Cosmological Distances

University of Liège
Faculty of Sciences

Department of Astrophysics, Geophysics and Oceanography

Cosmological Distances: Calculation of distances in cosmological models with small-scale inhomogeneities and their use in observational cosmology

Author:
Phillip Helbig

Supervisor:
Dr. Christian Barbier

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in Space Science in the
Space sciences, Technologies and Astrophysics Research group
Department of Astrophysics, Geophysics and Oceanography

September 2021

Members of the jury:

Dr. Christian BARBIER	Promotor, University of Liège
Dr. Ludovic DELCHAMBRE	Secretary, University of Liège
Prof. Marc-Antoine DUPRET	President, University of Liège
Prof. Rolf STABELL	University of Oslo
Prof. Jean SURDEJ	Co-promotor, University of Liège
Prof. Michel TYTGAT	Free University of Brussels

To my wife and children, for letting me have some fun.

I have yet to see any problem, however complicated, which, when you looked at it in the right way, did not become still more complicated.

—Poul Anderson

Hofstadters's law: It always takes longer than you expect, even when you take into account Hofstadter's law.

—Douglas R. Hofstadter

Author's note

This book was written as a doctoral thesis and, in accordance with the tradition at the University of Liège, contains some previously published papers as is, which have their original page numbers. The numbering of the pages of the new material in this thesis take those into account, *i.e.* the difference in the page numbers before and after an included paper is one more than the number of pages in the included paper. The bibliography at the end of this thesis refers to citations in the new text; each included paper has its own list of references; of course, there is some overlap between the two.

Preface

I obtained my *Diplom* (the rough equivalent of a master's degree, though including a one-year thesis) in physics with a minor in astronomy from the University of Hamburg in 1993. The thesis, entitled (in translation—the original is in German) *Determination of cosmological parameters through the redshift statistics of gravitational lenses*, was written at the Hamburg Observatory, under the supervision of the late Sjur Refsdal. Both the thesis and the overall mark of the *Diplom* were *sehr gut*, usually translated as *magna cum laude*.

I later worked at Jodrell Bank and the Kapteyn Institute, where most of my work was concerned with gravitational lensing, often within the context of the CERES network, a European Union project coordinated by Ian Browne at Jodrell Bank, which resulted in several papers, usually with several co-authors. In order to avoid overlap with that work, to give this thesis a clear focus, and to make clear what is my own work, with two exceptions (Chaps. 3 & 5, both referring to work done in Hamburg) all of the included papers are single-author papers, though in the case of the exceptions it is fair to say that I did the lion's share of the work. Also, with two exceptions (Chaps. 6 & 7) the work was done before (Chaps. 3–5) or after (Chaps. 1–2 and Chaps. 8–9) the time I was employed by CERES. The paper included in Chap. 2 was originally written especially for this thesis (as, of course, were Chaps. 1 & 10); I submitted it to *The Open Journal of Astrophysics* and what appears here now is the published version, incorporating comments from the referees. Like much of my work, the papers included in this thesis could be classified as 'theoretical observational cosmology'.

Abstract

In cosmology, one often assumes that the universe is homogeneous and isotropic. While originally a simplifying assumption, today there is observational evidence that this is a good approximation in our Universe on scales above a few hundred megaparsecs. This approximation is often used when calculating various distances as a function of redshift, even though the scales probed by a beam of light are much smaller than the scale of homogeneity. Since our Universe is obviously not homogeneous and isotropic on small scales, it is at least conceivable that this could affect distance calculation.

Two models have been proposed in order to take such small-scale inhomogeneities into account in a relatively simple way. One, due to Zel'dovich, involves a two-component universe where one component is smoothly distributed and the other in clumps, with the assumption that, when calculating distance from redshift, light propagates far from all clumps. Under those assumptions, one can derive a second-order differential equation for the distance. This is a simple *ansatz* but it is not obvious how valid it is. Another approach, originally due to Einstein and Straus but developed with regard to cosmological-distance calculation by Kantowski, involves removing material from a spherical region of an otherwise smooth universe and redistributing it inside this sphere (*e.g.* as a point mass at the centre, as a shell at the boundary, or in a more complicated manner). This *ansatz* is more difficult for calculations, but is an exact solution of the Einstein equations, so there is no question about its validity (how realistic such a mass distribution is as a model of our Universe is a separate question). Long after both had been investigated in detail, Fleury showed that they are equivalent at a well controlled level of approximation.

After a review of the history of those two approaches, I present my own work in this area: an efficient numerical implementation for the solution of the most general form of the differential equation (*i.e.* arbitrary values of λ_0, Ω_0, and the homogeneity parameter η, the last indicating the fraction of matter distributed smoothly), a discussion of the uncertainty in distance calculation due to uncertainty in the value of η, the effect of η on the calculation of H_0 from gravitational-lens time delays, the effect of η on the separation between images in a gravitational-lens system, and the effect of η on the determination of λ_0 and Ω_0 from the m–z relation for Type Ia supernovae—including evidence that observations indicate that, in our Universe, the standard distance is a good approximation, even though small-scale inhomogeneities can be appreciable, probably because the Zel'dovich model does not accurately describe our Universe.

Résumé

En cosmologie, on suppose souvent que l'univers est homogène et isotrope. Bien qu'il s'agisse à l'origine d'une hypothèse simplificatrice, il est aujourd'hui prouvé grâce à l'observation qu'il s'agit d'une bonne approximation dans notre Univers, à des échelles plus grand que quelques centaines de megaparsecs. Cette approximation est souvent utilisée pour calculer de diverses distances en fonction du décalage vers le rouge, même si les échelles sondées par un faisceau lumineux sont beaucoup plus petites que l'échelle d'homogénéité. Étant donné que notre Univers n'est évidemment pas homogène et isotrope à petite échelle, il apparaît concevable que cela pourrait affecter le calcul des distances.

Deux modèles ont été proposés afin de prendre en compte ces inhomogénéités à petit échelle de manière relativement simple. L'un, dû à Zel'dovich, propose un univers à deux composants où l'une est distribuée de manière lisse et l'autre sous forme de petites concentrations, avec l'hypothèse que, lors du calcul de la distance en fonction au décalage vers le rouge, la lumière se propage loin de toutes ces concentrations de matière. Adoptant ces hypothèses, on peut dériver une équation différentielle du second ordre pour la distance. Il s'agit d'une approche simple, mais sa validité n'est pas démontrée. Une autre approche, due à l'origine à Einstein et Straus mais développée par Kantowski en ce qui concerne le calcul de la distance cosmologique, consiste à retirer de la matière d'une région sphérique d'un univers par ailleurs lisse et à la redistribuer à l'intérieur de cette sphère (par exemple comme une masse ponctuelle au centre, comme une coque à sa frontière, ou de toute autre manière plus compliquée). Cet approche est plus difficile pour les calculs, mais elle conduit à une solution exacte des équations d'Einstein, donc il n'y a aucun doute sur sa validité (le réalisme d'une telle distribution de masse comme modèle de notre Univers est un autre question). Bien aprés que les deux modèles aient été étudiés en détail, Fleury ont montré qu'ils sont équivalents à un niveau d'approximation bien contrôlé.

Après une revue de l'histoire de ces deux approches, je présente mon propre travail dans ce domaine: une implémentation numérique efficace du solution de la forme la plus générale de l'équation différentielle (*i.e.* valeurs arbitraires de λ_0, Ω_0, et le paramètre d'homogénéité η, ce dernier indiquant la fraction de matière distribué de manière lisse). J'aborde une discussion sur l'incertitude du calcul de la distance due à l'incertitude sur la valeur de η, l'effet de η sur le calcul de H_0 à partir du décalage temporel observé pour certains mirages gravitationnelles, l'effet de η sur la séparation entre les images produites par une lentille gravitationnelle, et l'effet de η sur la détermination de λ_0 et Ω_0 de la relation m–z pour supernovae de Type Ia. Les observations semblent aussi indequer que, dans notre Univers, la distance standard est une bonne approximation, même si les inhomogénéités à petite échelle peuvent être appréciables, probablement parce que le modèle Zel'dovich ne décrit pas avec précision notre Univers.

Contents

Author's note ix

Preface xi

Abstract xiii

Résumé xv

I Background 1

1 Introduction 3
1.1 Kinematics: homogeneous and isotropic universes 4
1.1.1 The Robertson–Walker metric 4
1.1.2 Different forms of the same equation 5
1.1.3 Observational quantities 5
1.2 Dynamics: general relativity 5
1.2.1 Cosmological parameters 5
1.2.2 Einstein's model: matter without motion 7
1.2.3 de Sitter's model: motion without matter 7
1.2.4 Friedmann models: wide variety 8
1.2.5 Lemaître: the primeval atom 8
1.3 Classical FLRW cosmology 9
1.4 Many distances 9
1.4.1 Proper distance 10
1.4.2 Angular-size distance 10
1.4.3 Luminosity distance 10
1.4.4 Proper-motion distance 11
1.4.5 Parallax distance 11
1.4.6 Light-travel–time distance 11
1.5 Calculation of distances in ideal FLRW models 12
1.5.1 $z \ll 1$ 12
1.5.2 Special cases 12
1.5.3 Series expansions 12
1.5.4 Numerical calculations 12
1.5.5 Elliptic integrals 13
1.6 Effect of small-scale inhomogeneities 13
1.7 This thesis 13

2 Distances in inhomogeneous models 15
2.1 Context 15
2.2 Helbig (2020) 17
2.3 Follow-up 44

II General theory 45

3 Calculating cosmological distances 47
3.1 Context . . . 47
3.2 Kayser, Helbig & Schramm (1997) . . . 49
3.3 Follow-up . . . 61

4 ANGSIZ User's Guide 63
4.1 Context . . . 63
4.2 Helbig (1996) . . . 64
4.3 Follow-up . . . 77

III General applications 79

5 Sub-galactic clumps at redshift 2.39 81
5.1 Context . . . 81
5.2 Helbig & Kayser (1996) . . . 82
5.3 Follow-up . . . 84

IV Gravitational lensing 85

6 Gravitational-lensing time delays 87
6.1 Context . . . 87
6.2 Helbig (1997) . . . 88
6.3 Follow-up . . . 94

7 The image-separation–redshift relation 95
7.1 Context . . . 95
7.2 Helbig (1998) . . . 97
7.3 Further details and follow-up . . . 101

V Type-Ia supernovae 109

8 The magnitude–redshift relation. I 111
8.1 Context . . . 111
8.2 Helbig (2015a) . . . 112
8.3 Follow-up . . . 123

9 The magnitude–redshift relation. II 125
9.1 Context . . . 125
9.2 Helbig (2015b) . . . 126
9.3 Follow-up . . . 131

VI Summary, conclusions, & supplementary material 133

10 Summary and conclusions 135

Acknowledgements 137

Bibliography 139

Part I

Background

Chapter 1

Introduction

Cosmology is the study of our Universe (and/or model universes—the (lack of) capitalization distinguishes the two) as a whole. That is much easier with certain simplifications, such as a universe which is homogeneous and isotropic, at least on large scales. While initially those were assumptions, known as the Cosmological Principle (*e.g.* Harrison, 2000), now those are observational facts. Isotropy is observed, most dramatically in the cosmic microwave background radiation (CMB), which is isotropic to one part in 100,000 (*e.g.* Smoot *et al.*, 1992). Homogeneity is not directly observed, since we can observe only that which which is on our backward lightcone[1] (and can get information on nearby objects *via* other means as well). That allows us to observe objects at a variety of distances but, due to the finite speed of light, at correspondingly different times. Thus, we cannot say whether or not the Universe is homogeneous at a particular instant in cosmic time, such as now. However, homogeneity follows from isotropy as long as we are not in a special position (*e.g.* Harrison, 2000).

A homogeneous and isotropic universe can be parameterized by a scale factor R, which in general depends on time. A fundamental quantity is the redshift

$$z = \frac{\lambda_0}{\lambda} - 1 \quad , \tag{1.1}$$

where λ is the emitted wavelength and λ_0 is the observed wavelength (in general, the subscript 0 is used to denote the present value of a time-dependent quantity). That is related to the values of the scale factor now and when the radiation which we observe now was emitted:

$$z = \frac{R_0}{R} - 1 \quad , \tag{1.2}$$

where R_0 is the value of the scale factor now and R the value of the scale factor when the light was emitted. The redshift increases from 0 for objects at our location to a maximum value, which can be ∞ if the universe had an arbitrarily small scale factor in the past, as is the case in big-bang models.

Observational cosmology consists of calculating the dependence of various observational quantities (*e.g.* apparent brightness) as a function of redshift for various cosmological parameters (*e.g.* the density parameter) and comparing with observations, which allows one to determine the cosmological parameters,

[1] If one thinks of a universe with two rather than three spatial dimensions, in the x and y directions, and time t in the z direction, then at a given time, such as now, we can observe objects at various distances the light from which left them at various times. Defining a 'time dimension' as ct where c is the speed of light, light moves at 45° angles, thus that which we can observe lies on a cone in this spacetime diagram.

which in turn determine quantities which are related to different types of distance, but also other quantities. Distance is proportional to redshift only for small redshift. At larger redshift, not only is that no longer the case, but also various distances have different dependencies on redshift. In everyday experience, distances measured *via* various methods are equal; that is not the case in cosmology since, in general, the universe neither has Euclidean geometry nor is static.

1.1 Kinematics: homogeneous and isotropic universes

1.1.1 The Robertson–Walker metric

A homogeneous and isotropic cosmological model (I use the terms 'cosmological model' and 'universe' interchangeably) is known as a Robertson–Walker model (Robertson, 1935, 1936; Walker, 1935, 1937, the latter paper by Walker is very often incorrectly cited as having been published in 1936). Expansion is described by a time-dependent scale factor. Other coordinates are fixed with time.[2] The corresponding metric is

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - R^2(t)\left(\frac{\mathrm{d}\sigma^2}{(1-k\sigma^2)} + \sigma^2\,\mathrm{d}\theta^2 + \sigma^2\sin^2\theta\,\mathrm{d}\phi^2\right) \quad , \tag{1.3}$$

where the symbols are defined as follows (with the corresponding units):

s 4-dimensional interval	[length]
c speed of light	[velocity]
t time	[time]
R scale factor	[length]
σ radial coordinate	[dimensionless]
k curvature constant	[dimensionless]
θ angular coordinate	[dimensionless]
ϕ angular coordinate	[dimensionless]

Note that no physics is required to derive this; it is the only metric which describes a homogeneous and isotropic universe which remains so even if it expands or contracts. By the same token, the metric says nothing about the dynamics, *i.e.* how the scale factor changes with time. One can define the parameters $r = \sigma R$ and χ as $\mathrm{arcsinh}(\sigma)$, σ, or $\arcsin(\sigma)$ for k equal to -1, 0, $+1$, respectively. The proper distance D^{P}, *i.e.* the distance one could measure with a rigid ruler instantaneously, is $R\chi$. Most distances used in cosmology depend more directly on r.

[2]That assumes that objects have no peculiar motion, but move away from or towards each other only as a result of the overall expansion or contraction. Such objects are called comoving.

1.1.2 Different forms of the same equation

There are several equivalent forms of the Robertson–Walker metric. One involves writing it in terms of r instead of σ:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \frac{R^2(t)}{R_0^2}\left(\frac{\mathrm{d}r^2}{\left(1-k\frac{r^2}{R_0^2}\right)} + r^2\,\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2\right) \quad . \tag{1.4}$$

Another involves χ instead of r and S_k defined as $\sinh(\chi)$, χ, or $\sin(\chi)$ for k equal to -1, 0, $+1$, respectively:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - R^2(t)\left(\mathrm{d}\chi^2 + S_\mathrm{k}^2(\chi)\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right) \quad . \tag{1.5}$$

Some authors use an opposite sign convention and/or use $\mathrm{d}\Omega^2 := \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$ (not to be confused with the density parameter Ω). Often, a dimensionless scale factor $a := R/R_0$ is used; in such cases the R can be included in S in Eq. (1.5) (and writing either $R^2\,\mathrm{d}\chi^2$ or $\mathrm{d}(D^\mathrm{P})^2$ for the first term in parentheses). There are several further variations. Occasionally, one sees the Robertson–Walker metric expressed in (locally) Cartesian coordinates:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - R^2(t)\left(\frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{1 + \frac{1}{4}k(x^2+y^2+z^2)}\right) \quad . \tag{1.6}$$

Care must be taken when comparing different formulations from different authors, since the same symbols can be used for different quantities and/or the same quantities can be denoted by different symbols; that sometimes happens with the same author within a few pages (*e.g.* Heacox, 2015)[3]. In my notation, Latin letters (except c, k, and t, which are too entrenched to change) have the dimension of length and Greek letters are dimensionless.

1.1.3 Observational quantities

What quantities does one actually observe?[4] Essentially, one observes angles (between objects or between parts of one object) and brightnesses. In particular, σ is a derived quantity from an observational point of view; $r = \sigma R$ is related to some of the distances discussed below.

1.2 Dynamics: general relativity

1.2.1 Cosmological parameters

This thesis is restricted to homogeneous and isotropic (at least on large scales) cosmological models based on general relativity (GR), so-called Friedmann–Lemaître–Robertson–Walker (FLRW) models[5], without pressure; here, I follow

[3] This is an excellent book, which I reviewed for *The Observatory* (Helbig, 2016b). While the changing definitions can be confusing, they do force one to think the examples through, though I don't know if that was the intention. In his equation (8.3) his r is my r; in equation (8.5) his r is my σ, and in equation (8.6) his r is my D^P.

[4] At some level, of course, all that one observes are counts of photons on a detector as a function of position, whether in the case of imaging or in the case of spectroscopy, but it is useful to think of astronomically relevant quantities as basic.

[5] That is a reference to early papers exploring the expansion history of the universe as a function of the cosmological parameters (Friedmann, 1922, 1924; Lemaître, 1927). Some argue

the exposition of Kayser, Helbig and Schramm (1997). The dynamics of the universe is given by the Friedmann equations

$$\dot{R}^2(t) = \frac{8\pi G\rho(t)R^2(t)}{3} + \frac{\Lambda R^2(t)}{3} - kc^2 \tag{1.7}$$

and

$$\frac{\ddot{R}(t)}{R(t)} = -\frac{4\pi G\rho(t)}{3} + \frac{\Lambda}{3} \quad , \tag{1.8}$$

where dots denote derivatives with respect to t, G is the gravitational constant, $\rho(t)$ the matter density (this thesis assumes negligible pressure), Λ the cosmological constant and the sign of k determines the curvature of the 3-dimensional space.

Introducing the usual parameters

$$\begin{aligned} H &= \frac{\dot{R}}{R} && \text{(Hubble parameter)} \\ \Omega &= \frac{8\pi G\rho}{3H^2} && \text{(density parameter)} \\ \lambda &= \frac{\Lambda}{3H^2} && \text{(normalized cosmological constant)} \end{aligned} \tag{1.9}$$

(Ω and λ are dimensionless and H has the dimension t^{-1}) we can use Eq. (1.7) to calculate

$$kc^2 = R^2H^2\,(\Omega + \lambda - 1) \quad , \tag{1.10}$$

so that

$$k = \mathrm{sign}\,(\Omega + \lambda - 1) \quad . \tag{1.11}$$

Since $R > 0$ we can write

$$R = \frac{c}{H}\frac{1}{\sqrt{|\Omega + \lambda - 1|}} \quad ; \tag{1.12}$$

that is the radius of curvature of the 3-dimensional space at time t. For $k = 0$ it is convenient to *define* the scale factor R_0 to be c/H_0 (though note that, in general, $R \neq c/H$, including the flat case where $R_0 = c/H_0$). The index 0 is used to denote the present value of a given quantity, fixed, as usual, at the time t_0 of observation.[6] The explicit dependence on t will be dropped for brevity. Taking matter conservation into account and using the present-day values, we have

$$\rho R^3 = \rho_0 R_0^3 \quad , \tag{1.13}$$

and so from Eqs. (1.7), (1.9), (1.10), and (1.13) follows

$$\dot{R}^2 = H_0^2 R_0^2 \left(\frac{\Omega_0 R_0}{R} + \frac{\lambda_0 R^2}{R_0^2} - (\Omega_0 + \lambda_0 - 1) \right) \quad . \tag{1.14}$$

that the term Friedmann–Robertson–Walker (FRW) is more appropriate, since Friedmann had actually explored the entire range of such models, while Lemaître, after rediscovering the models explored by Friedmann, usually discussed a particular type, namely a spatially closed model with a positive cosmological constant. However, Lemaître was responsible for reviving interest in those models, and also discussed the related astrophysics, whereas Friedmann's work is more mathematical. Lemaître was also the first to calculate a value for what is now known as the Hubble constant H_0 (Lemaître, 1927). Thus, it seems appropriate to include Lemaître in the acronym.

[6] Note that this work is concerned with the calculation of distances from redshift. We are not concerned with a change in redshift during a period of observations, so-called redshift drift (*e.g.* Sandage, 1962; McVittie, 1962; Rüdiger, 1980; Lake, 1981; Loeb, 1998).

Since below we want to discuss distances as functions of the cosmological redshift z, by making use of the facts that

$$z = \frac{R_0}{R} - 1 \tag{1.15}$$

and that R_0 is fixed, we can use Eq. (1.14) to get

$$\mathrm{d}z = \frac{\mathrm{d}z}{\mathrm{d}R}\dot{R}\,\mathrm{d}t = -H_0(1+z)\sqrt{Q(z)}\;\mathrm{d}t \quad , \tag{1.16}$$

where

$$Q(z) = \Omega_0(1+z)^3 - (\Omega_0 + \lambda_0 - 1)(1+z)^2 + \lambda_0 \quad . \tag{1.17}$$

Note: *Here, the* $\sqrt{}$ *sign should be taken to signify the* positive *solution, except that* sign $\sqrt{Q(z)} = \mathrm{sign}(\dot{R})$ *always.*

1.2.2 Einstein's model: matter without motion

The first cosmological model based on GR was the static model proposed by Einstein (1917). In a static model, there is no cosmological redshift, but of course all the quantities which are otherwise dependent on redshift do exist. Also, λ_0 and Ω_0 are infinite since $H_0 = 0$, though the density ρ and the cosmological constant Λ have finite values. Einstein originally had set $\Lambda = 0$ and found that no static solutions existed, so he introduced Λ in order to have a static solution. That was not completely *ad hoc*, as in their general form the Einstein equations contain Λ, as indeed does the classical Poisson equation (*e.g.* Nowakowski, 2001). Allegedly, Einstein later referred to his introduction of the cosmological constant as the "biggest blunder of my life", though for a long time the only written source for that was Gamow (1956, 1970), whom some regard as an unreliable narrator. Sources differ on whether Einstein actually said (or believed) that. O'Raifeartaigh *et al.* (2018) give references for both sides, including newer written ones, and suggest a third alternative: it was a blunder, but the blunder was not the cosmological constant *per se*, but rather "his failure to consider the stability of his static cosmology of 1917", agreeing with Weinberg (2005). On the other hand, Barrow (2017) claims that since Einstein referred to signing the letter to Roosevelt encouraging the Manhattan project as "the one great mistake in my life", he would not have used similar wording to refer to the cosmological constant. However, considering that one is in the realm of science and the other in the realm of politics, that does not seem impossible.

1.2.3 de Sitter's model: motion without matter

The second cosmological model was that of de Sitter (1917c) (based to some extent on some earlier papers published in English but in the Netherlands (de Sitter, 1917a,b)): $\rho = 0$ and $\Lambda > 0$. This was sometimes referred to as 'motion without matter', as opposed to Einstein's model of 'matter without motion'. In the original presentation, de Sitter used coordinates which appeared to be static, perhaps influenced, as Einstein explicitly wrote that he (Einstein) was, by the expectation that the Universe is static at large scales (Weinberg, 2005). (In general, the presentation is quite different from the way in which the de Sitter model is thought of today, the latter being exponentially expanding Euclidean space, though of course mathematically equivalent. Lanczos (1922) was the first to discuss the de Sitter model in coordinates in which the scale factor is a function of time.)

Interestingly, de Sitter (1916c) was one of the first, perhaps *the* first, to write a semipopular description of GR in English, around the same time as he presented a technical account (de Sitter, 1916a,b) and addressed the problem of inertia (de Sitter, 1917a); this paper was also the first to introduce the de Sitter model, which he called 'system B' in contrast to Einstein's static universe 'system A'; further details on this model and other aspects of GR followed (de Sitter, 1917b,c), the former paper having the title 'On the curvature of space', which, apart from being in English rather than German, is the same as that of a later paper by Friedmann (1922), 'Über die Krümmung des Raumes'. He also wrote an early paper on 'distance, magnitude, and related quantities in an expanding universe', *i.e.* what later became known as observational cosmology (de Sitter, 1934).

1.2.4 Friedmann models: wide variety

While Einstein and de Sitter had introduced one cosmological model each, Friedmann (1922, 1924) considered the full range of FLRW models. Although published in one of the leading journals, his work was largely ignored, perhaps because he died young without a chance to promote it.

1.2.5 Lemaître: the primeval atom

Lemaître rediscovered Friedmann's work, but the emphasis was different. First, while Friedmann's work was largely mathematical, Lemaître's was more physical, discussing the very early Universe, which he dubbed the 'primeval atom' and imagined as a very large atomic nucleus which radioactively decayed.[7] At first, Lemaître favoured a model, also favoured by Eddington, which started at $t = -\infty$ arbitrarily close to the static Einstein model then, as a result of the instability of Einstein's model, began to expand, asymptotically approaching the de Sitter model for $t = +\infty$ (as do all models which expand forever and have a positive cosmological constant) (Lemaître, 1927, 1931a), though whether such instabilities can occur as a perturbation of the static Einstein model and if so whether they lead to expansion was far from clear (Lemaître, 1931b). Later, Lemaître favoured a big-bang model with an origin in the finite past (Lemaître, 1931c,d,e).[8] In particular, he preferred a model which, due to fine-tuning between the density of matter and the cosmological constant, goes through a period of very slow expansion between decelerating and accelerating phases. That allows the age of the universe to be larger than the Hubble time $1/H_0$, arbitrarily long for arbitrarily fine tuning. (Our Universe is of the same general type, though the quasi-static phase is neither very long nor very static, but rather just an inflection point; when Lemaître put forward his model, it was believed that the Hubble constant is much larger, which in a more conventional model, in particular one with no cosmological constant, would be too young, *e.g.* younger than the Earth.)

[7] Interestingly, if, ignoring gravity, the entire observable Universe could be compressed to a sphere with the density of nuclear matter, the resulting object would fit comfortably within the Solar System.

[8] At just 457 words, Lemaître (1931c) was concise; nevertheless, the importance of that contribution can be gauged by the fact that when the journal *General Relativity and Gravitation* reprinted it as a 'Golden Oldie', it was accompanied by an 18-page editorial (Luminet, 2011).

1.3 Classical FLRW cosmology

By 'classical cosmology', I refer to the study of FLRW models, both exact and approximate ones; in the latter case, it is assumed that the large-scale kinematics and dynamics are FLRW, with any inhomogeneities being essentially test particles, but also with the possibility that local inhomogeneities can affect the propagation of light (but do not affect the dynamics, hence no back-reaction). Building on that are the ideas of the hot big bang, primordial nucleosynthesis, the cosmic microwave background, the formation and evolution of structure, and so on—interesting topics all, but beyond the scope of this thesis.

Einstein, de Sitter, and Eddington all have cosmological models named after them (in Eddington's case Lemaître's orginal model), as do Lanczos (1922)[9] (a model with $\Omega_0 = 0$ and a positive cosmological constant which contracts from ∞ to a finite radius then expands) and Milne ($\Omega_0 = \lambda_0 = 0$; actually, this is the relativistic equivalent of a model proposed by Milne (1935)); Einstein and de Sitter (1932) have a joint model ($\Omega_0 = 1$ and $\lambda_0 = 0$ and hence $k = 0$); Friedmann and Lemaître have an entire set of models named after them. (Sometimes, models without a cosmological constant are dubbed Friedmann models, and those with a cosmological constant Friedmann–Lemaître models, though that is incorrect, as Friedmann explored all possibilities. It was Lemaître (and Eddington), though, who emphasized the importance of the cosmological constant.) For more on the classification of FLRW models, see Chap. 4 and the excellent paper by Stabell and Refsdal (1966).

Thus, by the early 1930s, the theory of FLRW models was in place, but observations were not yet good enough to decide which best describes our Universe. In addition, the physics of those models was investigated by the likes of Lemaître and Tolman, the latter even considering inhomogeneous models (Tolman, 1934). It was clear that, at least in principle, observations could determine which FLRW we live in or, equivalently, the values of H_0, λ_0, and Ω_0, primarily by comparing the observed distance (usually the angular-size distance or luminosity distance) as a function of redshift z to predictions of various models (see below). At the latest when Rindler (1956) cleared up the confusion concerning the concepts of different types of horizons, the theory of classical cosmology was complete. Good overviews of the history relativistic cosmology (including, but not limited to, the FLRW models) can be found in books by Barrow (2012) and Harwit (2013)[10] and in the *Oxford Handbook of the History of Modern Cosmology* (Kragh and Longair, 2019), while a good review of the history of GR has been given by Ferreira (2014) (see also my corresponding reviews of those books: Helbig, 2013, 2014b, 2019, 2014a).

1.4 Many distances

There are various definitions of distance that all correspond in the static Euclidean case, but in general are different. This work concentrates on the angular-size distance and the closely related luminosity distance; other distances are mentioned for completeness; see Kayser, Helbig and Schramm (1997, Chap. 3 in this thesis) for more details.

[9] Lanczos also discovered a solution of the Einstein field equations involving a rigid cylindrical arrangement of dust particles, later rediscovered by van Stockum and hence known as van Stockum dust; this model has closed timelike curves.

[10] While the emphasis of the book lies elsewhere, despite or perhaps because of that, the summary of the history of relativistic cosmology is excellent.

1.4.1 Proper distance

Proper distance D^{P} is shortest distance between two points, as measured by a rigid ruler. In the cosmological context, one must also specify the time of the measurement, *e.g.* for an object at redshift z, when the light was received at the observer (the usual case) or when the light was emitted. (The former is a factor of $(1+z)$ larger than the latter.) That is equivalent to $R\chi$, where R of course in general depends on time. $R_0\chi$ is often termed the co-moving distance. This distance corresponds most closely to the everyday idea of distance, and is the basic distance in GR.

1.4.2 Angular-size distance

If one knows the linear size ℓ of an object, then the angle θ under which it appears defines the angular-size distance: $D = \ell/\theta$. Since the corresponding triangle retains its shape as the universe expands, that is equivalent to the proper distance at the time the light was emitted in the case of a spatially flat universe; if the universe is not flat, then curvature effects play a role as well, since the rays of light from opposite ends of the object converge more or less than they would in the flat-universe case. The curvature effects mean that the angular-size distance is $r/(1+z)$ instead of $D^{\mathrm{P}}/(1+z)$ (the proper distance at the time of emission).

1.4.3 Luminosity distance

In static Euclidean space, the observed flux F of radiation decreases as the square of the distance, so one can define the luminosity distance

$$D^{\mathrm{L}} = D^{\mathrm{L}}_0\sqrt{\frac{F_0}{F}} \quad , \tag{1.18}$$

where F_0 is the flux at some fiducial distance D^{L}_0. Alternatively,

$$D^{\mathrm{L}} = \sqrt{\frac{L}{4\pi F}} \quad , \tag{1.19}$$

where L is the luminosity. In the cosmological case, the redshift increases the distance: since the luminosity is power, *i.e.* energy per time, the redshift decreases the energy of photons by a factor of $1/(1+z)$; in addition, the number of photons per time is reduced by the same factor, giving a factor of $1/(1+z)^2$ in flux or $(1+z)$ in distance. Without that effect, in a flat universe D^{L} would be equal to the proper distance at the time the light was received (because the corresponding angle is at the source and not at the observer as with the angular-size distance). Thus, in the cosmological context, in a flat universe the luminosity distance is greater than the proper distance at the time the light is observed by a factor of $(1+z)$; if the universe is not flat, then curvature effects play a role as well, since the rays of light diverge more or less than they would in the flat-universe case. The distance is essentially r due to curvature effects, but increased by a factor of $(1+z)$ (due to the factor of $1/(1+z)^2$ in flux mentioned above). In contrast to the angular-size distance D, the relevant quantity is r and not $r/(1+z)$ because the relevant angle is at the source (the luminosity distance is inversely proportional to the square root of the fraction of light we receive, which depends on the distance to the source *now*). That means that $D^{\mathrm{L}} = D(1+z)^2$ (Etherington, 1933), which implies that the surface brightness is proportional to $(1+z)^{-4}$. (The above applies to the bolometric luminosity.

