Chapter 2
Distance Measurement and Cosmography

Combining observation with mathematics can result in unexpected ways of probing the world—small and large. Examples start in Antiquity when Thales measured the height of a pyramid, using the length of its shadow at the very moment when the shadow cast by a vertical rod was as long as the rod itself. It is also told that he inferred the distance of a boat from the shore, without stretching any tape measure between them. Astronomers have made analogous things throughout history when constructing the cosmic distance ladder.

2.1 Cosmic Distance Indicators

It is common knowledge in our neighbourhood that of two similar shining objects the more distant one looks fainter and smaller, and has a smaller parallax shift. In Euclidean space, an object's *luminosity distance*, *angular size distance* and *parallax distance* have the same value when the same unit of length is used. These familiar kinds of distances are also used to construct the distance ladder beyond the Solar System, inside and even outside of our Galaxy, finally reaching the realm where a novel kind of distance measure appears: the *redshift* of light. The empirical Hubble law indicates that a higher redshift corresponds to a larger distance. Here "distance" is still a somewhat misty concept of remoteness.

With a cosmological model, the redshift and other distances are linked to the *metric distance* contained in the theoretical framework of Riemannian space. It can be said to correspond to the distance obtained with metersticks put one after the other between the observer and an object. In non-Euclidean spaces, as considered in large-scale physics, different distances have different values, giving a way to find out the geometry of space.

The Distance Ladder This concept refers to the steps making possible the measurement of progressively larger distances (see, e.g., Webb 1999).

One may say that *a distance indicator is a method where an astronomical object is placed in 3D space so that its properties observed through space agree with what*

Y. Baryshev, P. Teerikorpi, *Fundamental Questions of Practical Cosmology*, Astrophysics and Space Science Library 383,
DOI 10.1007/978-94-007-2379-5_2,

we know about such objects, their constituents, and the propagation of light. A good distance indicator would restrict the object's position in a narrow range around the true distance. For finding good distance indicators the diversity of cosmic phenomena and the knowledge and imagination of the cosmographer should meet. Thus numerous methods for "placing an object in space" have been invented, each requiring different kinds of measurements at the telescope. The most basic is to measure angles. Then there is the flux of light, and one may also measure its time variation and spectrum at different wavelengths.

2.1.1 Geometry and Photometry in Euclidean Space

In Antiquity, the term "parallax" meant the shift in the direction of a celestial body as looked from two places on Earth. In his *Almagest* Ptolemy (ca AD 150) stated that "none of the stars has a noticeable parallax (which is the only phenomenon from which distances can be derived)". One sees a sound method, but the observational means are insufficient. In fact, only after Copernicus was there strong motivation to search for stellar parallaxes, as a crucial test for proving the Earth's motion and as a distance indicator.

The Triangulation If the length of the baseline d is known in needed units and the measured peak angle in the triangle is θ, then the distance r is

$$r = \frac{d}{2\tan(\theta/2)} \approx \frac{d}{\theta}. \tag{2.1}$$

One may use two kinds of baselines. The baseline may be either "here" like the astronomical unit AU (the Earth–Sun distance) in the annual parallax method or "out there", like the radius of a galaxy or some other object.

The local baseline ($2\times$AU) is obtained in physical units using Kepler's Third Law together with distances to nearby planets and asteroids, measured with radar techniques nowadays. So a stellar parallax angle can be expressed as a distance in cm (see Appendix A for useful numerical data).

In the stellar realm in our Galaxy, the triangulation appears in various forms. The Solar System moves relative to the local standard of rest (about 13 km/s or 2.8 AU per year), so one can expand the local baseline much beyond the 2 AU of the annual parallax method and derive "secular parallaxes" (leading to a mean distance to a class of stars). Another related method derives "statistical parallaxes".

However, when the baseline is "out there", we have a problem: how to know its actual length? This critical question arises often in astronomy. Ideally, one should have a *standard rod*, a class of objects whose linear size has a small scatter around an average value. From the measured angular size of a standard rod, one can derive its angular size distance using Eq. (2.1).

The expanding nebula method, a special kind of triangulation, was suggested and first applied by Lundmark in the 1920s for the Crab nebula (Trimble 1973). Knowing the expansion velocity of the nebula in km/s (from the Doppler shift), one can calculate its distance from its transversal proper motion μ in arc seconds per year derived from the angular size and the date of explosion:

$$V_{\mathrm{r}} = 4.74\mu r, \tag{2.2}$$

where the expansion velocity V_{r} is in km/s and the distance r is in parsecs.

For the Milky Way, our distance from the Galactic centre, R_0, is fundamental. It is also important for extragalactic distances through its impact on the calibration of stellar standard candles. Earlier, the round number $R_0 \approx 10$ kpc was often used, but for years it has been known that the distance is shorter, around 8 kpc. In a summary of the results, Eisenhauer et al. (2003) reported a geometric measurement to the centre with an uncertainty of about 5%. They combined astrometric and spectroscopic measurements of the star S2 orbiting the massive black hole candidate in the Galactic centre. The solution for the best-fit Kepler orbit gave the needed parameters to calculate the distance $R_0 = 7.94 \pm 0.38$ kpc (the black hole mass turned out to be about 3.6×10^6 solar masses).

Triangulation in Nearby Extragalactic Space Water masers, observed in interstellar clouds, are strong and compact radio sources. Their positions can be measured accurately with Very Long Baseline Interferometry (VLBI) techniques. A few nearby galaxies have known water maser sources.

Brunthaler et al. (2005) made true the dream of van Maanen in the days of the Great Debate: to detect the angular rotation of M33, in order to infer its distance. Comparing the angular rotation rate, as measured with VLBI from water masers on opposite sides of the galaxy, with the rotation speed and inclination, they derived a distance of 730 ± 168 kpc. This comes from the measured angular motion of 30.8 ± 4 μarcsec/year (in RA) and the velocity of 106 ± 20 km/s from the rotation model, together with Eq. (2.2).

Miyoshi et al. (1995) observed megamasers close to the centre of the spiral galaxy NGC 4258. They could deduce the radius (0.13 pc) of the disk rotating around the massive black hole candidate from the rotational speed (1080 km/s) and centripetal acceleration (9.5 km/s/yr) of its edge. Then the angular radius (4.1 mas) gave the distance of 6.4 Mpc (revised to 7.2 ± 0.3 Mpc by Herrnstein et al. 1999).[1]

The Photometric Method: Standard Candles Before annual parallaxes were measured by Bessel and others in the 1830s, brightness had been recognized as a possible distance indicator. James Gregory (1638–1675) assumed that stars are other suns observed through transparent space where the inverse square law of light flux from a point source works (introduced in Kepler's *Optics*). Newton applied it to Sirius, using the Sun as the calibrator. He inferred a distance of one million solar distances (twice the true one) and was thus well aware of the remoteness of stars.

[1] A variant of this method applied to M33 (Argon et al. 2004) resulted in the distance 800 ± 180 kpc.

Consider an object radiating isotropically with the power output (luminosity in erg/sec) L. Then at the distance r it produces a flux f (erg/sec/cm^2) through the surface of the sphere having the total area $4\pi r^2$:

$$f = \frac{L}{4\pi r_{\mathrm{L}}^2}. \tag{2.3}$$

If one knows the luminosity L and measures the flux f, one can thus derive the *luminosity distance* r_{L}. It is good to underline that the photometric method is also geometric in nature. One essentially measures—in a special way—the solid angle made by the cone whose 1 cm^2 bottom is at the Earth and the peak is at the centre of the object in question.

Though we indicated above the units erg, sec, and cm, the method of standard candles is usually relative, i.e. one obtains the distance to an object relative a nearby calibrator. The distance modulus $m - M$ often appears in such calculations. It is related to the luminosity distance r_{L} as

$$m - M = 5 \log r_{\mathrm{L}}/10 \text{ pc}. \tag{2.4}$$

This formula has its origin in the $1/r^2$ law in Euclidean space. We remind the non-astronomer reader that the apparent magnitude is connected to the observed flux f as $m = -2.5 \log f + \text{const.}$ (e.g., Karttunen et al. 2006).

Some history: The familiar unit light-year was used before parsec which was suggested by Herbert H. Turner (1861–1930) and apparently first mentioned in text by Frank W. Dyson in 1913. The concept of absolute magnitude M was defined by the Dutch astronomer Jacobus C. Kapteyn (1851–1922) in 1902, as the apparent magnitude m which a star would have if moved to a distance corresponding to a parallax of $0.1''$, i.e. the distance of 10 pc (Hughes 2006).

