansion, and thus further increased density contrast. Provided the latter is large enough, the expansion of the mass shell can come to a hold (at time $t_{\rm max}$, say), and the radius will decrease from there on: the shell is collapsing. The symmetry of the equation of motion with respect to $t \to -t$ then predicts that the collapse of the mass shell to very small radii takes the same time as the expansion, so that the shell collapses at time $t_{\rm coll} = 2t_{\rm max}$. If the mass distribution was exactly symmetric, the collapse would indeed proceed to basically a single point; however, small-scale inhomogeneities of the matter distribution will deflect the matter particles from their radial orbit, thereby enhancing the density fluctuations, and very quickly the orbits of particles become randomized. During this process the mass overdensity will virialize, and this process takes place with a time scale comparable to $t_{\rm coll}$ (violent relaxation; Lynden-Bell 1967; Binney & Tremaine 1987). The final state is then a spherical halo of (dark) matter in nearly virial equilibrium.

In an EdS model, the various parameters of this model can be calculated analytically: in order for the collapse to occur before the present time, $t_{\rm coll} \le t_0$, the linearly extrapolated mean density contrast δ_0 of the mass shell must satisfy $\delta_0 \ge \delta_{\rm c} = 3(12\pi)^{2/3}/20 \simeq 1.69$, and the condition that the collapse happened before redshift z is that $\delta_0 \ge \delta_{\rm c}(1+z)$. The mean density of the virialized halo is $\langle\rho\rangle = 18\pi^2\rho_{\rm cr}(1+z_{\rm coll})^3 \simeq 178(1+z_{\rm coll})^3\rho_{\rm cr}$. Note that $(1+z_{\rm coll})^3\rho_{\rm cr}$ is the critical density of the EdS Universe at redshift $z_{\rm coll}$. Hence, the mean overdensity of a virialized halo is of order 200 times the critical density of the Universe at the time of formation. For other cosmological parameters, these numbers change, but have been calculated (e.g., Eke et al. 1996). Given the idealization of the spherical model, one often defines the virial radius $r_{\rm vir}$ of a dark matter halo to be the radius within which the mean density is 200 times the critical density of the Universe.

Next, consider the linear density field at some early time being smoothed with a top-hat filter of comoving radius R [see (128)]; this corresponds to a mass inside the filter scale of $M = 4\pi R^3\rho_{\rm m0}/3$. A peak with density contrast $\delta_R \ge \delta_{\rm c}(1+z)$ in this smoothed density field will then collapse before redshift z to a virialized halo of mass M. Given that the linear density fluctuations are assumed to be Gaussian, one can calculate the abundance of peaks exceeding a certain threshold, and therefore the abundance of halos of a given mass (determined by the filter scale R) that form before redshift z. This then yields the Press–Schechter mass formula for the comoving density $n(M,z)\,{\rm d}M$ of halos of mass within ${\rm d}M$ of M at redshift z,

$$n(M,z) = -\frac{2\Omega_{\rm m}\rho_{\rm cr}}{\sqrt{2\pi}M}\,\frac{\delta_c(z)}{\sigma^2(R)}\,\frac{{\rm d}\sigma(R)}{{\rm d}M}\,\exp\left(-\frac{\delta_{\rm c}(z)}{2\sigma^2(R)}\right)\,, \qquad (133)$$

where the radius R is related to the mass by the equation given above, $\sigma(R)$ is given by (129), and $\delta_{\rm c}(z)$ is the linearly extrapolated density contrast needed

for a mass shell to collapse before redshift z. The mass spectrum of halos behaves approximately as a power law for masses $M \lesssim M_*(z)$, where $M_*(z)$ is the mass scale at redshift z at which the density field becomes non-linear; it is defined implicitly through $\sigma^2(R_*)D_+^2(z) = 1$ [cf. (129), with the linear power spectrum at redshift z being $D_+^2(z)$ times the one today]. For masses above $M_*(z)$, the mass function decreases exponentially. The redshift evolution of the mass function depends on the cosmological parameters: in low-density Universes, the evolution with redshift is slower than in an EdS Universe. Thus, for a given abundance of cluster-mass halos today, the expected number of massive clusters at high redshift is much smaller for an EdS model than for a low-density Universe (see Fig. 14). The normalization of the matter power spectrum through the local cluster abundance, as discussed in Sect. 6.1, is based on the prediction of the Press–Schechter function (133) or variants thereof.

Comparison of the mass function with those obtained from N-body simulations (see Fig. 17), in which halos can be identified using a variety of techniques, leads to the conclusion that, although the Press–Schechter formula provides a very useful approximation of the halo abundance, it slightly overpredicts the number of halos with mass $\lesssim M_*(z)$ and underpredicts those with $M \gtrsim M_*(z)$. Various refinements to the original Press–Schechter approach have been conducted, including the collapse of ellipsoidal mass overdensities (Sheth & Tormen 1999). Jenkins et al. (2001) provided an accurate fit to the halo abundance obtained from their numerical models; it is very similar to the one obtained by Sheth & Tormen, and shares the simplicity of the Press–Schechter formula.

The Press-Schechter approach can also be generalized to include statistical information about the merger history of dark matter halos (e.g., Bond et al. 1991; Lacey & Coles 1993). These merger histories form the starting point for semi-analytic models of galaxy formation and evolution (e.g., Kauffmann et al. 1993, 1994); see Figs. 17 and 18.

The 'universal' density profile. From the numerical simulations, one can investigate the density profile of typical dark matter halos. Navarro, Frenk & White (1997; hereafter NFW) found that the density, averaged over spheres, of dark matter halos is described by a 'universal' profile given by

$$\rho(r) = \frac{\delta_{\rm c}\rho_{\rm cr}(z)}{(r/r_{\rm s})(1+r/r_{\rm s})^2} \,, \tag{134}$$

which is shallower than isothermal (r^{-2}) near the halo center and steeper than isothermal for $r \gtrsim r_{\rm s}$. The *virial radius* is denoted by r_{200} and is the radius inside which the mean mass density of the halo equals $200\rho_{\rm cr}(z)$, where $\rho_{\rm cr}(z) = \frac{3H^2(z)}{8\pi G}$ is the critical density of the Universe at the redshift of the halo. Hence, r_{200} immediately yields the mass of the halos, $M = 200\rho_{\rm cr}(z)\,4\pi r_{200}^3/3$. The ratio of the virial radius r_{200} and the scale