In practice, observations are often made within a finite spectral band, reducing the flux by another factor of $(1+z)$ since the observed band corresponds to a smaller band at the source; thus the surface brightness is proportional to $(1+z)^{-5}$ and the signal-to-noise ratio to $(1+z)^{-10}$. Also, the so-called K-correction must be applied to take into account any deviations of the spectrum from the flat-spectrum case.)

1.4.4 Proper-motion distance

Although proper motion is rarely observed in a cosmological context, the proper-motion distance D^{PM} is equivalent to $D(1+z)$, where D is the angular-size distance. In the static Euclidean case, where all distances are equivalent, one can think of the proper-motion distance as the angular-size distance but with ℓ replaced by vt, where v is a known velocity and t is the time taken. Since the redshift increases all time intervals by $(1+z)$ (not just the interval defined by the frequency of light), in the cosmological case the proper-motion distance is $(1+z)$ times larger than the angular-size distance. (In the flat-space case, that is the same as the proper distance D^{P}.)

1.4.5 Parallax distance

For completeness I include the parallax distance D^{π}; this is even rarer in cosmology. $D^{\pi} = D^{\mathrm{PM}}/\sqrt{1-k\sigma^2}$. One might think that D^{π} should be the the same as the proper-motion distance, but with the relevant angle at the source rather than the observer, since no redshift effects are involved. However, the actual angle measured, β, is at the observer, and the angle at the object is *computed* (in the small-angle approximation, which is correct) to be $\pi/2-\beta$.[11] However, that assumes Euclidean space; the extra factor takes spatial curvature into account. That was already noted by de Sitter (1917c).

1.4.6 Light-travel–time distance

All of the above distances can easily be converted into one another, since only factors of $(1+z)$ and the functions sin and sinh and their inverse functions (to convert from D^{P} to r and *vice versa*) are needed. The light-travel–time distance is not so easily convertible. However, it is closely related to the proper distance. The latter is given by

$$D^{\mathrm{P}}{}_{xy} = \frac{c}{H_0}\int_{z_x}^{z_y} \frac{\mathrm{d}z}{\sqrt{Q(z)}} \quad , \tag{1.20}$$

where $Q(z)$ is defined above in Eq. (1.17), whereas the former is given by

$$ct_{xy} = \frac{c}{H_0}\int_{z_x}^{z_y} \frac{\mathrm{d}z}{(1+z)\sqrt{Q(z)}} \quad . \tag{1.21}$$

Since the proper distance *now* would be equivalent to ct if the universe were static, the factor (1+z) in the integrand takes the expansion of the universe into account. (Note that one can always define the distance between objects at two arbitrary redshifts; one does not have to be 0 (corresponding to the observer). In

[11] Note that in the common case of the parallax of a star due to the motion of the Earth around the Sun, by convention the Sun–Earth distance is used as the baseline; the corresponding parallax angle is thus half of the one which would correspond to the diameter of Earth's orbit, which might be used in practice in order to achieve greater precision.

practice, that is done most often for the angular-size distance within the context of gravitational lensing, since, in addition to the distances to the deflector (lens) and to the source, D_d and D_s, respectively, the distance between the deflector and the source, D_ds, plays a role.)

1.5 Calculation of distances in ideal FLRW models

The history of distance calculation is worth a work in itself; here, I give only a *very* brief overview.

1.5.1 $z \ll 1$

If the redshift z is small ($z \ll 1$), then all distances are approximately the same; $D = (c/H_0)z$ (H_0 is usually given in the units km/s/Mpc). In this case, the recession velocity is given by cz. For larger redshifts, in general the linearity of distance with z breaks down, different definitions of distance (see above) result in different values for the same redshift, and the recession velocity is no longer given by cz. (In particular, it is *not* given by the relativistic Doppler formula (Harrison, 1993; Helbig, 2017); rather, it is given by $H_0 D^\mathrm{P}$, and can become arbitrarily large.)

1.5.2 Special cases

There are analytic formulae for certain special cases. Particularly well known is the one found by Mattig (1958) for $\lambda_0 = 0$. A compendium of all known solutions is given by Kayser, Helbig and Schramm (1997, appendix B); see Chap. 3.

1.5.3 Series expansions

For arbitrary values of λ_0 and Ω_0, series expansions were important as long as redshifts were relatively low and/or computing time relatively expensive. Of historical interest is the deceleration parameter $q := -(\ddot{R}R)/(\dot{R}^2) \equiv \ddot{R}/(RH^2) \equiv \Omega/2 - \lambda$, in particular $q_0 = \Omega_0/2 - \lambda_0$, since q_0 is the first non-trivial term in such expansions; Sandage (1970) once described cosmology as 'a search for two numbers', namely H_0 and q_0. (Note that $q = \Omega/2 - \lambda$ holds in general and thus the sign of q indicates whether the universe is accelerating, decelerating, or neither at the time Ω and λ have their corresponding values; in particular, the sign of q_0 indicates whether the universe is accelerating, decelerating, or neither at the present time. The definition of q_0 includes a minus sign because the assumption when the definition was first made was that the Universe would be decelerating, and one wanted the parameter to be positive.)

1.5.4 Numerical calculations

One can of course compute solutions numerically. That was done for a large range of cosmological models by Refsdal, Stabell and de Lange (1967), whose publication contains results for 100 cosmological models, 10 values of Ω (they used $\sigma := \Omega/2$, not to be confused with σ as defined above) and 10 values of q; in most cases, each model has one page of results. Various quantities, including distances, related quantities, and the age of the universe, were calculated for several redshifts.

1.5.5 Elliptic integrals

The general solution for distance as a function of redshift (and for $R(t)$) is given by elliptic integrals. This has long been known, but many authors simply noted this fact without giving any details: "…may be integrated in terms of elliptic functions, but it will be sufficient for our purposes to carry out a qualitative integration" (Bondi, 1960); "…[l]ike Bondi we shall here only give a qualitative solution…" (Stabell and Refsdal, 1966). The first clear and detailed exposition was given by Feige (1992).

1.6 Effect of small-scale inhomogeneities

All of the remarks on distances above assume that the universe is exactly homogeneous, at least as far as light propagation is concerned. It is certainly the case that our Universe is not exactly homogeneous, and it is conceivable that small-scale inhomogeneities affect light propagation even if the large-scale kinematics and dynamics are FLRW. In the homogeneous case, one uses the FLRW model as determined by the cosmological parameters λ_0 and Ω_0 to calculate the proper distance now of an object at a given redshift z. In the case of a spatially flat universe, that corresponds to $D(1+z)$, where D is the angular-size distance (or, alternatively, to $D^{\mathrm{L}}/(1+z)$, where D^{L} is the luminosity distance). If the universe is not spatially flat, curvature effects must be taken into account; see Sect. 1.4 for details. Matter in the universe influences the distance, in this method of calculation, only in that it influences the dynamics and hence the expansion history.

Zel'dovich (1964a) (English translation: Zel'dovich, 1964b) developed an alternative method for calculating the angular-size distance, based on a second-order differential equation where one term describes the expansion of the universe and another the convergence due to matter in the beam. Historically, that method has been important because it allows one to calculate the distance for the case that the density of matter in the beam is less than (or perhaps even if it is more than) the mean density of the universe. This thesis concerns the history of that idea and its applications, and more-detailed descriptions of my own work in this area. However, the method can of course also be used for distance calculation in the homogeneous case.

1.7 This thesis

Part I includes this introduction, which puts distance calculation into the general cosmological context, and Chap. 2, which is a review of the ZKDR[12] distance, *i.e.* distance calculated assuming an FLRW background to describe the geometry and expansion history, but with small-scale inhomogeneities which affect light propagation and hence distances determined by angles; that was written explicitly for this thesis but before submission published (incorporating some suggestions from the referees) in *The Open Journal of Astrophysics* (Helbig, 2020a). Part II concerns the work of Kayser, Helbig and Schramm (1997), which provides an overview of cosmological distances and discusses a general second-order differential equation and its numerical implementation for

[12]ZKDR is an acronym introduced by Santos and Lima (2006) and referring pioneers in the field of distance calculation in inhomogeneous universes, namely Zel'dovich, Kantowski, Dyer, and Roeder, though I take the 'D' to refer to Dashevskii as well, my criterion for being part of the acronym being having (co-)authored at least two papers on this topic, at least one of which was published within ten years of the first paper on this topic (Zel'dovich, 1964a).

their calculation, as well as a user's guide (Helbig, 1996) for technical details. Part III, consisting of just Chap. 5, demonstrates the potential danger of assuming that the Universe is completely homogeneous. Part IV discusses two different situations where the degree of homogeneity can be important in gravitational lensing: Chap. 6 discusses time delays between different images of the same source (which can be used to measure the Hubble constant H_0) (Helbig, 1997), and Chap. 7 discusses the separation between images of the same source as a function of the inhomogeneity and the cosmological parameters (Helbig, 1998a,b). Part V concerns the m–z relation for Type Ia supernovae: Chap. 8 shows how inhomogeneity can affect the measurement of cosmological parameters and, *vice versa*, how one can determine the inhomogeneity parameter η from observations of the m–z relation for Type Ia supernovae (Helbig, 2015a). Observations indicate that $\eta \approx 1$, *i.e.* the Universe (behaves as if it) is essentially homogeneous; Chap. 9 discusses a method to determine whether that is the case just on average, as first pointed out by Weinberg (1976), or whether each line of sight essentially traverses a fair sample of the Universe, so that $\eta \approx 1$ for each object observed (Helbig, 2015b).

Chapter 2

Calculation of distances in cosmological models with small-scale inhomogeneities and their use in observational cosmology: a review

2.1 Context

This thesis is concerned with the calculation of distances in cosmological models with small-scale inhomogeneities. The first paper on this topic was written about 57 years ago (Zel'dovich, 1964a,b). I started my master's thesis work at the Hamburg Observatory in 1992, almost precisely halfway between the start of research in this field and now (and, since I was born in 1964, at the halfway point of my life up until now.). My papers on this topic (Helbig, 1996; Helbig and Kayser, 1996b; Kayser, Helbig and Schramm, 1997; Helbig, 1997, 1998a,b, 2015a,b) collected in this thesis, as well as conference presentations with no additional material (Helbig, 2015c, 2016a), were all born out of necessity: the development of the general code (Helbig and Kayser, 1996a; Helbig, 1996; Kayser, Helbig and Schramm, 1997) and applying it to specific problems (Helbig and Kayser, 1996b; Helbig, 1997, 1998a,b, 2015a,b). When working on those, I knew the literature enough to know that I was doing something new, and cited some of the major papers to provide some basic context. For this thesis, I wanted to cite a general review of the topic, in order to present my work in a broader context. Since I found none, I wrote one myself (Helbig, 2020a), which is included in this chapter. It is published in *The Open Journal of Astrophysics*, which is a new, online-only[1] refereed journal. It is of similar quality as longer-established journals such as *Monthly Notices of the Royal Astronomical Society*, *The Astrophysicsl Journal*, and *Astronomy & Astrophysics*. Since many in the fields of cosmology, astrophysics, and astronomy read papers mainly on arXiv, it uses arXiv as a distribution mechanism, rather than re-inventing the wheel by setting up its own. The webpage of the journal, `https://astro.theoj.org/`,

[1] Note that even some traditional journals are now online only.

contains short descriptions, abstracts, and metadata of the papers as well as a link to arXiv, where the paper itself can be read. Papers are listed there only after they have been accepted *via* the normal refereeing process.

Version December 30, 2019
Preprint typeset using LaTeX style openjournal v. 09/06/15

CALCULATION OF DISTANCES IN COSMOLOGICAL MODELS WITH SMALL-SCALE INHOMOGENEITIES AND THEIR USE IN OBSERVATIONAL COSMOLOGY: A REVIEW

Phillip Helbig
Institut d'Astrophysique et de Géophysique (Bât B5c)
Université de Liège
Quartier Agora
Allée du 6 août, 19C
B-4000 Liège 1 (Sart-Tilman)
Belgium

(Received September 12, 2019; Revised November 17, 2019; Accepted December 23, 2019)
Version December 30, 2019

ABSTRACT

The Universe is not completely homogeneous. Even if it is sufficiently so on large scales, it is very inhomogeneous at small scales, and this has an effect on light propagation, so that the distance as a function of redshift, which in many cases is defined *via* light propagation, can differ from the homogeneous case. Simple models can take this into account. I review the history of this idea, its generalization to a wide variety of cosmological models, analytic solutions of simple models, comparison of such solutions with exact solutions and numerical simulations, applications, simpler analytic approximations to the distance equations, and (for all of these aspects) the related concept of a 'Swiss-cheese' universe.

Subject headings: cosmology: theory – cosmological parameters – distance scale – large-scale structure of Universe – cosmology: observations – cosmology: miscellaneous

1. INTRODUCTION

The Universe is not completely homogeneous; if it were, there would be no observers and no objects to be observed. Nevertheless, distances are often calculated as a function of redshift as if that were the case, at least as far as light propagation is concerned. Whether this is a good approximation depends at least on the angular scale involved. The simplest more refined model retains the background geometry and expansion history of a Friedmann–Robertson–Walker (FRW) model but separates matter into two components, one smoothly distributed comprising the fraction η of the total density and the other $(1-\eta)$ consisting of clumps, and considers the case where light from a distance object propagates far from all clumps (this is equivalent to the case of negligible shear). Over a period of more than 50 years, various authors have described more-general versions of this approximation with regard to cosmology, found analytic solutions, discussed similar approximations, compared it with exact solutions and with brute-force numerical integration based on the gravitational deflection of matter along and near the line of sight, examined the assumptions involved, applied it to various cosmological and astrophysical problems, and developed simple analytic approximations both for more-exact solutions (the latter based on more-complicated analytic formulae or on numerical integration) and for numerical simulations. While there is no doubt that such an approximation is valid for a universe with the corresponding mass distribution, recent work indicates that our Universe is not such a universe, but rather one in which the 'standard distance' (*i.e.* calculated under the assumption of strict homogeneity) is valid, even for small angular scales, at least to a good approximation.

Further refinements of this approximation are not discussed here, *e.g.* weak gravitational lensing with non-negligible shear, strong gravitational lensing[1], or inhomogeneities so appreciable that they influence the large-scale geometry and/or expansion history of the universe ('back reaction'). Similarly, extensions to the FRW models, such as some sort of 'dark energy' other than the cosmological constant, are not considered; neither are those which violate the Cosmological Principle, *e.g.* ones in which we are within a large void, Lemaître–Tolman–Bondi models, *etc.* I also omit wrong results or misleading conclusions unless they have been often cited without all of the community noticing the mistake (either there was no correction or the correction has been ignored).

The order is chronological in the sense that I discuss all papers on the first topic to appear, then all on the second topic, and so on.

I refer to the distance calculated based on the above approximation as the *ZKDR* distance, a term introduced by Santos & Lima (2006) and referring to Zel'dovich, Kantowski, Dyer, and Roeder, though I take the 'D' to refer to Dashevskii as well, my criteria for being part of the acronym being having (co-)authored at least two papers on this topic, at least one of which was published within ten years of the first paper on this topic (Zel'dovich 1964a,b).

In gravitational lensing, it is clear that the approximation of a completely homogeneous universe with regard to light propagation cannot be valid, since otherwise there would be no gravitational lensing. Perhaps for this reason, the ZKDR distance has been used more in gravitational lensing than in other fields. Since α is almost universally used to denote the gravitationl-lensing bend-

phillip.helbig@doct.uliege.be, helbig@astro.multivax.de

[1] While some cases of strong gravitational lensing are discussed, in most cases these are not concerned with the influence of the lensing effect on the distance; rather, the approximations discussed are used to calculate the distances involved, the strong lensing being calculated explicitly.

ing angle, Kayser *et al.* (1997), hereafter KHS, adopted η instead of the more confusing α or $\tilde{\alpha}$ used by some other authors; since then, some authors other than KHS have also used η instead of α or $\tilde{\alpha}$ for the inhomogeneity parameter. In the following, I will use the notation of KHS except occasionally when explicitly referring to equations in the works of other authors, who use various and sometimes confusing notation schemes—in particular, using z for anything other than redshift in a paper on cosmology is very confusing (see Tab. 1).

2. ZEL'DOVICH (1964)

Zel'dovich (1964b, hereafter Z64)[2] started the tradition; many (not only) today might find his paper somewhat idiosyncratic, difficult to follow, and wrong in parts, but he introduced a simple and useful basic idea: local inhomogeneities in the distribution of matter can lead to significantly different angular-size and luminosity densities from those derived under the assumption of a perfect FRW model.

2.1. *Summary*

The first attempt to calculate distances in a universe with small-scale inhomogeneities is, as far as I know, that of Z64. This begins a tradition of calculating distances in a more realistic universe, namely one with small-scale inhomogeneities, but where the large-scale dynamics is given by an FRW model. In other words, it is a perturbed FRW model: The zeroth-order approximation for cosmology, which is actually quite good (Green & Wald 2014), is that the Universe is described by a Robertson–Walker metric (Robertson 1935, 1936; Walker 1935, 1937, the latter paper by Walker is very often incorrectly cited as having been published in 1936) which is a purely descriptive kinematic idea with no physics content, merely the characterization of a homogeneous and isotropic universe, and that the expansion history is given by one of the models explored by Friedmann (1922, 1924) (hence FRW), which are in turn based on relativistic cosmology as introduced by Einstein (1917). Occasionally, the term Friedmann–Lemaître–Robertson–Walker (FLRW) is used to include a reference to Lemaître (1927); while he made important contributions to cosmology, none of them went beyond the work of Friedmann (1922, 1924) with respect to the metric. 'It is assumed that…the amount of matter removed is small and the general motion is not affected.' The main model considered is 'a flat Friedman [sic] model with pressure equal to zero'. In modern notation, $\Omega_0 = 1$ and $\lambda_0 = 0$. The physical model assumes that all matter exists in galaxies[3] and that distant objects are seen between galaxies, *i.e.* such distant objects 'do not have galaxies within the cone subtended by them at the observer'. (The cone is often referred to as the beam.)

After the standard angular-size distance[4] is derived *via* a differential equation for the separation between two light rays, the deflection due to a point mass, equation (12), is used to calculate the deviation from the completely homogeneous case when the beam is devoid of matter. This equation is generalized to a uniform density distribution to calculate the total deflection, which is towards the outside since the removal of matter in the beam formally corresponds to negative mass. This leads to a differential equation which in turn leads to the expression for the angular-size distance in the Einstein–de Sitter model in the empty-beam case, denoted by f_1 in Z64. It is noted that this function 'increases monotonically right up to the [particle] horizon ($\Delta = 1$) where it reaches the value 2/5'. The value 2/5 is exact, but the right-hand side of the unnumbered equation between equations (21) and (22), 1600, is too precise (though the correct value rounded to four digits is 1599, much closer than in the case discussed in Sect. 2.2).

It is noted that 'the calculation can be repeated for the case when $\rho \neq \rho_c$, *i.e.*, for a hyperbolic or closed universe', though this is given (without derivation) only for the 'limiting case $\rho \to 0$, Milne model'. Equations (23) and (24), for the Milne and Einstein–de Sitter models respectively, are of course special cases of the formulae derived by Mattig (1958), who gives a simple formula for $\Omega_0 = 0$, essentially the same as that of Z64, and a more complicated formula for $\Omega_0 > 0$ (though assuming $\lambda_0 = 0$); see KHS, equations (B24) and (B25). In the case of $\Omega_0 = 0$, the value of η doesn't matter. Equation (25),

$$f_1 = 2/5[1 - (1 - \Delta)^{5/2}] \quad ,$$

is, in modern notation,

$$\frac{H_0}{c} D^{\mathrm{A}} = \frac{2}{5}\left[1 - (z+1)^{-\frac{5}{2}}\right]$$

(*cf.* KHS, equation (B15 (II))).

2.2. *Remarks*

There are several strange things about this paper. First, the 'remarkable feature' that the angular-size distance has a maximum is noted. Second, it is claimed that this 'is caused by the curvature of space due to the matter filling the universe', which is strange because later in the paper the main model considered is a *flat* universe, *i.e.* one with no spatial curvature. Third, it is pointed out that the maximum 'occurs only when there is matter within the cone subtended by the object at the point of observation'. Fourth, for a modern reader, the notation is extremely bizarre; Tab. 1 shows the equivalents in modern notation of the quantities used. I now discuss each of these in turn.

What is remarkable about the fact that the angular-size distance has a maximum at some redshift? In modern notation, the angular-size distance D^{A} is, by definition, l/θ, where l is the physical projected length of the object and θ the angle which it subtends, *i.e.* the angle at the observer formed by light rays from both ends of the object.[5] The triangle made by the object and the light rays retains its shape as the universe expands. Thus, ignoring curvature effects for the moment,

[2] This discussion follows the English translation of the Russian original (Zel'dovich 1964a).

[3] More precisely, that the mass of the intergalactic medium can be neglected compared to the mass of matter contained in galaxies.

[4] His equation (7) assumes the Einstein–de Sitter model (introduced at the start of the appendix as the 'flat Friedman [*sic*] model with pressure equal to zero'); a casual reader might think that it applies more generally.

[5] See the paper by KHS for definitions of various cosmological distances which are consistent, use modern notation, and deviate as little as possible from the approximate consensus.

TABLE 1
NOTE THAT t, t_0, c, ρ ARE THE SAME IN THE Z64 AND MODERN NOTATIONS. Z64 DISTINGUISHES BETWEEN Θ AND Θ_1 FOR THE CASES $\eta = 1$ AND $\eta = 0$, RESPECTIVELY, THOUGH IN BOTH CASES THE QUANTITY IS THE *observed* ANGULAR SIZE. SIMILARLY, Θ_2 IS THE OBSERVED ANGULAR SIZE IN THE CASE OF STRONG GRAVITATIONAL LENSING. EXCEPT IN THE CASE OF f_0, QUANTITIES DEPENDENT ON THE COSMOLOGICAL MODEL ASSUME THE EINSTEIN–DE SITTER MODEL.

Z64 notation	modern notation
Θ	θ
Θ_1	θ
Θ_2	θ
Δ	$1-(1+z)^{-1}$
$(1-\Delta)^{-1}-1$	z
ω_1	ω_0
ω_0	ω
r	l or ℓ
f	$Hc^{-1}D^{\mathrm{A}}$ $(\eta = 1)$
f_1	$Hc^{-1}D^{\mathrm{A}}$ $(\eta = 0)$
f_0	$Hc^{-1}D^{\mathrm{A}}$ (Milne model)
κ	G
H	H_0
R	D^{A} $(\eta = 1)$
R_1	D^{A} $(\eta = 0)$
Mps	Mpc

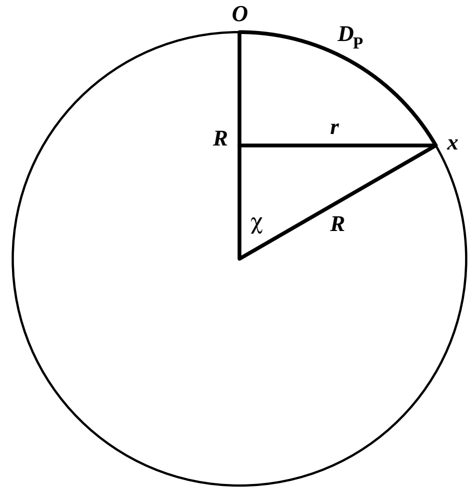

FIG. 1.— Although the corresponding definitions are valid for models with k of 0 and -1 as well, easiest to visualize are distance definitions for the case $k = +1$. The universe can be thought of as a curved three-dimensional space, corresponding to the circle. Two dimensions are hence suppressed, so that the two dimensions in the plane of the figure can show the universe and its spatial curvature. R is the scale factor of the universe, as usual chosen to correspond to the radius of curvature. The observer is located at the top of the circle at O and observes an object located at x. D^{P}, the length of the arc, is the proper distance to that object. For $\eta = 1$, the angular-size and luminosity distances (as well as other distances not discussed here such as the proper-motion distance and parallax distance) depend on $r = R\sin(\chi)$ in a relatively simple manner (see KHS). Note that χ is constant in time; one can use it or $\sigma = r/R$, which is also constant in time, as the basis for a so-called co-moving distance.

the angular-size distance is the proper distance to the object at the time the light was emitted: The proper distance D^{P} (sometimes written D_{p} or D_{P}) is the distance which one could, in a *gedankenexperiment*, measure with a rigid ruler instantaneously (such that the distance does not change during the measurement due to the expansion of the universe). As such, it changes with time due to the expansion of the universe. Often, the co-moving distance is defined as the proper distance at the present time. Thus, the proper distance at a different time is simply the current proper distance divided by $(1 + z)$, the time being that when the light of an object with redshift z was emitted. This agrees with the definition used by many authors, such as Berry (1986), who defines it as 'the distance measured with a standard rod or tape, in a reference frame where the events occur simultaneously'. Beware that sometimes the same distance is denoted by different symbols, *e.g.* d_{prop} by Weinberg (1972), L by Harrison (1993), d by Sandage (1995), D by Davis & Lineweaver (2004), d_{p} by Heacox (2015), d_p by Ryden (2017), and sometimes also by different names, though it is clear from the discussion that the same distance as that called the proper distance by Weinberg (1972) is being discussed, *e.g.* 'distance between two fundamental particles at time t' (D_1) by Bondi (1961), 'tape-measure distance' (L) by Harrison (2000), 'instantaneous physical distance' by Carroll (2019); the term 'line-of-sight comoving distance' is also sometimes used, as opposed to the 'transverse comoving distance', which is very confusingly called the 'angular size distance' by Peebles (1993), who uses the term 'angular diameter distance' for what is called the angular-size distance by almost everyone else—indeed, the two terms are usually considered to be equivalent; the transverse comoving distance is the same as the proper-motion distance; see KHS.

At small redshifts, as the redshift increases, the object was farther away (in proper distance) when the light was emitted, thus the angular-size distance increases with redshift. However, at large redshifts, light was emitted when the proper distance was small, long ago, but, due to the more rapid expansion of the universe in the past, is reaching the observer just now. Thus, at large redshifts, the angular-size distance, being the proper distance when the light was emitted, is small. This explains the 'remarkable' maximum. Another way of thinking of this is that the angular-size distance approaches zero as z approaches 0, but also as z approaches ∞, because the scale factor R (see Fig. 1) approaches 0 in such cases; in other words, the maximum in the angular-size distance depends on a finite particle horizon. Of course, not all cosmological models have a finite particle horizon and those that don't also have no maximum in the angular-size distance. This applies only to the standard distance, *i.e.* assuming complete homogeneity. For the ZKDR distance, it is of course possible that there is no maximum in the angular-size distance even though the universe has a particle horizon.

The above explanation is exact in a spatially flat universe, thus contradicting the claim that the maximum is somehow caused by the curvature of space. With spatial curvature, the angular-size distance corresponds not to the proper distance when the light was emitted, but rather to the coordinate distance r, defined as the product of the scale factor R and $\sin(\chi)$, χ, or $\sinh(\chi)$ for k equal to $+1$, 0, or -1, *i.e.* positive, zero, or negative spatial curvature, respectively; $\chi = D^{\mathrm{P}}/R$ (see Fig. 1.) This is analogous to the correction applied due to the curvature of the surface of the Earth when calculating the length along a parallel of latitude from the difference in longitude betweem the ends; the length (in the limit of small θ) is not $D^{\mathrm{P}}\theta$ but rather $R\sin(\chi)\theta$, where χ is D^{P}/R, D^{P} being the distance measured along the surface of the Earth ('as the crow flies') and R is the radius of the Earth (assumed to be perfectly spherical). (Note that $\chi = \pi/2 - \phi$, where ϕ is the geographic latitude, if we think of the observer as being at the north pole.) Thus, this distance at first increases with increasing D^{P}, though more slowly than in the flat case, reaches a maximum at the equator, then decreases to zero at the opposite pole.

Fig. 1 illustrates various distances. One can see that for small χ, D^{P} and r are approximately the same (exactly so in the limit $D_{\mathrm{P}} = r = 0$). When χ reaches 90 degrees, r (and hence the angular-size distance) reaches its maximum. For larger χ, the angular-size distance decreases, reaching 0 for $\chi = 180°$. It then increases again, reaching its maximum again at $\chi = 270°$, then decreases again, reaching 0 at $\chi = 360°$. The maximum value of χ depends on the cosmological model. Light travels along the circle from x to O. In an expanding universe, R was smaller when the light was emitted, hence, the distance defined *via* light-travel time is smaller than D^{P}, while they coincide in a static universe. The ratio R_0/R_{e}, the scale factor now compared to the scale factor when the light was emitted, is equal to $1 + z$. Distances related to r depend on η, while D^{P} and the distance defined *via* light-travel time do not (the latter at least to a very good approximation.) (More precisely, the angular-size distance and other distances can be calculated relatively easily from r for $\eta = 1$. For $\eta \neq 1$, r still exists, but the relation between r and the angle defining the distance is changed, so the distance can no longer be simply calculated from r.) For $\eta = 1$, since $D^{\mathrm{A}} = r/(1+z)$, it is clear that for $z = \infty$ the angular-size distance must be zero.

That is another mechanism for the presence of a maximum in the angular-size distance. Consider first a static universe with positive spatial curvature (the Einstein model) and an observer at the 'pole'. For increasing proper distance, the angular size of a standard rod first decreases (*i.e.* the angular-size distance increases) up to a minimum at the 'equator', then increases again, becoming infinite at the opposite 'pole'. (This can continue indefinitely, with the angular size decreasing again as the proper distance further increases until the 'equator' is reached (but at the 'opposite side'), then increasing again until the object returns back to the observer, then decreasing again during the second loop around the universe, and so on.) Of course, in a static model there is no redshift, but there are quasi-static models where the universe expands very slowly. Large differences in proper distance correspond to small differences in redshift, and hence small differences in the scale factor at the time the light was emitted. If light is received from an object near the opposite 'pole', it will obviously have a much smaller angular-size distance than one near the 'equator', even though the scale factor was only slightly smaller when the light was emitted in the former case (thus it will have a slightly larger redshift). (Our Universe never went through such a quasi-static phase, so the first effect is more important in practice.)

As noted above, the claim that the maximum is due to the curvature of space is strange, as it can exist in a flat universe; in particular, it exists in the first model considered in the paper, the Einstein–de Sitter model, which is spatially flat. (Perhaps he meant 'spacetime' rather than 'space'; it is not a wrong translation, since the original also has the Russian word for 'space' and not 'spacetime'.) The third point is more interesting: a maximum exists only if the beam is not empty. Since Z64 seemed surprised that the maximum exists, while I have shown above that it is perfectly natural to expect it, perhaps a better formulation is that the maximum disappears in the empty-beam case. I return to this in Sect. 3.1.

We are so used to the redshift z as the principle observable quantity and proxy for distance that the use of $\Delta = 1 - 1/(1+z)$ is rather confusing. It does have the interesting property, though, that it ranges from 0 at the observer to a maximum of 1 for light emitted at the big bang. It follows from the simple definition given by equation (8) that $\Delta = (\omega_0 - \omega_1)/\omega_0$, but note that '$\omega_1$ is the frequency of light received by the observer at time t_0 and ω_0 the frequency of light emitted by the object at time t'. Usually, '0' refers to the time of reception, but in any case various quantities almost always have the same indices to refer to the same times.[6] This is not a misprint, though, since other formulae which follow from this definition can be shown to be equivalent to more-familiar formulae; *e.g.* equation (9) corresponds to equation (B24) in the paper by KHS for $\Omega_0 = 1$, *i.e.* the angular-size distance in the (completely homogeneous) Einstein–de Sitter model, denoted by f in Z64; in modern notation, this is $cH^{-1}2((1+z)^{-\frac{1}{2}} - (1+z)^{-\frac{3}{2}})$. There is something of a misprint in equation (2) of Z64: the character before the exponent '2' in the denominator, which looks like a dagger in a slanted font, should be t, and it is not part of the exponent, *e.g.* correct is

$$\rho = \frac{1}{6\pi\kappa t^2} \quad ;$$

perhaps the lower part of the t has not been printed; this is correct in the Russian orginal (Zel'dovich 1964a). (Equation (2) in Z64 follows from the standard definition $\Omega = (8\pi G\rho)/(3H^2)$ for $\Omega = 1$ and using equation (3) to express H in terms of t.)