2.1.2 Relative and Absolute/Physical Methods

When using a "standard rod" or a "standard candle" as a distance indicator, in both cases one must ascribe a value to the object whose distance is measured (length or luminosity). For this there are two possibilities:

- Calibration based on nearby objects of the same kind, whose distances are known using other distance indicators.
- Derivation of the linear size or luminosity for the object directly from observation and physical theory, with no information about its distance.

The first variant requiring nearby counterparts is "relative". The latter type of methods bypasses the local distance ladder, and may be described as "physical" (e.g., Sandage et al. 2006) or "absolute" (Nikiforov 2004) or "one-step distance method" (Jackson 2007).

Relative Methods The very first application of the standard candle method used the calibrator "here", the Sun. Hence, Gregory's method had 1 AU as a natural distance unit. For several distance indicators of this kind, the size or the luminosity are inferred from a relation $X = ap + b$, where p is an observable parameter whose value can be measured without knowing the distance. Examples are the period–luminosity relation of Cepheid stars, the rotational velocity—luminosity (TF) relation for spiral galaxies, and the decay time—maximum luminosity relation for Supernovae Ia.[2]

The parameter p can refer to quite different observations and time-scales. A spiral galaxy remains in its position in the TF-relation for uncounted generations of astronomers, and observations of its maximum rotation velocity and magnitude may be repeated at will. Cepheid observations are limited to relatively nearby galaxies and require a sufficient time interval, so that different pulsational periods can be observed. Supernova explosions cannot be predicted, and when they occur, one should start following the process before the peak luminosity is reached.

Absolute/Physical Methods Relative methods rely on the calibrator distances. Physical methods result in distances directly expressed in physical units (cm). Such methods are rare at large extragalactic distances. One example is the Sunyaev-Zeldovich method for distant rich galaxy clusters which utilizes the interaction of the hot cluster gas and the cosmic background radiation (Sect. 4.3.1).

Redshifts and Distances Thanks to the regular universe, the measured redshift provides an estimate of relative distances (the Hubble law). It is quite different from the indicators mentioned above. In the assumed Friedmann model, (or any other sufficiently developed cosmological model) the redshift links other kinds of distances to the fundamental metric distance. For instance, locally calibrated standard candles lead to luminosity distances d_L, which can be measured independent of the cosmological model. Then, if one assumes that a flat Friedmann model describes the universe, one gets from the luminosity distance (in Mpc) the metric distance (in Mpc) just by dividing it by the redshift factor $1 + z$.

The procedure of distance measurement may explicitly involve a cosmological model. For instance, if one needs a time interval occurring at the object, one transforms it to the rest-frame by dividing the observed time by $1 + z$, when the Friedmann model is assumed. This happens, e.g., when the decline rate is used to correct the peak luminosity of a supernova of type Ia.

[2] Also the Faber-Jackson and Fundamental Plane (Davis–Djorkovski) methods for elliptical galaxies, utilizing relations between the luminosity and stellar velocity dispersion (FJ), and the effective size, surface brightness and velocity dispersion (DD) and a modification by Dressler et al. (1987) belong to this category.

Table 2.1 Some stellar distance indicators

Type of star	Pop	Absolute V mag	Range for V = 27
RR Lyrae variables	II	0.8	2 Mpc
TRGB (Red Giants)	II	−2.5	5 Mpc
Cepheid variables	I	−5 (for $P \approx 20$ d)	25 Mpc
Supernovae type Ia	I, II	≈ -19.5	$z < \sim 2$, ($\sim$ 10 Gpc)

2.2 Lower Rungs in the Distance Ladder

Lundmark (1919, 1920) made an early determination of the distance to the Andromeda galaxy using novae (the result was 175 000 pc). He also suggested the use of the brightest stars in galaxies: "If it were possible to determine the apparent magnitude at which the main body of giants in a spiral nebula begin to appear separated, it would give us an additional means of estimating its distance".

In his 1929 classic Hubble introduced the brightest blue stars as distance indicators. He identified these from the photographs of galaxies up to about half distance to the Virgo cluster. They usefully indicated relative distances, though Sandage (1958) showed that these objects actually were HII regions, bright ionized hydrogen clouds around blue stars. Calibrated by Hubble using true stars in nearby galaxies, those indicators gave too short distances.

2.2.1 The Brightest Giant Stars in Galaxies

But is there a sharp upper limit in the luminosities of normal stars? From the HR diagram for the solar neighbourhood, Sandage and Tammann (1974) concluded that the brightest blue stars surpass $M_B = -9$ mag. In other galaxies $M(1)$ depends on the galaxy luminosity (the number of stars), roughly as expected if the stellar luminosity function is exponential at the bright end. Indeed, if the luminosity function of a class of stars does not go abruptly to zero at a high luminosity, it is clear that in a large sample of those stars the extreme one is usually more luminous than in a small sample. This makes blue supergiants problematic as distance indicators.

The trend between $M(1)$ and the host galaxy magnitude M_{gal} depends on the steepness of the bright end of the luminosity function. In some cases the apparent magnitude $m(1)$ of the brightest star and the host magnitude m_{gal} can be used to make a distance estimate. If we know the dependency $M(1) = aM_{gal} + b$, then $(1-a)\mu = -(am_{gal} + b)$, where $\mu = m(1) - M(1)$ is the unknown distance modulus. If the slope $a \approx 1$, one cannot solve for the distance.

The brightest red supergiants ($\langle M(1) \rangle \approx -8.0$ mag) also show a trend with the host galaxy luminosity, though weaker than the blue stars (Sandage 1984). They have been used to derive distances to a number of nearby galaxies with an accuracy of ≈ 0.3 mag in the distance modulus.

Instead of extremal stars within a broad class, one may consider inherent features in the luminosity function, such as a maximum or a discontinuity. These maintain their location (absolute magnitude) in samples of different sizes, when there is no selection depending on the absolute magnitude.

The TRGB Method The Hubble Space Telescope could measure the luminosity distribution of bright red stars in many galaxies (earlier studied, e.g., by Sandage 1971), making a new way of using red giants possible: the Tip of the Red Giant Branch. The modern TRGB method was developed especially by Madore, Freedman and associates (e.g., Lee et al. 1993; Sakai et al. 1997) and is now an important way to measure distances in the local volume within about 10 Mpc (e.g., Karachentsev et al. 2007), where the Population I Cepheid method is not applicable to elliptical galaxies (cf. Table 2.1). The TRGB method is also efficient in using the HST telescope time, as one can measure distances out to 7 Mpc with a single orbit (a single-epoch observation at two wavelengths is enough).

The "tip" refers to a sudden discontinuity in the luminosity function of the stars in the red giant branch of the HR diagram and it is observed at the absolute magnitude $M_I \approx -4$ in the I photometric band (around 8200 Å). This feature is understood as marking the core helium flash of old, low-mass stars (less than ~ 1 solar mass) which evolve up the red giant branch, but very quickly change their physical characteristics upon ignition of helium.

The TRGB I-magnitude has been shown to be quite stable, only slightly sensitive (~ 0.1 mag) to age 2–15 Gyr and metallicity between $-2.2 < [Fe/H] < -0.7$ dex (the range for the Galactic globular clusters). The statistical problem affecting the use of extremes in a population is reduced and it is regarded that with the TRGB method one reaches an accuracy of 0.1–0.2 mag for the distance modulus, requiring more than ~ 100 red giants detected in the one-magnitude bin below the tip.

The calibration in the I-band is based on Galactic globular clusters whose distances have been determined using the RR Lyrae variable stars. Tabur et al. (2009) have made a direct geometrical calibration of the TRGB K-band magnitude in the solar neighbourhood using parallaxes measured by the HIPPARCOS satellite.

2.2.2 Cepheid Pulsating Stars

On the cosmic distance ladder the classical Cepheid variables continue to have a special role. They are young objects of Population I found in spiral and irregular galaxies. Their predecessors in the Main Sequence are stars two times heavier than the Sun. With their (V) absolute magnitudes in the range centered around -4 (for $P = 10$ d), they can be used up to $m_{\rm lim} - M \approx 28.5$ ($r \approx 5$ Mpc) with earth-bound telescopes ($m_{\rm lim} \approx 24.5$) and $m_{\rm lim} - M \approx 31$ (≈ 25 Mpc) from space ($m_{\rm lim} \approx 27$).

Some Physics A simple argument concerning the physics of the PL relation was outlined by Sandage (1958; also Madore and Freedman 1991)—such understanding helps one to formulate the law, to add relevant parameters, and perhaps to recognize a reason for possible discrepant results.