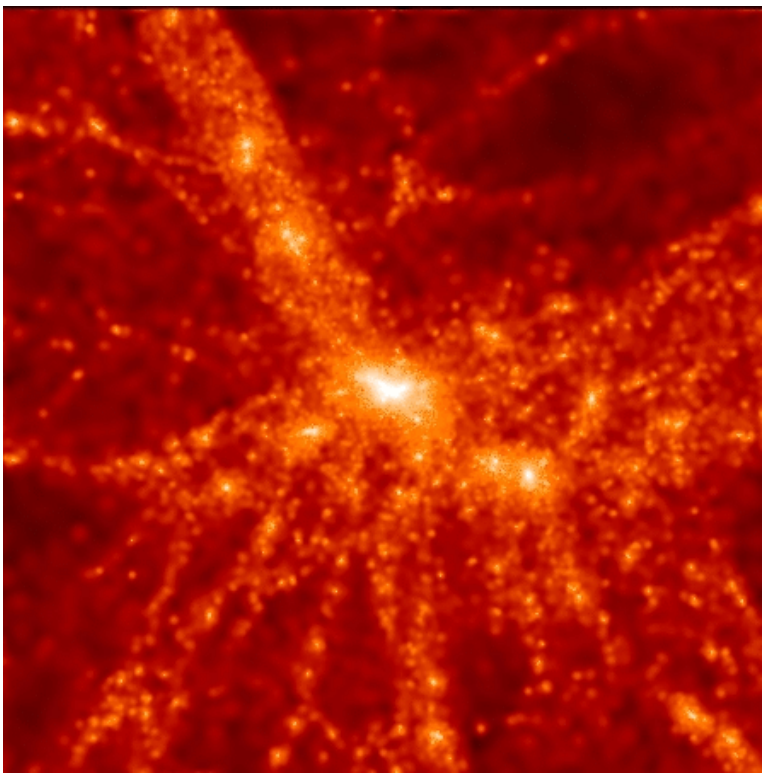


Fig. 17. One of the clusters obtained from N-body simulations. Shown is the dark matter distribution at redshift $z = 0$ in a region of $21 \times 21 \times 8 \left(h^{-1}\mathrm{Mpc}\right)^3$. The strongly structured mass in and around the cluster is clearly visible; it has been formed through successive mergers of subclusters and groups (taken from the GIF collaboration; Kauffmann et al. 1999a,b)

radius r_s is called the *concentration parameter* $c = r_{200}/r_\mathrm{s}$. From the definition of r_{200}, the parameter δ_c can be related to the concentration parameter through

$$\delta_\mathrm{c} = \frac{200}{3} \frac{c^3}{\ln(1+c) - c/(1+c)} \,. \tag{135}$$

NFW found that the concentration parameter depends on the mass of the halo; it is smaller for higher-mass halos. Takada & Jain (2002) found for the concentration parameter the dependence $c = c_0(1+z)^{-1}[M/M_*(z=0)]^{-\beta}$, with $c_0 \sim 10$ and $\beta \sim 0.2$.

There is no general agreement on the true 'universality' of the NFW profile; different groups obtain slightly different profile slopes for $r \to 0$, whereas the behavior $\rho \propto r^{-3}$ for large r is reproduced by other teams as well.

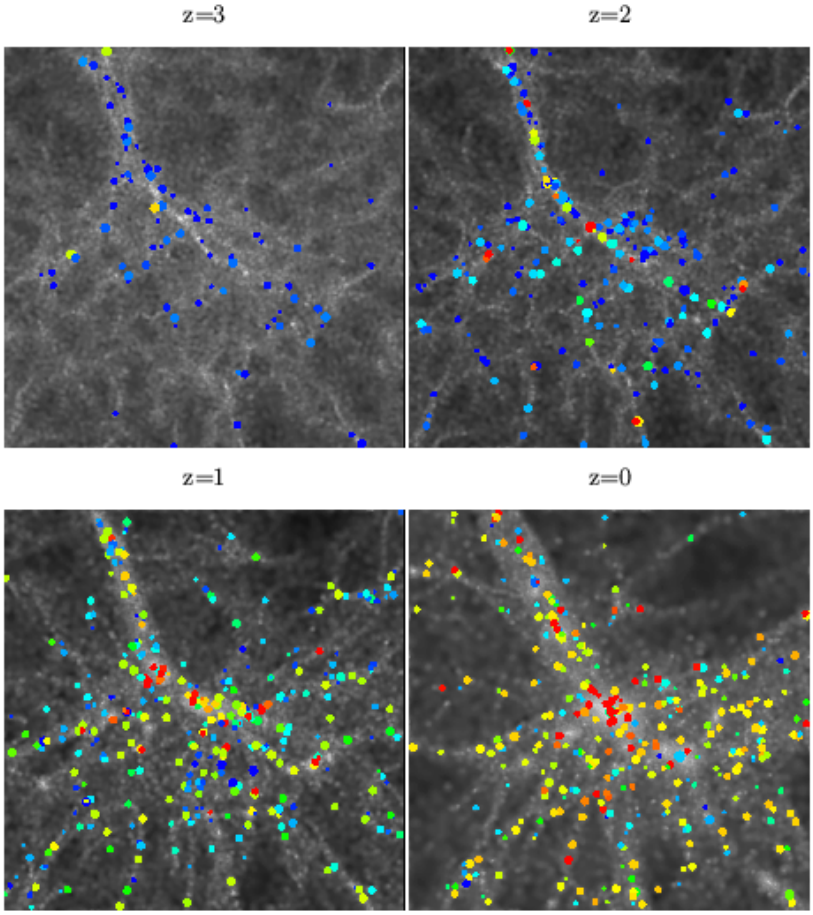


Fig. 18. The redshift evolution of the galaxy distribution, obtained by semi-analytic modeling based on the dark matter distribution as obtained from N-body simulation. The matter distribution is shown as grey scales and for $z = 0$ is the same as that shown in Fig. 17. Colors indicate the mean age of the stellar population in these galaxies, with blue (red) indicating a young (old) population; red galaxies are seen to be preferentially found in clusters. At high redshifts, there are of course no old stellar populations (taken from the GIF collaboration; Kauffmann et al. 1999a,b)