The bulk of the paper is the appendix (two and a half pages), which contains all the equations. The first one-and-one-half pages are essentially a non-mathematical summary, but also include several interesting points.

Figures 3 and 4 are never referred to in the text (neither in the translation nor in the original). Figure 3 seems to show the case in which the two ends of an object are multiply imaged, while figure 4 seems just to show the definition of an angle. Note that figure 6 is incorrect in that it appears that f_1 has a maximum for $\Delta < 1$; the angular-size distance *never* has a maximum for $\eta = 0$ (see Sect. 3).

The intergalactic medium is said to contain neutrinos and gravitons. Interestingly, gravitons have a rest-mass

[6] There is perhaps *some* justification for using the subscript 0 to correspond to the time of emission when the frequency is discussed, since times earlier than the present correspond to a *higher* frequency, though shorter wavelengths and smaller quantities based on length; this probably creates more confusion than it avoids, though.

of 0 and neutrinos were believed to as well when the paper was written. Such particles thus correspond to a different equation of state ($w = 1/3$), though in this case that is irrelevant since the density is assumed to be negligible. If the density of such matter is not negligible, then matters become more complicated. In general, the term ρ above is $\rho + p$, where p is the pressure. In the case of ordinary matter ('dust'), $p = 0$, hence ρ is sufficient. In the case of the cosmological constant, which can be thought of as a perfect fluid with $\rho = -p$, the two terms cancel; the only effect of the cosmological constant on the ZKDR distance is due to its effect on the expansion history of the universe. Other equations of state can in principle be taken into account in the ZKDR *ansatz* by including the corresponding $\rho + p$ terms, but in such cases the concept of a single parameter η would be inappropriate, since one would not expect the various components to clump in the same manner.

Gravitational lensing is mentioned for the case when there is a galaxy within the cone, confusingly citing Fritz Zwicky (see below in this section). In this case, it is noted that no general expression can be derived, but that (the equivalent of) the angular-size distance as a function of z is given by a weighted mean, though this is not defined, much less derived. Though worded somewhat confusingly, it is pointed out that a mass outside the cone acts, to first order, as a pure-shear gravitational lens, distorting though not changing the area subtended by the object and (due to the conservation of surface brightness in gravitational lensing[7], implicitly assumed here) thus also not changing the apparent magnitude.

His equation (11) looks suspicious because the right-hand side of 1200 appears to be too round a number. To the same precision, the correct value is 1184. (A more precise value is 1184.365. Of course, this much precision is not needed, but usually all quoted figures are correct. If two significant figures are sufficient, then 1.2×10^3 would make more sense.) Units, not explicitly mentioned, are Mpc. (Note that the units of H are 'km/sec · Mps', normally written 'km/s/Mpc' or 'km/(s·Mpc)' or '$\mathrm{km\,s^{-1}\,Mpc^{-1}}$'.)

As mentioned above, it is noted that 'the function f_1 increases monotonically right up to the [particle] horizon ($\Delta = 1$) where it reaches the value 2/5'. However, the plot of this function in figure 6 clearly shows a maximum for $\Delta < 1$, after which the value decreases somewhat.

The reference Zwicky (1937c) is wrongly assigned the year 1927. That reference contains only a very short and general discussion on 'nebulae as gravitational lenses' and does not address the phenomena mentioned in the text. It does say that a more detailed description will be provided in *Helv. Phys. Act.*, but that is not the paper in that journal mentioned in reference 4 (Zwicky 1933), a paper in German on various aspects of the redshift of extragalactic nebulae, which doesn't mention gravitational lensing at all; among other things, Zwicky points out that the dispersion of velocities of galaxies in the Coma cluster indicates that the density of dark matter must be at least 400 times that of luminous matter—and, of course, this was written before Zwicky (1937c). Zwicky (1937a,b) are the (two short) papers which discuss nebulae as gravitational lenses, both cited by Zwicky (1937c).

2.3. *Discussion*

Z64 presented analytic formulae for the angular-size distance for three cosmological models: $\Omega_0 = 1$ and $\lambda_0 = 0$ (Einstein–de Sitter) for the values $\eta = 1$ (standard distance) and $\eta = 0$ (the main result of that work), as well as for $\Omega_0 = 0$ and $\lambda_0 = 0$ (the general-relativistic equivalent of the Milne model; since the density is 0 in this case, the value of η doesn't matter).

Z64 alerted people to the fact that the standard distances, which assume complete homogeneity, are perhaps not appropriate, and demonstrated that effects due to a universe with small-scale inhomogeneities can be appreciable. He also introduced the idea of calculating the effect as a negative gravitational-lens effect, based on simplifying assumptions rather than calculating it for an analytically soluble (but perhaps less realistic) case.

3. DASHEVSKII & ZEL'DOVICH (1965)

Dashevskii & Zel'dovich (1965, hereafter DZ65)[8] derived an expression for the angular-size distance for the case of a completely empty beam for arbitrary values of Ω_0 ($\lambda_0 = 0$ is still assumed). Compared to Z64, it is more general with respect to the large-scale cosmological model. They noted that the expression does not have a maximum.

3.1. *Summary*

As noted above, Z64 claimed that the maximum in the angular-size distance (in the case of the Einstein–de Sitter model studied) 'is caused by the curvature of space due to the matter filling the universe'. This is somewhat dubious, since the Einstein–de Sitter model is spatially flat. DZ65 have a perhaps somewhat better formulation, claiming that the 'effect depends on the bending of light rays by matter present within the light cone' and assert that 'it follows from this that for objects in whose light cone there is by chance no matter there should be no minimum angular diameter right up to the [particle] horizon'. The claims are true, but one can ask whether their explanation is the best one. (As will be discussed in Sect. 4.1, there is *always* a maximum as long as the beam is not completely empty, though the emptier the beam, the higher the redshift of the maximum.)

In addition to the wider range of cosmological models considered, DZ65 derive the expression *via* a different, though equivalent, route. No analytic solutions are presented, but f and f_1 are plotted as functions of Δ for a few values of Ω_0, and for a few more values of Ω_0, Δ_{max} (the value of Δ at which the maximum in the angular-size distance for $\eta = 1$ occurs) and the values of f at Δ_{max} and f_1 at $\Delta = 1$ are tabulated. In addition, there is a column for $\Omega_0 = \infty$, where $\Delta_{\mathrm{max}} = 0.25$,

[7] Since gravitational lensing conserves surface brightness, magnification (increase in area) implies amplification (increase in apparent brightness, *i.e.* energy per time from the source received at the observer). In this sense, the terms are interchangeable. However, one or the other term can be more appropriate depending on the phenomenon discussed, *e.g.* 'amplification' when discussing the change in apparent magnitude of a lensed source and 'magnification' when discussing the size of an extended source. In the case of the number of sources in a certain range of apparent magnitude in a given area of sky, both effects play a role, and whether there is an increase or decrease depends on the luminosity function.

[8] This discussion follows the English translation of the Russian original (Dashevskii & Zel'dovich 1964).

$f(\Delta_{\max}) = 0.65/\sqrt{\Omega}$ and $f_1(\Delta = 1) = 1.18/\sqrt{\Omega}$. This is not mentioned in the text, but is apparently an approximation for $\Omega_0 \gg 1$. I have checked this numerically and found that their approximation answers pretty nearly. (Of course, $\Delta_{\max}$ also depends on Ω_0, though less sensitively than $f(\Delta_{\max})$ and $f_1(\Delta = 1)$.)

Note that an analytic solution, though a rather complicated one, for the case $\lambda_0 = 0$ and $\eta = 0$ does exist, first derived by Dyer & Roeder (1972); *cf.* KHS, equation (B15). For $\lambda_0 = 0$ and $\eta = 1$, the formulae derived by Mattig (1958) apply.[9]

Several interesting features are pointed out in the text and/or are obvious from the figure (if Ω_0 is not mentioned, then the effect is independent of the value of Ω_0):

- The angular-size distance for $\eta = 0$ increases monotonically with redshift.
- The angular-size distance for $\eta = 0$ is less than the light-travel–time distance $c(t_0 - t)$ and larger than the angular-size distance for $\eta = 1$ (at least for $\lambda_0 = 0$).
- The angular-size distance for $\eta = 0$ has its maximum value at $z = \infty$.
- For $\eta = 0$, $\mathrm{d}D/dz = 0$ at $z = \infty$.
- The angular-size distance for $\eta = 1$ has a maximum at $z < \infty$.
- The value of the maximum of the angular-size distance for $\eta = 1$ increases with decreasing Ω_0.
- The redshift of the maximum of the angular-size distance for $\eta = 1$ increases with decreasing Ω_0.
- The angular-size distance for $\eta = 1$ is 0 at $z = \infty$.
- Both for $\eta = 0$ and $\eta = 1$, the value of D^{A} at any redshift increases with decreasing Ω_0.
- For given values of Ω_0 and z, D^{A} for $\eta = 0$ is always larger than D^{A} for $\eta = 1$.

DZ65 end with remarks on the 'validity of the method proposed in the paper', the validity being guaranteed by the fact that they 'are adding small effects in the linear region'.

3.2. *Remarks*

The title is also confusing, since there is no paper with a similar title but with 'I' instead of 'II'. It is clear from the first sentence, though, that Paper I is Z64.

The theme of confusing notation continues. What Z64 called r, DZ65 call z. While r is often used for a length of some sort, this is less common for z. Of course, the fact that z is normally used for the redshift adds to the confusion. What Z64 called Θ, DZ65 call ϕ. DZ65 adopt the usual convention of using the suffix 0 to denote the present time, in this case the time of observation and the time the radiation reaches the observer. Hence, what Z64 called ω_1, DZ65 call ω_0, and what Z64 called ω_0, DZ65 call ω_t.

Criticizing Wheeler (1958), DZ65 note that the claim that the maximum occurs only in the case of a spatially closed universe is wrong.

I have calculated the values in their table 1, but in two cases find different values, namely 0.42 (0.421) instead of 0.40 for $f(\Delta_{\max})$ for $\Omega_0 = 1/10$, and 0.24 (0.237) instead of 0.23 for $f_1(\Delta = 1)$ for $\Omega_0 = 10$. I suspect that the former is a misprint while the latter could be as well, or possibly due to roundoff error in a less accurate numerical calculation.

3.3. *Discussion*

DZ65 presented an integral for the angular-size distance for cosmological models with $\lambda_0 = 0$ but arbitrary Ω_0 for $\eta = 0$ and compared the corresponding distances to those with $\eta = 1$. Although no analytic solution was presented, DZ65 extended to $\eta = 0$ the idea of calculating distances for various values of Ω_0 (though still setting $\lambda_0 = 0$). Around the same time, much more extensive numerical calculations were done by Refsdal *et al.* (1967), only for $\eta = 1$ but for several values of Ω_0 and λ_0.

4. DASHEVSKII & SLYSH (1966)

Dashevskii & Slysh (1966, hereafterDS66)[10] generalized the method of Z64 and DZ65 to the more realistic case that the beam is not completely empty, but only for the Einstein–de Sitter model.

4.1. *Summary*

The empty-beam case is criticized as being too unrealistic, as there will always be some intergalactic matter; this will mean that there will always be a maximum in the angular-size distance. DS66 derive, in their equation (2), the second-order differential equation which is the basis for all further work in this field

$$\ddot{z} - \frac{\dot{a}}{a}\dot{z} + 4\pi G\rho_g z = 0 \quad , \tag{1}$$

'which determines the linear distance $z(t)$ between rays', with $\rho_g = \alpha\rho$ (the subscript g refers to the *smooth* component, considered as a '***gas*** at zero pressure that fills all space uniformly' [my emphasis], the rest of the 'matter being concentrated in discrete galaxies'); a is the scale factor and G the gravitational constant. Compared to Z64 and DZ65, they allow α (in the notation of KHS, η) to take an arbitrary value $0 \leq \alpha \leq 1$; η is thus completely general. The cosmological model is implicit in the term $\dot{a}/a$, in principle allowing one to study any cosmological model in which $\dot{a}/a$ can be calculated, but DS66 then restrict themselves to the Einstein–de Sitter model for the subsequent discussion, presenting a completely analytic solution for the angular-size distance for this cosmological model, namely the first unnumbered equation in DS66, which is a generalization of equation (10) in Z64.

DS66 point out that, for arbitrary $0 < \eta \leq 1$, the angular-size distance has a maximum at finite z and the angular-size distance goes to 0 for $z = \infty$. Also, the

[9] Mattig gave formulae for $\Omega_0 > 0$ and for $\Omega_0 = 0$. Not only does one need two formulae, but the formula for $\Omega_0 > 0$ is numerically difficult for $\Omega_0 \approx 0$ (Peacock 1999). Both can be avoided *via* a more complicated formula which covers both cases (Terrell 1977).

[10] This discussion follows the English translation of the Russian original (Dashevskii & Slysh 1965).

smaller the fraction of homogeneously distributed matter, *i.e.* the smaller η, the higher the redshift of this maximum. Without proof, it is stated that this result also holds in the case of non-zero pressure.

4.2. *Remarks*

It is not clear why equation (3) is the last numbered equation; perhaps because the following equations are not referred to in the text (but, like the others, are of course part of the text). Also confusing is the expression $0 \leq \alpha \leq 1.1 \leq k \leq 5$, which should be $0 \leq \alpha \leq 1$, $1 \leq k \leq 5$. As in Z64, $\tilde{f}_1$[11], *i.e.* the angular-size distance for $\eta = 0$, is incorrectly shown as having a maximum at finite z (a mistake also made by DZ65, though barely perceptibly; in all cases, these are probably due to the figures having been drawn by hand). Also, there should be no inflection in the dashed curve.

4.3. *Discussion*

The generalization to an arbitrary value of η is obvious; less obvious is the relatively simple analytic solution for arbitrary η for the Einstein–de Sitter model.

5. OTHER PAPERS I

(Being discussed in an 'other papers' section does not imply that the paper lacks quality or influence; quite the opposite, in fact. Rather, these sections discuss papers which are not directly relevant to the main theme of this review, but nevertheless played some role in it.)

Kristian & Sachs (1966) discuss what I like to call 'theoretical observational cosmology' for very general (*i.e.* anisotropic, inhomogeneous) cosmological models, not necessarily based on general relativity (GR) (of which the FRW models—homogeneous and isotropic models based on GR—are special cases), mainly for inhomogeneities on the scale of 10^9 light-years or more (with small-scale inhomogeneities considered to be smoothed out, *i.e.* in some sense the reverse of the assumptions above). Many results, after 'straightforward, though somewhat tedious' calculations, are given in terms of series expansions. A key result is that the relation

$$d\,A = r^2 d\Omega$$

(in their notation), where '$d\,A$ is the intrinsic cross-sectional area of a distant object; r is a measured quantity, the "corrected luminosity distance," defined by equation (19); and $d\Omega$ is the meaured solid angle subtended by the distant object' is very general and holds in all cosmological models, whether or not they are based on GR. At the time, observations were not good enough that one could be sure that the Universe is actually very well described by an FRW model, hence the emphasis on generality and discussion of possible observations which could be used to determine the many more parameters than those needed to specify an FRW model.

Bertotti (1966) cites Z64 and DZ65 (erroneously making Dashevskii an author of Z64 as well), but considers not just the increase in the angular-size distance as compared with the standard FRW case, but also the decrease (corresponding to amplification) due to the gravitational-lens effect, both strong lensing and weak lensing, *i.e.* 'the small, but distance-dependent, brightening caused by near galaxies' which leads to a 'statistical spread in luminosity', shown to be proportional to $(D^{\mathrm{A}})^3$ for small distances. The main result is an expression for apparent luminosity as a function of redshift, noting that, in the inhomogeneous case, the first correction is quadratic in redshift and produces a dimming, but for higher values of z the brightening due to gravitational lensing becomes more important. That expression is for arbitrary Ω_0[12] and arbitrary η (called f), *i.e.* the case considered by DS66[13] but expressed as a series expansion. It is also noted that, to first order, the correction to the Euclidean relation to the expression for the number of sources brighter than a given apparent luminosity does not depend on η.

Gunn (1967a) also examined statistical fluctuations due to gravitational lensing, but in position, not apparent magnitude. This was done in more detail by Fukushige & Makino (1994), who pointed out that 'the distance between nearby photons grows exponentially because the two rays suffer coherent scatterings by the same scattering object'. Gunn (1967b) extended the discussion to fluctuations in apparent magnitude. Feynman, in a colloquium at Caltech, had discussed a scenario similar to that discussed by Z64, concentrating on the effects on angular diameters, apparently not realizing that apparent magnitude would also be affected. For the topic of this review, the most important result is the realization that, for large-enough redshifts, average luminosities and angular sizes will be the same as in the strictly homogeneous case, because not all lines of sight can be underdense, though there will be a scatter in their values compared to those in a strictly homogeneous universe. Babul & Lee (1991) discussed Gunn's formalism in more modern notation, adopting some simplification and deriving some new analytic results. Although only the Einstein–de Sitter model was considered (with—as extreme positions—a spectrum of mass fluctuations derived from CDM and a white-noise spectrum), their conclusions probably apply more generally, namely that the dispersion in amplification due to large-scale structure is negligible, while that on small scales depends strongly on the nature of the distribution.

Refsdal (1970) also discussed changes in the apparent luminosity and shape of distant light sources due to intervening inhomogeneities, but using a numerical ray-tracing approach rather than the more analytic methods of the works discussed above. (As would become clear later, this allows the effect of very concentrated masses, *e.g.* stars, to be taken into account, as well as general fluctuations due to galaxies and large-scale structure. In other words, it can handle strong lensing as well.) Ray-tracing simulations were done for a static flat universe (with all the mass in point massees, *e.g.* $\eta = 0$), but the results were generalized to an interesting collection

[11] It is unclear why D66 use $\tilde{f}$ while Z64 and DZ65 use f.

[12] As was the custom at the time, this was written in terms of q_0, *i.e.* $q_0 = \Omega_0/2$ under the assumption $\lambda_0 = 0$. The reason that q_0—in general, $q_0 = \Omega_0/2 - \lambda_0$ (or, as was common at the time, $q_0 = \sigma_0 - \lambda_0$, where $\sigma_0 = \Omega_0/2$)—was used is that q_0 is, after H_0, the next-higher term in series expansions of observational quantities as a function of redshift (*e.g.* Hoyle & Sandage 1956)

[13] Since Bertotti (1966) was submitted around the time that DS66 appeared, presumably the former was derived independently of the latter and *vice versa*.

of cosmological models: Einstein's static universe[14], two models with $\lambda_0 = 0$ ($\Omega_0 = 0.3$ and $\Omega_0 = 2$), and a model with $\Omega_0 = 0.4$ and $\lambda_0 = 1.7$ (a spatially closed model which will expand forever with an antipode at $z \approx 4$).

In retrospect, one conclusion was very prescient:

> An interesting aspect of the problem is the possibility of using the effect to obtain information on the mass distribution in the Universe. Even if the effect is not observable after some systematic efforts to detect it, one should be able to determine upper limits on the number of condensed and massive objects in the Universe.

Press & Gunn (1973) pointed out that (at least for $\lambda_0 = 0$) if Ω_0 is due mainly to compact objects, then the probability is high that a distant source will be multiply imaged, independently of the mass of the objects (which does, of course, set the scale of the image separation). (At the time, it was not clear that most of Ω_0 consists of non-baryonic matter, and, since arguments against a substantial density of intergalactic gas had been presented, it seemed natural to look for the missing matter in compact objects.) A more detailed analysis shows that the lack of dependence on the mass is exact, while the image separation has a weak dependence on Ω_0.[15] In contrast to the other papers in this section and that in the next section, the emphasis is on detecting the scattering masses, not the influence of those masses on observable properties of the sources. Nevertheless, the ZKDR distance was used, in particular the extreme empty-beam case, with the lensing effect of individual clumps explicitly taken into account.

6. KANTOWSKI (1969)

Kantowski (1969, hereafter K69) took a somewhat different approach, using Swiss-cheese models (Einstein & Straus 1945, 1946). These are arguably less realistic than the approximation used in the papers discussed above, since in these models clumps of matter are surrounded by voids with $\rho = 0$. However, since these models are exact solutions of the Einstein field equations, the validity of approximations used to calculate the angular-size distance is not an issue (though, of course, one can question the validity of this approximation to the distribution of matter).

6.1. *Summary*

'The Swiss-cheese models are constructed by taking a Friedmann model ($p = \Lambda = 0$), randomly removing comoving spheres from the dust, and placing Schwarzschild masses at the "center" of the holes.' K69 makes five realistic assumptions in order to facilitate calculations: the Schwarzschild radii of the clumps are very small compared to their opaque radii, the size of the Swiss-cheese hole is much larger than the opaque radius, the change in the scale factor of the universe is negligible during the time it takes light to cross a hole, there are enough holes so that the change in the scale factor is negligible during the time between interactions with two holes, and the mass density of the opaque clumps is independent of the clump (though not all clumps have the same mass).

K69 calculated the bolometric luminosity, which is inversely proportional to the square of the luminosity distance. Since the luminosity distance is larger than the angular-size distance by a factor of $(1 + z)^2$, it is easy to compare his results with those discussed above. At least under the assumptions mentioned above, the relation between scale factor and redshift, $1 + z = R_0/R$ (R is the scale factor of the universe and, as usual, the subscript 0 denotes the currecnt value), holds to a high degree of accuracy. (The previous discussions mentioned above essentially assume, though correctly, that this is the case.)

K69 used the optical scalar equations (Sachs 1961) as the starting point for his calculations, as did Gunn (1967a). These describe the expansion, shear, and twist of the cross section of a beam of light due to the gravitational effect of matter on the beam, and are a special case of the Raychaudhuri equation (Raychaudhuri 1955).

Historically, when observational cosmology was done with objects at low redshift, cosmology was 'a search for two numbers' (Sandage 1970), $H_0 := \dot{R}/R$ (giving the scale) and the deceleration parameter $q_0 = -\ddot{R}R/\dot{R}^2 \equiv -\ddot{R}/(RH^2) = \Omega_0/2 - \lambda_0$ (describing the first higher-order effects). K69 points out that in the case that most matter is in clumps (*i.e.* $\eta \approx 0$), a real value of $q_0 = 2.2$ would, were one to wrongly assume the standard distance, appear as $q_0 = 1.5$. This foreshadows later work, for example as discussed in Sect. 27, stressing the importance of taking inhomogeneities into account in classical observational cosmology, at least as long as a significant fraction of matter is in clumps and the Universe is similar to the approximations used to calculate distances in such a case.

6.2. *Remarks*

Both the approach of K69 and that discussed in the previous sections have clumps of matter embedded in a smooth distribution of matter. However, because the Swiss-cheese approach of K69 has the clumps surrounded by voids, the mass of the clumps being equal to the mass removed from the voids, the density of the smooth component is the same as the overall density ρ, whereas in the previous approach the density of the smooth component is $\eta\rho$, while that of the clumps is $(1 - \eta)\rho$. Thus, light propagating outside the holes will propagate exactly as in the completely homogeneous case; the defocussing occurs only because the beam crosses some holes.

6.3. *Discussion*

K69 has been very influential, having at the time of writing more than 180 citations, almost as many as the first three papers discussed above together, although these papers are easier to understand and good enough for most purposes; the fact that the latter have fewer citations is perhaps due to their having been published in (a translation of) a Russian journal.

[14] Because it is static, λ_0 and Ω_0 are infinite, since $H_0 = 0$ ($\lambda_0 = \Lambda/(3H_0^2)$ and $\Omega_0 = 8\pi G\rho/(3{H_0}^2)$). Both the density ρ and the cosmological constant Λ are positive, $\Lambda = 4\pi G\rho$.

[15] Since the cross section for strong lensing is proportional to the mass of the lens, the distribution of masses does not matter. The weak dependence on Ω_0 is due to the effect of Ω_0 on the angular-size distance.

7. DYER & ROEDER (1972)

Dyer & Roeder (1972, hereafter DR72) discussed the completely empty-beam case; despite starting out with an expression for arbitrary Ω_0 and λ_0 (using the standard notation at the time with $\sigma_0 = \Omega_0/2$ and $q_0 = \sigma_0 - \lambda_0$), results were presented for $\sigma_0 = q_0$, *i.e.* $\Lambda = 0$ (and hence $\lambda_0 = 0$).

7.1. *Summary*

For an integral expression for the angular-size distance, analytic solutions are presented for the three cases $\Omega_0 < 1$, $\Omega_0 = 1$, and $\Omega_0 > 1$; only the much simpler solutions for $\Omega_0 = 1$ (Z64) and $\Omega_0 = 0$ (Mattig 1958) (see also Z64) were previously known. As was also pointed out by DZ65, there is no maximum in the angular-size distance for $\eta = 0$. The famous result of Etherington (1933),

$$D^{\mathrm{L}} = (1+z)^2 D^{\mathrm{A}} \quad , \tag{2}$$

is invoked to note that an empty beam leads to a lower apparent luminosity which, as discussed by Kantowski (1969), leads one to underestimate q_0 if a completely homogeneous universe is assumed; their example has a real value of $q_0 = 1.82$ which, if calculated assuming a completely homogeneous universe, results in the value $q_0 = 1.40$. Kantowski (1969) had a real value of $q_0 = 2.2$ being interpreted as $q_0 = 1.5$. The exact numbers are not important; the point is that, to first order, the ZKDR distance is larger than in the standard case, which is also the case for a lower value of q_0. But this is only to first order; with higher-redshift data, the two effects are not degenerate. It is also shown that, while the difference between the ZKDR distance and the standard distance is non-negligible, there is little difference between the ZKDR distance and that obtained by numerical integration in a corresponding Swiss-cheese model (which, as mentioned above, is not an exactly equivalent model).

7.2. *Remarks*

Compared to the papers discussed above, especially the first three, there is much less emphasis on physical models and more on mathematical results. Also, comparisons are done between a relatively simple formula and a more involved numerical integration based on a more complicated mass distribution.

7.3. *Discussion*

Dyer & Roeder (1972) covered the same ground as DZ65, but more thoroughly, presenting an analytic solution.

The distance for an empty or partially filled beam has become known as the Dyer–Roeder distance, although various aspects had been discussed before. This is probably due to the fact that the corresponding papers were published in a major English-language journal, used standard notation, and were more concerned with results than with theory. Dyer and Roeder were certainly responsible for putting the topic on the agenda of many astronomers. However, for the reasons outlined above, I refer to this distance as the ZKDR distance.

8. DYER & ROEDER (1973)

Dyer & Roeder (1973, hereafter DR73) can be seen as a combination of DZ65 and DS66, *i.e.* Ω_0 and η are both arbitrary (though $\lambda_0 = 0$ is still assumed). For the general case, they derive a hypergeometric equation, and present explicit solutions for $\eta = 0$, $2/3$, and 1 as well as $\Omega_0 = 0$ (the second one being new).

8.1. *Summary*

As in DR73, general discussion is narrowed down by setting $\lambda_0 = 0$ before explicit solutions are presented. Second-order differential equations for both the angular-size distance and the luminosity distance are derived, though of course once one has a solution one can use the Etherington reciprocity relation to simply derive one from the other. Using a substition, these are converted to hypergeometric equations.

The special case $\eta = 1$ is the solution derived by Mattig (1958) while that for $\eta = 0$ is that derived by DR72. New is a solution for $\eta = 2/3$, which is given for the luminosity distance. For $\Omega_0 = 1$, one has the solution derived by DS66, which is given for the angular-size distance. Differentiation of that equation leads to an expression for the maximum in the angular-size distance, showing that as η goes from 1 to 0, the redshift of this maximum goes from 1.25 to ∞. The point first made by Z64, that the maximum is due to matter in the beam, is emphasized. (Note, however, that an arbitrarily small η will lead to a maximum, though at arbitrarily large z.) They suggest comparing observations with calculations for each of the three values of η for which there is an analytic solution, given the lack of knowledge about intergalactic matter. Finally, as in DR72, they note that calculations for Swiss-cheese models (interestingly, including $\lambda_0 \neq 0$) confirm that this is a good approximation, *i.e.* 'the mass deficiency in the beam is in general much more important than the gravitational-lens effect for reasonable deflectors', at least for 'redshifts in the range of interest'.

8.2. *Remarks*

There is a huge literature on hypergeometric functions, and many well known functions, including many used in phyics, are special cases, but in general it is not possible to reduce hypergeometric functions to (combinations of) standard functions which are easily and efficiently calculated, either analyically or numerically. As such, the fact that the distance equations are hypergeometric equations is interesting, but (except for the analyically soluble special cases) of little practical use.

8.3. *Discussion*

DR73 can be seen as a combination of DZ65 and DS66, *i.e.* Ω_0 and η are both arbitrary (though $\lambda_0 = 0$ was still assumed).

Starting with the Einstein–de Sitter model, Z64 had investigated $\eta = 0$, presenting an analytic solution (as well as one for $\Omega_0 = 0$, in which the value of η irrelevant since there is no matter). DZ65 had expanded this to arbitrary Ω_0, though no analytic solution was presented. DS66 had returned to the Einstein–de Sitter model, but allowed η to be arbitary. DR72 had covered the same ground as DZ65, but presented an analytic solution. Finally, DR73 addressed the most general case so far, with both Ω_0 and η as free parameters, and presented analyic results (most already known) for special cases.

9. DYER & ROEDER (1974)

Dyer & Roeder (1974, hereafter DR74) extended Swiss-cheese models to include cases where $\lambda_0 \neq 0$ and showed that the distances so computed correspond well to those based on previous work (DR72, DR73).

9.1. *Summary*

After a short review of previous work on the topic, the method of Kantowski (1969) is extended to $\lambda_0 \neq 0$. Essentially, $\lambda_0 \neq 0$ affects the expansion history of the universe but nothing else; in particular, $R_0/R = 1 + z$ still holds. A second-order differential equation for (a quantity simply related to) the angular-size distance is presented, but no solution is given. It is noted that a 'series solution about $z = 0$ can be obtained', but the emphasis is on calculating the correction factor relative to the homogeneous model of the same mean density; there is a series expansion for this, but it breaks down by the time the redshift has become high enough for the effect to be interesting, so results have to be calculated numerically. With regard to distortion of the beam, they show that a beam retains its elliptical cross section, though orientation and ellipticity can change.

In Swiss-cheese models, the structure of the clumps must be taken into account, but for realistic assumptions (assuming that the clumps model galaxies), 'the calculations indicate that... the distance–redshift relations do not differ signifancly from the "zero-shear" relations discussed in [DR72 and DR73]. Similarly, the distortion effect has been found to be negligible in the range of redshifts observable at present, being at most a few percent.' Although the Swiss-cheese models are perhaps unrealistic in that real galaxies are not usually surrounded by a region of lower than average density, they do show the potentially real effect that there is a dispersion in the distance calculated from redshift which increases with redshift. (In Sect. 19 it is discussed how important this is for our Universe.) Another important result is that the dependence of the distance–redshift relation on Ω_0 is increased for $\eta \approx 0$, thus reducing the precision obtainable in practice. Previous conclusions mentioned above that decreasing η means that observations interpreted assuming that $\eta = 1$ will underestimate q_0 are repeated.

9.2. *Remarks*

Calculations involving the Swiss-cheese models are inherently statistical in nature and more complicated than those based on approximations. The models are even arguably less realistic. However, they are important because, being exact solutions to the Einstein equations, one does not have to worry about approximations. The fact that results are very similar to those based on simpler assumptions is encouraging, and provides justification for using the simpler approach. It could of course be the case that this approach is too simple for the real Universe, but in that case a Swiss-cheese model would also probably be too unrealistic.

9.3. *Discussion*

DR74 is interesting because it presents for the first time distance–redshift relations in a universe with arbitrary Ω_0, λ_0, and η. However, not only because the calculations are based on Swiss-cheese models, no closed formulae are given.