The Stefan's law connects the luminosity, radius, and effective surface temperature of a star: $L = 4\pi R^2 \sigma T_{\mathrm{eff}}^4$, thus mapping all stars on the magnitude-temperature diagram, with stars of a constant radius defining a straight line. Such a line is also one of constant volume, so moving along it the average density of stars changes, as there is a luminosity–mass relation. A result from classical mechanics states that the radial pulsational period of a star $P \propto 1/\sqrt{\rho}$. Here ρ is the average density of the star, hence *the pulsational period of a star tells us about its density, and, at the end, about its luminosity*. For the bolometric (total luminosity) absolute magnitude, we may write the PL relation as:

$$\langle M_{\mathrm{bol}} \rangle = A \log P + B \langle \log T_{\mathrm{eff}} \rangle + \mathrm{const.} \tag{2.5}$$

In practice, instead of the total flux one observes a Cepheid through different wavelenght bands. Each has its own PL slope and zero-point. A major advance has been the near-infrared photometry, complementing the measurements in the B and V systems. In longer wavelengths (1) the extinction and reddening of the light by the dust are reduced, (2) the scatter of the PL relation is smaller, and (3) the amplitude of magnitude variation is smaller.

The PL Relation As the colour is the natural third parameter in the PL relation, reflecting the surface temperature, the precept behind the use of a simple PL relation is as follows: If one assumes that the calibrating sample of Cepheids has the same properties as the distant sample on which one applies the relation, the colour term may be considered as the same for all galaxies. The PLC relation is thus transformed into a simpler PL relation.

Soon after the discovery of the PL relation by Henrietta S. Leavitt (Leavitt and Pickering 1912), Hertzsprung applied it to distance determination. He and then Shapley used statistical parallaxes for stars in our Galaxy to derive the zero-point of the relation. In the modern situation there are several routes to the calibration. Assuming a universal PL relation, the slope of the PL relation has been usually taken from observations in the Large Magellanic Cloud, while the zero-point comes from our Galaxy, together with independent distance measurements of the LMC (see e.g. Fouqué et al. 2007). For example, the PL relation for the LMC adopted by the HST Key Project (for the distance modulus $\mu_{\mathrm{LMC}} = 18.5$) was $M_V = -2.76 \log P - 1.46$ (Freedman et al. 2001).

The HIPPARCOS astrometric satellite measured many parallaxes for Cepheids in our Galaxy. However, the large distances cause large parallax errors, which make it difficult to derive the PL zero-point in a proper statistical manner. The future GAIA astrometric space observatory will open here new possibilities.

The PL relations may be systematically different in different galaxies. It is known that later galaxy types (like the Large Magellanic Cloud) have a lower metallicity than earlier types (like our Galaxy) (Paturel et al. 2002), and the relations in the Galaxy and in the LMC may be different (Tammann et al. 2003). Hence, the Cepheid properties could be different due to metallicity of the interstellar gas from which

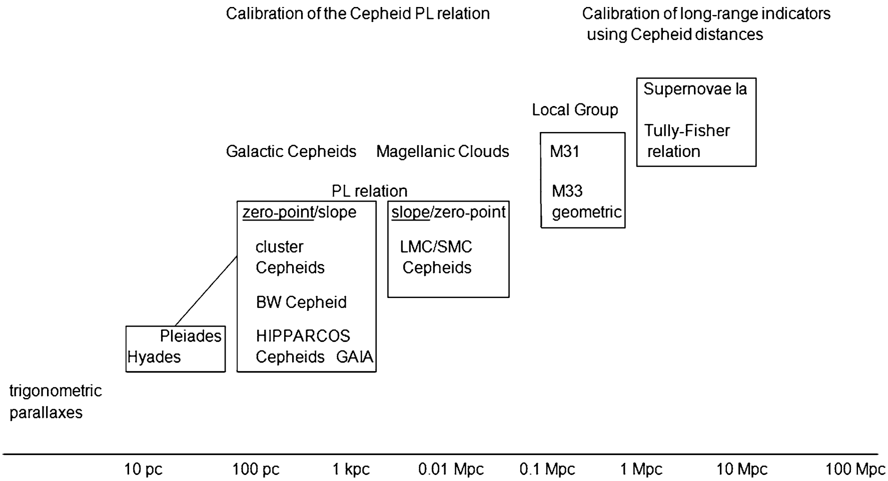

Fig. 2.1 Parts of the cosmic distance ladder involving Cepheids, the TF method, and supernovae Ia. The PL relation is calibrated in the Galaxy and in the Magellanic Clouds and checked in nearby galaxies. It is used to measure distances to galaxies and groups within about 20 Mpc, where the TF-relation and the SNe Ia luminosities are calibrated

these stars were initially formed. Such differences could influence the determination of the distance scale and the value of the Hubble constant (we return to this point in Chap. 4).

The absolute magnitude being predicted by the PL relation, the distance modulus can be derived from the mean apparent magnitude. It should be corrected for the sum of the extinctions in our own Galaxy and in the host galaxy. This is possible by using two different photometric bands (V and I; Madore and Freedman 1991). Then the corrected distance modulus μ_0 is derived from the separate moduli μ_V and μ_I as

$$\mu_0 = \frac{R_V}{R_V - R_I}\mu_I - \frac{R_I}{R_V - R_I}\mu_V . \tag{2.6}$$

The factors R_V and R_I give the ratio of the total extinction to the reddening E_{B-V} for the photometric band (V or I). They depend on the extinction law. For instance, if in round numbers $R_V \approx 3$ and $R_I \approx 2$, one has $\mu_0 \approx 3\mu_I - 2\mu_V$.

2.3 Long-Range Indicators

A common route to large extragalactic distances (as outlined in Fig. 2.1) is to use the PL relation for Cepheids to derive distances for a set of galaxies that serve to calibrate the luminosity of a long-range distance indicator. Such indicators can extend the distance ladder from the local scales of ~ 10 Mpc to scales of ~ 100 Mpc (e.g., the TF relation) or even 1000 Mpc or more (Type Ia supernovae).

2.3.1 The Tully–Fisher Relation: The Method of Rotating Galaxies

The Tully-Fisher (TF) relation is a widely applied distance indicator, with samples of thousands of spiral galaxies. Tully and Fisher (1977) discussed the relation between the 21 cm emission line of the neutral hydrogen HI, widened by the rotation of a galaxy, and the absolute magnitude.[3] The width of the HI line serves as a measure of the maximum rotational velocity $V_{\max}$. It is related to the mass of the galaxy and, hence, to its luminosity. One may get an idea of the expected relation between $V_{\max}$ and the absolute magnitude by assuming that the mass-to-luminosity ratio M/L and the surface mass density M/r^2 are constant. Then from $GM/r^2 = V_{\max}^2/r$ one can derive $L \propto V_{\max}^4$ and in terms of magnitudes $\log V_{\max} = -0.1M+$ constant, roughly as observed.

The rotation curves of spiral galaxies usually show a flat part at large distances from the centre. This is generally interpreted as indicating a dark matter halo. To understand well the TF method, we should know how the luminous and dark matter are distributed inside galaxies (Theureau et al. 1997a) and how far away from the centre the horizontal rotation is reached for different galaxy types etc. The result about the universality of galactic surface densities (Donato et al. 2009; Gentile et al. 2009), both baryonic and dark matter, within one dark halo scale-length may bring more order to this complex field.

The Slope Different TF relations appear in the literature, depending on the photometric band, but also on how they were derived and from what data. The zero-point b poses a separate problem, requiring a sample of calibrating galaxies with known distances (the zero-point may also be normalized to a value of the Hubble constant, if derived from field galaxies or galaxy clusters with known (cosmological) recession velocity). We give a few examples of the slope a in some bands ($M = a \log V_{\max} + b$). The B-band: -5.82, -7.97 (Theureau et al. 1997b; Sakai et al. 2000); the I-band: -7.6, -9.24 (Giovanelli et al. 1997; Sakai et al. 2000); the H-band: -11.03 (Sakai et al. 2000).

In the first B-magnitude slope field galaxies (with distances from the Hubble law) were used and the linear fit is made with keeping all errors in the calculated absolute magnitude M. In the second one (the HST Key Project), the fit was made on cluster galaxies and takes into account errors in both M and $\log V_{\max}$. Because of the fairly large scatter in the B relation (≈ 0.5 mag), the two kinds of fits result in rather different slopes (cf. Fig. 2.2).