The NFW gravitational lens. The gravitational lensing properties of the NFW profile has been discussed by Bartelmann (1996), Golse & Kneib (2002) and others. By projecting (134) along the line-of-sight, the surface mass density can be written as

$$\kappa(\theta) = \kappa_k f(\theta/\theta_{\rm s}) \ , \tag{136}$$

where $\theta_{\rm s} = r_{\rm s}/D_{\rm d}^{\rm ang}$ is the angular scale radius,

$$\kappa_k = \frac{2 r_{\rm s} \delta_{\rm c} \rho_{\rm cr}(z)}{\Sigma_{\rm cr}} \tag{137}$$

the characteristic surface mass density, and

$$f(x) = \frac{1}{x^2 - 1} \left[1 - \mathcal{F}(x) \right] \ , \tag{138}$$

where

$$\mathcal{F}(x) = \frac{\mathrm{acosh}(1/x)}{\sqrt{1 - x^2}} \ ; \quad \mathcal{F}(x) = \frac{\mathrm{acos}(1/x)}{\sqrt{x^2 - 1}} \tag{139}$$

for $x < 1$ and $x > 1$, respectively; taking the limit $x \to 1$ in both cases yields $f(1) = 1/3$. Similarly, the mean surface mass density $\bar{\kappa}(\theta)$ inside θ is given by $\bar{\kappa}(\theta) = \kappa_k h(\theta/\theta_{\rm s})$, with

$$h(x) = \frac{2}{x^2} \left[\mathcal{F}(x) + \ln\left(\frac{x}{2}\right) \right] \ , \tag{140}$$

with $h(1) = 2[1 - \ln(2)]$. The absolute value of the shear is then, as usual, $\gamma = \bar{\kappa} - \kappa$. The surface mass density diverges logarithmically as $\theta \to 0$; therefore, the NWF lens is critical and thus has a tangential and radial critical curve. Since the deflection angle $\alpha = \theta \bar{\kappa}$ is a smooth function also at $\theta = 0$, the NFW lens produces either one or three images, i.e. the peculiarities of the SIS model do not occur here.

6.3 The concordance model

The past few years have seen great advances in the determination of the cosmological parameters, and the progress is continuing. At present, a set of cosmological parameters can be defined which seems to be in accord with all cosmological observations. Particularly notable is the fact that for each of the parameters there are at least two very different methods for its determination. Here we briefly mention the major results which led to the current concordance model, excluding the results from gravitational lensing, that will be described in the later chapters.

The main observational results which led to the current set of cosmological parameters came from the following sources:

- Anisotropies in the CMB. The CMB is nearly isotropic, but there are temperature fluctuations of order $\Delta T/T \sim 10^{-5}$ superimposed on the isotropic field (plus, there is the dipole anisotropy reflecting our peculiar motion). The *primary anisotropies* are due to density, temperature, and potential inhomogeneities at the time of recombination, together with corresponding peculiar velocities of the matter at this epoch (see Hu & Dodelson 2002 for a recent review on the physics of the CMB

anisotropies). Furthermore, anisotropies can be generated and modified during the propagation of the light from $z \sim 1100$ to today, causing *secondary anisotropies*. The angular power spectrum of these anisotropies depends on basically all cosmological parameters; therefore, their measurements in recent years have yielded a wealth of cosmological constraints. Measurements before the release of the first WMAP data were summarized and analyzed by Wang et al. (2003). The breathtaking results obtained by WMAP (Bennett et al. 2003; Spergel et al. 2003) have confirmed earlier measurements, but with substantially smaller uncertainties.

- Light element abundances. During the first three minutes after the Big Bang, the Universe was hot and dense enough to form the lightest chemical elements. Their primordial abundances depend on the baryon density $\Omega_{\rm b}$ of the Universe, as well as on the number of neutrino species which determine the expansion rate in the radiation-dominated phase. Whereas the primordial abundance of helium-4 (about 25% by mass) is fairly insensitive to $\Omega_{\rm b}$, the deuterium abundance is a very strong function of $\Omega_{\rm b}$. In the past few years, observations of intergalactic clouds in the form of the Lyman-α absorption lines in high-redshift quasars have yielded determinations of the deuterium-to-hydrogen abundance (e.g., Tytler et al. 2000). These measurements are extremely valuable, since this intergalactic material is thought to be fairly unprocessed chemically, and thus still reflects the primordial abundance ratios.
- Type Ia supernovae. This type of supernova explosions is believed to origin from the white dwarfs which just exceed their maximum possible (Chandasekhar) mass; hence, they all would have essentially the same explosion energy, which makes them excellent candidates for standard candles. In fact, although their maximum luminosity shows an intrinsic spread, this variation is correlated with the characteristic width of the light curve, which has been used for an empirical correction of the maximum luminosity; after this correction, the remaining spread in their peak luminosities is very small. No redshift evolution in their intrinsic properties (such as rest-frame colors or spectra) has been found. Hence, by measuring the flux at maximum light of SN Ia at different redshifts, one can measure the luminosity distance as a function of z; on the other hand, the luminosity distance depends on the cosmological parameters $\Omega_{\rm m}$ and Ω_Λ. Two teams have systematically searched for high-redshift supernovae (Schmidt et al. 1998; Perlmutter et al. 1999), and constructed the redshift–distance relation from their events, extending up to $z \sim 1$. By now, many SN Ia have been found even with redshifts $\gtrsim 1$ (e.g. Riess et al. 2004); the analysis of their brightness shows the expected behavior for a Universe which is currently accelerating, but has be decelerating before $z \sim 0.7$, as expected in a model with $\Omega_\Lambda \sim 0.7$, $\Omega_{\rm m} \sim 0.3$, where the transition to vacuum domination has occurred rather recently.