10. OTHER PAPERS II

Roeder (1975a) applied the work of DR73 to the data of Sandage & Hardy (1973), concluding that the value obtained for q_0 depends both on assumptions about (in)homogeneity and on galaxy evolution and suggesting $q_0 > 0.5$ if the conclusion of Gott *et al.* (1974) is assumed, namely that $\eta \approx 0$.

Roeder (1975b) applied the conclusions of DR73 to a claim by Hewish *et al.* (1974) that there is a lack of small-diameter sources at the largest redshifts, whereby they assume the standard angular-size distance. If $\eta < 1$, then the angular-size distance is larger than otherwise, and if one wrongly assumes $\eta = 1$, then one will underestimate the true physical size of the source. Thus, an inhomogeneous Universe is not a possible explanation of that claim; rather, it would exacerbate the problem.

11. FURTHER SOLUTIONS (ANALYTIC AND NUMERICAL) OF THE ZKDR DISTANCE

11.1. *Kayser, Helbig & Schramm (1997)*

Increasingly general equations (EQ), analytic solutions (AS), and numerical calculations (NC) had been presented in the 1960s and 1970s (AS implies EQ) (all but the last two below discussed above):

Zel'dovich (1964b) (Z64): $\eta = 0$ for the Einstein–de Sitter model (AS)

Dashevskii & Zel'dovich (1965) (DZ65): $\eta = 0$ for arbitrary Ω_0, $\lambda_0 = 0$ (NC)

Dashevskii & Slysh (1966) (DS66): $0 \leq \eta \leq 1$ for the Einstein–de Sitter model (AS)

Kantowski (1969)(K69): Swiss-cheese models: $0 \leq \eta \leq 1$, arbitrary Ω_0, $\lambda_0 = 0$ (EQ, NC)

Dyer & Roeder (1972) (DR72): $\eta = 0$ for arbitrary Ω_0, $\lambda_0 = 0$ (AS)

Dyer & Roeder (1973) (DR73): $0 \leq \eta \leq 1$ for arbitrary Ω_0, $\lambda_0 = 0$ (hypergeometric EQ); AS for $\eta = (0, 2/3, 1)$

Dyer & Roeder (1974) (DR74): Swiss-cheese models: $0 \leq \eta \leq 1$, arbitrary Ω_0 and λ_0 (EQ, NC)

Dyer & Roeder (1976): $\eta > 1$ (heuristic, not exact, approach)

Kantowski *et al.* (1995): $0 \leq \eta \leq 1$ for arbitrary Ω_0, $\lambda_0 = 0$ (hypergeometric EQ); AS for arbitrary values of η

The only expression available for $\lambda_0 \neq 0$ was a complicated differential equation derived by Dyer & Roeder (1974), but for Swiss-cheese models. No closed solution was presented. Of course, it can be integrated numerically. However, it is rather cumbersome, and the terms do not have an obvious physical interpretation like those in the differential equations of Z64 and DS66. While it was appreciated that Swiss-cheese models are in some sense equivalent to the ZKDR distance derived *via* the Zel'dovich method, this was not shown strictly until much later (Fleury 2014). Thus, between the work of Dyer & Roeder (1976) and Kantowski *et al.*

(1995), work on the ZKDR distance concentrated mostly on understanding the approximation, applications (both in more-traditional cosmology and in gravitational lensing), and, to some extent, more-realistic models (this field would come into its own only later, when computer power allowed more-complicated scenarios to be investigated). However, the development of the basic ZKDR distance picked up again later. Kayser (1985) derived a differential equation for the angular-size distance in the style of Z64, DS66, and DR73, but for $0 \leq \eta \leq 1$ and arbitrary values of λ_0 and Ω_0, which he integrated numerically *via* standard but basic means. Kayser *et al.* (1997) saw a need for an efficient numerical implementation of that equation, which is the most general equation for the ZKDR distance under the standard assumptions that the universe is a (just slightly) perturbed FRW model (*i.e.* no pressure, no dark energy more complicated than the cosmological constant, no back reaction, only Ricci (de)focussing; even today, there is no evidence that the first three are not excellent approximations, and the fourth is as well in many cases). Also, no efficient general implementation existed for the standard ($\eta = 1$) distance.[16] Thus, a description of the differential equation derived by Kayser (1985) and the efficient numerical implementation—using the Bulirsch–Stoer method in FORTRAN (see Helbig 1996, for technical details)—evolved to include a general description of various types of cosmological distances and a compendium of analytic solutions, probably the first time all this information had been presented in a uniform notation. Despite being a numerical (though very efficient) implementation, it is only a factor of ≈ 3 slower than elliptic-integral solutions for $\eta = 0$ or $\eta = 2/3$ (Rollin Thomas, personal communication); for $\eta = 1$, the factor is ≈ 20 (Kantowski *et al.* 2000). Of course, a comparison can be done only for those cases where elliptic-integral solutions exist, but the numerical-integration time for the differential equation, valid for all values of λ_0, Ω_0, and η, is essentially the same whether or not an elliptic-integral or analytic solution exists. (Analytic solutions are of course faster than elliptic-integral solutions, which can be described as semi-numerical or semi-analytic; in general, the elliptic-integral solutions do not work if there is an analytic solution (an exception being the expression for light-travel time in a flat universe).)

11.2. *Kantowski and collaborators*

Kantowski, with collaborators, had returned to the topic of distance calculation in locally inhomogeneous cosmological models (Kantowski *et al.* 1995), coincidentally around the same time that I was writing the code for KHS. Although partially motivated by the m–z relation for Type Ia supernovae, further progress was made regarding the theory. Kantowski (1998) used the Swiss-cheese formalism to derive an analytic expression for the ZKDR distance using Heun functions, valid for arbitrary λ_0, Ω_0, and η. Kantowski *et al.* (2000) gave analytic expresssions using elliptic integrals for arbitrary λ_0, Ω_0, and $\eta = (0, 2/3, 1)$, corresponding to $\nu = (2, 1, 0)$; see Eq. (3) in Sect. 27. For the flat-universe case, there are simpler expression involving associated Legendre and hypergeometric functions; these were given by Kantowski & Thomas (2001). Kantowski (2003) pointed out that the general case can be expressed via the Lamé equation, which can be solved *via* Weierstrass elliptic integrals for $\nu = (2, 1, 0)$. While not directly related to the ZKDR distance, but related mathematically, note that Thomas & Kantowski (2000) also expressed the age–redshift relation (related to lookback time and light-travel–time distance) via incomplete Legendre elliptic integrals, but only for $\lambda_0 > 0$.[17] To summarize:

Kantowski (1998): arbitrary λ_0, Ω_0, and η using Heun functions

Kantowski *et al.* (2000): arbitrary λ_0, Ω_0, and $\eta = (0, 2/3, 1)$ (corresponding to $\nu = (2, 1, 0)$) using elliptic integrals

Kantowski & Thomas (2001): flat but otherwise arbitrary; associated Legendre and hypergeometric functions

Kantowski (2003): Lamé equation for the general case, Weierstrass elliptic integrals for $\eta = (0, 2/3, 1)$ (corresponding to $\nu = (2, 1, 0)$)

Thomas & Kantowski (2000): age–redshift *via* incomplete Legendre elliptic integrals

12. TESTING THE APPROXIMATION

Unlike the Swiss-cheese model, the ZKDR distance is an approximation based on various assumptions. While it is reasonably clear that it must be correct in the appropriate limit (*i.e.* the light propagates *very* far from all clumps, the fraction of mass in clumps is negligible so that it is clear that an FRW model is a good approximation, *etc.*), it is not immediately clear how good the approximation is in a more realistic scenario. One way to test this is to compare the ZKDR distance to an explicit numerical calculation, namely following photon trajectories through a mass distribution produced by a cosmological simulation. Some of this work will be mentioned below in Sect. 28. Watanabe & Tomita (1990), building on work by Futamase & Sasaki (1989), solved directly the equations of null geodesics and explicitly calculated the shear. Only the Einstein–de Sitter model was considered, and the explicit calculations were compared to the ZKDR distance for $\eta = 1$ and $\eta = 0$. The former is the better fit for the *average* distance, but it was assumed that mass is transparent, so this result essentially follows from flux conservation (Weinberg 1976). Kasai *et al.* (1990) carried out a similar study, noting that, as expected, the distance–redshift relation depends on angular scale, with the standard ($\eta = 1$) distance appropriate for large angles and the ZKDR distance (in the limiting case, $\eta = 0$) for small angles, a conclusion also arrived at by Linder (1998). His numerical result was demonstrated analytically by Watanabe & Tomita (1991). Similar results were found by Giblin *et al.* (2016b), who used a

[16] Despite having been known for decades that the standard distance can be calculated *via* elliptic integrals, this was almost never done, results being presented as a 'qualitative integration' (*e.g.* Bondi 1961) or calculated numerically (*e.g.* Refsdal *et al.* 1967).

[17] The history of the use of elliptic integrals to calculate cosmological quantities is interesting in itself, but is beyond the bounds of this review.

much more realistic model of the mass distribution, based on state-of-the-art simulations ('the first numerical cosmological study that is fully relativistic, non-linear and without symmetry') (Giblin *et al.* 2016a; Mertens *et al.* 2016). They stressed the scatter in the distance for a given redshift, which generally increases with redshift and is also dependent on the line of sight. Nakamura (1997) numerically investigated the effect of shear on the angular-size distance in a linearly perturbed FRW model and found it to be negligible, thus justifying the ZKDR distance. (For the Einstein–de Sitter model, an analytic result was presented.) Okamura & Futamase (2009), while not setting out to test the ZKDR distance, found that a universe with the halo-mass function of Sheth & Tormen (1999) is, remarkably, well approximated by the ZKDR distance with the η parameter calculated from their model. Busti *et al.* (2013) compared the ZKDR distance to other approximations: the weak-lensing approximation with uncompensated density along the line of sight, the flux-averaging approximation, and a modified ZKDR distance which allows for a different expansion rate along the line of sight. This work is interesting for its analysis of the underlying issues (essentially assumptions about the mass distribution and how this affects light propagation, different approximations corresponding to different assumptions) and its combination of detailed theory and application to real data—the Union2.1 sample, also used by Helbig (2015a) and Yang *et al.* (2013).

13. WEINBERG (1976)

Weinberg (1976) pointed out that in a locally[18] inhomogeneous universe in which gravitational deflection by individual clumps is taken into account, the conventional distance formulae remain valid *on average* as long as the clumps are sufficiently small, while for galactic-size clumps, this depends on the selection procedure and redshift of the source.

13.1. *Summary*

For a locally inhomogeneous universe, the *average* apparent luminosity (for the case $\lambda_0 = 0$, but this is true in general) is given by the conventional formula, *e.g.* that due to Mattig (1958), rather than the empty-beam formula, *e.g.* that investigated by Dyer & Roeder (1972). The reason is clear: the empty-beam formula 'leaves out the gravitational deflections caused by occasional close encounters with clumps near the line of sight'. Moreover, '[t]hese gravitational deflections produce a shear which on the average has the same effect in the optical scalar equation as would be produced in a homogeneous universe by the Ricci tensor term'.

The special case of $q_0 \ll 1$ is considered, in which the average number of clumps close enough to the line of sight to produce an appreciable deflection is of order q_0 for $z \approx 1$ (Press & Gunn 1973). Even in this case, where multiple deflections can be ignored, the standard formula is appropriate when considering the average distance. The decrease in the luminosity distance due to gravitational lensing cancels the increase due to the empty-beam formula.

This result is generalized to models with arbitrary q_0 and transparent intergalactic matter *via* a simple argument: due to flux conservation, the conventional distance must hold, on average; not all lines of sight can be underdense, and occasional lines of sight with strong amplification due to gravitational lensing exactly balance the larger number of underdense lines of sight. However, this ignores the selection effect that there can be no opaque clump between the source and the observer. If the clumps are dark stars, the conventional distance formula is a very good approximation, but only marginally so for galaxy-size clumps. The important quantity is the radius of avoidance, which could lead to the empty-beam distance being more appropriate at low redshifts and the conventional formula at high redshifts.[19] Details depend on selection effects: perhaps distant objects are observed (by accident or by design) on lines of sight which avoid clumps (and hence absorption); on the other hand, amplification bias might cause objects which have been gravitationally amplified to be observed preferentially.

The empty-beam distance is nevertheless useful since it gives a lower limit on the apparent luminosity (for a given absolute luminosity) at a given redshift. In general, there is a scatter in luminosity distance, comparable to the difference between the empty-beam and filled-beam formulae. Also, it is noted that the standard distance should be used to calculate the mean inverse-square luminosity distance, not the mean luminostiy distance itself. Weinberg speculates that this might be part of the reason for the difference in apparent luminosity between quasars at the same redshift.

13.2. *Remarks*

Weinberg (1976) is not concerned with developing the theory of the ZKDR distance; in fact, he doesn't go beyond DZ65. Rather, the emphasis is on understanding the validity of the approximation, its domain of applicability, and its use in a statistical context.

13.3. *Discussion*

This paper has been cited many times, perhaps because Weinberg is well known, but probably mainly because it is clear and to the point. Not until much later were more-detailed analyses presented.

14. OTHER PAPERS III

Wardle & Pottash (1977) discussed the effect of the ZKDR distance on the angular sizes of quasars, noting 'that the median angular size in fact decreased with redshift faster than expected in any Friedmann cosmology. This implied that there was a deficiency of sources of large *linear* size at high redshifts' [emphasis in the original].[20] A cosmological model with $\eta < 1$ could at least partially explain this.

Wagoner (1977) discussed determining q_0 from the m–z relation for supernovae, noting in passing that the Dyer–Roeder distance can be used. While not dwelling on the

[18] In the context of the ZKDR distance, 'locally' means 'on small scales', not some local inhomogeneity near us in an otherwise (more) homogeneous universe.

[19] This is a special case of $\eta = \eta(z)$, discussed by KHS.

[20] As mentioned above, Hewish *et al.* (1974) arrived at the opposite conclusion.

question of distance calculation, the paper is one of the first to advocate determining cosmological parameters from the m–z relation for supernovae rather for galaxies, deemed to be worth pursuing mainly because of the lack of knowledge about galaxy evolution.

Ellis (1980) noted that the uncertainty in η needs to be considered when attempting to derive cosmological parameters from observations. Ellis would later return to this topic many times.

15. THE END OF AN ERA

The work by Weinberg (1976) marks a turning point, for two related reasons. First, the theory is now more or less complete; future work would be concerned with refinements. Second, the development of theory is now secondary to applications, at least in terms of numbers of papers. The three papers mentioned in Sect. 14 are in some sense obvious consequences of the theory as known when they were written; most future work would be more limited in scope but also more detailed. As such, it makes sense to switch from the mainly chronological discussion presented until now to a discussion based on topic. (Nevertheless, some chronology is retained: topics are presented in the order of their appearance, and the discussion of each topic is roughly chronological. The order of the topics is based not on the average age of the papers, but rather on the time of publication of the first one.) Though some build on somewhat earlier work (some of which has been mentioned above), most of these topics were investigated after the work of Weinberg (1976).

Before doing so, however, the influential work of Canizares (1982) deserves special mention. Building on the work of Press & Gunn (1973), who had concentrated on the production of multiple images by compact objects, he discussed other observational effects. As such, this work belongs more in the gravitational-lensing camp than in the light-propagation camp. It also appeared at a time which saw a rapid increase in the number of papers devoted to these two topics. Obviously, the discovery of the first gravitational-lens system by Walsh *et al.* (1979) played a role as far as gravitational lensing itself was concerned; but probably because gravitational lensing forces one to think about the degree of homogeneity between source and observer, many studies were done which looked at further applications of the ZKDR distance, and, somewhat later, refinements to and extensions of the basic theory were investigated.

The next 16 sections, discussing various applications of the ZKDR distance, are chronological with respect to the first paper discussed in each. These are followed by a section discussing analytic approximations; the final section is a summary.

16. FLUX CONSERVATION 1

Weinberg (1976) pointed out that the standard distance formula, *e.g.* assuming $\eta = 1$, must hold *on average* if lenses are transparent and there are no selection effects. This is due to flux conservation. Dyer & Roeder (1981a) considered the effect of a finite source size in gravitational lensing, concluding that, all else being equal, η increases with the size of the source. (The fact that almost all beams are underdense and hence the average magnification is less than 1 is offset by the occasional strong-lensing event.) The important quantity is not the size of the source *per se*, but rather the size of the source relative to the clumps; as already mentioned by Weinberg (1976), one could think of η increasing with redshift since, due to structure formation, matter was more uniform at high redshift. The fact that the angular size of the beam also increases with redshift (the base of the cone is at the source; the apex at the observer) is an additional effect in the same direction. This was made more explicit by Dyer & Roeder (1981b), who showed that, '[i]n the weak-field approximation, the net amplification resulting from small amplifcations due to many small spherical deflectors bending light at their perimeters corresponds to the Ricci amplification where the source and observer are located well outside the lens'. Ehlers & Schneider (1986) question several assumptions regarding the derivation of the ZKDR distance. Subsequent work has shown these doubts to be misplaced; provided that the universe has a 'ZKDR-style' mass distribution, the ZKDR distance is appropriate. When calculating the probabilities of a source being lensed, however, they point out that a random line of sight is not an average line of sight. Rather, what is random is the position of a source on the celestial sphere. They conclude that lensing probabilities had thus been underestimated. This conclusion was arrived at considering flux conservation for an ensemble of lenses; Hamana (1998) showed that it holds for individual beams also (see Sect. 22). The general idea when considering averages is that most lines of sight are underdense and this is offset by the occasional strong-lensing event. In other words, the fact that the average amplification is 1 depends on the existence of an ensemble. On the other hand, a transparent lens neither creates nor absorbs photons. Avni & Shulami (1988) showed by an explicit calculation that this also holds for a single, isolated Schwarzschild gravitational lens; the usual amplification for small impact parameters is exactly compensated by de-amplification for large impact parameters.

Around the same time, Peacock (1986) noted that the solution given by Dyer & Roeder (1973) for arbitrary Ω_0 and η (but $\lambda_0 = 0$) is mathematically valid for $\eta < 25/24$, although $\eta > 1$ is unphysical, since this would imply that light propagates along a uniformly overdense tube. Nevertheless, this can be used as a rough model for gravitational lensing (see also Dyer & Roeder 1976). More importantly, Peacock (1986) generalized the result of Weinberg (1976) to arbitrary Ω_0. (As far as I know, no-one has repeated this calculation for arbitrary λ_0). He also agrees that the conclusion of Ehlers & Schneider (1986) that a more exact treatment reveals that lensing probabilities had been underestimated, but points out that their final result is not very useful since any difference between it and previous estimates becomes significant only at large optical depth, where the single-lens approximation breaks down. (Nevertheless, it still holds that previous estimates had underestimated the lensing probability.)

Fang & Wu (1989) pointed out that flux conservation can be used as a constraint when evaluating various approximations used in calculating the probability of lensing. Isaacson & Canizares (1989) compared the approach of Press & Gunn (1973) to that of Ehlers & Schneider (1986) in the Einstein–de Sitter model, finding that the former approach can be made to agree with the latter 'by adjusting the average magnification along a random line of sight so as to conserve flux'. Jaroszyński & Paczyński (1996) considered flux conservation within the context of microlensing (in which case $\eta = 0$ is appropriate, as long as any smooth mass distribution is ignored, since the lensing effect is taken into account explicitly in the microlensing calculation, as opposed to $\eta \approx 1$ which would be appropriate if one considered the average effect for a source size larger than that of the lenses). They pointed out that in addition to the redistribution of flux, there is another redistribution of energy because some observers see an additional redshift, some an additional blueshift.

17. KIBBLE $ LIEU (2005)

Kibble & Lieu (2005) also contributed significantly to the understanding of flux conservation in the context of the ZKDR distance; so much so that they deserve their own section. They showed analytically that, under very general conditions (including arbitrary shapes of clumps and strong lensing), the average *reciprocal* magnification in a clumpy universe is the same as that in a homogeneous universe, as long as the clumps are uncorrelated. The reciprocal magnification has the advantage that it goes to zero rather than infinity on the caustics (regions of—for a point source—infinite magnification), and so is more useful in the strong-lensing case. They also discussed various measures of magnification and the circumstances in which they are appropriate.

An important distinction is whether one averages over a set of sources on the unperturbed celestial sphere, or whether one averages over all lines of sight: 'If one part of the sky is more magnified,... the corresponding area of the constant-z surface will be smaller, so fewer sources are likely to be found there. In other words, choosing a source at random will give on average a smaller magnification or larger angular-size distance.' This is related to whether it is the mean magnification or the mean reciprocal magnification that is the same as in the homogeneous case. In the weak-lensing case, both are. In the strong-lensing case, it is the magnification which averages to 1 over the celestial sphere, the random-source average—the case implicitly considered by Weinberg (1976)—, however strong lensing effects are, while it is the *reciprocal* magnification which averages to 1 over all lines of sight, again however strong the lensing effects are. As a corollary, the random-source average of the *total* magnification of *unresolved* images is the same as in the homogeneous case, while for *resolved images* it can be significantly different, essentially because there can be more than one image of a given source.

Another distinction is between the angular-size distance and the so-called area distance (though both distances can be applied to both lengths and areas) as introduced by Ellis *et al.* (1998). If strong lensing is involved, *i.e.* multiple images (whether resolved or not) are present, then the magnification can be defined as negative for images of odd parity; sometimes, the angular-size distance itself is considered to be negative in such cases. (This is also the case for an object located at a coordinate distance χ between $n\pi$ and $2n\pi$, where n is an integer, because the rays defining the angle in the definition of the angular-size distance (see Sect. 2.2) cross between source and observer.) Such areas are counted negatively when calculating the average angular-size distance; if the absolute values are used, the corresponding distance is the area distance, which is thus always larger than the angular-size distance. The area distance is thus appropriate if one is interested in the total number of images within a given area of sky or their average magnification; the angular-size distance is appropriate if one is interested in the total number of distinct sources (say, when multiple images are not resolved) or their average magnification.

The work of Kibble & Lieu (2005) is also important because it is analytic (though some assumptions are made, which in practice are always fulfilled to a very good approximation: the surface of constant z is the same as the surface of constant affine parameter; shear vanishes when light is propagating far from all clumps; the clumps are widely separated, slowly moving, and randomly distributed). Their work confirms that of Weinberg (1976), which is based on energy conservation, when averaging over the celestial sphere (*i.e.* the source is random), and also considers the case of averaging over lines of sight.

18. FLUX CONSERVATION 2

Wang (2000) suggested that flux conservation justifies the use of the standard distance in the analysis of the m–z relation for Type Ia supernovae and performed such flux averaging by combining data in redshift bins, pointing out that this reduces systematic uncertainties from effects such as weak lensing, while Barber (2000) claimed that weak-lensing effects are about an order of magnitude larger than previously found (and hence probably need to be taken into account more explicitly). On the other hand, Wang (2005) found only marginal evidence for weak-lensing effects in the m–z relation for Type Ia supernovae.

Even if the *mean* magnification is 1, due to the skewness of the distribution, the *median* magnification is < 1. Clarkson *et al.* (2012) pointed out that most narrow-beam lines of sight are significantly underdense, even for beams as thick as 500 kpc. On the other hand, they also point out that this does not necessarily lead to a increase in apparent magnitude (*i.e.* dimming) if one drops the assumption that inhomogeneities can be modelled as perturbations on a uniformly expanding background, a point also emphasized by Bolejko & Ferreira (2012); see also Bagheri & Schwarz (2014).

Although the basic idea of flux conservation is clear (and there are obvious caveats such as non-transparent matter), exact treatments can be very complicated and have led to confusion, much of which has been cleared up by Kaiser & Peacock (2016): Weinberg (1976) is essentially right, though one needs to keep in mind the distinction between magnification and reciprocal magnification as discussed above in connection with Kibble & Lieu (2005). Since $\eta \sim \kappa$, where κ is the convergence, and $\mu \sim (1-\kappa)^{-1}$, the relation is linear only in the limit of vanishing deviations, though approximately linear for the small deviations considered here.[21] Non-linear functions of the conserved quantity μ must be handled with care. For example, the average angular-size distance $\langle D^{\mathrm{A}} \rangle$, and hence the average luminosity distance $\langle D^{\mathrm{L}} \rangle$, is biased

[21] See Wang *et al.* (2005) and Bolejko (2011) for details on the relationship between the inhomogeneity parameter η and the convergence κ.

even in the case of $\langle\mu\rangle = 1$. This can be considered one aspect of the averaging problem: we are interested in the average values of the cosmological parameters determined by observers throughout the universe, but can at best average observations over several lines of sight. See also Bonvin *et al.* (2015), who point out that the ensemble average and the directional average do not commute; 'observing the same thing in many directions over the sky is not the same thing as taking an ensemble average'; this is a restatement of the result of Kibble & Lieu (2005).

19. GALAXY CLUSTERS AND FURTHER WORK ON SWISS-CHEESE MODELS

Rubin *et al.* (1973) noted a non-random distribution of radial velocities on the sky for a sample of galaxies, later known as the Rubin–Ford effect, and discussed various possible explanations, though none involving gravitational lensing in any form. Karoji & Nottale (1976) confirmed the effect with two samples of galaxies chosen from the literature, discussed a number of possible causes, and tentatively concluded that 'light emitted by distant galaxies are [*sic*] redshifted when passing through clusters of galaxies or distant sources are more luminous when seen through intermidiate clusters of galaxies which could act as gravitational lenses'. Similar work was done by Nottale & Vigier (1977). Dyer & Roeder (1976) tried to explain the Karoji–Nottale effect via $\eta > 1$. On the one hand this is straightforward: $\eta < 1$ implies that there is less matter in the beam than for a random line of sight, so $\eta > 1$ would imply that there is more. On the other hand, this situation violates the assumptions under which the ZKDR distance is calculated, so the applicability is somewhat questionable. In any case, the conclusion was that accounting for the effect of gravitational lensing by clusters of galaxies in this manner cannot explain the Karoji–Nottale effect. Swiss-cheese models were also used to estimate the effect of inhomogeneities on the CMB (*e.g.* Dyer 1976; Nottale 1984), but this strays too far from the main topic of this article.

Nottale (1982a), in the spirit of Kantowski (1969), developed a more complicated but exact-solution model; the question is then how realistic it is physically, rather than whether the approximations are valid. While the Swiss-cheese model of Kantowski (1969) had holes consisting of completely empty voids with the mass removed from the void concentrated at the centre, and the corresponding Schwarzschild volume considered opaque, Nottale (1982a) had a more realistic model where the mass removed from the hole forms a Friedmann model of higher density than that surrounding the hole; importantly, the matter at the centre of the hole is transparent. Between the two Friedmann solutions is a Schwarzschild solution. The main conclusion here is that there is a change in the observed redshift of objects seen through such a cluster. Nottale (1982b) examined the perturbation of the magnitude–redshift relation in that model, deriving an expression for the change in magnitude dependent on the cosmological model (H_0,q_0), η, the cluster radius, the cluster redshift, and the source redshift; typical values of those parameters result in 'some tenths of magnitude'. Nottale (1983) studied this model with respect to 'the effects intrinsic to a cluster, *i.e.* the purely gravitational perturbations on redshift and magnitude (or equivalently diameter) for sources situated *in* a cluster, with respect to exterior sources' [emphasis in the original]. Nottale & Hammer (1984) investigated this in more detail, examining the amplification of light from distant sources by a transparent lens *via* an exact solution of the optical scalar equations (Sachs 1961). Nottale & Chauvineau (1986) used this formalism to calculate the global Ricci amplification by multiple gravitational lenses, noting that it usually differs significantly from the product of individual amplifications (an approximation valid only if all amplifications are small).

Sato (1985) continued working with the Swiss-cheese paradigm, finding that the modification is third order in Hr_b/c for redshift and first order for apparent luminosity, where r_b is the radius of a void (Swiss-cheese hole). Dyer & Oattes (1988) examined the dispersion of observational quantities such as magnitudes (related to the luminosity distance) in a Swiss-cheese model, emphasizing a fundamental limit 'due to the "fuzzy" structure of the perceived past null cone' and selection effects due to the skewness of the distribution of observational quantities (even though the means are the same as for FRW). Brouyzakis *et al.* (2008) arrived at similar results, noting even 'inhomogeneities with sizes of order 10 Mpc or larger' cannot lead to 'dispersion and bias of cosmological parameters derived from the supernova data' large enough 'to explain the perceived acceleration without dark energy, even when the length scale of the inhomogeneities is comparable to the horizon distance'. Clifton & Zuntz (2009) investigated the effect of large-scale structure on the Hubble diagram *via* a Swiss-cheese model. Kostov (2010) examined flux conservation in the sense of averaging over all lines of sight in Swiss-cheese models, with exact, non-perturbative calculations including all non-linear effects.

Fleury *et al.* (2013b) suggested that the well known 'tension' between *Planck* and the m–z relation for Type Ia supernovae (see *e.g.* Conley *et al.* 2011, for Type Ia supernovae data) could be relieved if the calculations are done with a Swiss-cheese model. This is because the CMB data have a typical angular scale of 5 arcmin while the typical angular size of a supernova is 10^{-7} arcsec. If the Swiss-cheese model is more appropriate, but a homogeneous model assumed, then one will underestimate Ω_0. Note that at the distances used to determine H_0, the effect of $\eta < 1$ is negligible (and would also go in the opposite direction: compared to $\eta = 1$, distances would be larger and hence the derived value of H_0 smaller[22]). Rather, Fleury *et al.* (2013b) pointed out that a lower η has, to first order, the same effect as a lower value of Ω_0 (or a higher value of λ_0).[23] Thus, incorrectly assuming $\eta = 1$ leads to an underestimate of Ω_0. If in fact $\eta < 1$, then the derived value of Ω_0 will be larger, while the value of H_0 changes only slightly. This reduces the tension between the values derived by *Planck* and the m–z relation for supernovae, though by changing the value of

[22] Odderskov *et al.* (2016), by examining the redshift–distance relation of mock sources in N-body simulations, concluded that local inhomogeneties cannot explain the tension. However, they were looking at the effect on H_0 itself.

[23] To first order in z, the luminosity distance dependes on z, to second order on q_0 ($\Omega_0/2 - \lambda_0$), and to third order on Ω_0 as well as q_0; thus, at relatively low redshift it is expected that Ω_0 is more important than λ_0 (*e.g.* Solheim 1966).

Ω_0 derived from the m–z relation. While the m–z relation still prefers higher values of H_0, there is no longer any serious discrepancy with the *Planck* results. This interesting result is a consequence of the very detailed Swiss-cheese calculations by Fleury *et al.* (2013a). Alas, as pointed out by Betoule *et al.* (2014), it appears that the low value of Ω_0 obtained by Conley *et al.* (2011) was due to a wrong calibration of the MegaCam zero points in the g and z bands and corrections to the MegaCam r and i filter bandpasses, thus the analysis by Fleury *et al.* (2013a) is in some sense no longer relevant (though one could turn it around and see the lack of tension in Ω_0 as evidence against such extreme Swiss-cheese models). Although interesting because they are exact solutions to the Einstein equation, Swiss-cheese models are today arguably mainly of historical interest. In particular, the redshift aspects should not be worrying, since they are merely one aspect of the integrated Sachs-Wolfe effect, which can be calculated for a CDM-like power spectrum, now known empirically to be a good approximation.

Fleury (2014) demonstrated with completely analytic arguments the equivalence of the ZKDR distance and that calculated from a certain class of Swiss-cheese models at a well controlled level of approximation. This had been known for a long time based on comparisons of numerical results, but of course an analytic proof is very important. Since the Swiss-cheese models are exact solutions of the Einstein equations, this means that there can be no problem using the ZKDR distance, as long as one makes the reasonable assumption that the mass at the centre of a Swiss-cheese hole is effectively opaque and reasonable assumptions about the order of magnitude of the mass and compactness of the clumps. (Of course, as discussed in Sect. 27, even if there can be no debate that the ZKDR distance is appropriate if a universe has the corresponding mass distribution, it is another question whether our Universe does indeed have such a mass distribution, even approximately.) He also stressed that the Etherington reciprocity relation (Eq. (2)) holds for *any* spacetime in which the number of photons is conserved, a point which is sometimes misunderstood. The present work is concerned with the theory and applications of the ZKDR distance, assuming that it is correct. Fleury (2014) has written the definitive paper on the justification of the ZKDR distance; it and references therein should be consulted for those interested in details.