The slope of the inverse relation ($\log V_{\max} = a'M + b'$) is derived assuming that all errors are in $\log V_{\max}$. For the B magnitude Ekholm et al. (1999b) and Tully and Pierce (2000) derive the slopes $-1/10$ and $-1/7.3$, respectively, while for the I magnitude Tully and Pierce give $-1/8.14$.

When one uses the TF relation for distance determination, different kinds of systematic errors appear depending on the chosen slope (direct, inverse), the nature of

[3]Even earlier Gouguenheim (1969) had inferred distances to six galaxies, using angular sizes and 21-cm line observations made with the Nancay radiotelescope.

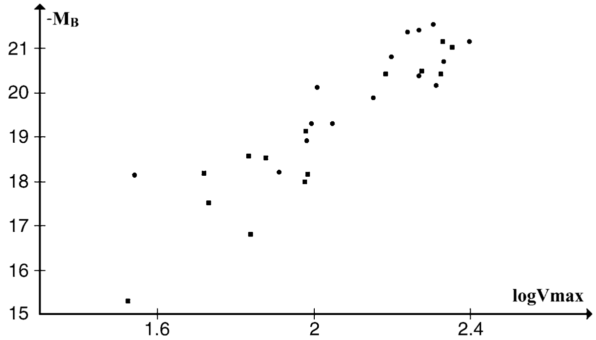

Fig. 2.2 Tully-Fisher (B magnitude) relation for nearby ($cz < 1000$ km/s) galaxies. The scatter is partly caused by distance errors. Based on calibration data in Theureau et al. (1997b)

the sample, and the situation (Chaps. 3 and 4). The systematic errors depend especially on the scatter in the relation. When one goes from shorter to longer wavelengths, the scatter decreases and may be quite small in the mid-IR range at 3.6 μm (± 0.12 mag as tentatively given by Freedman and Madore 2010).

Corrections When working with the TF relation, one has to apply several corrections, most importantly the inclination correction, the Hubble type dependence of the zero-point, and the galactic extinction correction.

Disc galaxies are viewed from different directions. One needs the inclination angle i (between the line of sight and the polar axis of the galaxy) to correct two quantities: (1) the width of the HI line (as it would be seen for $i = 90°$), and (2) the magnitude (or diameter) of the galaxy (as seen for $i = 0°$ or "face-on"). The angle i is inferred from the axis ratio $R = D/d$ of the galaxy image, where D and d are the lengths of the major and minor axes, respectively, defined up to a fixed surface brightness (isophotal sizes). The influence of inclination on the magnitude is a topic with decades of history. It is interestingly connected to the degree of transparency of galaxy discs (the greater the opacity, the larger is the change in the magnitude when the disk is viewed from different directions; e.g., Kankare et al. 2009).

Bottinelli et al. (1995) derived the following correction for B-magnitude:

$$m_i = m_0 + a_{\text{incl}}(R) = -2.5 \log(k + (1-k)R^{2C(1+0.2/K)-1}). \tag{2.7}$$

Here R is the axis ratio and k may be expressed as a function of the Hubble type t as $k = 0.754 \cdot 10^{-0.2t}$. The parameter K expresses how the apparent diameter changes with the surface brightness: $K = \partial \log D / \partial \log \mu$. The constant value $K = 0.094$ is used for disk galaxies. C is defined as $C = \partial \log D / \partial \log R$. The value $C = 0.04$ has been adopted in the LEDA extragalactic data base.

The *type dependence* is such that for a fixed $V_{\max}$ early type galaxies (Sa) are fainter than late type galaxies (or for a fixed absolute magnitude early types rotate more quickly than later types). This may be understood as due to the higher M/L ratio of the latter, as was modelled by Theureau et al. (1997a). The effect may be put into the zero-point of the TF relation.

The galactic extinction correction a_{gal} is usually taken (e.g., in the LEDA database) from Schlegel et al. (1998), based on infrared dust emission. This measure of extinction gives an average value in the direction of a galaxy and one should not forget possible deviations. There is evidence for high dust extinction in some directions even at high Galactic latitudes (e.g., Teerikorpi 1981a; Teerikorpi and Kotilainen 1989; Berdyugin and Teerikorpi 2002).

2.3.2 The Peak Luminosity for Ia Supernovae

Lundmark (1946) envisioned, on the basis of five supernovae observed in five galaxies, that supernovae seem to have a small scatter in their maximum luminosity so that they "seem to furnish an excellent distance indicator" and "will enable us to reach rather far out in the Metagalaxy". In fact, supernovae as a whole form a heterogenous blend of different origins and absolute magnitudes at their maxima. However, on the basis of their spectra one may group them into a few classes. Zwicky and Minkowski classified supernovae into Types I and II. Later Type I was divided into Ia, Ib, and Ic (all these originating from quite different processes).

Type Ia Luminosity The Ia supernovae are the most luminous of all supernovae, and they can be identified from the lack of hydrogen lines in their spectra. They form a rather uniform class of stellar explosions, in that more luminous objects have slower decline-rates (as was noticed by Pskovskii 1977). This simple behaviour allows them to be calibrated as standard candles. The peak luminosity is about -19.5 in both B and V magnitude bands, depending on the decline rate of the light curve.

The SNIa events that occur both in spiral and elliptical galaxies, have provided the sharpest Hubble relation along a wide range of distances (cf. Fig. 1.3). A drawback in supernova explosions is their unpredictability and even rareness in galaxies. Systematic search programmes are necessary so that one may catch a supernova before it reaches the maximum luminosity. We will discuss their use for cosmological purposes in later chapters (the Hubble constant in Chap. 4, the magnitude–redshift test in Chap. 8).

Lundmark (1919, 1920) had been the first to realize that there are novae and much more luminous supernovae in connection with his study of the distance of M31, with an "incredible foresight and imagination" as praised by Fritz Zwicky (Johnson 1961), who in 1934 together with Walter Baade connected supernovae with the death of massive stars. In 1925 Lundmark termed the giants and dwarfs among novae as "upper and lower class", with an upper class nova reaching a luminosity comparable to that of the whole host nebula).

2.3.3 Morphological Galaxy Classes

It would be fine if just by looking at the photo of a galaxy one could tell how big it is. Is the luminosity of a galaxy written on its appearance? This was asked by Van den Bergh (1960), after he noted that galaxies may differ in luminosity by a factor of about 10 000. He indeed found a correlation: the stronger its spiral structure is, the more luminous is a spiral galaxy. His luminosity classes range from I (supergiant) to V (dwarf galaxies), with intermediate cases (such as II–III). In the program "Steps toward the Hubble constant" by Sandage and Tammann in the 1970s, one goal was to calibrate the luminosity of the giant ScI galaxies, then use it as a standard candle. They found that the scatter within a luminosity class leads to rather inaccurate individual distances, and easily leads to a systematic error in distance determinations for a galaxy sample.

In connection with the morphology of galaxies it is interesting to mention that a half of S0 and S galaxies possess an inner ring structure (r), which is easy to observe. This finding led Buta and de Vaucouleurs (1983) to propose a distance indicator based on the size of the ring. The potentials of this method may not yet have been fully exploited (Teerikorpi 1986).

The Sosies Georges Paturel (1984) introduced the method of *sosies* into the distance ladder, whereby distant "clones" of nearby calibrators are searched for.[4] The idea is: if the absolute magnitude M depends on a measurable parameter p, but one does not know reliably the relation, one may restrict the study to such distant galaxies that have their values of p close to those of some nearby calibrators. The number of parameters p is in practice larger than one, including the morphological type and other more objectively measurable parameters. Van den Bergh's luminosity classes also define a kind of sosies, involving the Hubble type and the luminosity class index (labeled *beauty index* by Sandage). De Vaucouleurs's luminosity index Λ_c was calculated as $(T + L_c)/10$, where T is the morphological type code and L_c is van den Bergh's luminosity class number.

2.4 The Concept of Metric Distance in Curved Space

We described above some examples of practical ways to estimate cosmic distances, leaving a few other ones to later chapters. The physical concept of distance is deep-seated in the steps farther and farther in space, where one is compelled to take into account geometries more general than Euclidean. In cosmology, world models describe how distance and its measurement appear in the physical geometry of the large-scale universe. In order to facilitate understanding the meanings of different coordinates, forms of metric, and distances, we take advantage of analogies with more easily visualized spaces.