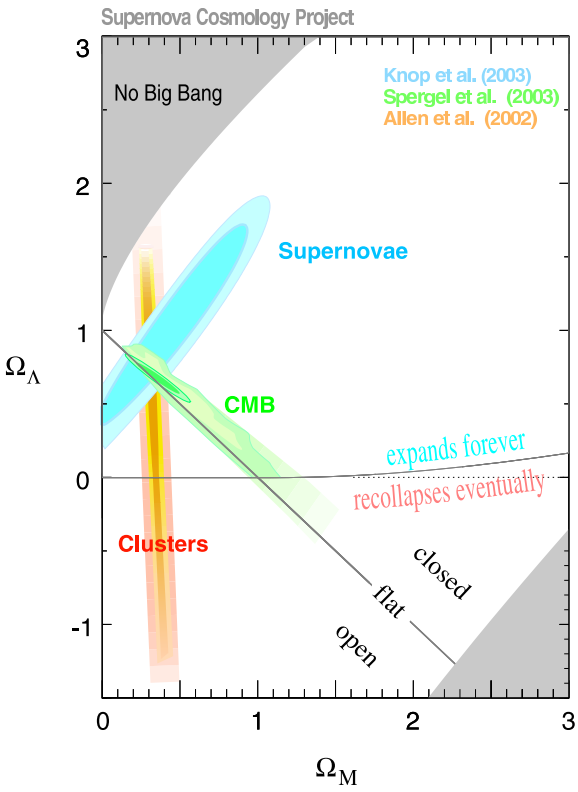


Fig. 19. This figure summarizes several constraints in the $\Omega_{\rm m} - \Omega_\Lambda$-plane, from the WMAP measurement of the CMB, the abundance of clusters, and the high-redshift SN Ia. Each of these cosmological test probe very different physics; nevertheless, they yield consistent results – this form the basis of the concordance model (taken from Knop et al. 2003)

- Large scale structure. The relation between the distribution of galaxies and the underlying dark matter distribution is not known; on the other hand, it is at least plausible that they follow each other closely. In particular on large spatial scales, say where the density field is still in the linear regime ($L \gtrsim 10h^{-1}\,$Mpc), one assumes that the number density fluctuations of galaxies is proportional to that of the dark matter, with the proportionality factor being called bias factor b. Then, the power spectrum of the galaxy distribution is b^2 times the matter power spectrum. Large two-dimensional and three-dimensional (i.e. redshift) galaxy surveys have recently been performed (in particular, the 2dF galaxy redshift survey; see e.g. Hawkins et al. 2003; the Sloan Digital Sky Survey, e.g, Tegmark et al. 2004) to construct the galaxy power spectrum. From that, the shape of the matter power spectrum, i.e., the shape parameter $\Gamma_{\rm spect}$, can be determined, among other parameters (e.g., Peacock 2003).
- Statistics of the Ly-α forest. The spectra of all QSOs show a dense ensemble of absorption lines shortward of the Ly-α emission line. These absorption lines are due to the inhomogeneous distribution of intergalactic hydrogen. At high redshift ($z \sim 3$) when these lines are observable in the optical spectrum of QSOs, the corresponding density field of the gas

is still in the linear regime. One therefore expects that the gas follows the underlying dark matter closely. The gas is in photoionization equilibrium with the UV radiation field, and obeys a simple temperature-density relation. Most of the unknown physical parameters can be put into a mean absorption which can be measured from the flux decrement across the Ly-α emission line. The statistics of the Ly-α absorbers therefore probes the corresponding matter density fluctuation spectrum. Recently, large samples of QSO spectra became available for this kind of analysis; see Kim et al. (2004) and Seljak et al. (2004) for recent results.

- Cosmology from galaxy clusters. Clusters provide a wealth of cosmological information: their abundance depends strongly on the normalization σ_8 of the power spectrum, as mentioned in Sect. 6.1, the evolution of their abundance with redshift probes the rate of growth of structure, which in turn depends on the density parameters, and their correlation function probes the shape of the power spectrum on large scales. In addition, clusters are so large that one can assume their baryon-to-mass ratio f_b being very similar to the cosmological mean of this ratio. Since the baryon contents of clusters can be measured from their X-rays, and their mass can be determined by X-rays, dynamics of their member galaxies and by gravitational lensing, this baryon fraction can be determined and yields $f_\mathrm{b} \approx 0.17$, with rather little scatter between clusters.

From these and several other methods, the set of cosmological parameters can be determined. It must be realized that the various parameters are correlated in a given data set, and sometime rather degenerate [such as seen in (132)]. Estimates of one parameter need to be obtained by marginalizing over the remaining ones, and the estimated error bars will depend on how many parameters were considered in the analysis. This is not the place to discuss these issues; we therefore present the currently 'best' estimates and approximate 1-σ error bars of the relevant parameters. In Fig. 19, some of the constraints on the density parameters are summarized, and Fig. 20 illustrates the concordance in the determination of the power spectrum from a large variety of different methods.

- The Hubble constant, as determined from the Cepheid distances within the Hubble Key Project (Freedman et al. 2001) and from combining CMB data with galaxy redshift surveys, is

$$H_0 \approx 71\,\mathrm{km\,s^{-1}\,Mpc^{-1}}\ ,\ \text{or}\quad h \approx 0.71 \pm 0.04, \tag{141}$$

  where the error from the CMB plus LSS is formally smaller than from the Hubble Key Project, by about a factor of two.
- From the deuterium abundance in QSO absorption lines, as well as from WMAP combined with LSS and Ly-α statistics, the baryonic density parameter is estimated to be