Peel *et al.* (2014, 2015) examined the effcts of inhomogeneities on distance measures in a Swiss-cheese model, concentrating on the distance modulus. Their model is more general because the holes are non-symmetric structures described by the Szekeres (1975) metric (in general inhomogeneous and anisotropic). This allows an exact description which includes non-trivial evolution of structure. Interestingly, the standard deviation for dispersions $\Delta\mu$ was found to be $0.004 \leq \sigma_{\Delta\mu} \leq 0.008$, smaller than the intrinsic dispersion of magnitudes of Type Ia supernovae.

Lavinto & Räsänen (2015) examined the CMB as seen through random Swiss cheese. Usually, 'closed' holes had been examined, *i.e.* an overdense centre surrounded by an underdensity. Lavinto & Räsänen (2015) examind 'open' holes as well, *i.e.* an underdense void surrounded by a thin overdense shell. This is arguably a better model of our Universe, though of course still an approximation. The size of the holes corresponds to galaxy clusters. There is no statistically significant systematic shift in the angular-diameter distance, with a 95-per-cent upper limit of $|\Delta D^{\mathrm{A}}/\bar{D}^{\mathrm{A}}| < 10^{-4}$, and larger values reported in the literature are shown to be due to selection effects.

Observed inhomogeneities in the CMB are caused by a combination of primordial inhomogeneities and the effects of inhomogeneities on light propagation. Since the relevant angular scales are much larger than those involved in the ZKDR distance, further discussion of CMB anisotropies is beyond the scope of the present work. Lavinto & Räsänen (2015), apart from presenting original results, also gave a good review of this topic and its connection to the ZKDR distance.

20. GRAVITATIONAL LENSING: TIME DELAYS

The basic observational quantities in a strong (*e.g.* multiple-image) gravitational lens system—angles, flux ratios—are dimensionless, except for the time delays between pairs of images (Refsdal 1964). This allows one to determine the Hubble constant from a measurement of the time delay, assuming a mass model for the lens. However, this is true only in the low-redshift limit; at higher redshift, the cosmological model plays a role (Refsdal 1966). The cosmological parameters Ω_0 and λ_0 are now known very well from cosmological tests other than gravitational-lensing time delays (*e.g.* *Planck* Collaboration 2014, 2016, 2019); one could thus assume them to be exactly known and use observations related to cosmological distances to determine η (*e.g.* Helbig 2015a).[24] Within the uncertainties as they were 35–40 years ago, for the angular-size distance, at low redshift the values of Ω_0 and λ_0 are more important, while η becomes more important at high redshift (*e.g.* figure 1 in KHS). Due to the different combination of angular-size distances, for lensing statistics the effect of η tends to cancel (*e.g.* Quast & Helbig 1999) while in the case of gravitational-lensing time delays the importance of η is enhanced even at lower redshift (*e.g.* Kayser & Refsdal 1983; Helbig 1997).

Kayser & Refsdal (1983) illustrated this dramatically for several world models with $\lambda_0 = 0$, comparing the $\eta = 1$ and $\eta = 0$ cases. For the double quasar 0957+561 (Walsh *et al.* 1979), the cosmological correction factor (which gives the influence of the cosmological model compared to the limiting low-redshift case) was calculated for σ_0 values ranging from 0 to 2 (corresponding to $0 \leq \Omega_0 \leq 4$) with q_0 values of 1.0, 0.5, 0.0, and -1 ($\lambda_0 = \sigma_0 - q_0$). Helbig (1997) repeated the exercise for arbitrary combinations of λ_0, Ω_0, and η, again showing the importance of η, which has become even more important now that the values of λ_0 and Ω_0 are so well known.

A somewhat more complicated model (not neglecting shear) was investigated by Alcock & Anderson (1985), for $\lambda_0 = 0$ (not stated but assumed) and Ω_0 values of 0 and 1, using two gravitational-lens systems as concrete examples. They stressed the fact that ignorance of the

[24] The data from these other tests cannot usefully constrain Ω_0, λ_0, and η simultaneously (Busti *et al.* 2012; Helbig 2015a), not even if one restricts the analysis to a flat universe; the same is true of similar tests involving the angular-size–redshift relation (Santos & Lima 2008).

mass distribution along the line of sight makes it difficult to determine the Hubble constant by this method, but also that, once the Hubble constant is known *via* other means, this method could be used to learn something about the mass distribution. Similar results were obtained by Watanabe *et al.* (1993).

Usually one thinks of the possibility of determining H_0 or, if H_0 is known, other cosmological parameters from a measured time delay and mass model for the lens. Narayan (1991) pointed out that the measurement actually gives one the angular-size distance between observer and lens (which, if the redshift of the lens is known, is easily converted into the Hubble constant). Of course, this depends on η, but since lens redshifts are usually low, the effect of η is limited.

Giovi & Amendola (2001) examined a more general quintessence model where, in addition to ordinary matter ('dust') there is a perfect fluid with equation of state $p = (\frac{m}{3} - 1)\rho$ with $0 \leq m < 3$. The case $m = 0$ corresponds to the cosmological constant while $m = 3$ corresponds to ordinary matter; $m < 2$ implies that the universe is accelerating (as long as the quintessence term dominates). However, only $k = 0$ models are considered. One might think that this is justified since the Universe does seem to be very close to being flat (*e.g.* *Planck* Collaboration 2014, 2016, 2019); however, such an interpretation usually assumes that $m = 0$. Nevertheless, all known analytic solutions within this framework are presented (except one which 'is so complicated that it is not worth reporting'). Other cases are calculated numerically. Including quintessence usually reduces the estimated value of H_0 compared to the standard $m = 0$ case. Marginalizing over Ω_0 and m for the time delays considered results in $H_0 = 71 \pm 6$ and $H_0 = 64 \pm 4$ km/s/Mpc for the cases $\eta = 0$ and $\eta = 1$, respectively. Considering the facts that there is no evidence at all for values of m other than 0 (the cosmological constant) and 3 (dust), apart from radiation with $m = 4$ which, however, is important only in the early Universe, and that $\eta = 1$ is obviously not correct (at least in the strict sense), I find it somewhat disconcerting that there are a large number of papers investigating the possible effects of quintessence on the interpretation of cosmological observations compared to the number which discuss the influence of η.

While the idea is simple in principle (Refsdal 1964), in practice many details need to be taken into account when determining H_0 from gravitational-lens time delays (especially if the uncertainties should be small enough to be competitive with other methods), such as measuring the time delay itself and determining realistic uncertainties (*e.g.* Biggs & Browne 2018) and constructing a realistic mass model for the lens (*e.g.* Wong *et al.* 2016; Rusu *et al.* 2019). At this level of detail, characterizing the density along the line of sight by a single parameter η, or even $\eta(z)$, is too coarse. Rather, one attempts to measure the mass distribution explicitly, by counting galaxies (*e.g.* Rusu *et al.* 2017) or using weak gravitational lensing (*e.g.* Tihhonova *et al.* 2018).

21. GRAVITATIONAL LENSING: AMPLIFICATION

Schneider (1984) showed that a general transparent mass distribution always leads to amplification of at least one image compared to the case of an $\eta = 0$ universe (*i.e.* compared to the case that the lens were absent, not compared to the case that its mass is smoothly distributed throughout the universe). Of course, this is not in contradiction with the result of Weinberg (1976) that there is no mean amplification compared to a *homogeneous* universe, a point also emphasized by Nottale & Hammer (1984, see Sect. 19) and Hammer (1985).

Of course, all discussion of the ZKDR distance involves (negative) amplification, and in general all gravitational lensing involves amplification. Gravitational lensing has a huge literature which is beyond the scope of the present work. Therefore, I discuss here only those aspects of gravitational lensing which are directly related to the ZKDR distance, are interesting for other reasons, or in which I was personally involved. One example of the last is a study (Zackrisson *et al.* 2003) which demonstrated that various claims (Hawkins 1993, 1996, 1997; Hawkins & Taylor 1997) that most dark matter must be in compact objects of about a solar mass—because this is assumed to be responsible for most of the long-term optical variability of QSOs *via* microlensing—cannot be correct. In short, while arguments were presented that many of the observations are not only compatible with microlensing but also have no other obvious explanation, there are nevertheless other observations which contradict this hyposthesis, in particular the distribution of amplifications.

22. GRAVITATIONAL LENSING: GENERAL

Alcock & Anderson (1986) qualitatively discussed the optical scalars—implying a model more complicated than the ZKDR distance—and the possibility to learn something about distribution of mass in the universe from the distance measures derived from gravitational-lens systems. (Often the reverse is done: one has some model to calculate the distance as a function of redshift, and uses this as input for modelling the lens system.) Perhaps because in the case of gravitational lensing it is obvious that there are small-scale inhomogeneities which affect light rays (*i.e.* the gravitational lenses themselves), the ZKDR distance and similar topics were discussed earlier and more often than in other areas, even though their role there could be just as important.

Lee & Paczyński (1990) investigated gravitational lensing by three-dimensional mass distributions, finding that 16 screens are a sufficiently good approximation. Their conclusion that 'the distribution of amplifications of single images is dominated by the convergence due to matter within the beam' and that '[t]he shear caused by matter outside the beam has no significant effect'—even in the case of strong lensing—increases one's confidence that the zero-shear ZKDR distance is a realistic approximation (at least in a universe with the corresponding mass distribution). Although their goal was not to test the ZKDR approximation, their work could be seen as an early comparison of the ZKDR distance with numerical simulations. Jaroszyński *et al.* (1990) numerically studied gravitational lensing in the Einstein–de Sitter model, also concluding that shear can be neglected but also that the filled-beam approximation ($\eta = 1$) appears to be justified, at least for strong lensing by galaxies or clusters of galaxies. However, 'the column density was averaged over a comoving area of approximately $(1h^{-1}\mathrm{Mpc})^2$', so this could be a self-fulfilling prophecy, together with the fact that they found no case of strong lensing. Nevertheless, it does seem to be the fact that 'the large-scale structure of the universe as it is presently known does not produce multiple images with gravitational lensing on a scale larger than clusters of galaxies'. The same conclusion, namely that Weyl focussing can be neglected compare to Ricci focussing, was also found by Hamana (1999) to apply to a universe modelled as randomly distributed isothermal objects. It thus appears that the ZKDR distance, which is based on a very simple model, is also valid in more-realistic models, confirming a result of Nakamura (1997) based on solving the optical-scalar equation for light passing through linear inhomogeneities in CDM models.

Seitz *et al.* (1994) and Seitz & Schneider (1994) derived the gravitational-lens equations in an 'on average' Friedmann universe, in particular one with the mass distribution (smooth component with clumps) used in the derivation of the ZKDR distance. This very detailed work is an analytic complement to the numerical investigations mentioned above regarding the effects of inhomogeneities on the propagation of light beams; in particular, necessary approximations are made clear, lending support to the idea that the ZKDR distance is an acceptable approximation.

Gravitational-lensing statistics (*e.g.* Turner *et al.* 1984; Fukugita *et al.* 1990, 1992; Falco *et al.* 1998; Kochanek 1993, 1996a,b; Kochanek *et al.* 1995; Quast & Helbig 1999; Helbig *et al.* 1999; Chae *et al.* 2002) is usually not concerned with η. Apart from the general neglect of η in observational cosmology, there are probably several reasons for this. First, such studies are usually concerned with all-sky surveys, so one might expect η to 'average out' to 1 (Weinberg 1976). Second, in the relevant combination of angular-size distances, the effect of η tends to cancel out (in contrast to the situation regarding time delays). Third, while selection effects are important in such analyses, selection effects due to the value of η are smaller than others. Fourth, any effect of η would, in practice, be degenerate with other effects. Covone *et al.* (2005) found that the expected number of gravitationally lensed quasars is a decreasing function of η; Castañeda & Valencia (2008) investigated strong lensing (by galaxy clusters) with $\eta = \eta(z)$ as a means of taking structure formation into account.[25]

Asada (1998), by contrast, assumed the validity of the ZKDR distance and used it to investigate how inhomogeneities affect observations of gravitational lenses, in particular bending angle, lensing statistics, and time delay. An interesting analytic result is that all three combinations of distances[26] involved in these phenomena are monotonic with respect to the clumpiness for all combinations of λ_0, Ω_0, and source and lens redshifts. The clumpiness decreases the bending angle and number of strong-lensing events and increases the time delay. (Of course, not all combinations are monotonic in η, but physically relevant ones are.) In the first two cases, decreasing η has the same effect as decreasing λ_0. In other words, using a value of η which is too large (such as the common assumption $\eta = 1$) would lead one underestimate the value of λ_0.[27] (In the conclusions, this is confusingly stated as 'the use of the DR distance always leads to the *overestimate* of the cosmological constant' [emphasis in the original]; of course, it is not an overestimate but rather the correct estimate if the correct value of η for the ZKDR distance is used.) More detail was provided by Tomita *et al.* (1999).

At almost the same time (publication was one month later) and completely independently, Helbig (1998) investigated not the common gravitational-lensing topics mentioned above, but rather the correlation between image separation and source redshift, in a reply to the work of Park & Gott (1997) who had noted a negative correlation. Helbig (1998) showed that decreasing η has the same effect as decreasing $K := \lambda_0 + \Omega_0 - 1$ (*i.e.* this effect is also monotonic in η); also, decreasing η reduces the differences between cosmological models characterized by λ_0 and Ω_0. The strong negative correlation reported by Park & Gott (1997), though, seems to be based on an unclean data sample and also is not statistically significant.

It had been known for some time (*e.g.* Schneider *et al.* 1992; Ehlers & Schneider 1986) that gravitational-lensing magnification as calculated using the standard distance is smaller than that using the ZKDR distance by a factor of the square of the ratio of the corresponding distances, a result derived by averaging magnifications over a number of sources and making use of flux conservation. Hamana (1998) showed that it is actually true not just on average but for each individual ray bundle as well.

23. MONTE-CARLO SIMULATIONS

Refsdal (1970) had studied numerically the propagation of light in an inhomogeneous universe (see Sect. 5). This technique was expanded by Schneider & Weiss (1988a,b). Pei (1993a,b) showed that, to a reasonable approximation, the effect of multiple lenses can be calculated by multiplying the individual amplifications. Tomita (1998) used N-body simulations with the CDM power spectrum in four cosmological models to investigate the behaviour of angular-diameter distances in inhomogeneous cosmological models, determining η for each pair of rays and investigating the mean and dispersion of η. Further studies along these lines (*e.g.* Premadi *et al.* 1998, 2001; Martel *et al.* 2002; Premadi *et al.* 2004, 2008) involving ray shooting

[25] Note that one expects η to increase with z for two reasons when the angular-size distance is concerned. First, structure formation implies that the universe is more homogeneous at higher redshift. Second, for a fixed angle at the observer, the physical size of the object observed increases with redshift (as long as the redshift is lower than that of the maximum in the angular-size distance), so one averages over a larger volume at higher redshift. Both effects exist for the luminosity distance as well.

[26] The combinations are $D_{\rm ds}/D_{\rm s}$, $D_{\rm d}D_{\rm ds}/D_{\rm s}$, and $D_{\rm d}D_{\rm s}/D_{\rm ds}$, respectively. The subscripts refer to the deflector (lens) and source. In the case of only one subscript, it is the second, the first being understood to refer to the observer. This is probably the most common notation. Other schemes explicitly write the first subscript when it refers to the observer as well, use 'l' instead of 'd' to refer to the lens (deflector), use capital letters, or some combination of these. The same subscripts are used to refer to the corresponding redshifts, *e.g.* $z_{\rm s}$, though sometimes $z_{\rm d}$ is used in the sense of a variable and $z_{\rm l}$ to refer to the redshift of an explicit gravitational lens.

[27] Note that this is opposite the effect in the m–z relation.

through N-body simulations with the explicit calculation of the paths of (bundles of) light rays, while interesting, are too far removed from the main topic of the present article for further discussion.

24. CLASSICAL COSMOLOGY: REDSHIFT–VOLUME RELATION

Omote & Yoshida (1990) examined the effect of statistical gravitational amplification on the cosmological redshift–volume test, in particular its influence on the derived value of Ω_0, using the extreme $\eta = 0$ model to examine the data of Loh & Spillar (1986), concluding that their derived value of Ω_0 is smaller, *i.e.* η and Ω_0 are positively correlated.[28] Of course, there are much better data today, Loh & Spillar (1986) neglected galaxy evolution, and so on; nevertheless, this work demonstrates the effect of η on the redshift–volume test.

25. CLASSICAL COSMOLOGY: MAGNITUDES

Wu (1998) suggested that interest in the ZKDR distance had subsided after Weinberg (1976) had shown that flux conservation implies that, on average, there is no amplification.[29] He then points out that the fact that the luminosity distances in the homogeneous and inhomogeneous cases are the same on average does not mean that apparent magnitudes are the same in both cases. This is illustrated with a simple model. More important than the model are the conclusions: because most lines of sight are underdense, compensated by the occasional large amplification, the apparent magnitude is essentially a random variable; also, the value of q_0 obtained depends on the value of η assumed, or, *vice versa*, one could use the m–z relation to determine η if the cosmological parameters are known with some degree of certainty.

Rose (2001) pointed out that the argument of Weinberg (1976) does not hold if the sphere centred on the observer is affected by the mass distribution, concluding that, in a perturbed FRW universe, 'more photons from a source at a given redshift' will be received than in an FRW universe, *i.e.* the sources are brighter. Somewhat confusingly, it is claimed that they 'therefore have a higher apparent magnitude', which is correct if 'higher' means 'brighter', but of course larger magnitudes correspond to fainter objects. However, this is a second-order effect; to first order, small deviations from homogeneity do not change the average magnification (Claudel 2000).

26. CLASSICAL COSMOLOGY: MAGNITUDE–NUMBER RELATION

Although going somewhat beyond the simple approximation of the ZKDR distance, Watanabe (1992, 1993) investigated the effects of an inhomogeneous universe on another classic cosmological test, namely the magnitude–number relation (see *e.g.* Sandage 1995, for details), also checking the validity of the assumptions used by Omote & Yoshida (1990) (see Sect. 24). These sorts of cosmological tests have gone out of fashion, primarily because the uncertainty in the evolution of the sources is too large, leaving the m–z relation for Type Ia supernovae, baryon acoustic oscillations (BAO), and the CMB as the most useful cosmological tests. It is not yet possible to calculate galaxy evolution from first principles, and observations of it have to be interpreted within the context of an assumed cosmological model, so now such classic tests are useful mainly as consistency checks.

27. CLASSICAL COSMOLOGY: MAGNITUDE–REDSHIFT RELATION

One of the most important advances in observational cosmology has been the application of the m–z relation to Type Ia supernovae.[30] In an influential paper, Colgate (1979) had suggested using the Hubble Space Telescope for that purpose. Goobar & Perlmutter (1995) discussed the feasability of such a programme, and were later involved in the Supernova Cosmology Project, which reported measurements of λ_0 and Ω_0 based on 42 supernovae (Perlmutter *et al.* 1999; Knop *et al.* 2003), a result confirmed and published slightly earlier by the High-z Supernova Search team (Riess *et al.* 1998; Schmidt *et al.* 1998). While there had been hints, based on joint constraints from several cosmological tests, not only that the cosmological constant is positive but also that it has such a value that the Universe is currently accelerating (Ostriker & Steinhardt 1995; Krauss & Turner 1995), the m–z relation for Type Ia supernovae was the first cosmological test which, by itself, confirmed such a value for λ_0. (Contrary to some claims, this test does not 'directly' measure acceleration in any meaningful sense, even if one does not adopt the extreme view that all that is ever 'really' measured in observational astronomy, whether in imaging or in spectroscopy, are photon counts as a function of position on a detector.) Perlmutter *et al.* (1999) also checked for the influence of η, using the FORTRAN code of KHS to compare the standard distance to that of two other models, one with $\eta = 0$ and the other with $\eta = \eta(\Omega_0)$, the latter based on the idea that all matter is in clumps for $\Omega_0 \leq 0.25$ and for $\Omega_0 \geq 0.25$ the fraction $0.25/\Omega_0$ is in clumps, thus $\eta = 0$ for $\Omega_0 \leq 0.25$, otherwise $\eta = 1 - 0.25/\Omega_0$. Their conclusion, based of course on their data at the time, is that significant differences occur only for models ruled out by other arguments, *i.e.* $\Omega_0 > 1$.

Kantowski *et al.* (1995), still using the soon-to-be-obsolete q_0-notation, had pointed out that η should be taken into account when discussing the m–z relation for Type Ia supernovae. They also presented an analytic solution for $\lambda_0 = 0$ but arbitrary Ω_0 and q_0, and introduced the parameter ν:

$$\eta = 1 - \frac{\nu(\nu + 1)}{6} \quad , \tag{3}$$

[28] Note that in the simpler case of the m–z relation, η and Ω_0 are negatively correlated. This is easy to understand, since both a higher value of η and a higher value of Ω_0 mean that more matter is in the beam. In the redshift–volume test, both the apparent magnitude and the volume (which is independent of η) are involved, the luminosity function plays a role, *etc.*, making the test much more complicated; (*e.g.* Sandage 1995). Also, all mass was assumed to be in point masses with regard to the gravitational-lens effect. Yoshida & Omote (1992) performed a similar study using the model of a spherical opaque lens, arriving at similar conclusions.

[29] This is not my impression. There was a slow trickle of papers up until about 1982, after which the number per year increased each year. This appears to be mainly data-driven, with a large increase after the measurement of the m–z relation for Type Ia supernovae.

[30] The m–z relation for Type Ia supernovae has spawned an extensive literature; in this review, I mention only those aspects of it directly concerned with the ZKDR distance. Many good reviews are available (Riess 2000; Leibundgut 2001; Schmidt 2002; Perlmutter & Schmidt 2003; Filippenko 2005; Leibundgut 2008).

due to the fact that there are analytic solutions for certain integer values of ν. Frieman (1996) disputed the importance of the effect, arguing that the Swiss-cheese model is not a valid model for the distribution of mass in the Universe, and that the uncertainty due to η would be smaller; Kantowski *et al.* (1995) disagree. Frieman (1996) emphasized the dispersion in the apparent magnitude of supernovae caused by a given mass distribution, rather than considering a range of η. A similar approach, with the aim of determining the density of compact objects, the properties of galaxy haloes, or estimating the uncertainty in the measurement of λ_0 and Ω_0, was taken up by many authors (*e.g.* Holz 1998; Seljak & Holz 1999; Metcalf & Silk 1999; Valageas 2000; Mörtsell *et al.* 2001; Minty *et al.* 2002; Amanullah *et al.* 2003; Payne & Birkinshaw 2004; Metcalf & Silk 2007; Dodelson & Vallinotto 2006; Martel & Premadi 2008; Yoo *et al.* 2008; Jönsson *et al.* 2010; Ben-Dayan & Takahishi 2016; Zumalacárregui & Seljak 2018). Rather than calculating the dispersion, one could also attempt to measure it indirectly due to the fact that the same matter fluctuations would cause weak lensing. However, the shear maps smoothed on arcminute scales are not of much use since an appreciable fraction of the lensing dispersion derives from sub-arcminute scales (Dalal *et al.* 2003). Another approach is to estimate the amplification from the matter visible along the line of sight; Jönsson *et al.* (2006, 2007, 2008) and Smith *et al.* (2014), building on ideas by Gunnarsson *et al.* (2006), found a tentative detection, *i.e.* a correlation between the computed and observed amplification (difference between the observed flux and that expected from the redshift in the concordance model). One can also turn this around, and use the observed matter distribution to estimate the amplification due to lensing and thus correct the observed flux (Jönsson *et al.* 2009).

Iwata & Yoo (2015) took a somewhat different approach, assuming a flat universe and taking Ω_0 from CMB measurements, then calculating $\eta(z)$ such that the cosmological parameters from the m–z relation for Type Ia supernovae agree; this was done for four different scenarios. This is complementary to the work of Helbig (2015a) (next paragraph) who, at almost exactly the same time, considered only constant η but for arbitrary FRW models, determining the value of η such that the m–z relation for Type Ia supernovae results in the same values for λ_0 and Ω_0 as those derived from the CMB.

Helbig (2015a) investigated the influence of η, noting that more and higher-redshift data had become available. While the data were not good enough to determine λ_0, Ω_0, and η simultaneously[31], the constraints in the λ_0–Ω_0 plane depend strongly on η. Only by assuming $\eta \approx 1$ does one recover the concordance-cosmology values of $\lambda_0 \approx 0.7$ and $\Omega_0 \approx 0.3$. Since these values are now known to high precision independently of the m–z relation for Type Ia supernovae (*e.g. Planck* Collaboration 2014, 2016, 2019), one can use the m–z relation for Type Ia supernovae to measure η. The result $\eta \approx 1$ agrees well with other tests to determine η from observations. (While no useful constraints are possible, the global maximum likelihood in the λ_0–Ω_0–η cube also indicates a high value of η.) Unknown to me at the time, very similar results, based on the same data, were obtained by Yang *et al.* (2013), Bréton & Montiel (2013), and, somewhat later, Li *et al.* (2015) (the latter two restricted to a flat universe). While perhaps not surprising, it is of course important in science for results to be confirmed by others working independently. Although they investigated a wider range of models, when restricted to standard FRW models, the results of Dhawan *et al.* (2018) are also consistent.

Since the observations indicate that $\eta \approx 1$, one can ask whether this is true 'on average' as discussed by Weinberg (1976), or whether each line of sight indicates $\eta \approx 1$. In the former case, one would expect a dispersion in the distance at high redshift. Indeed, the scatter does increase with redshift, but so do the observational uncertainties. Since their quotient is independent of redshift, this indicates that each line of sight indicates $\eta \approx 1$, in other words that all lines of sight fairly sample the mass distribution of the Universe[32] (Helbig 2015b). Note that Holz & Linder (2005) find a scatter (calculated theoretically) approximated by a Gaussian with standard deviation $\sigma_{\mathrm{eff}} = 0.088z$ (in flux) or $\sigma_{\mathrm{eff,m}} = 0.093z$ (in magnitudes). However, as discussed by Helbig (2015b), the observed increase in scatter with redshift seems primarily due to observational uncertainties in addition to the theoretically calculated scatter sometimes incorporated into those uncertainties.

28. MORE-DETAILED MODELS

Holz & Wald (1998) developed a generalization of the Swiss-cheese approximation by including all mass explicitly (thus there is no smoothed-out 'cheese' component), requiring the mass within a given spherical region (corresponding to a hole in the Swiss-cheese approach) to be equal to that of the background FRW model only on average, and dropping the requirement of spherical symmetry. In addition, rather than having a fixed mass distribution and calculating the trajectories of photons within it, the mass distribution along a given trajectory is calculated on the fly. Also, no opaque-radius cutoff is imposed. Such a model is clearly more realistic than that of Zel'dovich or a Swiss-cheese model, and leads to a distribution of apparent luminosities at a given redshift. In principle, the shape of such a distribution can be used to determine both the background FRW model and the fraction of matter in compact objects. While there are a few highly amplified sources (which, due to flux conservation, there must be, in order to compensate for the fact that most sources are de-amplified), most of the distribution can be thought of as η varying with position on the sky. As expected, if thought of in terms of η, η increases with redshift, as the higher the redshift, the more likely it is that a typical trajectory crosses a fair sample of the universe.

Bergström *et al.* (2000) generalized the method of Holz & Wald (1998) by allowing for different types of flu-

[31] This would imply the somewhat dubious assumption that η is independent of both redshift and the line of sight. Of course, more-realistic models could take such effects into account, but obviously the data would not be able to constrain them since even the simpler model with a constant η could not be constrained.

[32] As discussed in Sects. 16–19, this does not imply that the Universe is effectively homogeneous, but rather that the distance calculated from redshift is approximately the same as that which would be calculated in an effectively homogeneous universe.

ids, possibly with non-vanishing pressure, instead of just dust, and by considering the NFW profile (Navarro *et al.* 1997) in addition to point masses and singular isothermal spheres as lenses (see also Goliath & Mörtsell 2000). Also, multiple imaging is taken into account. This is thus an even more complicated and thus more realistic model of the universe. As a consistency check, their results for empty cells and cells with a homogeneous dust component were compared with results obtained from the code of KHS for $\eta = 0$ and $\eta = 1$, respectively. For a variety of cosmological models, the discrepancy was less than 1 per cent up to $z = 10$. This is a further justification that the ZKDR distance is an excellent approximation provided that the mass in the universe is distributed according to the assumptions underlying the ZKDR distance. They also found analytic approximations which are very good representations of various observable quantities, such as magnification distributions.

Mörtsell (2002) used essentially the same scheme to investigate the relation between η and the fraction of compact objects. By definition, $1 - \eta$ is the fraction of compact objects $f_{\rm c}$ in the pure ZKDR case, *i.e.* only de-amplification due to underdensity and no amplification due to gravitational lensing. As expected, taking lensing into account results in $1 - \eta < f_{\rm c}$. Interestingly, for a variety of cosmological models ($(\Omega_0, \lambda_0) = (0.3, 0.6), (0.2, 0.0), (1.0, 0.0)$), for redshifts between 0 and 3, and for various models of the mass distribution (homogeneous and point masses, NFW profiles and point masses), the relation is approximated very well by $1 - \eta \approx 0.6 f_{\rm c}$.

Some authors have claimed that that a universe with *large-scale* inhomogeneities could appear as if it has a positive cosmological constant when in fact it doesn't, either because the m–z relation mimics that of an accelerating model (*e.g.* Alnes *et al.* 2006; Garfinkel 2006) and/or because the inhomogeneities produce accelerations without a cosmological constant (*e.g.* Kai *et al.* 2007). However, Vanderveld *et al.* (2006) present evidence against these claims. Also, while in principle one can reproduce an arbitrary m–z relation with an *ad hoc* mass distribution, there are two arguments against this, other than the fact that it is *ad hoc*—or, equivalently, of all possible m–z relation which could be produced, it just so happens that one is produced which is not only explicable with 1920s cosmology, but also where the derived parameters agree with those determined by other means—: there is no believable route to explaining the CMB observations, and we are required to be at or near the centre of a large and approximately spherical region. Those topics go beyond the scope of this article, so I don't discuss them further here. However, it has even been claimed that this is possible in a Swiss-cheese universe (*e.g.* Marra *et al.* 2007, 2008). Vanderveld *et al.* (2008) showed, however, that this is not the case if the voids have a random distribution.

Flanagan *et al.* (2012) used a variant of the method of Holz & Wald (1998) to calculate the distribution of magnitude shifts, but using a simplified Swiss-cheese model for the mass distribution. Flanagan *et al.* (2013) extended this with a more refined Swiss-cheese model: the mass removed to make the voids is distributed on shells surrounding the holes in the form of randomly located NFW haloes and in the interior of the holes (either smoothly distributed or as randomly located haloes).

Hada & Futamase (2014) carried out a similar exercise, concentrating on the difference between the magnitude–redshift relation in a homogeneous universe and that in an inhomogeneous universe (with a mass distribution given by the non-linear matter power spectrum), as well as its dispersion, taking into account the blocking effect by collapsed objects and examining the resulting uncertainty in Ω_0 (≈ 0.4) and the equation of state w (≈ 0.04), all in a flat universe.

The work by Giblin *et al.* (2016a,b) and Mertens *et al.* (2016) has been mentioned above in Sect. 12; a similar approach was adopted by Bentivegna & Bruni (2016). Detailed discussion of such work is of course beyond the scope of this review, which concentrates on the use of the ZKDR distance as opposed to the standard distance when calculating distance from redshift for a given cosmological model. Nevertheless, for present purposes such works are interesting because they allow for comparison between the ZKDR distance and much more realistic simulated matter distributions, making it possible to see how well the ZKDR *ansatz* approximates reality. However, such simulations are still not entirely free of approximations: those above are fully relativistic but use the fluid approximation, while a different approach was adopted by Adamek *et al.* (2016), which does not rely on the fluid approximation, but on the other hand is based on a weak-field expansion of GR. Which approach is better of course depends on what one wants to study. It is perhaps surprising that a simple equation such as Eq. (1) agrees so well with results from numerical ray tracing through ΛCDM simulations, at least if one allows the additional freedom of $\eta(z)$ and a certain stochastic element depending on the individual line of sight ($\eta(\alpha, \delta)$). Somewhat similarly, the FRW metric was originally a simplifying assumption, made in order that at least some results could be obtained with the limited methods of calculation available at the time. Now, however, it is an observational fact, as demonstrated by observations of the CMB and the large-scale structure of the Universe, that our Universe is in fact very close to an FRW model (Green & Wald 2014).