2.4.1 Universes of Constant Curvature Within E^3

We introduce main geometrical concepts in curved space so that the reader will see how Riemannian geometry enters cosmology and permits one to interpret astro-

[4] From *Larousse Classique*: Sosie—Personne avant une ressemblance parfaite avec une autre. Or—"Sosie—A person with a perfect resemblance to another."

nomical observations in terms of the geometry of the universe. For this purpose we first study the mathematics of two-dimensional homogeneous and isotropic spaces embedded in 3D Euclidean space E^3. Usually this subject is presented by differential geometry without referring to any external higher-dimensional embedding space. However, besides helping one to imagine and learn geometrical concepts, the embedding space may actually be essential for the physical concepts of distance, surface, and volume.

Systems of Coordinates for a 2D Sphere in E^3 Consider a spherical surface $\mathcal{S}^2$ in the 3D Euclidean space E^3 and suppose that a 2D inhabitant, surveyor by profession, lives in this space. He cannot feel the third dimension, similarly as the fourth spatial dimension is not present for us. The surveyor is able to study his 2D universe by measuring distances with sticks and ropes or observing light sources scattered all around space. The light travels along the geodesics ("straight lines") of this spherical universe. The external 3D observer in E^3 sees that the trajectory of light is a great circle representing the shortest distance between two points as measured on the sphere.

To see what it means to measure lengths, angles, and fluxes in the 2D spherical space, we introduce four coordinate systems (Fig. 2.3). Three of these are defined in the embedding space E^3 and one in the space $\mathcal{S}^2$ itself. The first three directly relate to the embedding space:

$$K_{\mathrm{Cart}} = \{x, y, z\}, \qquad K_{\mathrm{cyl}} = \{\rho, \phi, z\}, \qquad K_{\mathrm{sph}} = \{R, \theta, \phi\}. \tag{2.8}$$

K_{Cart} is the Cartesian coordinate system, and x, y, z are its ordinary coordinates. These can be employed only in flat Euclidean space where it is possible to extend the local orthogonality over the whole space (for instance, any two parallel lines have everywhere the same distance between them).

K_{cyl} is the cylindrical coordinate system, where ρ is the polar radius on the XY plane, ϕ is the azimuthal angle, and z is the Cartesian coordinate orthogonal to the XY plane. K_{sph} is the familiar spherical system in E^3, so that R is the radial distance, θ is the polar angle, and ϕ is the azimuthal angle. These *external* coordinate systems are of course related by

$$x = R\sin\theta\cos\phi = \rho\cos\phi; \qquad y = R\sin\theta\sin\phi = \rho\sin\phi; \qquad z = R\cos\theta. \tag{2.9}$$

Let us suppose that in the spherical space $\mathcal{S}^2$ the 2D observer at the point P can measure the angle between the light rays coming from the points A and B and also the distances to these light sources. Such a being would use a rope to measure the metric distance u from the stake pounded into the "ground" at point P. To measure the angle Φ he fixes a "zero-meridian" and the direction of increasing angle. In this manner the fourth coordinate system is generated, *internal* for the 2D spherical space:

$$K_{\mathrm{int}} = \{u, \Phi\}. \tag{2.10}$$

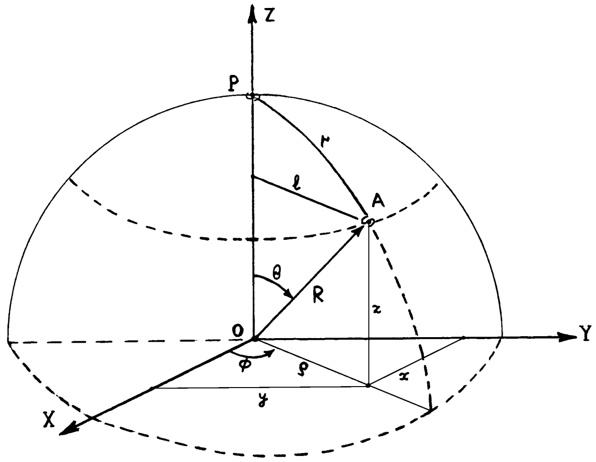

Fig. 2.3 Coordinates for 2D spherical space embedded in 3D Euclidean space

The coordinate u is the length of the arc PA of the great circle of the sphere and Φ is the angle which has the same value as the azimuthal angle ϕ in the embedding 3D Euclidean space. In his world the surveyor might, we guess, term u the polar radius and Φ the azimuthal angle.

Our one goal is to derive expressions that link the internal measurements by the 2D surveyor to the measurements made in the embedding E^3 space.

Metric Tensor and Metric Distance in E^3 The element of spatial distance dl in n-dimensional Riemannian space has the form

$$dl^2 = \gamma_{\alpha\beta} dx^\alpha dx^\beta, \tag{2.11}$$

where $\gamma_{\alpha\beta}$ is the metric tensor of the space and (the vector) dx^α gives the differentials of the coordinates $\{x^\alpha\} = \{x^1, x^2, \ldots, x^n\}$.

What is the practical importance of the metric tensor? The n-dimensional observer, who knows the mathematical expression for this tensor (i.e. how it depends on the chosen coordinate system), immediately knows all geometrical properties of his space. In particular, the metric tensor $\gamma_{\alpha\beta}$ defines the fundamental *metric distance* between any two points in that space: it is the integral of the distance element along the geodesic line connecting these points. One may identify the distance element dl with a local, very small standard rod, and the metric distance essentially means the length measured with such an unchangable "cm". When one applies this geometry to the real world, one must assume the existence of rigid rods.

In Euclidean 3D space the metric tensor assumes the simplest form when the Cartesian coordinates $\{x^\alpha\} = \{x^1, x^2, x^3\} = \{x, y, z\}$ are used:

$$\gamma_{\alpha\beta} = \mathrm{diag}(1, 1, 1). \tag{2.12}$$

Here we employ the convenient expression diag(1, 1, 1) for the 3×3 matrice where the diagonal components $\gamma_{11}, \gamma_{22}, \gamma_{33}$ are equal to 1 and all other components are zero. For such a metric tensor the distance element in the embedding space E^3 has the Pythagorean form

$$dl^2 = dx^2 + dy^2 + dz^2. \tag{2.13}$$

In another coordinate system the metric tensor and the distance element will have a different form. For example, in the spherical coordinates of the system $K_{\rm sph}$ with $\{x^\alpha\} = \{R, \theta, \phi\}$ the metric tensor is

$$\gamma_{\alpha\beta} = \mathbf{diag}(1, R^2, R^2 \sin^2\theta). \tag{2.14}$$

Hence the distance element will be

$$dl^2 = dR^2 + R^2 d\theta^2 + R^2 \sin^2\theta d\phi^2. \tag{2.15}$$

Basic Geometrical Measurements in E^3 From the metric tensor one can calculate the metric distance between two fixed points in 3D space, and the result is *independent* of the choice of the coordinate system. For instance, the distance between the origin O and the point A is derived to be

$$l_{OA} = \int_O^A dl = \int_0^{R_A} dR = R_A = \sqrt{x_A^2 + y_A^2 + z_A^2}. \tag{2.16}$$

Though the end result is trivial, this calculation illustrates how in the step from the integral of dl (valid in any coordinate system) to the integral of dR one has chosen the spherical system and hence the distance element (2.15). This coordinate system is convenient, because R varies along the straight line connecting the points O and A, and the other coordinates θ and ϕ remain constant, hence the differentials $d\theta = d\phi = 0$. The metric distance l_{OA} thus calculated is indeed the shortest distance between O and A, because the geodesic curve in Euclidean space is a straight line.

We intentionally describe applications of the metric tensor for the simple Euclidean space. This helps one to grasp the role that the metric tensor has in curved spaces. Also, Euclidean geometry is the basis for the study of the local universe and the standard with which to compare observations from the deep universe where curved space may become measurable.

Several other familiar geometrical properties of E^3 can be rigorously derived from the fundamental metric tensor $\gamma_{\alpha\beta}$. The sum Σ of the angles a, b, c of a triangle, the length $l_{\rm circ}$ of the circumference of a circle of radius R, and the surface area $A_{\rm sph}$ of a sphere of radius R are, respectively,

$$\Sigma = a + b + c = \pi, \qquad l_{\rm circ} = 2\pi R, \qquad A_{\rm sph} = 4\pi R^2. \tag{2.17}$$

The volume element for the metric tensor $\gamma_{\alpha\beta}$ having the determinant γ is

$$dV = \sqrt{\gamma} dx^1 dx^2 dx^3 = dx dy dz = R^2 \sin\theta dR d\theta d\phi. \tag{2.18}$$

It gives the volume $V_{\rm sph}$ of a sphere of radius R as $V_{\rm sph} = (4\pi/3)R^3$.