$$h^2\Omega_\mathrm{b} \approx 0.023 \pm 0.002\ , \tag{142}$$

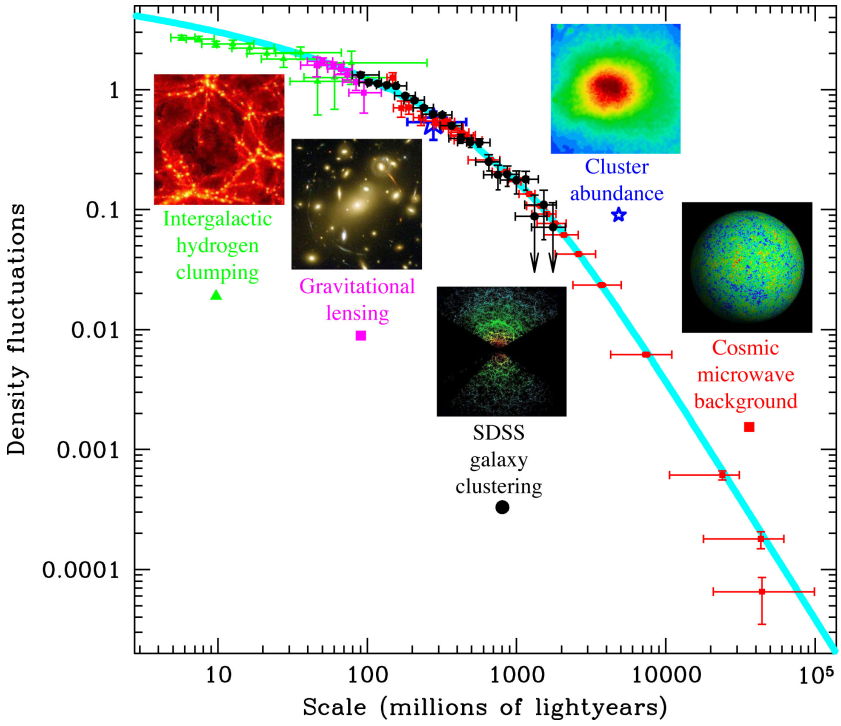


Fig. 20. The power spectrum of density fluctuations, as obtained from a variety of measurement methods. The CMB anisotropies measure the fluctuations on the largest spatial scale. Next come measurements from the clustering properties of galaxies, as obtained in galaxy redshift surveys. The cluster abundance provides a measurement of the fluctuation power near $\sim 10h^{-1}$ Mpc, i.e. close to the scale where σ_8 is defined. Cosmic shear and the statistical properties of the Ly-α forest provide reliable measurements on small spatial scales (taken from M. Tegmark's homepage, based on Tegmark et al. 2004)

where we give a slightly larger error than quoted in some recent papers. Again, consistent results are obtained from totally different methods.

- The CMB anisotropies constrain the total density of the Universe to be very close to unity,
$$\Omega_{\rm m} + \Omega_\Lambda \approx 1.02 \pm 0.02 \; ; \tag{143}$$
combining the results from SN Ia studies with the evolution of the cluster abundance, a similar conclusion is obtained, though with a larger error estimate.
- The supernovae projects yield a joint constraint on the density parameters as is shown in Fig. 19, and hence by themselves require, a non-zero cosmological constant.

- The 2dF and SDSS galaxy redshift surveys determine the shape parameter; in particular, from the 2dFGRS one finds

$$\Gamma_{\rm spect} \approx \Omega_{\rm m} h \approx 0.18 \pm 0.02 \,, \tag{144}$$

  while the SDSS yield a slightly larger value.
- CMB anisotropies, together with the LSS, yield a value for the matter density parameter of

$$\Omega_{\rm m} = 0.29 \pm 0.04 \,, \tag{145}$$

  a value that is in excellent agreement from cluster abundance evolution, the determination of the shape parameter $\Gamma_{\rm spect}$ from the LSS, and the baryon fraction $f_{\rm b}$ in clusters, when combined with the baryon density $\Omega_{\rm b}$ and the value for the Hubble constant.
- The CMB anisotropy also determines the slope of the primordial density fluctuation spectrum, which turns out to be close to the Harrison–Zeldovich value, $n \approx 0.98 \pm 0.02$.
- Perhaps the parameter with the largest discrepancies between different methods is the normalization of the power spectrum; we quote here the value from Seljak et al. (2004), obtained by combining WMAP with the SDSS galaxy redshift survey and the Ly-α forest analysis,

$$\sigma_8 = 0.89 \pm 0.04 \,. \tag{146}$$

- The shape of the power spectrum, as shown in Fig. 20, is sufficiently well determined to rule out any significant contribution of Hot Dark Matter to the energy budget of the Universe. Translated into an upper bound on the sum of neutrino masses, this constraint reads

$$\sum m_\nu \lesssim 0.5\,{\rm eV} \,, \tag{147}$$

  an upper limit that is better by a factor of about 10 for the electron neutrino, and tremendously much better for the other two neutrino species, than those obtained from laboratory measurements.

6.4 Challenges

One cannot finish a section on cosmology without pointing out the impressive developments that we are currently witnessing, and some of their implications. The concordance model that we have described in the previous section is indeed a remarkable achievement, if one considers the huge variety of methods and processes that have entered the determination of its parameters. There was no a priori guarantee that all of this would fit together. Constraints obtained from nuclear physics about one minute after the Big Bang are in agreement with those from the distribution of galaxies in the local Universe!

This concordance model has a large impact on other branches of physics, most noticeably particle physics. The tight constraints on neutrino masses obtained from cosmology is the most obvious example of a strong connection between these two branches of fundamental science. Even more important are the clear astronomical and cosmological evidences on the existence of non-baryonic dark matter which most likely is in the form of some as yet unknown species of elementary particles. This challenge has triggered a large number of underground experiments for a direct search for these dark matter particles. On the other hand, the next generation of particle accelerators may cross the energy threshold where new physics will be discovered. Many elementary particle physicist would put their eggs into the basket labeled Super-Symmetry, a theory which would provide a 'natural' candidate for the Cold Dark Matter particle, the neutralino.

Whereas one sees a possible solution to the Dark Matter problem, the situation is quite different with respect to vacuum energy, or the cosmological constant, or Dark Energy – this variety of names already tells a lot about how well this dominant component of the Universe is understood. Simple estimates of the density of vacuum energy from quantum field theory are barely 120 orders of magnitudes off, and we do not know why. It is not understood why the vacuum energy density is essentially zero (compared to the simple estimates), but in addition, why it is then not *exactly* zero. A constant vacuum energy density is not the only 'model' discussed for the Dark Energy; different equations of state $p = w\,\rho\,c^2$ cannot be ruled out by the current data, except that $w \lesssim -0.7$ at the present epoch. Obtaining tighter constraints on the equation of state of the Dark Energy from astronomical observations is probably the only way to investigate it empirically. The existence of this component to the cosmic energy budget arguably provides the largest challenge to fundamental physics, and its solution will almost certainly involve a unification of the laws of gravity and quantum mechanics – the long-sought theory of quantum gravity.

The concordance model also has made inflation a part of the standard model. Invented some 20 years ago, inflation provides a solution to the flatness problem (why the Universe has a total density parameter that is within an order of magnitude around unity), the horizon problem (why the CMB temperature on two opposite sides of the sky is the same within $\sim 10^{-5}$), and the apparent absence of magnetic monopoles and other topological defects. The model implies that the Universe underwent an early phase of exponential expansion, some 10^{-32} s after the Big Bang, before a phase transition ('reheating') brought it back on track for normal Friedmann expansion. In this model, the initial density fluctuations in the Universe are quantum fluctuations, inflated to macroscopic scales in the exponential expansion phase. The predictions of inflation, including that the Universe is nearly perfectly flat ($—\Omega_{\rm m} + \Omega_\Lambda - 1| \ll 1$) and that the primordial fluctuation spectrum is very close to the Harrison–Zeldovich form, $1 - n \ll 1$, have been impressively

verified with the recent cosmological observations. Probing the physics of inflation, e.g. through the presence of primordial gravitational waves which would leave an observable imprint on the polarization of the CMB, is one of the challenges of future cosmological studies.