29. WEAK GRAVITATIONAL LENSING

Weak gravitational lensing is normally defined as gravitational lensing without multiple images. If the source can be resolved, then information can be gleaned from the distortion of the image. In such a case, however, if the source is at a cosmological distance, $\eta \approx 1$ (because the distance implies a large physical extent near the source, averaging over the matter distribution, and because it appears that, at large redshift, distances behave as if $\eta \approx 1$, as noted in Sect. 27). Relevant for the ZKDR distance with respect to weak lensing is thus weak lensing of point sources.[33] Some aspects of this are

[33] An example of strong lensing of resolved sources are multiple images of background galaxies lensed by clusters of galaxies. An example of strong lensing of point sources are multiple images of QSOs; here, η can play a role since it influences the distance calculated from redshift, which in addition to the lens model is important for the time delay (see Sect. 20). Microlensing can be thought of as a combination of weak and strong lensing, depending on the impact parameter, though since the source is not resolved, one observes only a change in apparent magnitude due to amplification.

discussed above in Sect. 21. This section is concerned particularly with weak lensing of standard candles.

Wang (1999) pointed out that weak lensing leads to a non-Gaussian magnification distribution of standard candles at a given redshift, due to the fact that η can vary with direction. One can thus think of our Universe as a mosaic of cones centred on the observer, each with a different value of η, where there is a unique mapping between η and the magnification of a source. Of course, since the ZKDR distance depends on Ω_0 and λ_0 as well as η, different cosmological models can lead to very different magnification distributions for the same matter distribution.[34] Wang (1999) derived an approximation for the ZKDR distance (see Sect. 32), and also treated η as a function of position on the sky, *i.e.* different lines of sight can have different values of η. This effective value of η depends not only on the amount of matter in the beam, but also on how it is distributed (though only the total amount in the beam is considered—the possibility that a significant fraction could be in point masses is not taken into account). An approximation to matter distribution at a given redshift is found *via* comparison with the results of Wambsganss *et al.* (1997), who used $\Omega_0 = 0.4$ and $\lambda_0 = 0.6$. She then calculated the distribution of η as well as the magnification distribution for standard candles, both for the same three different redshifts 0.5, 2, and 5. Also, for the same matter distribution, the probability of magnification was calculated for the same three redshifts and three different cosmological models: $(\Omega_0,\lambda_0) = (1,0)$, $(0.2,0)$, and $(0.2,0.8)$.

Wang *et al.* (2002) extended this idea to a universal probability-distribution function for the reduced convergence which can be directly computed from Ω_0 and λ_0, well approximated by a three-parameter stretched Gaussian distribution, where the three parameters depend only on the variance of the reduced convergence; in other words, all possible weak-lensing probability distributions can be well approximated by a one-parameter family, which was normalized *via* the simulations of Wambsganss *et al.* (1997). The reduced convergence is the same as the direction-dependent η used by Wang (1999). Fitting formulae were presented for thre fiducial cosmological models: $(\Omega_0, \lambda_0, h, \sigma_8) = (1.0, 0.0, 0.5, 0.6), (0.3, 0.7, 0.7, 0.9), (0.3, 0.0, 0.7, 0.85)$.

Williams & Song (2004) took the opposite approach: assuming that the standard distance ($\eta = 1$) is correct, they found that bright SNe are preferentially found behind regions (5–15 arcmin in radius) that are overdense in the foreground due to $z \approx 0.1$ galaxies, the difference between brightest and faintest being about 0.3–0.4 mag. (In other words, the fact that bright supernovae are preferentially found behind overdense regions indicates that the standard distance is incorrect.) The effect, significant at > 99 per cent, depends on the amount and distribution of matter along the line of sight to the sources but not on the details of the galaxy-biasing scheme.

In a very detailed work, Kainulainen & Marra (2009) studied the effects of weak gravitational lensing caused by a stochastic distribution of dark-matter haloes, restricted to flat FRW models and examining those with $\Omega_0 = 0.28$ (close to the current concordance model) and $\Omega_0 = 1$ (the Einstein–de Sitter model) as representative examples. In particular, they calculated the difference between the distance in their model and the ZKDR distance for $\eta = 0.5$ and $\eta = 0$ for these two models, finding a maximum relative error of only 0.06 for the extreme case of the empty-beam Einstein–de Sitter model at $z = 1.6$ (the upper limit of their redshift range). This is yet another example of the proof of the validity of the assumptions underlying the ZKDR distance. Their main goal was to compute the probability-distribution function and the most likely value of the lens convergence along arbitrary photon geodesics as a function of their model parameters.

30. CLASSICAL COSMOLOGY: GENERAL

In an interesting but somewhat confusingly written paper, Yu *et al.* (2011) use the m–z relation and the angular-size–redshift relation (based on data from the literature) to determine Ω_0 and η in flat cosmological models (and the equation-of-state parameter w—confusingly referred to as ω—and η for flat models with $\Omega_0 = 0.28$). Of course, H is in general a function of z, but this is not something which is measured directly.[35] Rather, $H(z)$ is calculated from the magnitude or angular size. Although not stated, presumably the reason is to be able to fit to both data sets simultaneously. Their results (1-σ uncertainties) $\eta = 0.80^{+0.19}_{-0.2}$ (with no prior on Ω_0) and $\eta = 0.93^{+0.07}_{-0.19}$ ($\Omega_0 = 0.26 \pm 0.1$) can be compared to $\eta = 0.75^{+0.15}_{-0.15}$ ($\lambda_0 = 0.72$ and $\Omega_0 = 0.28$, *i.e.* the confordance-model values) obtained by Helbig (2015a) using only the m–z relation for Type Ia supernovae (see Sect. 27). Although not directly comparable, and keeping in mind that one would expect the supernova data to indicate a lower value of η due to the smaller beam size, the general trend is clear: observational data indicate a relatively high value of η. There are two possible explanations. First, this could be the averaging mentioned by Weinberg (1976), skewed to slightly lower values because of selection effects. Second, the physical model on which the ZKDR distance is based is wrong, but in our Universe the m–z relation is similar to that in a high-η ZKDR universe (Peel *et al.* 2014; Helbig 2015b) (see Sect. 27).

Busti & Santos (2011) pointed out that the procedure used by Yu *et al.* (2011) to calculate $H(z)$ is not consistent, because the equation relating $H(z)$ and the angular-size distance is valid only for $\eta = 1$. Santos *et al.* (2008) had done a similar analysis to that of Yu *et al.* (2011)

[34] Note that her claim that Perlmutter *et al.* (1999) 'assumed a smooth universe' is somewhat misleading. While they did not consider a direction-dependent η, they did compare the extreme cases of $\eta = 1$ and $\eta = 0$ as well as the case of an Ω_0-dependet η (*i.e.* galaxies assigned to clumps and the rest of the matter distributed smoothly, which implies an increase in η with increasing Ω_0), in all cases using the code of KHS.

[35] There is a range of directness in measurement. At one level, all that is ever measured in astronomy is number of photons as a function of position on a detector, which can be related to apparent magnitude for a conventional exposure or as the intensity of a spectrum in the case of spectroscopy. Everything else is interpretation. Nevertheless, it makes sense to say that one can directly measure redshift, magnitude, and angular size, and, one step less concrete, that one can measure λ_0 and Ω_0 *via* the derived parameters (assuming some framework, such as FRW). Despite some claims to the contrary, no cosmological test directly measures acceleration; this is calculated from the cosmological parameters obtained. Similarly, $H(z)$ is a calculated quantity.

using supernova data, concluding that $\eta > 0.42$ (2σ). Adding the $H(z)$ data used by Yu *et al.* (2011) of course improves the constraints, resulting in $0.66 \le \eta \le 1.0$ (2σ) with the best fit at $\eta = 1$, a broadly similar result. Note that Helbig (2015a) also finds the best-fit value $\eta = 1$ if λ_0 and/or Ω_0 are constrained. Thus, while Yu *et al.* (2011) did indeed make a mistake, the fact that $\eta \approx 1$ means that it didn't appreciably affect their main result.

While there is no evidence that our Universe is not well described by an FRW model, it is important to test for deviations from this assumption. One possibility is to test the Copernican Principle by looking for a redshift dependence of the curvature parameter (Clarkson *et al.* 2008); another is to express Ω_0 in terms of observable quantities, resulting in an expression which must hold at all redshifts (Sahni *et al.* 2008; Zunkel & Clarkson 2008). Busti & Lima (2012) pointed out that these tests implicitly assume that the universe is assumed to be homogeneous and isotropic on all scales (in other words, the 'RW' is assumed; the idea is to test the 'F' part of FRW), and showed that using the ZKDR distance leads to false positives for these tests (*i.e.* the Copernican Principle appears to be violated when in fact it is not). Busti & Lima (2012) also rewrite the ZKDR equation so that η is given as a function of observable quantities, allowing one to reconstruct $\eta(z)$ from observations for a general ΛCDM model. Such an $\eta(z)$ can also mimic the behaviour of model with $\eta = 1$ but with $w \neq -1$, *i.e.* some form of dark energy other than a cosmological constant.

Inhomogeneous cosmological models definitely affect light propagation. Whether they affect the expansion rate of the universe is still debated. Green & Wald (2014) claimed that there is no evidence that an FRW model is not a good description of the Universe on essentially all scales (except perhaps the extremely small scales encountered in, for example, the m–z relation for Type Ia supernovae).

31. CLASSICAL COSMOLOGY: ANGULAR DIAMETERS

One of the basic cosmological tests is the 'standard rod' test, *i.e.* the comparison of the angular size as a function of redshift of an object of given size to the theoretical expection derived from the angular-size–redshift relation, which in turn depends on the the cosmological parameters. (By the same token, the calculation of the physical size from the observed angular size depends on the cosmological model, and on η.) Although a classic test, no useful constraints have been derived from it—except in the cases of the CMB and BAO, though here the corresponding physical lengths are so large that the ZKDR distance plays no role (*e.g.* Lewis & Challinor 2006)—primarily because of the difficulty in finding a standard rod. Nevertheless, some progress can be made. For example, Alcaniz *et al.* (2004), assuming a Gaussian prior $\Omega_0 = 0.35 \pm 0.07$ in a flat universe, found the best fit at $\Omega_0 = 0.35$ and $\eta = 0.8$ (consistent with the results mention in Sect. 27).

Araújo & Stoeger (2009) point out the interesting, long-known, but generally unappreciated fact that, for a flat universe, the redshift at which the maximum of the angular-size distance occurs is a direct measure of λ_0, independently of H_0. For a non-flat universe, knowledge of the redshift of the maximum and H_0 allows one to determine both Ω_0 and λ_0. Note, however, that this depends on the assumption that $\eta = 1$.

Also, Chen & Ratra (2012) examined constraints from the angular sizes of galaxy clusters, both for general FRW models and for two classes of flat models with different types of dark energy. Their conclusion is still valid today: such constraints are approximately as restrictive as those based on gammay-ray–burst apparent-luminosity data, strong–gravitational-lensing measurements, or the age of the Universe, but less so than those from BAO or the m–z relation for Type Ia supernovae (or the CMB). Nevertheless, as an independent constraint, the fact that they are compatible with other data strengthens our confidence in the concordance model.

32. ANALYTIC APPROXIMATIONS

In general, analytic solutions of the ZKDR distance are very complicated. Moreover, there are analytic solutions only for special values of λ_0, Ω_0, or η.

Wang (1999) presented an approximation for the ZKDR distance as a polynomial in η with coefficients which depend on redshift and the cosmological parameters, the latter via the fact that the coefficients depend on the distance calculated for given values of Ω_0 and λ_0 for η values of 0, 0.5, 1, and 1.5. Note that $\eta = 1.5$ is in conflict with the assumptions under which the ZKDR distance is derived; nevertheless, this can be valid from a heuristic point of view (*e.g.* Lima *et al.* 2014). Of course, $\eta > 1$ everywhere is impossible, but could be valid if η depends on the line of sight. In that case, however, one should think of it as an average along the line of sight, *i.e.* a particular line of sight might, by chance, have an above-average amount of matter along it. If this were constant, it would imply an extremely long structure aligned with the line of sight, which would not be compatible with an approximate FRW model.

Demianski *et al.* (2003)[36] found 'an approximate analytic solution... which is simple enough and sufficiently accurate to be useful in practical applications'. It is not clear how useful this is, though. It was apparently discovered more or less by accident and has no theoretical basis. As such, it is not clear *a priori* in which cases it is a good approximation, so one needs to test it against an (at least numerically) exact solution, in which case one might just as well use the better solution.[37]

Also, the numerical implementation of KHS is, in most cases, only a factor of 3 or so slower than the elliptic-integral solution (of course, one can compare only in those cases where such solutions exist; the numerical implementation knows no special cases—and is valid for all values of the input parameters, using the same algorithm for all—and the speed depends only weakly on the input parameters), so there doesn't seem to be a real need for approximate solutions; even if such an approximation is faster than the elliptic-integral solution (and valid for all input parameters), the elliptic-integral solution is 'almost analytic' and reasonably fast, so a factor of 3 for a general and accurate numerical implementation is not a big disadvantage in practice. Though restricted to $k = 0$ and $\eta = 1$, similar remarks apply to the work of Pen (1999).

[36] See also Demianski *et al.* (2000) which is the precursor, but longer and substantially different in places.

[37] Lewis Carroll, in one of his less famous books, describes a map with a scale of 1:1, but it was easier to just use the real Earth than the map (Carroll 1893).

33. SUMMARY

The basis of observational cosmology is calculating the dependence of some observational quantity—usually related to some distance—on redshift for a variety of cosmological models, then determining the corresponding cosmological parameters *via* finding the model which gives the best fit to the data. Small-scale inhomogeneities can affect the relation between redshift and distance, thus it at least needs to be investigated whether results depend on the amount of inhomogeneity. Zel'dovich (1964b) introduced a simple model for such small-scale inhomogeneities and an analytic solution (for the Einstein–de Sitter model) for the extreme case, namely that light propagates through completely empty space, all of the matter being located in clumps outside the beam. Subsequent work generalized that model to other cosmological models and/or intermediate degrees of inhomogeneity (later known as the ZKDR distance, after the initials of the most influential pioneers), investigated a similar approach involving so-called Swiss-cheese models (not necessarily more realistic, but exact solutions of the Einstein equations) which were later shown to correspond to the Zel'dovich (1964b) model in a well defined way, investigated assumptions in the models and their effects (*e.g.* whether the clumps are transparent, if averaging whether the average is taken over the celestial sphere or over all lines of sight, *etc.*), compared the results of the models with exact solutions or numerical simulations, and developed approximations to various distance formulae. Approximations are no longer needed, now that computing power has increased and an efficient numerical implementation is available for the general case (Kayser *et al.* 1997).

Most of the theory was complete by the middle of the 1970s. The discovery of the first gravitational-lens system in 1979 revived interest in this topic: since gravitational lenses obviously require an inhomogeneous universe, in such cases the assusmption of a completely homogeneous universe with regard to light propagation becomes more obvious. Until the middle of the 1990s or so, effects of inhomogeneities were not that important in observational cosmology, for two reasons. First, the uncertainty in the cosmological parameters was large, comparable to (with respect to the effect on the distance as a function of redshift) variation in the inhomogeneity parameter η. Second, most observations were at low redshift, whereas η is a higher-order effect compared to the first- and second-order parameters H_0 and q_0 (see also equation (8) in Kantowski 1998). The use of Type Ia supernovae for the m–z relation extended observations to higher redshift. Also, both this test as well as others had constrained the cosmological parameters to a degree that the effect of η could no longer be ignored, which led to another revival of interest.

In general, the effect of η depends on angular scale: large angular scales correspond to a fair sample of the universe within the beam, while this is not necessarily the case for small angular scales. Since supernovae have an angular scale of about 10^{-7} arcsec, which is very small, one would perhaps expect to see effects of η in the m–z relation for Type Ia supernovae. However, many independent investigations come to the conclusion that $\eta \approx 1$, not just on average, as is to be expected, at least under certain assumptions (Weinberg 1976), but also for each individual line of sight. The reason for that is probably that the Zel'dovich (1964b) model is incorrect in the sense that it is not a good approximation for our Universe: no-one doubts that the ZKDR distance is correct in a universe with a mass distribution well modelled by that on which the idea of the ZKDR distance is based, but apparently that is not our Universe. In other words, most of the matter is not outside the beam, even for very narrow beams, but rather even such very narrow beams fairly sample the Universe. Note that $\eta \approx 1$ does not necessarily imply that matter is distributed homogeneously within the beam; it just implies that the distance as calculated from redshift is approximately the same as if that were the case. In reality, such a beam will traverse voids with less than average density, but also regions (corresponding to the filaments and sheets of large-scale structure) with much higher than average density. Although this violates the assumptions on which the ZKDR distance is based, nevertheless in practice such a mass distribution results in distance as a function of redshift very close to the standard distance, *i.e.* that optained by assuming that the universe, at least with regard to light propagation, is completely homogeneous.

Acknowledgements

I thank both anonymous referees for thorough reviews of the first draft and resulting helpful comments, which have been incorporated into the final version. I'm particularly grateful to Maria C. McEachern (John G. Wolbach Library, Smithsonian Astrophysical Observatory) for supplying me with copies of the original Russian articles by Zel'dovich, Dashevskii, and Slysh. This research has made use of NASA's Astrophysics Data System Bibliographic Services. The figure was made with help of the GRAL software package written by Rainer Kayser.

REFERENCES

Adamek J., Daverio D., Durrer R., Kunz M., 2016, J. Cosmol. Astropart. Phys., 2016, 053
Alcaniz J.S., Lima J.A.S., Silva R., 2004, Int. J. Mod. Phys. D, 13, 1309
Alcock C., Anderson N., 1985, ApJ, 291, L29
Alcock C., Anderson N., 1986, ApJ, 302, 43
Alnes H., Amarzguioui M., Grøn Ø., 2006, Phys. Rev. D, 73, 083519
Amanullah R., Mörtsell E., Goobar A., 2003, A&A, 397, 819
Araújo M.E., Stoeger W.R., 2009, MNRAS, 394, 438
Asada H., 1998, ApJ, 401, 473
Avni Y., Shulami I., 1988, ApJ, 332, 113
Babul A., Lee M.H., 1991, MNRAS, 250, 407
Bagheri S., Schwarz D.J., 2014, J. Cosmol. Astropart. Phys., 2014, 073
Barber A.J., 2000, MNRAS, 318, 195
Ben-Dayan I., Takahishi R., 2016, MNRAS, 455, 552
Bentivegna E., Bruni M., 2016, Phys. Rev. Lett., 116, 251302
Bergström L., Goliath M., Goobar A., Mörtsell E., 2000, A&A, 358, 13
Berry M.V., 1986, Cosmology and Gravitation. Adam Hilger, Bristol
Bertotti B., 1966, Proc. Roy. Soc. Lond. A, 294, 195
Betoule M., *et al.*, 2014, A&A, 568, A22
Biggs A.D., Browne I.W.A., 2018, MNRAS, 476, 5393
Bolejko K., 2011, MNRAS, 412, 1937

Bolejko K., Ferreira P.G., 2012, J. Cosmol. Astropart. Phys., 2012, 003
Bondi H., 1961, Cosmology. Cambridge University Press, Cambridge (UK)
Bonvin C., Clarkson C., Durrer R., Martens R., Umeh O., 2015, J. Cosmol. Astropart. Phys., 2015, 040
Bréton N., Montiel A., 2013, Phys. Rev. D, 87, 063527
Brouyzakis N., Tetradis N., Tzavara E., 2008, J. Cosmol. Astropart. Phys., 2008, 008
Busti V., Holanda R.F.L., Clarkson C., 2013, J. Cosmol. Astropart. Phys., 2013, 020
Busti V.C., Lima J.A.S., 2012, MNRAS, 426, L41
Busti V.C., Santos R.C., 2011, Res. Astron. Astrophys., 11, 637
Busti V.C., Santos R.C., Lima J.A.S., 2012, Phys. Rev. D, 85, 103503
Canizares C.R., 1982, ApJ, 263, 508
Carroll L., 1893, Sylvie and Bruno Concluded. Macmillan and Co., London
Carroll S.M., 2019, Spacetime and Geometry: An Introduction to General Relativity. Cambridge University Press
Castañeda L., Valencia D.M., 2008, In Oscoz A., Mediavilla E., Serra-Ricart M., eds., Spanish Relativity Meeting - Encuentros Relativistas Españoles, ERE2007: Relativistic Astrophysics and Cosmology, EDP Sciences, vol. 30 of EAS Publications Series, pp. 295–298
Chae K.H., *et al.*, 2002, Phys. Rev. Lett., 89, 151301
Chen Y., Ratra B., 2012, A&A, 543, A104, jul.
Clarkson C., Bassett B., Lu T.H.C., 2008, Phys. Rev. Lett., 101, 011301
Clarkson C., Ellis G.F.R., Faltenbacher A., Maartens R., Umeh O., Uzan J.P., 2012, MNRAS, 426, 1121
Claudel C.M., 2000, Proc. Roy. Soc. Lond. A, 456, 1455
Clifton T., Zuntz J., 2009, MNRAS, 400, 2185
Colgate S.A., 1979, ApJ, 232, 404
Conley A., *et al.*, 2011, ApJS, 192, 1
Covone G., Sereno M., de Ritis R., 2005, MNRAS, 357, 773
Dalal N., Holz D.E., Chen X., Frieman J.A., 2003, ApJ, 585, L11
Dashevskii V.M., Slysh V.J., 1965, Astronomicheskii Zhurnal, 42, 863
Dashevskii V.M., Slysh V.J., 1966, SvA, 9, 671
Dashevskii V.M., Zel'dovich Y.B., 1964, Astronomicheskii Zhurnal, 41, 1071
Dashevskii V.M., Zel'dovich Y.B., 1965, SvA, 8, 854
Davis T.M., Lineweaver C.H., 2004, PASA, 21, 97
Demianski M., de Ritis R., Marino A.A., Piedipalumbo E., 2000, Approximate angular diameter distance in a locally inhomogeneous universe with a nonzero cosmological constant. arXiv:astro-ph/0004376
Demianski M., de Ritis R., Marino A.A., Piedipalumbo E., 2003, A&A, 411, 33
Dhawan S., Goobar A., Mörtsell E., 2018, J. Cosmol. Astropart. Phys., 2018, 024
Dodelson S., Vallinotto A., 2006, Phys. Rev. D, 74, 063515
Dyer C.C., 1976, MNRAS, 175, 429
Dyer C.C., Oattes L.M., 1988, ApJ, 326, 50
Dyer C.C., Roeder R.C., 1972, ApJ, 174, L115
Dyer C.C., Roeder R.C., 1973, ApJ, 180, L31
Dyer C.C., Roeder R.C., 1974, ApJ, 189, 167
Dyer C.C., Roeder R.C., 1976, Nature, 260, 764
Dyer C.C., Roeder R.C., 1981a, ApJ, 249, 290
Dyer C.C., Roeder R.C., 1981b, Gen. Relativ. Gravit., 13, 1157
Ehlers J., Schneider P., 1986, A&A, 168, 57
Einstein A., 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, pp. 142–152
Einstein A., Straus E.G., 1945, Rev. Mod. Phys., 17, 120
Einstein A., Straus E.G., 1946, Rev. Mod. Phys., 18, 148
Ellis G.F.R., 1980, In Ehlers J., Perry J.J., Walker M., eds., Ninth Texas Symposium on Relativistic Astrophysics, New York Academy of Sciences, New York, vol. 336, pp. 130–160
Ellis G.F.R., Bassett B.A.C.C., Dunsby P.K.S., 1998, Class. Quant. Grav., 15, 2345
Etherington I.M.H., 1933, Philosophical Magazine, 15, 761
Falco E., Kochanek C.S., Muñoz J.A., 1998, ApJ, 494, 47
Fang L., Wu X., 1989, Chinese Physics Letters, 6, 233
Filippenko A.V., 2005, In Sion E.M., Vennes S., Shipman H.L., eds., White dwarfs: cosmological and galactic probes, Springer, Dordrecht, vol. 332 of Astrophysics and Space Science Library, pp. 97–133
Flanagan É.É., Kumar N., Wasserman I., Vanderveld R.A., 2012, Phys. Rev. D, 85, 023510
Flanagan É.É., Kumar N., Wasserman I., 2013, Phys. Rev. D, 88, 043004
Fleury P., 2014, J. Cosmol. Astropart. Phys., 2014, 054
Fleury P., Dupuy H., Uzan J.P., 2013a, Phys. Rev. D, 87, 123526
Fleury P., Dupuy H., Uzan J.P., 2013b, Phys. Rev. Lett., 111, 091302
Friedmann A.A., 1922, Zeitschrift für Physik, 1, 377
Friedmann A.A., 1924, Zeitschrift für Physik, 21, 326
Frieman J.A., 1996, Comments Astrophys., 18, 323
Fukugita M., Futamase T., Kasai M., 1990, MNRAS, 246, 24
Fukugita M., Futamase K., Kasai M., Turner E.L., 1992, ApJ, 393, 3
Fukushige T., Makino J., 1994, ApJ, 436, L111
Futamase T., Sasaki M., 1989, Phys. Rev. D, 40, 2502
Garfinkel D., 2006, Class. Quant. Grav., 23, 4811
Giblin Jr. J.T., Mertens J.B., Starkman G.D., 2016a, Phys. Rev. Lett., 116, 251301
Giblin Jr. J.T., Mertens J.B., Starkman G.D., 2016b, ApJ, 833, 247
Giovi F., Amendola L., 2001, MNRAS, 325, 1097
Goliath M., Mörtsell E., 2000, Phys. Lett. B, 486, 249
Goobar A., Perlmutter S., 1995, ApJ, 450, 14
Gott III J.R., Gunn J.E., Schramm D.N., Tinsley B., 1974, ApJ, 194, 543
Green S.R., Wald R.M., 2014, Class. Quant. Grav., 31, 234003
Gunn J.E., 1967a, ApJ, 147, 61
Gunn J.E., 1967b, ApJ, 150, 737
Gunnarsson C., Dahlén T., Goobar A., Jönsson J., Mörtsell E., 2006, ApJ, 640, 417
Hada R., Futamase T., 2014, J. Cosmol. Astropart. Phys., 2014, 042
Hamana T., 1998, Prog. Theor. Phys., 99, 1085
Hamana T., 1999, MNRAS, 302, 801
Hammer F., 1985, A&A, 152, 262
Harrison E.R., 1993, ApJ, 403, 28
Harrison E.R., 2000, Cosmology, the Science of the Universe. Cambridge University Press, Cambridge (UK), 2nd edn.
Hawkins M.R.S., 1993, Nature, 366, 242
Hawkins M.R.S., 1996, MNRAS, 278, 787
Hawkins M.R.S., 1997, Hunting down the universe : the missing mass, primordial black holes, and other dark matters. Addison-Wesley, Reading, Mass.
Hawkins M.R.S., Taylor A.N., 1997, ApJ, 482, L5
Heacox W.D., 2015, The Expanding Universe: A Primer on Relativistic Cosmology. Cambridge University Press, Cambridge (UK)
Helbig P., 1996, `ANGSIZ` user's guide. arXiv:astro-ph/9603028v3
Helbig P., 1997, In Helbig P., Jackson N., eds., Golden Lenses, University of Manchester, NRAL, Jodrell Bank, `http://www.astro.multivax.de:8000/ceres/workshop1/proceedings.html`
Helbig P., 1998, MNRAS, 298, 395
Helbig P., 2015a, MNRAS, 451, 2097
Helbig P., 2015b, MNRAS, 453, 3975
Helbig P., Marlow D.R., Quast R., Wilkinson P.N., Browne I.W.A., Koopmans L.V.E., 1999, A&AS, 136, 297
Hewish A., Readhead A.C.S., Duffett-Smith P.J., 1974, NAT, 252, 657
Holz D.E., 1998, ApJ, 506, L1
Holz D.E., Linder E.V., 2005, ApJ, 631, 678
Holz D.E., Wald R.M., 1998, Phys. Rev. D, 58, 063501
Hoyle F., Sandage A.R., 1956, PASP, 68, 301
Isaacson J.A., Canizares C.R., 1989, ApJ, 336, 544
Iwata K., Yoo C.M., 2015, Europhysiscs Letters, 109, 39001
Jaroszyński M., Paczyński B., 1996, Acta Astronomica, 46, 361
Jaroszyński M., Park C., Paczyński B., Gott III J.R., 1990, ApJ, 365, 22
Jönsson J., Dahlén T., Goobar A., Gunnarsson C., Mörtsell E., Lee K., 2006, ApJ, 639, 991
Jönsson J., Dahlén T., Goobar A., Mörtsell E., Riess A., 2007, J. Cosmol. Astropart. Phys., 2007, 002