2.4.2 Metric Tensor and Distance Element in $\mathcal{S}^2$

Assume that the surveyor within his 2D spherical space is located at the "North Pole" P of the sphere that also belongs to the 3D embedding Euclidean space. P is the point where the Z-axis intersects the sphere, and it is the origin of the internal coordinate system K_{int}.

Distance in Spherical (External) Coordinates A sphere is defined in the spherical coordinates K_{sph} simply by the equation $R = \text{constant}$. In Cartesian coordinates K_{cart} this becomes

$$x^2 + y^2 + z^2 = R^2, \tag{2.19}$$

where x, y, z are the orthogonal coordinates, or the lengths of the projections of the radius vector $\vec{R} = \{x^\alpha\} = \{x, y, z\}$.

Because the sphere is a 2D manifold, any position P on it is given by two independent variables. The metric tensor and the distance element of this sphere can be expressed by these very coordinates.

By differentiating (2.19) we express the differential of z as $dz = -(xdx + ydy)/\sqrt{R^2 - x^2 - y^2}$. This allows us to eliminate z from the distance element of the embedding space $dl^2 = dx^2 + dy^2 + dz^2$, and we may write the distance element on the sphere in the Cartesian coordinates of E^3 as

$$dl^2 = dx^2 + dy^2 + \frac{(xdx + ydy)^2}{R^2 - (x^2 + y^2)}. \tag{2.20}$$

This form utilizes external Cartesian coordinates only. The following step is to express dl in the other two coordinate systems in E^3, and then to link these "external forms" to the internal coordinate system.

Distance in Cylindrical (External) Coordinates Recognizing that $x^2 + y^2 = \rho^2$ and using the azimuthal angle ϕ one may transform the distance element in Cartesian coordinates (2.13) into another form in cylindrical coordinates K_{cyl}:

$$dl^2 = \frac{d\rho^2}{1 - \frac{\rho^2}{R^2}} + \rho^2 d\phi^2. \tag{2.21}$$

Here R is the radius of the sphere in the embedding space E^3. For these coordinates $x^\alpha = (\rho, \phi)$ the metric tensor is

$$\gamma_{\alpha\beta} = \mathrm{diag}\left(\frac{1}{1 - \frac{\rho^2}{R^2}}, \rho^2\right). \tag{2.22}$$

In the spherical coordinate system K_{sph} one may express ρ using R and θ as $\rho = R\sin\theta$. Then the distance element is

$$dl^2 = R^2 d\theta^2 + R^2 \sin^2\theta d\phi^2, \tag{2.23}$$

where the range of θ is $[0, \pi]$ and that of ϕ is $[0, 2\pi]$. In this case

$$\gamma_{\alpha\beta} = \mathrm{diag}(R^2, R^2 \sin^2\theta). \tag{2.24}$$

Distance in Internal Coordinates These expressions for the distance element dl (2.20, (2.21), (2.23) and the metric tensor $\gamma_{\alpha\beta}$ (2.22), (2.24) employ the embedding Euclidean space and are quite inconvenient for the practical 2D surveyor. Therefore, it is time to show how these geometrical quantities are related to the internal coordinate system K_{int}. The internal system is simply related to the spherical and cylindrical systems:

$$u = R\theta = R \arcsin\frac{\rho}{R}\Phi = \phi. \tag{2.25}$$

The surveyor, who naturally uses the internal (u, Φ) system, will find that the distance element has the following form:

$$dl^2 = du^2 + R^2 \sin^2\frac{u}{R} d\Phi^2 \tag{2.26}$$

hence the fundamental metric tensor of the 2D spherical space is expressed in the surveyor's (u, Φ) coordinates as

$$\gamma_{\alpha\beta} = \mathrm{diag}\left(1, R^2 \sin^2\frac{u}{R}\right). \tag{2.27}$$

These expressions written in *internal* coordinates contain the *external* radius R. This is interesting: in principle, the surveyor can recognize that the world has a non-Euclidean geometry and a finite area (2D volume). The radius R can be found from measurements of a circle (see below).

From our external viewpoint the rope that the surveyor uses for distance measurements is an inseparable part of his universe and cannot be straightened through the external space. In mathematical language this means that in order to find the metric distance on the sphere we must use the distance element in some of its "external" forms (Eqs. (2.20), (2.21), (2.23)).

A simple example is the metric distance u_A between the pole P, which is the origin of the internal coordinate system, and the point A:

$$u_A = \int_P^A dl = \underbrace{\int_0^{\rho_A} \frac{d\rho}{\sqrt{1 - (\rho/R)^2}}}_{K_{\mathrm{cyl}}} = R \arcsin\frac{\rho_A}{R} = \underbrace{\int_0^{\theta_A} R d\theta}_{K_{\mathrm{sph}}} = R\Theta_A. \tag{2.28}$$

This calculation shows that the metric distance does not depend on the chosen coordinate system (K_{cyl} or K_{sph}). The internal surveyor uses his rope to determine the metric distance u_A, the length along a great circle of the sphere.

2.4.3 Basic Geometrical Properties of $\mathcal{S}^2$

The Gaussian curvature K of the spherical space is determined by the radius R of the sphere in E^3. A celebrated formula derived by Carl Friedrich Gauss expresses the (scalar) curvature K via the fundamental metric tensor. It yields in the present case the constant value

$$K = 1/R^2. \tag{2.29}$$

Hence, $\mathcal{S}^2$ is a Riemannian space of constant positive curvature.

Geometrical Measurements in $\mathcal{S}^2$ In a triangle ABC, whose sides are parts of three great circles, the sum of the angles a, b, c was found by Gauss to be

$$\Sigma = a + b + c = \pi + K \times \sigma, \tag{2.30}$$

where K is the curvature of the sphere and σ is the *area* of the triangle.

The length l_{circ} of the circumference of a circle of radius u is

$$l_{\mathrm{circ}} = \int_0^{2\pi} R \sin\theta d\phi = 2\pi R \sin\frac{u}{R}. \tag{2.31}$$

In this calculation we use external spherical coordinates and consider the circle with its centre in the polar point P. The needed distance element is given by (2.23). For the chosen circle the coordinate θ is constant, i.e. $d\theta = 0$, and the integration over ϕ yields 2π. For the last step we recall that $u = \theta R$. Note that here we compare the internal length u in cm with the external length R in the same centimetres.

It is useful to write out the first two terms of the Taylor series of sinus, in order to see how the length l_{circ} differs from the Euclidean circumference:

$$l_{\mathrm{circ}} = 2\pi u\left[1 - \frac{1}{6}\left(\frac{u}{R}\right)^2 + \cdots\right]. \tag{2.32}$$

When $u \ll R$, the small relative deflection from Euclidean is $\frac{\Delta l}{l} \approx -\frac{1}{6}\frac{u^2}{R^2}$.

The 2D volume element dA of the spherical space $\mathcal{S}^2$ with the determinant γ of the metric tensor (2.27) is

$$dA = \sqrt{\gamma} dx^1 dx^2 = R \sin\frac{u}{R} du d\phi. \tag{2.33}$$

Then the 2D volume A_{circ} of the circle of radius u is obtained as

$$A_{\mathrm{circ}} = \int_0^{2\pi}\int_0^u dA = 2\pi R^2\left(1 - \cos\frac{u}{R}\right) = \pi u^2\left[1 - \frac{1}{12}\left(\frac{u}{R}\right)^2 + \cdots\right]. \tag{2.34}$$

The 2D volume as measured by the internal surveyor is the area of the polar cap as seen by the external observer. When the surveyor makes only local measurements ($u \ll R$), the volume is close to the area of a circle of the same radius in Euclidean plane. The total volume of space is finite, in fact (2.34) tells that it equals $4\pi R^2$, simply the area of a sphere in E^3.

2.5 Physical Measurements in Spherical and Lobachevskij Spaces

Before going to the 3D curved spaces on which the expanding Friedmann models are based in Chap. 7, we inspect how the imagined surveyor will make observations from his fixed position at the point P.

2.5.1 *Angular Sizes, Fluxes, and Number Counts in $\mathcal{S}^2$*

Consider a small rigid rod at the metric distance u from the surveyor. During its motion in $\mathcal{S}^2$, the length of the rod does not change ("rigidity").

Angular Size and Metric Distance Then what is the angular size of the rod, if its length is $dl = d$ and it lies perpendicularly to the line of sight?