7 Final remarks

These notes are an extended version of two introductory lectures given at the beginning of the Saas-Fee course; they were intended to bring the participants up to speed on topics on which much of the rest of the course rested. Overlap with some of the later material was unavoidable, but given the different character and temperament of the three lecturers, maybe even not undesired. Some of this overlap will surely be present in these write-ups; hopefully, our readers do not mind.

Acknowledgments. I enjoyed the week of lecturing at Les Diablerets a lot. I thank Georges Meylan and his colleagues Philippe Jetzer and Pierre North for organizing this school so efficiently and smoothly. Furthermore, I thank my fellow lecturers Chris Kochanek and Joe Wambsganss for helpful comments on this manuscript, discussions, great company and good spirits. Thanks to Marusa Bradac for detailed comments on this manuscript. This work was supported by the German Ministry for Science and Education (BMBF) through the DLR under the project 50 OR 0106, by the German Ministry for Science and Education (BMBF) through DESY under the project 05AE2PDA/8, and by the Deutsche Forschungsgemeinschaft under the project SCHN 342/3–1.

References

Alcock, C., Akerloff, C.W., Allsman, R.A. et al. 1993, Nat 365, 621
Arp, H. 1987, *Quasars, redshifts, and controversies*, Interstellar Media, Berkeley
Aubourg, E., Bareyre, P., Brehin, S. et al. 1993, Nat 365, 623
Bacon, D.J., Refregier, A.R. & Ellis, R.S. 2000, MNRAS 318, 625
Bardeen, J.M., Bond, J.R., Kaiser, N. & Szalay, A.S. 1986, ApJS 304, 15
Barnothy, J.M. 1965, AJ 70, 666
Bartelmann, M. 1996, A&A 313, 697
Bartelmann, M. & Schneider, P. 2001, Phys. Rep. 340, 291
Bennett, C.L., Halpern, M., Hinshaw, G. et al. 2003, ApJS 148, 1
Binney, J. & Tremaine, S. 1987, *Galactic Dynamics*, Princeton University Press
Blandford, R.D. & Narayan, R. 1986, ApJ 310, 568
Blandford, R.D. & Narayan, R. 1992, ARA&A 30, 311
Bond, J.R. 1996, in: *Cosmology and large-scale structure*, Schaeffer, R., Silk, J., Spiro, M. & Zinn-Justin, J. (eds.), Les Houches, Session LX, Elsevier:North Holland, p. 469
Bond, J.R., Cole, S., Efstathiou, G. & Kaiser, N. 1991, ApJ 379, 440
Bourassa, R.R. & Kantowski, R. 1975, ApJ 195, 13

Bourassa, R.R., Kantowski, R. & Norton, T.D. 1973, ApJ 185, 747
Brainerd, T.G., Blandford, R.D., & Smail, I. 1996, ApJ, 466, 623
Broadhurst, T. & Lehár, J. 1995, ApJ 450, L41
Browne, I.W.A., Wilkinson, P.N., Jackson, N.J.F. et al. 2003, MNRAS 341, 13
Buchert, T. & Ehlers, J. 1993, MNRAS 264, 375
Burke, B.F., Lehàr, J. & Conner, S.R. 1992, in: *Gravitational Lenses*, R. Kayser, T. Schramm & L. Nieser (eds.), (Springer-Verlag: Berlin), Lecture Notes in Physics, 406, p.237
Burke, W.L. 1981, ApJ 244, L1
Canizares, C.R. 1982, ApJ 263, 508
Canizares, C.R. 1981, Nat 291, 620
Carroll, S.M., Press, W.H. & Turner, E.L. 1992, ARA&A 30, 499
Chang, K. & Refsdal, S. 1979, Nat 282, 561
Chang, K. & Refsdal, S. 1984, A&A 132, 168
Chwolson, O. 1924, AN 221, 329
Cooke, J.H. & Kantowski, R. 1975, ApJ 195, L11
Courbin, F., Saha, P. & Schechter, P.L. 2002, in: *Gravitational Lensing: An Astrophysical Tool*, F. Courbin & D. Minniti (eds.), (Springer-Verlag: Berlin), Lecture Notes in Physics 608, p. 1
Einstein, A. 1936, Sci 84, 506
Eisenstein, D.J. & Hu, W. 1998, ApJ 496, 605
Eke, V.R., Cole, S. & Frenk, C.S. 1996, MNRAS 282, 263
Etherington, I.M.H. 1933, Philos. Mag. 15, 761
Falco, E.E., Gorenstein, M.V. & Shapiro, I.I. 1985, ApJ 289, L1
Fort, B. & Mellier, Y. 1994, A&AR 5, 239
Fort, B., Prieur, J.L., Mathez, G., Mellier, Y. & Soucail, G. 1988, A&A 200, L17
Freedman, W.L., Madore, B.F., Gibson, B.K. et al. 2001, ApJ 553, 47
Frenk, C.S., White, S.D.M., Bode, P. et al. 1999, ApJ 525, 554
Friedmann, A. 1922, Z. Phys. 10, 377
Golse, G. & Kneib, J.-P. 2002, A&A 390, 821
Gorenstein, M.V., Cohen, N.L., Shapiro, I.I. et al. 1988a, ApJ 334, 42
Gorenstein, M.V., Falco, E.E. & Shapiro, I.I. 1988b, ApJ 327, 693
Hamilton, A.J.S., Kumar, P., Lu, E. & Matthews, A. 1991, ApJ 374, L1
Harvanek, M., Stocke, J.T., Morse, J.A. & Rhee, G. 1997, AJ 114, 2240
Hawkins, E., Maddox, S., Cole, S. et al. 2003, MNRAS 346, 78
Hewitt, J.N., Turner, E.L., Schneider, D.P., Burke, B.F., Langston, G.I. & Lawrence, C.R. 1988, Nat 333, 537
Hu, W. & Dodelson, S. 2002, ARA&A 40, 171
Hu, E.M., Cowie, L.L., McMahon, R.G. et al. 2002, ApJ 568, L75
Ibata, R.A., Lewis, G.F., Irwin, M.J., Lehár, J. & Totten, E.J. 1999, AJ 118, 1922
Impey, C.D., Falco, E.E., Kochanek, C.S. et al. 1998, ApJ 509, 551
Irwin, M.J., Ibata, R.A., Lewis, G.F. & Totten, E.J. 1998, ApJ 505, 529
Irwin, M.J., Webster, R.L., Hewett, P.C., Corrigan, R.T. & Jedrzejewski, R.I. 1989, AJ 98, 1989
Jain, B., Mo, H.J. & White, S.D.M. 1995, MNRAS 276, L25
Kaiser, N., Wilson, G. & Luppino, G. 2000, astro-ph/0003338
Jenkins, A., Frenk, C., White, S. et al. 2001, MNRAS 321, 372
Kauffmann, G., Guiderdoni, B. & White, S.D.M. 1994, MNRAS 267, 981
Kauffmann, G., White, S.D.M. & Guiderdoni, B. 1993, MNRAS 264, 201