Jönsson J., Kronborg T., Mörtsell E., Sollerman J., 2008, A&A, 467, 463
Jönsson J., Mörtsell E., Sollerman J., 2009, A&A, 493, 331
Jönsson J., Dahlén T., Hook I., Goobar A., Mörtsell E., 2010, MNRAS, 402, 526
Kai T., Kozaki H., Nakao K., Nambu Y., Yoo C., 2007, Prog. Theor. Phys., 117, 229
Kainulainen K., Marra V., 2009, Phys. Rev. D, 80, 123020
Kaiser N., Peacock J.A., 2016, MNRAS, 455, 4518
Kantowski R., 1969, ApJ, 155, 89
Kantowski R., 1998, ApJ, 507, 283
Kantowski R., 2003, Phys. Rev. D, 68, 123516
Kantowski R., Thomas R.C., 2001, ApJ, 561, 491
Kantowski R., Vaughan T., Branch D., 1995, ApJ, 447, 35
Kantowski R., Kao J.K., Thomas R.C., 2000, ApJ, 545, 549
Karoji H., Nottale L., 1976, NAT, 259, 31
Kasai M., Futamase T., Takahara F., 1990, Phys. Lett. A, 147, 97
Kayser R., 1985, Helligkeitsänderung durch den statistischen Gravitationslinseneffekt. Ph.D. thesis, University of Hamburg
Kayser R., Refsdal S., 1983, A&A, 128, 156
Kayser R., Helbig P., Schramm T., 1997, A&A, 318, 680
Kibble T.W.B., Lieu R., 2005, ApJ, 632, 718
Knop R.A., *et al.*, 2003, ApJ, 598, 102
Kochanek C.S., 1993, ApJ, 419, 12
Kochanek C.S., 1996a, ApJ, 466, 638
Kochanek C.S., 1996b, ApJ, 473, 595
Kochanek C.S., Falco E.E., Schild R., 1995, ApJ, 452, 109
Kostov V., 2010, J. Cosmol. Astropart. Phys., 2010, 001
Krauss L.M., Turner M.S., 1995, Gen. Relativ. Gravit., 27, 1137
Kristian J., Sachs R.K., 1966, ApJ, 143, 379
Lavinto M., Räsänen S., 2015, J. Cosmol. Astropart. Phys., 2015, 57
Lee M.H., Paczyński B., 1990, ApJ, 357, 32
Leibundgut B., 2001, ARA&A, 39, 67
Leibundgut B., 2008, Gen. Relativ. Gravit., 40, 221
Lemaître G., 1927, Annales de la Societé Scientifique de Bruxelles, 47, 49
Lewis A., Challinor A., 2006, Phys. Rep., 429, 1
Li Z., Ding X., Zhu Z.H., 2015, Phys. Rev. D, 91, 083010
Lima J.A.S., Busti V.C., Santos R.C., 2014, Phys. Rev. D, 89, 067301
Linder E., 1998, ApJ, 497, 28
Loh E.D., Spillar E.J., 1986, ApJ, 307, L1
Marra V., Kolb E.W., Matarrese S., Riotto A., 2007, Phys. Rev. D, 76, 123004
Marra V., Kolb E.W., Matarrese S., 2008, Phys. Rev. D, 77, 023003
Martel H., Premadi P., 2008, ApJ, 673, 657
Martel H., Premadi P., Matzner R., 2002, ApJ, 571, 17
Mattig W., 1958, Astron. Nachr., 284, 109
Mertens J.B., Giblin Jr. J.T., Starkman G.D., 2016, Phys. Rev. D, 93, 124059
Metcalf R.B., Silk J., 1999, ApJ, 519, L1
Metcalf R.B., Silk J., 2007, Phys. Rev. D, 98, 071302
Minty E.M., Heavens A.F., Hawkins M.R.S., 2002, MNRAS, 330, 378
Mörtsell E., 2002, A&A, 382, 787
Mörtsell E., Goobar A., Bergström L., 2001, ApJ, 559, 53
Nakamura T.T., 1997, PASJ, 49, 151
Narayan R., 1991, ApJ, 378, L5
Navarro J.F., Frenk C.S., White S.D.M., 1997, ApJ, 490, 493
Nottale L., 1982a, A&A, 110, 9
Nottale L., 1982b, A&A, 114, 261
Nottale L., 1983, A&A, 118, 85
Nottale L., 1984, MNRAS, 206, 713
Nottale L., Chauvineau B., 1986, A&A, 162, 1
Nottale L., Hammer F., 1984, A&A, 141, 144
Nottale L., Vigier J.P., 1977, NAT, 268, 608
Odderskov I., Koksbang S.M., Hannestad S., 2016, J. Cosmol. Astropart. Phys., 2016, 001
Okamura T., Futamase T., 2009, Prog. Theor. Phys., 122, 511
Omote M., Yoshida H., 1990, ApJ, 361, 27
Ostriker J.P., Steinhardt P.J., 1995, Nature, 377, 600
Park M.G., Gott III J.R., 1997, ApJ, 489, 476
Payne T., Birkinshaw M., 2004, MNRAS, 348, 581
Peacock J.A., 1986, MNRAS, 221, 113
Peacock J.A., 1999, Cosmological Physics. Cambridge University Press, Cambridge
Peebles P.J.E., 1993, Principles of Physical Cosmology. Princeton University Press, Princeton
Peel A., Troxel M.A., Ishak M., 2014, Phys. Rev. D, 90, 123536
Peel A., Troxel M.A., Ishak M., 2015, Phys. Rev. D, 92, 029901
Pei Y., 1993a, ApJ, 403, 7
Pei Y., 1993b, ApJ, 404, 436
Pen U.L., 1999, ApJS, 120, 49
Perlmutter S., Schmidt B.P., 2003, In Weiler K., ed., Supernovae and Gamma-Ray Bursters, Springer, Lecture Notes in Physics, pp. 195–2017
Perlmutter S., *et al.*, 1999, ApJ, 517, 565
Planck Collaboration, 2014, A&A, 571, A16
Planck Collaboration, 2016, A&A, 593, A13
Planck Collaboration, 2019, A&A, arXiv:1807.06209
Premadi P., Martel H., Matzner R., 1998, ApJ, 493, 10
Premadi P., Martel H., Matzner R., Futamase T., 2001, ApJS, 135, 739
Premadi P., Martel H., Matzner R., 2004, ApJ, 611, 1
Premadi P., Martel H., Matzner R., 2008, ApJ, 673, 657
Press W.H., Gunn J.E., 1973, ApJ, 185, 397
Quast R., Helbig P., 1999, A&A, 344, 721
Raychaudhuri A., 1955, Phys. Rev., 98, 1123
Refsdal S., 1964, MNRAS, 128, 307
Refsdal S., 1966, MNRAS, 132, 101
Refsdal S., 1970, ApJ, 159, 357
Refsdal S., Stabell R., del Lange F.G., 1967, Mem. R. Astron. Soc., 71, 143
Riess A.G., 2000, PASP, 112, 1284
Riess A.G., *et al.*, 1998, AJ, 116, 1009
Robertson H.P., 1935, ApJ, 82, 284
Robertson H.P., 1936, ApJ, 83, 187
Roeder R.C., 1975a, ApJ, 196, 671
Roeder R.C., 1975b, Nature, 255, 124
Rose H.G., 2001, ApJ, 560, L15
Rubin V.C., Ford Jr. W.K., Rubin J.S., 1973, APJ, 183, 111
Rusu C.E., *et al.*, 2017, MNRAS, 467, 4220
Rusu C.E., *et al.*, 2019, accepted by MNRAS, arXiv:1905.09338
Ryden B., 2017, Introduction to Cosmology. Cambridge University Press, Cambridge, 2nd edn.
Sachs R., 1961, Proc. Roy. Soc. Lond. A, 264, 309
Sahni V., Shafieloo A., Starobinsky A.A., 2008, PRD, 78, 103502
Sandage A.R., 1970, Physics Today, 23, 34
Sandage A.R., 1995, In Binggeli B., Buser R., eds., The Deep Universe, Springer, Berlin
Sandage A.R., Hardy E., 1973, APJ, 183, 743
Santos R.C., Lima J.A.S., 2006, ZKDR distance, angular size and phantom cosmology. arXiv:astro-ph/0609129
Santos R.C., Lima J.A.S., 2008, Phys. Rev. D, 77, 083505
Santos R.C., Cunha J.V., Lima J.A.S., 2008, Phys. Rev. D, 77, 023519
Sato H., 1985, Prog. Theor. Phys., 73, 649
Schmidt B.P., 2002, Class. Quant. Grav., 19, 3487
Schmidt B.P., *et al.*, 1998, ApJ, 507, 46
Schneider P., 1984, A&A, 140, 119
Schneider P., Weiss A., 1988a, ApJ, 327, 526
Schneider P., Weiss A., 1988b, ApJ, 330, 1
Schneider P., Jürgen Ehlers, Falco E.E., 1992, Gravitational Lenses. Springer-Verlag, Heidelberg
Seitz S., Schneider P., 1994, A&A, 287, 349
Seitz S., Schneider P., Ehlers J., 1994, Class. Quant. Grav., 11, 2345
Seljak U., Holz D.E., 1999, A&A, 351, L10
Sheth R.K., Tormen G., 1999, MNRAS, 308, 119
Smith M., *et al.*, 2014, ApJ, 780, 24
Solheim J.E., 1966, MNRAS, 133, 321
Szekeres P., 1975, Comm. Math. Phys., 41, 55
Terrell J., 1977, Amer. J. Phys., 45, 869
Thomas R.C., Kantowski R., 2000, Phys. Rev. D, 62, 103507
Tihhonova O., *et al.*, 2018, MNRAS, 477, 5657
Tomita K., 1998, Prog. Theor. Phys., 100, 77
Tomita K., Asada H., Hamana T., 1999, Prog. Theor. Phys. Supp., 133, 155
Turner E.L., Ostriker J.P., Gott III J.R., 1984, ApJ, 284, 1
Valageas P., 2000, A&A, 354, 767

Vanderveld R.A., Flanagan É.É., Wasserman I., 2006, Phys. Rev. D, 74, 023506
Vanderveld R.A., Flanagan É.É., Wasserman I., 2008, Phys. Rev. D, 78, 083511
Wagoner R.V., 1977, APJ, 214, L5
Walker A.G., 1935, Quart. J. Math., 6, 81
Walker A.G., 1937, Proceedings of the London Mathematical Society (Series 2), 42, 90
Walsh D., Carswell R.F., Weymann R.J., 1979, Nature, 279, 381
Wambsganss J., Cen R., Xu G., Ostriker J.P., 1997, ApJ, 475, L81
Wang Y., 1999, ApJ, 525, 651
Wang Y., 2000, ApJ, 536, 351
Wang Y., 2005, J. Cosmol. Astropart. Phys., 2005, 0005
Wang Y., Holz D.E., Munshi D., 2002, ApJ, 572, L15
Wang Y., Tenbarge J., Fleshman B., 2005, ApJ, 624, 46
Wardle J.F.C., Pottash R., 1977, In Papagiannis M.D., ed., Eighth Texas Symposium on Relativistic Astrophysics, New York Academy of Sciences, New York, vol. 302, p. 605
Watanabe K., 1992, Prog. Theor. Phys., 87, 367
Watanabe K., 1993, Vistas in Astronomy, 37, 531
Watanabe K., Tomita K., 1990, ApJ, 355, 1
Watanabe K., Tomita K., 1991, ApJ, 370, 481
Watanabe K., Sasaki M., Tomita K., 1993, ApJ, 394, 38
Weinberg S., 1972, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity. John Wiley & Sons, New York
Weinberg S., 1976, ApJ, 208, L1
Wheeler J.A., 1958, In La Structure et l'Évolution de l'Univverse, Onzième Conseil de Physique Solvay, Editions R. Stoops, Brussels, pp. 124–141
Williams L.L.R., Song J., 2004, MNRAS, 351, 1387
Wong K.C., *et al.*, 2016, MNRAS, 465, 4895
Wu X., 1998, A&A, 239, 29
Yang X., Yu H.R., Zhang T.J., 2013, J. Cosmol. Astropart. Phys., 2013, 007
Yoo C., Ishihara H., Nakao K., Tagoshi H., 2008, Prog. Theor. Phys., 120, 961
Yoshida H., Omote M., 1992, ApJ, 388, L1
Yu H.R., Lan T., Wan H.Y., Zhang T.J., Wang B.Q., 2011, Res. Astron. Astrophys., 11, 125
Zackrisson E., Bergvall N., Marquart T., Helbig P., 2003, A&A, 408, 17
Zel'dovich Y.B., 1964a, Astronomicheskii Zhurnal, 41, 19
Zel'dovich Y.B., 1964b, SvA, 8, 13
Zumalacárregui M., Seljak U., 2018, Phys. Rev. Lett., 121, 141101
Zunkel C., Clarkson C., 2008, Phys. Rev. Lett., 101, 181301
Zwicky F., 1933, Helv. Phys. Acta, 6, 110
Zwicky F., 1937a, Phys. Rev., 51, 290
Zwicky F., 1937b, Phys. Rev., 51, 679
Zwicky F., 1937c, ApJ, 86, 217

2.3 Follow-up

Since it is relatively new, there has not been enough time for this paper to have had much impact, though it does already have at least three citations (excluding self-citations). I hope that it provides a good introduction to, and historical survey of, the topic and will prove useful.

A few months later, I published an abridged version, with a bit more emphasis on why it turned out that it is usually acceptable to assume $\eta \approx 1$ (Helbig, 2020c).

Part II

General theory

Chapter 3

A general and practical method for calculating cosmological distances

3.1 Context

When I started my *Diplom*[1] thesis work at the Hamburg Observatory, in the group of the late Sjur Refsdal, I intentionally chose a topic where it was not clear from the start that programming would be involved. As it turned out, programming was necessary (and, as soon as I got a taste of it, I was happy that it had turned out that way). In particular, I needed to calculate the angular-size distance for cosmological models with arbitrary λ_0 and Ω_0, and also arbitrary η. There was no standard package to do this.[2] Kayser (1985) had derived the second-order differential equation for the angular-size distance for cosmological models with arbitrary λ_0 and Ω_0, and also arbitrary η, but had only a rudimentary numerical implementation (using Simpson integration). Since that had been published only in his doctoral thesis, we thought that it would be a good idea to write it up for a paper and, since I was working on a better numerical implementation, make that code publicly available in conjunction with the paper. To put things into context, the paper also summarizes the FLRW models, illustrates various definitions of distance, and discusses the numerical implementation. Relegated to appendices are a discussion of the symmetry properties (in particular the generalization of the Etherington (1933) reciprocity relation), known analytic solutions, and the calculation of the volume element.

The paper is based on the equation derived by Rainer Kayser (1985). I wrote the paper and wrote the FORTRAN77 code (FORTRAN90 compilers were not yet available). Thomas Schramm was involved in the planning and discussion of the paper, especially the mathematical aspects. Although I was officially a student of Refsdal, I worked most closely with Rainer and benefitted greatly from almost daily discussions with him and Tom.

This paper was published at a time when *Astronomy & Astrophysics* pub-

[1] The German *Diplom* degree, now essentially obsolete as a result of the Bologna process, was the rough equivalent of a master's degree, but included a one-year thesis, for a total nominal time of five years.

[2] Around the same time, Feige (1992) published a paper dealing with distance calculation *via* elliptic integrals for arbitrary λ_0 and Ω_0, but only for $\eta = 1$. His paper was more of a cookbook, and he didn't make the code publicly available, but he let me have it on request. It is written in C but not in a portable fashion.

lished some content only electronically. In particular, the appendices of this paper were published only electronically. I posted the full version to arXiv, and a full version is also available on my own web pages. In order to include official versions as much as possible in this thesis, the first part of the paper is the PDF file as provided by ADS (not including the appendices). (As far as I know, PDF is not available from *A&A* for papers this old.) The appendices are still available *via* anonymous ftp: `cdsarc.u-strasbg.fr/A+A/318/680`. I have retrieved the gzipped POSTSCRIPT file available there. Since the text is cut off at the top of each page, I have added the POSTSCRIPT command "0 -72 translate" at the start of each page in the POSTSCRIPT file, then converted this to PDF using GHOSTSCRIPT. As with the other papers, PDFTK was used to combine official PDF versions of the papers with PDF (produced from LaTeX *via* POSTSCRIPT) from the rest of the thesis.

Astron. Astrophys. 318, 680–686 (1997)

ASTRONOMY
AND
ASTROPHYSICS

A general and practical method for calculating cosmological distances*

Rainer Kayser[**1], **Phillip Helbig**[***1], **and Thomas Schramm**[†1,2]

[1] Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg-Bergedorf, Germany
[2] Rechenzentrum, Technische Universität Hamburg-Harburg, Denickestraße 17, D-21071 Hamburg-Harburg, Germany

Received 19 February 1996 / Accepted 22 April 1996

Abstract. The calculation of distances is of fundamental importance in extragalactic astronomy and cosmology. However, no practical implementation for the general case has previously been available. We derive a second-order differential equation for the angular size distance valid not only in all *homogeneous* Friedmann-Lemaître cosmological models, parametrised by λ_0 and Ω_0, but also in *inhomogeneous* 'on-average' Friedmann-Lemaître models, where the inhomogeneity is given by the (in the general case redshift-dependent) parameter η. Since most other cosmological distances can be obtained trivially from the angular size distance, and since the differential equation can be efficiently solved numerically, this offers for the first time a practical method for calculating distances in a large class of cosmological models. We also briefly discuss our numerical implementation, which is publicly available.

Key words: cosmology: theory – methods: numerical – cosmology: distance scale – gravitational lensing

1. Introduction

The determination of distances is one of the most important problems in extragalactic astronomy and cosmology. Distances between two objects X and Y depend on their redshifts z_x and z_y, the Hubble constant H_0, the cosmological constant λ_0, the density parameter Ω_0 and the inhomogeneity parameter η.[1] Usually, smaller distances are determined by the traditional 'distance ladder' technique and larger distances are calculated from the redshift, assuming some cosmological model. Since the redshift is for most purposes exactly measurable, knowledge of or assumptions about two of the factors (a) Hubble constant, (b) other cosmological parameters and (c) 'astronomical distance' (i.e. ultimately tied in to the local distance scale) determines the third. In this paper we discuss distances given the Hubble constant H_0, the redshifts z_x and z_y and the cosmological parameters λ_0, Ω_0 and η. Traditionally, a simple cosmological model is often assumed for ease of calculation, although the distances thus obtained, and results which depend on them, might be false if the assumed cosmological model does not appropriately describe our universe. A general method allows one to look at cosmological models whether or not they are easy-to-calculate special cases and offers the possibility of determining cosmological distances which are important for other astrophysical topics once the correct cosmological model is known.

We stress the fact that the inhomogeneity can be as important as the other cosmological parameters, both in the field of more traditional cosmology and in the case of gravitational lensing, where, e.g. in the case of the time delay between the different images of a multiply imaged source, the inhomogeneity cannot be neglected in a thorough analysis (Kayser & Refsdal 1983). For an example involving a more traditional cosmological test, Perlmutter et al. (1995) (see also Goobar & Perlmutter (1995)) discuss using supernovae with $z \approx 0.25$–0.5 to determine q_0; for z near the top of this range or larger, the uncertainty due to

Send offprint requests to: P. Helbig

* All three appendices are *only* available in electronic form at the CDS via anonymous ftp to `ftp://cdsarc.u-strasbg.fr/pub/A+A/` (130.79.128.5) or via `http://cdsweb.u-strasbg.fr/Abstract.html`

** rkayser@hs.uni-hamburg.de
*** phelbig@hs.uni-hamburg.de
† schramm@tu-harburg.d400.de

[1] When discussing the distance between *two* objects, one can always make a coordinate transformation such that the contribution from the θ and ϕ terms in Eq. (1) vanish. Then one simply needs the redshifts and cosmological parameters in order to determine the distance between them. When discussing the distances between several objects, for example QSOs with α, δ and z as coordinates, this is no longer possible. In many cases, however, suitable geometrical approximations can be made so that the most complicated part of the problem is essentially a determination of a distance between two objects. This point is further discussed in Sect. 5.

our ignorance of η is comparable with the other uncertainties of the method.

The plan of this paper is as follows. In Sect. 2 the basics of Friedmann-Lemaître cosmology are briefly discussed; this also serves to define our terms, which is important since various conflicting notational schemes are in use. (For a more thorough discussion using a similar notation see, e.g., Feige (1992).) Sect. 3 defines the various distances used in cosmology. In Sect. 4 our new differential equation is derived. Similar efforts in the literature are briefly discussed. Sect. 5 briefly describes our numerical implementation and gives the details on how to obtain the source code for use as a 'black box' (which however can be opened) for use in cosmology and extragalactic astronomy. The symmetry properties of the angular size distance, analytic solutions and methods of calculating the volume element are addressed in three appendices.

2. Basic theory

Considering for the moment *homogeneous* Friedmann-Lemaître cosmological models, we can write the familiar Robertson-Walker line element:

$$ds^2 = c^2 \mathrm{d}t^2 - R^2(t) \times \left(\frac{\mathrm{d}\sigma^2}{(1 - k\sigma^2)} + \sigma^2 \mathrm{d}\theta^2 + \sigma^2 \sin^2\theta \mathrm{d}\phi^2 \right), \quad (1)$$

where the symbols are defined as follows (with the corresponding units):

s	4-dimensional interval	[length]
c	speed of light	[velocity]
t	time	[time]
R	scale factor	[length]
σ	radial coordinate	[dimensionless]
k	curvature constant	[dimensionless]
θ	angular coordinate	[dimensionless]
ϕ	angular coordinate	[dimensionless]

The dynamics of the universe is given by the Friedmann equations

$$\dot{R}^2(t) = \frac{8\pi G\rho(t)R^2(t)}{3} + \frac{\Lambda R^2(t)}{3} - kc^2 \quad (2)$$

and

$$\frac{\ddot{R}(t)}{R(t)} = -\frac{4\pi G\rho(t)}{3} + \frac{\Lambda}{3}, \quad (3)$$

where dots denote derivatives with respect to t, G is the gravitational constant, $\rho(t)$ the matter density (this paper assumes negligible pressure), Λ the cosmological constant and the sign of k determines the curvature of the 3-dimensional space.

Introducing the usual parameters

$$\begin{aligned} H &= \frac{\dot{R}}{R} && \text{(Hubble parameter)} \\ \Omega &= \frac{8\pi G\rho}{3H^2} && \text{(density parameter)} \quad (4) \\ \lambda &= \frac{\Lambda}{3H^2} && \text{(normalised cosmological constant)} \end{aligned}$$

(Ω and λ are dimensionless and H has the dimension t^{-1}) we can use Eq. (2) to calculate

$$kc^2 = R^2 H^2 (\Omega + \lambda - 1), \quad (5)$$

so that

$$k = \mathrm{sign}\,(\Omega + \lambda - 1). \quad (6)$$

Since $R > 0$ we can write

$$R = \frac{c}{H} \frac{1}{\sqrt{|\Omega + \lambda - 1|}}; \quad (7)$$

this is the radius of curvature of the 3-dimensional space at time t. For $k = 0$ it is convenient to *define* the scale factor R to be c/H. In the following the index 0 will be used to denote the present value of a given quantity, fixed, as usual, at the time t_0 of observation.[2] The explicit dependence on t will be dropped for brevity. Taking matter conservation into account and using the present-day values, we have

$$\rho R^3 = \rho_0 R_0^3 \quad (8)$$

and so from Eqs. (2), (4), (5) and (8) follows

$$\dot{R}^2 = H_0^2 R_0^2 \left(\frac{\Omega_0 R_0}{R} + \frac{\lambda_0 R^2}{R_0^2} - (\Omega_0 + \lambda_0 - 1) \right). \quad (9)$$

Since below we want to discuss distances as functions of the cosmological redshift z, by making use of the facts that

$$z = \frac{R_0}{R} - 1 \quad (10)$$

and that R_0 is fixed, we can use Eq. (9) to get

$$\mathrm{d}z = \frac{\mathrm{d}z}{\mathrm{d}R}\dot{R}\mathrm{d}t = -H_0(1+z)\sqrt{Q(z)}\,\mathrm{d}t, \quad (11)$$

where

$$Q(z) = \Omega_0(1+z)^3 - (\Omega_0 + \lambda_0 - 1)(1+z)^2 + \lambda_0. \quad (12)$$

Note: *Throughout this paper, the* $\sqrt{}$ *sign should be taken to signify the* positive *solution, except that* sign $\sqrt{Q(z)}$ = sign($\dot{R}$) *always.*

3. Distance measures

3.1. Distances defined by measurement

In a static Euclidean space, one can define a variety of distances according to the method of measurement, which are all equivalent.

[2] Note that this paper is concerned with the calculation of distances from redshift. We are not concerned with a change in redshift with t_0.

3.1.1. Angular size distance

Let us consider at position y two light rays intersecting at x with angle θ. If l is the distance between these light rays, it is meaningful to define the angular size distance D_{xy} as

$$D_{xy} = \frac{l}{\theta}, \tag{13}$$

since an object of projected length l at position y will subtend an angle $\theta = l/D_{xy}$ (for small θ) at distance D_{xy}.

3.1.2. Proper motion distance

The proper motion distance is similar to the angular size distance, except that l is given by vt, where v is the tangential velocity of an object and t the time during which the proper motion is measured.

3.1.3. Parallax distance

Parallax distance is similar to the proper motion distance, except that the angle π is at y instead of x, so that we have

$$D^{\pi}_{xy} = \frac{l}{\pi}. \tag{14}$$

In the canonical case, $l = 1$ AU.

3.1.4. Luminosity distance

Since the apparent luminosity L of an object at distance D is proportional to $1/D^2$, one can define the luminosity distance as

$$D^{\mathrm{L}} = D^{\mathrm{L}}_0 \sqrt{\frac{L_0}{L}}, \tag{15}$$

where L_0 is the luminosity at some fiducial distance D^{L}_0.

3.1.5. Proper distance

By proper distance D^{P} we mean the distance measured with a rigid ruler.

3.1.6. Distance by light travel time

Finally, from the time required for light to traverse a certain distance, one can define a distance D^{c} by

$$D^{\mathrm{c}} = ct \tag{16}$$

where t is the so-called look-back time.

3.2. Cosmological distances

3.2.1. General considerations

In a static Euclidean space, which was used above when defining the distances through a measurement description, these distance measures are of course equivalent. In the general case in cosmology, where the 3-dimensional space need not be flat ($k = 0$) but can be either positively ($k = +1$) or negatively ($k = -1$) curved, and where the 3-dimensional space is scaled by $R(t)$, not only do the distances defined above differ, but also (in the general case) $D_{xy} \neq D_{yx}$. The definitions are still applicable, but different definitions will result in different distances.

In reality, of course, the universe is neither perfectly homogeneous nor perfectly isotropic, as one assumes when deriving Eq. (1). However, as far as the usefulness of the Friedmann equations in determining the global dynamics is concerned, this appears to be a good approximation. (See, for example, Longair (1993) and references therein for an interesting discussion.) The approximation is certainly too crude when using the cosmological model to determine distances as a function of redshift, since the angles involved in such cases can have a scale comparable to that of the inhomogeneities. In this paper, we assume that these inhomogeneities can be sufficiently accurately described by the parameter η, which gives the fraction of homogeneously distributed matter. The rest $(1 - \eta)$ of the matter is distributed clumpily, where the scale of the clumpiness is *by definition* of the same order of magnitude as the angles involved.

For example, a halo of compact MACHO type objects around a galaxy in a distant cluster would be counted among the homogeneously distributed matter if one were concerned with the angular size distance to background galaxies further away, but would be considered clumped on scales such as those important when considering microlensing by the compact objects themselves. Since we don't know exactly how dark matter is distributed, different η values can be examined to get an idea as to how this uncertainty affects whatever it is one is interested in. If one has no selection effects, then, due to flux conservation, the 'average' distance cannot change (Weinberg 1976); η introduces an additional uncertainty when interpreting observations. It is generally not possible to estimate this scatter by comparing the cases $\eta = 0$ and $\eta = 1$, since, depending on the cosmological parameters and the cosmological mass distribution, not all combinations are self-consistent. For instance, if one looks at scales where galaxies are compact objects, and the fraction of Ω_0 due to the galaxies is x, then η *must* be $\leq (1 - x)$.

We further assume that light rays from the object whose distance is to be determined propagate sufficiently far from all clumps. (See Schneider et al. (1992) – hereafter SEF – for a more thorough discussion of this point.) Compared to the perfectly homogeneous and isotropic case, the introduction of the η parameter will influence the angular size and luminosity distances (as well as the proper motion and parallax distances) since these depend on angles between light rays which are influenced by the amount of matter in the beam, but not the proper distance and only negligibly the light travel time. The last two distances are discussed briefly in Sect. 3.2.2 and in Appendix B3 and B6.

Since there is a simple relation between the angular size distance and the luminosity distance (Sect. 3.2.2) which also holds for the inhomogeneous case (see Appendix A), for the general case it suffices to discuss the angular size distance, which we do in Sect. 4.

3.2.2. Relationships between different distances

Without derivation[3] we now discuss some important distance measures, denoting the redshifts of the objects with the indices x and y. Due to symmetry considerations (see Appendix A)

$$D_{yx} = D_{xy}\left(\frac{1+z_y}{1+z_x}\right), \tag{17}$$

where the term in parentheses takes account of, by way of Eq. (10), the expansion of the universe. It is convenient, in keeping with the meaning of angular size distance, to think of the expansion of the universe changing the angle θ in Eq. (13) and not l, if one identifies l as the (projected) size of an object. The angle is defined at the time when the light rays intersect the plane of the observer. Thus D_{xy} with the observer at $x = 0$ defines what one normally thinks of as an angular size distance. On the other hand, D_{xy} and D_{yx} with x in general $\neq 0$ can be important in, for example, gravitational lensing.[4]

Although the angle between the rays (at the source) at the time of reception of the light is important for the luminosity distance, this distance is *not* simply D_{yx}, since in the cosmological case the observed flux is obtained by multiplying the 'non-redshifted flux' by the factor $(1+z_x)^2/(1+z_y)^2$. One factor of $(1+z_x)/(1+z_y)$ occurs because a given wavelength is increased by $(1+z_y)/(1+z_x)$, which reduces the flux correspondingly; an additional factor of $(1+z_x)/(1+z_y)$ occurs because the arrival rate of photons is also decreased. Therefore, since D^{L} is inversely proportional to the square root of the (observed, 'redshifted') flux the luminosity distance is

$$D^{\mathrm{L}}_{xy} = D_{yx}\left(\frac{1+z_y}{1+z_x}\right). \tag{18}$$

From this and Eq. (17) follows the relation

$$D^{\mathrm{L}}_{xy} = D_{xy}\left(\frac{1+z_y}{1+z_x}\right)^2. \tag{19}$$

This means that the surface brightness of a 'standard candle' is $\sim (1+z)^{-4}$, a result independent of the cosmological model parameters, including η.[5] (This result also holds for the inhomogeneous case, since Eq. (17) still holds (see Appendix A) and the additional factor due to the expansion of the universe (given by the term in parentheses in Eq. (18)) is of course present in the inhomogeneous case as well.)

Of course, this applies only to the *bolometric* luminosity. Observing in a finite band introduces two corrections. The so-called K-correction as it is usually defined today (see, e.g., Coleman et al. (1980) or, for an interesting and thorough discussion, Sandage (1995)) takes account of these, both of which come from the fact that the observed wavelength interval is redshifted compared to the corresponding interval on emission. This means that, first, for a flat spectrum, less radiation is observed, because the bandwidth at the observer is $(1+z)$ times larger than at the source. Second, the spectrum need not be flat, in which case additional corrections based on the shape of the spectrum have to be included.[6] Thus,

$$m = M + 5\log\left(\frac{D^{\mathrm{L}}[\mathrm{pc}]}{10\ \mathrm{pc}}\right) + K \tag{20}$$

where m is the apparent magnitude, M the absolute magnitude, D^{L} is the luminosity distance and K is the K-correction as defined in Coleman et al. (1980). Perhaps more convenient is

$$m = M + 5\log D^{\mathrm{L}} + K + N \tag{21}$$

where N is a normalisation term: $N = -5$ for D^{L} in units of 1 pc, $N = 25$ for D^{L} in units of 1 Mpc and $N = x - 5\log h$ for D^{L} in units of the Hubble length[7] c/H_0, where

$$x = 5\log\left(\frac{\text{Hubble length}}{1\ \mathrm{pc}}\right) - 5 \approx 42.384$$

and h is the Hubble constant in units of $100\,\mathrm{km/s/Mpc}$. In practice one has to add terms to correct for various sources of extinction and consider the fact that M is the absolute magnitude of the object when the light was emitted, which of course could be different from the present M of similar objects at negligible redshift.

The light travel time (or lookback time) $t_{xy} = t_x - t_y$ between z_x and z_y (where $t_x = t(z_x) > t_y = t(z_y)$) is given by the integration of the reciprocal of Eq. (11):

$$t_{xy} = \int_{z_y}^{z_x}\left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^{-1}\mathrm{d}z = \frac{1}{H_0}\int_{z_x}^{z_y}\frac{\mathrm{d}z}{(1+z)\sqrt{Q(z)}}, \tag{22}$$

[3] See, e.g., Feige (1992) Berry (1986) or Bondi (1961) for a more general discussion. What we present in the rest of this section is not new, but is important in order to clarify the notation. The results are obvious from the definitions introduced above.

[4] Although not useful in cosmology or extragalactic astronomy, for completeness we mention the fact that the proper motion distance is equivalent to D_{yx} and the parallax distance is equivalent to $D_{yx}/\sqrt{1-k\sigma^2}$.

[5] Thus, a 'surface brightness test' can in principle show that cosmological redshifts are due to the expansion of the universe and not to some other cause. See, e.g., Sect. 6 in Sandage (1995).

[6] Since the observed objects generally evolve with time, and redshifted objects are necessarily observed as they were when the radiation was emitted, some authors include an evolutionary term in the K-correction. Still other authors prefer to absorb one or more of these terms into the definition of the luminosity distance. Our luminosity distance is a bolometric distance based on the geometry and includes the unavoidable dimming due to the redshift. Our K-correction takes account of both effects of a finite bandwidth. Evolutionary effects are considered separately from distances.

[7] For example, as given by our numerical implementation; see Sect. 5

where the minus sign from Eq. (11) is equivalent to the swapped limits of integration on the right-hand side so that the integral gives $t_x - t_y$ instead of $t_y - t_x$, making the light travel time increase (for $\dot{R} > 0$) with z; thus $D^{\mathrm{c}}_{xy} = ct_{xy}$.

Since the proper distance would be the same as D^{c} *were there no expansion*, the former can be calculated by multiplying the integrand in Eq. (22) by $c(1 + z)$. Thus

$$D^{\mathrm{P}}_{xy} = \frac{c}{H_0}\int\limits_{z_x}^{z_y} \frac{\mathrm{d}z}{\sqrt{Q(z)}}. \tag{23}$$

This gives the proper distance at the present time. Since D^{P} scales linearly with the expansion of the universe, the proper distance at some other time can be obtained by dividing Eq. (23) with $(1 + z_i)$, where z_i is the redshift at the corresponding time. For homogeneous ($\eta = 1$) cosmological models,[8] the propagation of light rays is determined by the global geometry, so that there is a simple relation between D^{P} and D and, thus, D^{L}. This is discussed in Sect. B3. Although not 'directly' observable, the proper distance is nevertheless important in cosmological theory, since it is the basic distance of general relativity. Although not useful as a distance, the light travel time is of course important when considering evolutionary effects.