The expression (2.26) for the distance element dl with $du = 0$, gives $dl = R \sin \frac{u}{R} d\Phi$. Note that $d\Phi$ is the difference of the Φ-coordinates of the ends of the rod of length $l = dl$. It is the desired angular size $\alpha = d\Phi$:

$$\alpha = \frac{d}{R \sin \frac{u}{R}} = \frac{d}{u_{\mathrm{ang}}}, \tag{2.35}$$

where the *angular size distance* is denoted by u_{ang} and equals

$$u_{\mathrm{ang}} = R \sin \frac{u}{R}. \tag{2.36}$$

When the metric distance u grows from 0 to its maximum value πR, the angular size distance u_{ang} first increases from 0 to R, and then starts to decrease and becomes 0 when the metric distance is at its maximum (Fig. 2.4). This fictive construction is made in order to have an analogy of the ordinary relation between angle and distance. Simply, the angular size distance is by definition equal to the Euclidean distance at which a rod in Euclidean plane *would* have the angular size that it actually is observed to have in $\mathcal{S}^2$.

Flux and Distance Suppose a distant "galaxy" at the point A in the spherical universe $\mathcal{S}^2$ (Fig. 2.3) emits "light" isotropically around it with the rate L [energy/time] (its luminosity). In this 2D world the circle centered about A plays the role of a sphere in E^3. Define the flux of the light flowing through the circle with radius u as the energy per time unit per length unit (cf. unit of area in E^3). Then the flux observed at the point A is

$$F = \frac{L}{l_{\mathrm{circ}}} = \frac{L}{2\pi R \sin \frac{u}{R}} = \frac{L}{2\pi u_{\mathrm{lum}}}. \tag{2.37}$$

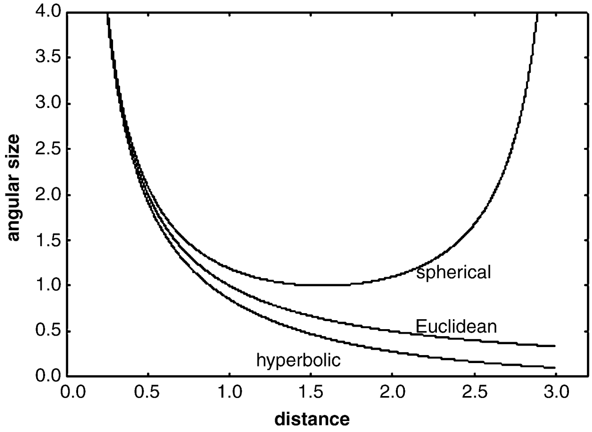

Fig. 2.4 Angular size–metric distance relation in different 2D spaces

Here appears the *luminosity distance* u_{lum}. It is such Euclidean metric distance at which the galaxy in Euclidean plane would produce the flux that is actually measured at P in $\mathcal{S}^2$. Note that in the 2D spherical universe the angular size distance = the luminosity distance. This is valid also in other regular *static* spaces, both 2D and 3D, but not in expanding spaces such as used in the Friedmann model (Chap. 7).[5]

Counts of Galaxies in $\mathcal{S}^2$ Suppose that $\mathcal{S}^2$ is filled by some objects and their number density around the observer is $n(u)$. Then the total number $N(u)$ of these objects within the radius u (metric distance) grows according to a definite number-radius relation. To calculate $N(u)$, one must integrate the number density over the volume element dA of $\mathcal{S}^2$ (Eq. (2.33)):

$$N(u) = \int_0^{2\pi} \int_0^u n(u) dA. \tag{2.38}$$

If the objects are uniformly distributed, i.e. $n(u) = n_0$, then:

$$N(u) = 2\pi R^2 \left(1 - \cos \frac{u}{R}\right) n_o.$$

As the spherical space has a finite "volume", the total number of these objects is finite, too, and is equal to $4\pi R^2 n_0$.

[5]Traditionally cosmologists speak about "luminosity distance". It might be more logical to term it "flux distance", emphasizing the observed quantity flux instead of the not directly observed luminosity. On the other hand, "luminosity" refers to the essential assumption that the object radiates isotropically with the rate L. With the angular size distance the analogous assumption is that the rod is always viewed perpendicularly.

2.5.2 Measurements in 2D Lobachevskij Space

We have devoted much attention to the simple 2D space with positive constant curvature $\mathcal{S}^2$. The other homogeneous and isotropic 2D space is the Lobachevskij space $\mathcal{L}^2$ (or the hyperbolic space). It is an infinite surface with negative constant curvature. However, this surface cannot be embedded as a whole in 3D Euclidean space. One could actually "see" only small local regions of $\mathcal{L}^2$ which remind one of a saddle. The negative curvature can be expressed in terms of the curvature radius as

$$K = -1/R^2.$$

This and other basic geometrical properties of $\mathcal{L}^2$ are obtained from the corresponding formulae of $\mathcal{S}^2$ by making everywhere the replacements[6]

$$R \Longrightarrow i\,R, \qquad \sin\theta \Longrightarrow \sinh\chi\,.$$

That is why it is sometimes said that the Lobachevskij space is a pseudospherical space with imaginary radius. Certainly, it is difficult to imagine. Fortunately, its mathematics is as simple as in $\mathcal{S}^2$. For example, the distance element for $\mathcal{L}^2$ in cylindrical coordinates is

$$dl^2 = \frac{d\rho^2}{1 + \frac{\rho^2}{R^2}} + \rho^2 d\phi^2, \tag{2.39}$$

where the polar radius ρ can have values from zero to infinity. Another form of dl is obtained by substituting $\rho = R\sinh\chi$ into Eq. (2.39). Then

$$dl^2 = R^2 d\chi^2 + R^2 \sinh^2\chi\, d\phi^2. \tag{2.40}$$

In the internal coordinates (u,Φ) of a surveyer in $\mathcal{L}^2$ the line element is

$$dl^2 = du^2 + R^2 \sinh^2\frac{u}{R} d\Phi^2, \tag{2.41}$$

where $u = R\chi$, hence the fundamental metric tensor of the Lobachevskij 2D space is expressed in internal polar coordinates as

$$\gamma_{\alpha\beta} = \mathrm{diag}\left(1,\, R^2 \sinh^2\frac{u}{R}\right). \tag{2.42}$$

The radius R refers to the pseudosphere in the external Euclidean space.

The sum of the angles of a triangle in $\mathcal{L}^2$ always fall shy of the Euclidean value π and the larger the triangle, the smaller the sum:

$$\Sigma = a + b + c = \pi - K \times \text{area}. \tag{2.43}$$

---

[6] We recall the hyperbolic functions: $\sinh x = (e^x - e^{-x})/2$, $\cosh x = (e^x + e^{-x})/2$.

The length of the circumsphere of a circle with radius u is given by

$$l_{\text{circ}} = 2\pi R \sinh \frac{u}{R}.$$

The angular size–distance relation for the observer in $\mathcal{L}^2$ is

$$\alpha = \frac{d}{R \sinh \frac{u}{R}} = \frac{d}{u_{\text{ang}}}.$$

This formula shows that the angular size of a "galaxy" in $\mathcal{L}^2$ monotonously decreases with increasing metric distance. Hence, different spaces E^2, $\mathcal{S}^2$, and $\mathcal{L}^2$ are characterized by clear differences in the behaviour of the angular size of a standard rod viewed at different metric distances (Fig. 2.3).

The volume of $\mathcal{L}^2$ is infinite, as may be seen from the volume up to u:

$$V(u) = 2\pi R^2 \left(\cosh \frac{u}{R}\right) - 1.$$

The fact that the volume (area) increases quicker than in the Euclidean plane may help one to understand why $\mathcal{L}^2$ cannot be embedded as a whole in $\mathcal{E}^3$—there is not enough space!

2.5.3 The Step to Three-Dimensional Curved Spaces

Before going to 3D spaces, we fix a convenient unified notation for the 2D spaces with constant curvature. This notation is easily transferred to the regular 3D spaces of modern cosmology.