Kauffmann, G., Nusser, A. & Steinmetz, M. 1997, MNRAS 286, 795
Kauffmann, G, Colberg, J.M, Diaferio, A. & White, S.D.M. 1999a, MNRAS 303, 188
Kauffmann, G, Colberg, J.M, Diaferio, A. & White, S.D.M. 1999b, MNRAS 307, 529
Kayser, R. & Refsdal, S. 1983, A&A 128, 156
Kayser, R., Refsdal, S. & Stabell, R. 1986, A&A 166, 36
Keeton, C.R. & Kochanek, C.S. 1998, ApJ 495, 157
Kim, T.-S., Viel, M., Haehnelt, M.G., Carswell, R.F. & Cristiani, S. 2004, MNRAS 347, 355
King, L.J., Jackson, N., Blandford, R.D. et al. 1998, MNRAS 295, L41
King, L.J., Browne, I.W.A., Marlow, D.R., Patnaik, A.R. & Wilkinson, P.N. 1999, MNRAS 307, 225
Klimov, Yu.G. 1963, Soviet Phys. Doklady 8, 119
Kneib, J.-P., Ellis, R.S., Santos, M.R. & Richard, J. 2004, ApJ 607, 697
Knop, R.A., Aldering, G., Amanullah, R. et al. 2003, ApJ 598, 102
Kochanek, C.S. 2004, in: Kochanek, C.S., Schneider, P., Wambsganss, J., *Gravitational Lensing: Strong, Weak & Micro*, Proceedings of the 33rd Saas-Fee Advanced Course, G. Meylan, P. Jetzer & P. North, eds. (Springer-Verlag: Berlin) (SL), also: astro-ph/0407232
Kochanek, C.S., Blandford, R.D., Lawrence, C.R. & Narayan, R. 1989, MNRAS 238, 43
Kolb, E.W. & Turner, M.G. 1990, *The early universe*, Addison-Wesley Publ. Comp., Redwood
Kundić, T., Colley, W.N., Gott, J.R. et al. 1995, ApJ 455, L5
Kundić, T., Turner, E.L., Colley, W.N. et al. 1997, ApJ 482,75
Lacey, C. & Cole, S. 1993, MNRAS 262, 627
Langston, G.I., Schneider, D.P., Conner, S. et al. 1989, AJ 97, 1283
Laplace, P.S. 1795, *Exposition du système du monde.*
Lehár, J., Hewitt, J.N., Roberts, D.H. & Burke, B.F. 1992, ApJ 384, 453
Liddle, A.R. & Lyth, D.H. 2000, *Cosmological Inflation and Large-Scale Structure*, Cambridge University Press
Liebes Jr., S. 1964, Phys. Rev. 133, B835
Lodge, O.J. 1919, Nat 104, 354
Lynden-Bell, D. 1967, MNRAS 136, 101
Lynds, R. & Petrosian, V. 1986, BAAS 18, 1014
Mao, S. 1992, ApJ 389, 63
Maoz, D., Bahcall, J.N., Schneider, D.P. et al. 1993, ApJ 409, 28
Mattig, W. 1958, AN 284, 109
Mellier, Y. 1999, ARA&A 37, 127
Michalitsianos, A.G., Dolan, J.F., Kazanas, D. et al. 1997, ApJ 474, 598
Narayan, R. & Bartelmann, M. 1999, in: *Formation of Structure in the Universe*, A. Dekel & J.P. Ostriker (eds.), (Cambridge University Press: Cambridge), p.360
Navarro, J., Frenk, C. & White, S. 1997, ApJ 490, 493
Navarro, J.F., Hayashi, E., Power, C. et al. 2004, MNRAS 349, 1039
Ohanian, H.C. 1983, ApJ 271, 551
Paczyński, B. 1986a, ApJ 301, 503
Paczyński, B. 1986b, ApJ 304, 1
Paczyński, B. 1987, Nat 325, 572