Unified Notation for E^2, $\mathcal{S}^2$, and $\mathcal{L}^2$ First, the curvature K is simply

$$K = \frac{k}{R^2}. \tag{2.44}$$

The curvature constant k is 1, 0, or -1 (for $\mathcal{S}^2$, E^2, and $\mathcal{L}^2$, respectively). R is the radius of curvature of the space in question. Now consider the dimensionless variable χ expressing the internal coordinate u relative to R:

$$\chi = \frac{u}{R}.$$

Then the polar radius is

$$\rho = Ra(\chi),$$

where $a(\chi)$ is a new variable:

$$a(\chi) = \begin{cases} \sin\chi, & \text{for } k = +1,\ \chi \in [0, \pi], \\ \chi, & \text{for } k = 0,\ \chi \in [0, \infty], \\ \sinh\chi, & \text{for } k = -1,\ \chi \in [0, \infty]. \end{cases} \tag{2.45}$$

The differential of $a(\chi)$ is then given by the simple formula $da(\chi) = \sqrt{1 - ka^2(\chi)}d\chi$. The distance element for all three spaces now becomes

$$dl^2 = R^2(d\chi^2 + a^2(\chi)d\phi^2) \tag{2.46}$$

and the fundamental metric tensor and the volume element are

$$\gamma_{\alpha\beta} = \mathrm{diag}(R^2, R^2a^2(\chi)), \qquad dA = Ra(\chi)d\chi d\phi.$$

Spaces $\boldsymbol{E^3}$, $\boldsymbol{\mathcal{S}^3}$, and $\boldsymbol{\mathcal{L}^3}$ The line element of homogeneous and isotropic 3D spaces is often written as follows:

$$ds^2 = R^2(d\chi^2 + I_k(\chi)^2(d\theta^2 + \sin^2\theta d\phi^2)). \tag{2.47}$$

χ, θ, ϕ are the spherical coordinates and $I_k(\chi) = \sin(\chi), \chi, \sinh(\chi)$ correspond to the curvature constants $k = +1, 0, -1$, respectively. The fundamental metric tensor is

$$\gamma_{\alpha\beta} = \mathrm{diag}(R^2, R^2I_k^2(\chi), R^2I_k^2(\chi)\sin^2\theta). \tag{2.48}$$

Note the similarity of this 3D line element to the internal line element of the 2D regular spaces, Eq. (2.46). It is also internal in the sense that all three coordinates (χ, θ, ϕ) are measured entirely within the 3D space.

Another form of the metric can be written in "cylindrical coordinates":

$$ds^2 = R^2\left(\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right). \tag{2.49}$$

Now the coordinate r is not internal—Rr may be interpreted as the angular size distance. Integrating over ds from the observer to the point in question at r, the metric distance u and this coordinate distance r become related as $u = R\sin^{-1} r$, Rr, and $R\sinh^{-1} r$, for $k = 1$, 0, and -1, respectively.

2.6 Practical Geometry

If one asks a mathematician: what is geometry?, the reply may be unexpected, but instructive. So Veblen and Whitehead (1932) wrote: a part of mathematics is called geometry because a sufficient number of competent people think "on emotional and traditional grounds" that it is a good name.

This simply means that for mathematics itself it is not interesting how geometrical concepts relate to real space. However, for physics this is central. Working with abstract things, mathematical space cannot be a fully adequate picture of real physical space. *Mathematical* concepts coordinate, distance, and curvature are not equivalent to their *physical* counterparts. A "singularity" and the break-down of physical laws in it may only mean the inadequacy of the mathematical scheme leading to the singularity.

2.6.1 Geometry and Physics

Two opposite views on geometry and physics were put forward by Poincaré (1902) and Einstein (1921) when writing about the philosophy of science.

Poincaré vs. Einstein For Poincaré, geometry does not deal with real things: its notions belong to an ideal world. Only geometry together with physics, the geometry-physics unity, is open to empirical study. One may first choose geometry arbitrarily and then find the rest of the unity, physical laws, so that these will not contradict experiments. Or one may go other way round, to start from physical laws and find the geometry. Poincaré thought that it is preferable to leave the simplest Euclidean geometry intact and to change the physical laws. Modern physics gives another good reason not to forget flat geometry: the energy-momentum conservation laws directly follow from the maximal symmetry of Euclidean and Minkowski spaces (Noether's theorem).

Einstein's practical geometry assumes that geometry can be the subject of empirical study. The question whether the geometry of the universe is Euclidean or not has a clear meaning and may be answered by experience. Hence it is natural to formulate physical laws within Riemann's geometry. This is the way of general relativity.

Experiment is the final court in physics, and if the predictions for all experiments are identical, then both ways finally lead to the same geometry-physics unity. However, if predictions differ, crucial tests are required. General relativity and field gravity are modern examples (Chaps. 5 and 6).

Embedding Space In mathematics, a geometry of space may be built without assuming any embedding high-dimensional space (that is why one speaks about the internal geometry of curved spaces). In physics the concept of embedding space makes sense and deserves to be studied. In fact, the expression of the distance element requires that the internal distance u and the external radius R be measured in the same units. Thus the external standard of length in E^3 as if induces the unit in spherical space.[7] For instance, local lengths may be expressed in terms of the radius of curvature. Apparently having this in mind, in a 1824 letter to German scholar

[7]How numbers and physical quantities with units differ was extensively discussed by physicist Bridgman (1936) and mathematicians Menger (1959) and Whitney (1968).

Taurinus, Gauss wrote *I sometimes joke that it would be good, if Euclidean geometry were not true, because then we would have a priori absolute measure of length...*

In flat (Euclidean) space the radius of curvature is infinite and does not determine any absolute unit common with that of the embedding space. But another property is now related to the definition of the rigid length unit: Euclidean space is self-similar on all scales (congruency), and there is no preferred length which could be chosen as the unit. So, in $\mathcal{S}^2$ one may determine the absolute size of a triangle by measuring the sum of its angles, but in $\mathcal{E}^2$ all triangles have $a + b + c = \pi$—by looking at a triangle you cannot tell its size. One might surmise that a true scale-invariant fractal could exist only in Euclidean space, if one requires that the scale cannot be determined from the geometrical appearance of the fractal.

2.6.2 Measuring the Curvature

Non-Euclidean geometries—as mathematical models—do not contain internal contradictions. Hence, real space might be non-Euclidean (already Lobachevskij and Gauss made attempts to measure space curvature). Such an experimental approach probes the link between physics and geometry: is the curvature measurable, when one uses real units and procedures of length measurement? "If space curvature is real, it must make a difference in something we can measure" (Sandage 1992).

Rigid Sticks and Curvature The 2D beings of a spherical universe could walk around their world with a number of steps and establish its finite size. Locally, they could also measure the curvature R, e.g., using the angular excess formula. In analogy, such a measurement could be made also in our 3D world. However, the possibility to measure the curvature depends on a crucial assumption: There should be in curved space a length unit that can in principle be transferred from one place to another without changing it.

How to define a rigid rod in curved space is not obvious. In Euclidean space the ends of a free rigid rod draw parallel straight lines (geodesic curves of E^3), and there are no tensions between the internal parts of the rod, all moving along geodesics. So it is possible to imagine the rod made from independent, freely moving material points. In fact, one can also understand the unit length as the distance between two freely moving particles that were put into motion with equal velocities, e.g., perpendicularly to the line through them. In this way the length unit may be transferred into any point in space. Such a "soft" (free motion) standard gives the same results of measurement as an "absolutely hard" (rigid) stick. Indeed, the rigid stick is something that resists all forces trying to change the distance between its end points when it moves in space.

Things are different in curved space. There the Euclidean straight lines are replaced by geodesics, the shortest routes connecting two points. Then the two procedures (rigid stick, free motion) give different results. If one has the intuitively appealing rigid meter, its end points do not move along geodesic curves. Then the measurement of curvature is possible and its radius may be expressed in terms of the rigid meter. Or vice versa, lengths may be given with the fundamental radius of curvature as a unit.

2.6.3 The Deeper Value of Practical Geometry

The usual approach to curved space in physics has adopted the procedure, where the rigid meterstick ignores the geodesics of physical space. Thus distance measurement in astronomy has special significance beyond just knowing "how far away" a star or a galaxy is (needed for all that complicated astrophysics): it also probes our fundamental concepts of practical geometry and tests the validity of the rigid body hypothesis in our attempts to infer the curvature of space via non-local observations.

In this chapter we have briefly discussed some cosmic distance indicators and different kinds of distances which naturally arise in practice, notably angular size and luminosity distances. When one makes first steps in Euclidean space, indicators of different distance types are expected to give identical results. This is what happens, say, in the local galaxy universe within 100–200 Mpc where the Tully-Fisher relation both in the size and magnitude mode gives similar values for the Hubble constant.

In deep space, which means distances of the order of ~ 1000 Mpc and more, it is expected that different distance types start to diverge and only a good cosmological model will correctly link the distances to the underlying metric distance. We will see in later chapters how this happens with the Friedmann model. Above we have illustrated theoretical counterparts of practical distances in metric spaces using simple 2D spaces.

http://www.springer.com/978-94-007-2378-8

Fundamental Questions of Practical Cosmology
Exploring the Realm of Galaxies
Baryshev, Y.; Teerikorpi, P.
2012, XVI, 332 p., Hardcover
ISBN: 978-94-007-2378-8