Paczyński, B. 1996, ARA& 34, 419
Padmanabhan, T. 1993, *Structure Formation in the Universe*, Cambridge University Press
Peacock, J.A. 1982, MNRAS 199, 987
Peacock, J.A. 1999, *Cosmological Physics*, Cambridge University Press
Peacock, J.A. 2003, astro-ph/0309240
Peacock, J.A. & Dodds, S.J. 1996, MNRAS 280, L19
Pearce, F.R., Jenkins, A., Frenk, C.S. 2001, MNRAS 326, 649
Peebles, P.J.E. 1993, *Principles of Physical Cosmology*, Princeton University Press
Pelló, R., Schaerer, D., Richard, J., Le Borgne, J.-F. & Kneib, J.-P. 2004, A&A 416, L35
Perlmutter, S., Aldering, G., Goldhaber, G. et al. 1999, ApJ 517, 565.
Petters, A.O., Levine, H. & Wambsganss, J. 2001, *Singularity Theory and Gravitational Lensing* (Birkhäuser: Boston) (PLW)
Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. 1992, *Numerical Recipes*, Cambridge University Press, Cambridge
Press, W.H. & Schechter, P. 1974, ApJ 187, 425
Refsdal, S. 1964a, MNRAS 128, 295
Refsdal, S. 1964b, MNRAS 128, 307
Refsdal, S. & Surdej, J. 1994, Rep. Prog. Phys. 56, 117
Renn, J., Sauer, T. & Stachel, J. 1997, Sci 275, 184
Riess, A.G.; Strolger, L.-G., Tonry, J. et al. 2004, ApJ 607, 665
Schild, R.E. 1990, AJ 100, 1771
Schmidt, B.P.; Suntzeff, N.B.; Phillips, M.M. et al. 1998, ApJ 507, 46.
Schneider, P. 1984, A&A 140, 119
Schneider, P. 1985, A&A 143, 413
Schneider, P. 1987, A&A 183, 189
Schneider, P. 1992, A&A 254, 14
Schneider, P. 2004, in: Kochanek, C.S., Schneider, P., Wambsganss, J., *Gravitational Lensing: Strong, Weak & Micro*, Proceedings of the 33rd Saas-Fee Advanced Course, G. Meylan, P. Jetzer & P. North, eds. (Springer-Verlag: Berlin) (WL)
Schneider, P., Ehlers, J. & Falco, E.E. 1992, *Gravitational Lenses* (Springer-Verlag: Heidelberg), 2nd printing 1999 (SEF)
Schneider, P. & Seitz, C. 1995, A& 294, 411
Schneider, P. & Weiss, A. 1987, A&A 171, 49
Schneider, P. & Weiss, A. 1992, A&A 260, 1
Schramm, T. 1990, A&A 231, 19
Seitz, S., Saglia, R.P., Bender, R., Hopp, U., Belloni, P. & Ziegler, B. 1998, MNRAS 298, 945
Seitz, S. & Schneider, P. 1992, A&A 265, 1
Seitz, S., Schneider, P. & Ehlers, J. 1994, Class. Quantum Gravity 11, 2345
Seljak, U., Makarov, A., McDonald, P. et al. 2004, astro-ph/0407372
Sheth, R. & Tormen, G. 1999, MNRAS 308, 119
Smith, R.E., Peacock, J.A., Jenkins, A. et al. 2003, MNRAS 341, 1311
Smooth, G.F. 1997, in: *The Cosmic Microwave Background*, C.H. Lineweaver, J.G. Bartlett, A. Blanchard, M. Signore & J. Silk (eds.), Kluwer Academic Press, p. 185
Soldner, J. 1804, *Berliner Astron. Jahrb. 1804*, p. 161
Soucail, G., Fort, B., Mellier, Y. & Picat, J.P. 1987, A&A 172, L14

Soucail, G., Mellier, Y., Fort, B., Mathez, G. & Cailloux, M. 1988, A&A 191, L19
Spergel, D.N., Verde, L., Peiris, H.V. et al. 2003, ApJS 148, 175
Springel, V., White, S.D.M., Tormen, G. & Kauffmann, G. 2001, MNRAS 328, 726
Stockton, A. 1980, ApJ 242, L141
Takada, M. & Jain, B. 2002, MNRAS 337, 875
Tegmark, M., Blanton, M.R., Strauss, M.A. et al. 2004, ApJ 606, 702
Tyson, J.A. 1981, ApJ 248, L89
Tyson, J.A. 1988, AJ 96, 1
Tyson, J.A., Valdes, F. & Wenk, R.A. 1990, ApJ 349, L1
Tytler, D., O'Meara, J.M., Suzuki, N. & Lubin, D. 2000, Phys. Rep. 333, 409.
Udalski, A., Szymanski, M., Kaluzny, J. et al. 1993, Acta Astron. 43, 289
van Waerbeke, L. & Mellier, Y. 2003, astro-ph/0305089
van Waerbeke, L., Mellier, Y., Erben, T. et al. 2000, A&A, 358, 30
Vanderriest, C., Schneider, J., Herpe, G., Chevreton, M., Moles, M. & Wlérick, G. 1989, A&A 215, 1
Viana, P.T.P., Nichol, R.C. & Liddle, A.R. 2002, ApJ 569, L75
Vietri, M. & Ostriker, J.P. 1983, ApJ 267, 488
Wallington, S. & Narayan, R. 1993, ApJ 403, 517
Walsh, D. 1989, in: *Gravitational lenses*, Moran, J.M., Hewitt, J.N. & Lo, K. Y. (eds.), Lecture Notes in Physics 330 (Springer-Verlag: New York), p.11
Walsh, D., Carswell, R.F. & Weymann, R.J. 1979, Nat 279, 381
Wambsganss, J. 2004, in: Kochanek, C.S., Schneider, P., Wambsganss, J., *Gravitational Lensing: Strong, Weak & Micro*, Proceedings of the 33rd Saas-Fee Advanced Course, G. Meylan, P. Jetzer & P. North, eds. (Springer-Verlag: Berlin) (ML)
Wang, X., Tegmark, M., Jain, B. & Zaldarriaga, M. 2003, PhRvD 68, 68l3001
Weinberg, S. 1976, ApJ 208, L1
Weymann, R.J., Latham, D., Angel, J.R.P. et al. 1980, Nat 285, 641
White, S.D.M., Davis, M., Efstathiou, G. & Frenk, C.S. 1987, Nat 330, 451
Williams, R.E., Blacker, B., Dickinson, M. et al. 1996, AJ 112, 1335
Wittman, D. 2002, in: *Gravitational Lensing: An Astrophysical Tool*, F. Courbin & D. Minniti (eds.), (Springer-Verlag: Berlin), Lecture Notes in Physics 608, p. 55
Wittman, D.M., Tyson, J.A., Kirkman, D., Dell'Antonio, I. & Bernstein, G. 2000, Nat 405, 143
Yee, H.K.C. 1988, AJ 95, 1331
Yee, H.K.C., Ellingson, E., Bechtold, J., Carlberg, R.G. & Cuillandre, J.-C. 1996, AJ 111, 1783
Young, P., Gunn, J.E., Kristian, J., Oke, J.B. & Westphal, J.A. 1980, ApJ 241, 507
Young, P., Gunn, J.E., Kristian, J., Oke, J.B. & Westphal, J.A. 1981, ApJ 244, 736
Zel'dovich, Ya.B. 1970, A&A 5, 84
Zwicky, F. 1937a, Phys. Rev. 51, 290
Zwicky, F. 1937b, Phys. Rev. 51, 679