For inhomogeneous models, where this relation between global geometry and local light propagation does not exist, another approach must be used, which takes account of both the expansion of the universe as well as the local propagation of light, when calculating angle-defined distances such as the angular size distance.

4. The general differential equation for the angular size distance

In a series of papers Zeldovich (1964), Dashevskii and Zeldovich (1965) and Dashevskii and Slysh (1966) developed a general differential equation for the distance between two light rays on the boundary of a small light cone propagating far away from all clumps of matter in an inhomogeneous universe:

$$\ddot{l} = -4\pi G\eta\rho\, l + \frac{\dot{R}}{R}\,\dot{l} \tag{24}$$

where η and ρ are functions of the time t (*not* the lookback time of Eq. 22). The first term can be interpreted as Ricci focusing due to the matter inside the light cone, and the second term is due to the expansion of space during the light propagation. We now have to transform this *time* dependent differential equation into a *redshift* dependent differential equation. From Eq. (11) we obtain[9]

$$\mathrm{d}t = -\left(H_0(1 + z)\sqrt{Q}\right)^{-1}\,\mathrm{d}z, \tag{25}$$

and thus

$$\frac{\mathrm{d}l}{\mathrm{d}t} = -H_0(1 + z)\sqrt{Q}\,\frac{\mathrm{d}l}{\mathrm{d}z} \tag{26}$$

and

$$\frac{\mathrm{d}^2 l}{\mathrm{d}t^2} = H_0^2(1 + z)\sqrt{Q}\frac{\mathrm{d}}{\mathrm{d}z}\left((1 + z)\sqrt{Q}\,\frac{\mathrm{d}l}{\mathrm{d}z}\right) \tag{27}$$

$$= H_0^2\left(\left((1 + z)Q + (1 + z)^2\frac{1}{2}\frac{\mathrm{d}Q}{\mathrm{d}z}\right)\frac{\mathrm{d}l}{\mathrm{d}z} + (1 + z)^2 Q\,\frac{\mathrm{d}^2 l}{\mathrm{d}z^2}\right). \tag{28}$$

Furthermore, since $R = R_0/(1 + z)$ (Eq. (10)), we obtain, using Eq. (25),

$$\frac{\mathrm{d}R}{\mathrm{d}t} = -H_0(1 + z)\sqrt{Q}\,\frac{\mathrm{d}R}{\mathrm{d}z}. \tag{29}$$

From the definition of Ω (Eq. (4)) and matter conservation (Eq. (8)) we obtain

$$4\pi G\rho = \frac{3}{2}\,H_0^2\Omega_0(1 + z)^3. \tag{30}$$

If we now insert Eqs. (26), (28), (29) and (30) into Eq. (24), sort the terms appropriately and cancel H_0^2, which appears in all terms, we obtain

$$Q\,l'' + \left(\frac{2Q}{1 + z} + \frac{1}{2}\,Q'\right)\,l' + \frac{3}{2}\,\eta\,\Omega_0(1 + z)\,l = 0\quad, \tag{31}$$

where a prime denotes a derivative with respect to redshift and from Eq. (12) follows

$$Q'(z) = 3\Omega_0(1 + z)^2 - 2(\Omega_0 + \lambda_0 - 1)(1 + z)\quad. \tag{32}$$

From the definition of the angular size distance (Eq. (13)) it is obvious that it follows the same differential equation as l:

$$\boxed{Q\,D'' + \left(\frac{2Q}{1 + z} + \frac{1}{2}\,Q'\right)\,D' + \frac{3}{2}\,\eta\,\Omega_0(1 + z)\,D = 0} \tag{33}$$

with special boundary conditions at the redshift z_x where the two considered light rays intersect. The first boundary condition is trivially

$$D = 0\quad\text{for}\quad z = z_x\quad, \tag{34}$$

[8] This includes *empty* models ($\Omega_0 = 0$); although η has no meaning here, the same arguments apply.

[9] This transformation causes problems if the integration interval contains a point where $\dot{R} = 0$ and thus $\sqrt{Q}$ changes sign. In this case the integration interval (t_x, t_y) has to be transformed into *two* integration intervals, namely $(z_x, z_{\max})$ and $(z_{\max}, z_y)$, where $z_{\max}$ is the redshift at $\dot{R} = 0$, with the boundary conditions for the second integration interval chosen appropriately.

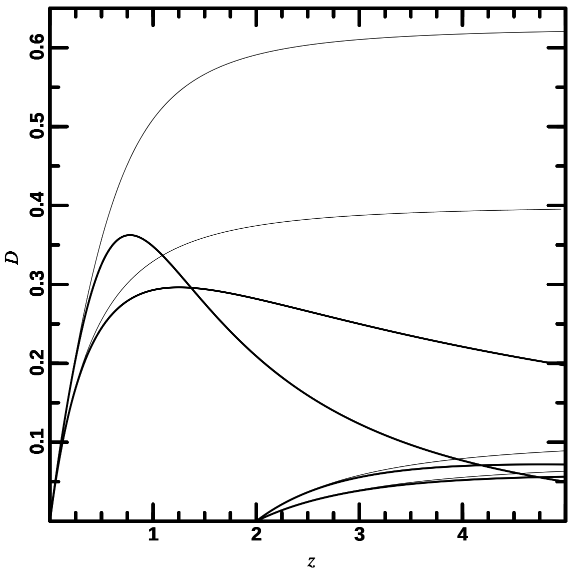

Fig. 1. The angular size distance from the observer ($z_1 = 0$) and from $z_1 = 2$ (lower right) as a function of the redshift z_2 for different cosmological models. Thin curves are for $\eta = 0$, thick for $\eta = 1$. The upper curves near $z = 0$ ($z = 2$ at lower right) are for $\lambda_0 = 2$, the lower for $\lambda_0 = 0$. $\Omega_0 = 1$ for all curves. The angular size distance D is given in units of c/H_0

and the second boundary condition follows from the Euclidean approximation for small distances, i.e.

$$\left.\frac{\mathrm{d}D}{\mathrm{d}t}\right|_{z=z_x} = c\,\mathrm{sign}(t_x - t_y), \qquad (35)$$

hence

$$D' = \frac{c}{H_0}\frac{1}{(1+z_x)\sqrt{Q(z_x)}}\,\mathrm{sign}(t_y - t_x) \quad \text{for} \quad z = z_x, \qquad (36)$$

where the sign has been chosen such that D is always > 0 locally. We denote these special solutions of Eq. (33) with $D_x(z)$, and, following the definition (Eq. (13)), the angular size distance of an object at redshift z_y is then given as

$$D_{xy} = D_x(z_y) \quad . \qquad (37)$$

Fig. 1 shows the influence of z, η and λ on the angular size distance, calculated using Eq. (33) with our numerical implementation.

For completeness we note that after the original derivation by Kayser (1985) an equivalent equation was derived by Linder (1988) which, however, is difficult to implement due to the cumbersome notation.

Special mention must be made of the so-called bounce models, which expand from a finite R after having contracted from $R = \infty$. (See, e.g., Feige (1992).) A glance at Eq. (10) shows that in these cosmological models there must be *four* distances for an (ordered) pair of redshifts. If we denote the distances by D_{12}, D_{14}, D_{34} and D_{32}, where 1(2) und 3(4) refer to $z_1(z_2)$ during the expanding (contracting) phase, then symmetry considerations dictate that $D_{12} = D_{34}$ and $D_{14} = D_{32}$ *as long as the dependence of η on z is the same during both phases*. In this case, there are *two* independent distances per (ordered) pair of redshifts. If this is not the case, the degeneracy is no longer present and there are *four* independent distances per (ordered) pair of redshifts.

5. Numerics and practical considerations

For the actual numerical integration of the differential equation, we have found the Bulirsch-Stoer method to be both faster and more exact than other methods such as Runge-Kutta. However, the conventional method of rational function extrapolation is rather unstable in this particular case; fortunately, using polynomial extrapolation solves the problem. Although programming the integration is rather straightforward in theory, in numerical practice considerable effort is needed to determine combinations of free parameters which work for all cases. We have tested the finished programme intensively and extensively, for example by comparing the results of calculations for $\eta = 1$ (the value of η plays no special role in the integration of the differential equation) with those in Refsdal et al. (1967) or given by the method of elliptical integrals as outlined in Feige (1992) and have used it in Kayser (1995), Helbig (1996) and Helbig & Kayser (1996). For a general discussion of various methods of integrating second-order differential equations, see Press et al. (1992). Those interested in technical details can read the comments in our source code and the accompanying user's guide.

Since H_0, in contrast to the other cosmological parameters, merely inversely scales the angular size distance, our routine actually calculates the angular size distance in units of c/H_0. This *dimensionless* quantity must be multiplied by c/H_0 (in whatever units are convenient) in order to obtain the actual distance. Other than reducing numerical overhead, this allows all distances to be calculated modulo c/H_0, which is convenient for expressing quantities in an H_0-independent manner. In practice, H_0 cancels out of many calculations anyway.

Apart from auxiliary routines which the user does not have to be concerned with, our implementation consists of four `FORTRAN77` subroutines. The first, `INICOS`, calculates z-independent quantities used by the other routines, some of which are returned to the calling programme. `ANGSIZ` calculates the angular size distance. Normally, η is used as a z-independent cosmological parameter, on an equal footing with λ_0 and Ω_0. If desired, however, the user can let `INICOS` know that a variable (that is, z-dependent) η is to be used; this is given by the function `VARETA`. We supply an example; the user can modify this to suit her needs. In particular, many different dependencies of η on z can be included, and a decision made in the calling programme about which one to use. This feature is also included in our example. `ANGSIZ` returns only the distance D_{12}; if one is interested in the other distances in the bounce models, our subroutine `BNGSIZ` returns all of these (though internally calculating only the independent distances, of course, depending on the dependence of η on z).

Due to the fact that not everyone has a Fortran90 compiler at his disposal, we have coded the routines in `FORTRAN77`. Only standard `FORTRAN77` features are used, and thus the routines should be able to be used on all platforms which support `FORTRAN77`. Since standard `FORTRAN77` is a subset of Fortran90, the routines can be used without change in Fortran90 as well.

With the exception of D^{c}, all distance measures can be easily transformed into one another. Thus, it suffices to calculate the angular size distance for a given case.[10]

When discussing the distance between two objects other than the observer, rather than between the observer and one object, in many cases one of two simplifying assumptions can be made:

$D(\Delta z) \ll D(\beta)$ In this case, the *proper distance* D^{P} *at the time of emission* between the two objects is $\beta D_{0x} \approx \beta D_{0y}$, where $\beta \ll 1$ is the angle in radians between the two objects on the sky.

$D(\beta) \ll D(\Delta z)$ In this case, the *angular size distance* between the two objects is D_{xy}.

$D(\Delta z)$ ($D(\beta)$) refers to the distance due to Δz (β) when setting β (Δz) equal to zero. In the first case, where the two objects are practically at the same redshift, one uses the angular size distance to this redshift to transform the observed difference in angular position on the sky into the *proper distance* between the two objects at the time of emission. This follows directly from the definition of the angular size distance. Since the distance between the objects is much less than the distance from the observer to the objects, the differently defined distances between the objects are for practical purposes degenerate. A practical example of this case would be the distance between individual galaxies in a galaxy cluster at large redshift. Naturally, one should use *one* redshift, say, of the cluster centre; the individual redshifts will in most cases be overlaid with the doppler redshift due to the velocity dispersion of the cluster, so the difference in *cosmological* redshifts is negligible. (Of course, the *present* distance would be a factor of $(1 + z)$ larger, due to the expansion of the universe, were the objects comoving and not, as in a galaxy cluster, bound.) In the second case, which is typical of gravitational lensing, the angles on the sky between, for example, source and lens, are small enough to be neglected, so that the angular size distance between the objects is determined by the difference in redshift. If neither of these assumptions can be made, any sort of distance between the two objects is probably of no practical interest. (Of course, there is the trivial case where the redshifts are all $\ll 1$ in which case one can simply use α, δ and cz/H_0 as normal spherical coordinates.)

[10] The proper distance, which is η-independent, can be calculated from the angular size distance assuming $\eta = 1$, by making use of the simple relation between proper distance and angular size distance in this case. The result holds of course for all values of η.

6. Summary

After discussing cosmological distances with an emphasis on practical distance measures for general use in cosmology and extragalactic astronomy, we have obtained a new differential equation, which gives the angular size distance for a class of 'on average' Friedmann-Lemaître cosmological models, that is, models described not only by λ_0 and Ω_0 but also by $\eta(z)$, which describes the clumpiness of the distribution of matter. We have also developed a practical numerical method of solving this equation, which we have made publicly available. Since the equation is valid for *all* cases, this offers for the first time an efficient means of calculating distances in a large class of cosmological models.

The numerical implementation (in `FORTRAN77`), user's guide and a copy of the latest version of this paper (including appendices) can be obtained from either of the following URLs:

http://www.hs.uni-hamburg.de/english/persons/helbig/Research/Publications/Info/angsiz.html

ftp://ftp.uni-hamburg.de/pub/unihh/astro/angsiz.tar.gz

Acknowledgements. It is a pleasure to thank O. Czoske, S. Refsdal and A. Smette for helpful discussions and comments on the manuscript.

References

Berry M. V., 1986, Cosmology and Gravitation. Adam Hilger, Bristol
Bondi H., 1961, Cosmology. Cambridge University Press, Cambridge
Coleman G. D., Wu C.-C., Weedman D. W., 1980, ApJS 43, 393
Dashevskii V. M. , Slysh V. J., 1966, Sov. Astr. 9, 671
Dashevskii V. M. , Zeldovich Y. B., 1965, Sov. Astr. 8, 854
Feige B., 1992, Astr. Nachr. 313, 139
Goobar A., Perlmutter S., 1995, ApJ 450, 14
Helbig P., 1996, Predicted lens redshifts and magnitudes. In: Kochanek C. S., Hewitt J. (eds.) Astrophysical Applications of Gravitational Lensing (IAU Symposium 173). Kluwer, Dordrecht
Helbig P., Kayser R., 1996, A&A 308, 359
Kayser R., Refsdal S., 1983, A&A 128, 156
Kayser R., 1985, doctoral thesis, University of Hamburg
Kayser R., 1995, A&A 294, L21
Linder E. V., 1988, A&A 206, 190
Longair M., 1958, QJRAS 34, 157
Perlmutter S., Pennypacker C. R., Goldhaber G., et al., 1995, ApJ 440, L41
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in FORTRAN. Cambridge University Press, Cambridge
Refsdal S., Stabell R., de Lange F. G., 1967, Mem. R. Astron. Soc. 71, 143
Sandage A., 1995, Practical Cosmology: Inventing the Past. In: Binggeli B., Buser R. (eds.) The Deep Universe. Springer, Berlin
Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer-Verlag, Heidelberg
Weinberg S., 1976, ApJ 208, L1
Zeldovich Y. B. , 1964, Sov. Astr. 8, 13

This article was processed by the author using Springer-Verlag LaTeX A&A style file *L-AA* version 3.

A&A manuscript no.
(will be inserted by hand later)

Your thesaurus codes are:
02 (12.03.4; 03.13.4 12.04.3; 12.03.2; 12.07.1)

ASTRONOMY
AND
ASTROPHYSICS
2.3.1997

A general and practical method for calculating cosmological distances

Rainer Kayser[*1], **Phillip Helbig**[**1] **and Thomas Schramm**[***1,2]

[1] Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg-Bergedorf, Germany
[2] Rechenzentrum, Technische Universität Hamburg-Harburg, Denickestraße 17, D-21071 Hamburg-Harburg, Germany

received ; accepted

A. Symmetry: The relation between D_{xy} and D_{yx}

The proof in this appendix follows closely the proof presented in Kayser (1985). For completeness we note that after the original derivation by Kayser (1985) an equivalent equation was derived by Linder (1988). We rewrite the differential equation, Eq. (33), for the angular size distance in the normal form:

$$a_2 D''(z) + a_1(z)\, D'(z) + a_0(z)\, D(z) = 0 \tag{A1}$$

with the coefficient functions

$$a_2(z) = Q(z) \tag{A2}$$

$$a_1(z) = \frac{2Q(z)}{1+z} + \frac{1}{2}\, Q'(z) \tag{A3}$$

$$a_0(z) = \frac{3}{2}\,\eta\,\Omega_0(1+z) \quad . \tag{A4}$$

Now let $D^{(1)}$ and $D^{(2)}$ be two solutions of Eq. (A1) which build a fundamental system, i.e. the Wronskian for these two solutions does not vanish:

$$W(z) = \begin{vmatrix} D^{(1)} & D^{(2)} \\ \frac{dD^{(1)}}{dz} & \frac{dD^{(2)}}{dz} \end{vmatrix} \neq 0 \quad \forall z \quad . \tag{A5}$$

Every solution D_i of Eq. (A1) can then be written as a linear combination of $D^{(1)}$ and $D^{(2)}$:

$$D_i = \alpha_i\, D^{(1)} + \beta_i\, D^{(2)}; \quad with\ \alpha_i,\, \beta_i = const \quad . \tag{A6}$$

The angular size distances are special solutions D_x of Eq. (A1) fulfilling the following boundary conditions:

$$D_x = 0 \quad for \quad z = z_x \tag{A7}$$

and

$$\frac{dD_x}{dz} = b(z_x) \quad for \quad z = z_x \tag{A8}$$

with

$$b(z_x) = \frac{c}{H_0}\left((1+z_x)\sqrt{Q(z_x)}\right)^{-1} sign(t_y - t_x) \quad , \tag{A9}$$

compare Eq. (36). From Eq. (A6) we obtain

$$0 = \alpha_i\, D^{(1)}(z_x) + \beta_i\, D^{(2)}(z_x) \tag{A10}$$

and

$$b(z_x) = \alpha_i\, \left.\frac{dD^{(1)}}{dz}\right|_{z=z_x} + \beta_i\, \left.\frac{dD^{(2)}}{dz}\right|_{z=z_x} \quad . \tag{A11}$$

These equations can easily be solved for α_i and β_i:

$$\alpha_i = \beta_i\, \frac{D^{(2)}(z_x)}{D^{(1)}(z_x)} \tag{A12}$$

$$\beta_i = \frac{b(z_x)\, D^{(1)}(z_x)}{W(z_x)} \tag{A13}$$

and inserting α_i and β_i back into Eq. (A6) we obtain for the special solutions D_x:

$$D_x(z) = \frac{b(z_x)}{W(z_x)}\Big(D^{(1)}(z_x)\, D^{(2)}(z) - D^{(1)}(z)\, D^{(2)}(z_x)\Big) \quad . \tag{A14}$$

If we now consider a second special solution D_y we find the relation

$$\frac{D_x(z_y)}{D_y(z_x)} = -\frac{b(z_x)}{b(z_y)}\,\frac{W(z_y)}{W(z_x)} \quad . \tag{A15}$$

The Wronskians can be calculated using Liouville's formula:

$$W(z) = W(z_0)\, \exp \int_z^{z_0} a_2(z)\, dz \quad , \tag{A16}$$

Send offprint requests to: P. Helbig
* rkayser@hs.uni-hamburg.de
** phelbig@hs.uni-hamburg.de
*** schramm@tu-harburg.d400.de

where z_0 is arbitrary. Thus

$$\frac{D_x(z_y)}{D_y(z_x)} = -\frac{b(z_x)}{b(z_y)} \exp \int_{z_y}^{z_x} \frac{a_1(z)}{a_2(z)}\, dz \qquad \text{(A17)}$$

and after inserting a_0, a_1 and a_2 from Eqs. (A2), (A3) and (A4) as well as $b(z_x)$ and $b(z_y)$ from Eq. (A9) and integration we finally obtain for the angular size distances (cf. Eq. (37)) the relation

$$\frac{D_{xy}}{D_{yx}} = \frac{1+z_x}{1+z_y} \quad . \qquad \text{(A18)}$$

B. Special cases

For certain special cases the differential equation can be simplified and sometimes analytically solved.

B.1. $\Omega_0 = 0$

A glance at Eq. (33) shows that for $\Omega_0 = 0$ the third term on the left hand side of Eq. (33) vanishes; one thus has a first order differential equation for D'. (Of course η has no meaning for $\Omega_0 = 0$.) Due to the fact that a vanishing Ω_0 also simplifies $Q(z)$, it is possible to calculate the angular size distance analytically. Since in this case the angular size distance is determined exclusively by global effects, one can use an approach based on global geometry.[1] Depending on the value of λ_0, one can use the following expression to calculate $\chi_{xy} = \chi(y) - \chi(x)$ (Feige 1992)

$$\chi(z) = \begin{cases} arccosh(\psi) & for \quad \lambda_0 < 0 \\ \ln(1+z) & for \quad \lambda_0 = 0 \\ arcsinh(\psi) & for \quad 0 < \lambda_0 < 1 \\ z & for \quad \lambda_0 = 1 \\ \arcsin(\psi) & for \quad \lambda_0 > 1 \end{cases} , \qquad \text{(B1)}$$

where $\psi := (1+z)\sqrt{\frac{|1-\lambda_0|}{|\lambda_0|}}$. The relationship between χ and the angular size distance D is

$$D_{xy} = \frac{R_0}{(1+z_y)} \begin{cases} \sinh\chi & for \quad k = -1 \\ \chi & for \quad k = 0 \\ \sin\chi & for \quad k = +1 \end{cases} , \qquad \text{(B2)}$$

as discussed below in Sect B.3.

[1] See the discussion in Sect. B.3.

B.2. $\eta = 0$

In the case $\eta = 0$ the third term on the left hand side of Eq. (33) vanishes; one thus has a first order differential equation for D'. Assuming $D' \neq 0$, Eq. (33) can be written as

$$\frac{D''}{D'} = -\frac{2}{1+z} - \frac{1}{2}\frac{Q'(z)}{Q(z)} \quad . \qquad \text{(B3)}$$

This equation can be solved in two steps. For D' we obtain

$$D' = \frac{c_1}{\sqrt{Q(z)}\,(1+z)^2} \qquad \text{(B4)}$$

and consequently for D

$$D = \int \frac{c_1}{\sqrt{Q(z)}\,(1+z)^2} + c_2 \quad . \qquad \text{(B5)}$$

The constants c_1, c_2 are determined by the appropriate boundary conditions (Eqs. (34) and (35)). We then find the solution (see also Schneider et al. (1992) – hereafter SEF – for an equivalent discussion with $\lambda_0 = 0$)

$$D_{xy} = \frac{c}{H_0}(1+z_x)(\omega(z_y) - \omega(z_x)), \qquad \text{(B6)}$$

where

$$\omega(z) = \int_0^z \frac{dz'}{(1+z')^2\sqrt{(1+z')^2(\Omega_0 z' + 1 - \lambda_0) + \lambda_0}}, \qquad \text{(B7)}$$

or, perhaps more convenient,

$$D_{xy} = \frac{c}{H_0}\,(1+z_x)\int_{z_x}^{z_y} \frac{dz}{(1+z)^2\sqrt{Q(z)}} \quad . \qquad \text{(B8)}$$

For $\lambda_0 = 0$ there is an analytic solution (see Sect. B.4).

B.3. $\eta = 1$

The case $\eta = 1$ has all matter distributed homogeneously. Due to homogeneity, the matter *locally* affecting the propagation of light is known when the *global* geometry is known, so that the 'classical' approach of relating global geometry to observable relations is a better approach than using (the simplified form of) Eq. (33). This approach offers an analytic solution. Here, we simply sketch the most important points; the interested reader can refer to Feige (1992) for a good description of this method.

The angular size distance in this case is

$$D_{xy} = R_y\sigma_{xy} = \frac{R_0\sigma_{xy}}{(1+z_y)}, \qquad \text{(B9)}$$

where σ is the radial coordinate in the Robertson-Walker metric (cf. Eq. (1)) and thus

$$D_{yx} = \frac{R_0\sigma_{xy}}{1+z_x} = D_{xy}\left(\frac{1+z_y}{1+z_x}\right), \qquad \text{(B10)}$$

since this angle is inversely proportional to R for constant σ and physical size. (The value of R at the time the light rays defining the angle intersect is important.)

Since σ is given by

$$\sigma = F(\chi) = \begin{cases} \sinh\chi & for\ k\ =\ -1 \\ \chi & for\ k\ =\ \ 0 \\ \sin\chi & for\ k\ =\ +1 \end{cases}, \tag{B11}$$

an expression for $\chi(z)$ is sufficient for calculating the angular size distance D (and of course the luminosity distance D^L (via Eq. (19)) and the 'coordinate distance' σ (via Eq. (B11)). In general, $\sigma_{xy} \neq \sigma_y - \sigma_x$; however, $\chi_{xy} = \chi_y - \chi_x$, so that

$$\sigma_{xy} = F(\chi_{xy}) \tag{B12}$$

where F is given by Eq. (B11). Using Eq. (23) one can calculate

$$\chi_{xy} = \frac{D^P}{R_0} = \frac{c}{H_0 R_0}\int\limits_{z_x}^{z_y} \frac{dz}{\sqrt{Q(z)}}. \tag{B13}$$

In the general case, Eq. (23) can be solved by elliptic integrals, as explained in Feige (1992). For the cases $\lambda_0 = 0$ and $\Omega_0 = 0$ the formulae using elliptic integrals break down; in these cases, easier analytic formulae, which fortunately exist, can be used. The case $\Omega_0 = 0$ has been discussed above. The case $\lambda_0 = 0$ will be discussed below. *Again, we stress that the differential equation derived in Sect. 4 is completely general and can be used in* all *cases.*

B.4. $\lambda_0 = 0$

For $\lambda_0 = 0$, there is in general no simpler solution. This case has been discussed by Dyer and Roeder for $\eta = 0$ (1972) and for general η values (1973). They point out the interesting result that the maximum in the angular size distance from $z_1 = 0$ to z_2 increases monotonically from 1.25 to ∞ as η decreases from 1 to 0. See also the discussion (with a differing notation!) in Sect. 4.5.3 in SEF. However, some solutions exist for special values of Ω_0 and η. The case $\Omega_0 = 0$ has been discussed in Sect. B.1 above; the value of η is of course irrelevant in this case. With the exception of Sect B.4.3 below, in the following we simply quote results from SEF in our notation.

B.4.1. $\lambda_0 = 0$ and $\eta = 0$

As discussed above, for $\eta = 0$ Eq. (33) is effectively a first order equation for D'. For $\lambda_0 = 0$ $Q(z)$ is sufficiently simplified to allow an analytic solution. Recalling Eq. (B6),

$$D_{xy} = \frac{c}{H_0}(1+z_x)(\omega(z_y) - \omega(z_x)),$$

Eq. (B7) simplifies to

$$\omega(z) = \int\limits_0^z \frac{dz'}{(1+z')^3\sqrt{\Omega_0 z' + 1}}, \tag{B14}$$

which has the solution:

$$\omega(z) = \begin{cases} \frac{3\Omega_0^2}{4(\Omega_0-1)^{\frac{5}{2}}} \arctan(\psi) + \frac{3}{4(\Omega_0-1)^2} \quad \times \\ \left(\frac{\left(\Omega_0 z + \frac{5\Omega_0}{3} - \frac{2}{3}\right)\sqrt{\Omega_0 z+1}}{(1+z)^2} - \frac{5\Omega_0}{3} + \frac{2}{3}\right) & (I) \\ \frac{2}{5}\left(1 - (\Omega_0 z + 1)^{-\frac{5}{2}}\right) & (II) \\ \frac{3\Omega_0^2}{4(1-\Omega_0)^{\frac{5}{2}}}\, arctanh(\psi) + \frac{3}{4(1-\Omega_0)^2} \quad \times \\ \left(\frac{\left(\Omega_0 z + \frac{5\Omega_0}{3} - \frac{2}{3}\right)\sqrt{\Omega_0 z+1}}{(1+z)^2} - \frac{5\Omega_0}{3} + \frac{2}{3}\right) & (III) \end{cases} \tag{B15}$$

with

$$\psi = \begin{cases} \left(\frac{\sqrt{\Omega_0 - 1}(1+\sqrt{\Omega_0 z+1})}{\Omega_0 - 1 + \sqrt{\Omega_0 z+1}}\right) & (I) \\ \left(\frac{\sqrt{1-\Omega_0}(1+\sqrt{\Omega_0 z+1})}{\Omega_0 - 1 + \sqrt{\Omega_0 z+1}}\right) & (III) \end{cases} \tag{B16}$$

and

$$\begin{array}{ll} \text{case I:} & \Omega_0 > 1 \\ \text{case II:} & \Omega_0 = 1\ . \\ \text{case III:} & 0 < \Omega_0 < 1 \end{array} \tag{B17}$$

Note that in SEF, the text at the top of page 137 is unclear—the expression in parentheses in the denominator of the first term $(\Omega - 1)$ for the $\Omega > 1$ case has to be replaced with $(1-\Omega)$ as well for $\Omega < 1$. Note also that $\Omega \equiv \Omega_0$ and that after page 131 $\lambda_0 = 0$ is always assumed.

B.4.2. $\lambda_0 = 0$ and $\Omega_0 = 1$

For $\Omega_0 = 1$ and $\lambda_0 = 0$ (the Einstein-de Sitter model) we have the solution

$$D_{xy} = \frac{c}{H_0}\frac{1}{2\beta}\left(\frac{(1+z_y)^{\beta-\frac{5}{4}}}{(1+z_x)^{\beta+\frac{1}{4}}} - \frac{(1+z_x)^{\beta-\frac{1}{4}}}{(1+z_y)^{\beta+\frac{5}{4}}}\right), \tag{B18}$$

where

$$\beta := \frac{1}{4}\sqrt{25 - 24\eta}. \tag{B19}$$

B.4.3. $\lambda_0 = 0$ and $\eta = 1$

For $\eta = 1$ the special case of the expression for $\chi(z)$ for $\lambda_0 = 0$ is (Feige 1992)

$$\chi(z) = -2\begin{cases} \arcsin\left(\sqrt{\frac{\Omega_0 - 1}{\Omega_0(1+z)}}\right) & (\Omega_0 > 1) \\ \sqrt{\frac{1}{1+z}} & (\Omega_0 = 1) \\ arcsinh\left(\sqrt{\frac{1-\Omega_0}{\Omega_0(1+z)}}\right) & (0 < \Omega_0 < 1) \end{cases}, \tag{B20}$$

where $\chi_{xy} = \chi(y) - \chi(x)$. (It is obvious that in the case $\lambda_0 = \Omega_0 = 0$ Eq. (B1) should be used.) From this, it is possible to obtain a general expression for the angular size distance (see, e.g., SEF):

$$D_{xy} = \frac{c}{H_0}\frac{2}{\Omega_0^2}(1+z_x)\left(R_1(z_y)R_2(z_x) - R_1(z_x)R_2(z_y)\right), \quad \text{(B21)}$$

with

$$R_1(z) = \frac{\Omega_0 z - \Omega_0 + 2}{(1+z)^2} \quad \text{(B22)}$$

and

$$R_2(z) = \frac{\sqrt{\Omega_0 z + 1}}{(1+z)^2}. \quad \text{(B23)}$$

For $z_x = 0$ and $z_y = z$ one gets for the angular size distance

$$D(z) = \frac{c}{H_0}\frac{2}{\Omega_0^2(1+z)^2} \times \left(\Omega_0 z - (2-\Omega_0)\left(\sqrt{\Omega_0 z + 1} - 1\right)\right). \quad \text{(B24)}$$

valid for $\Omega_0 > 0$. For $\Omega_0 = 0$ one obtains

$$D = \frac{c}{H_0}\frac{z\left(1+\frac{z}{2}\right)}{\left(1+z\right)^2} \quad \text{(B25)}$$

(Multiplying Eq. (B24) or Eq. (B25) with $(1+z)/R_0$ results in the respective expression for σ as a function of redshift as first derived by Mattig (1958). See also Sandage (1995), Sect. 1.6.3). In this case, the volume element given by Eq. (C4) reduces to

$$dV = 16\pi R_0^3 \frac{(\Omega_0 z - (2-\Omega_0)(\sqrt{\Omega_0 z + 1} - 1)^2)}{\Omega_0^4 (1+z)^3 \sqrt{\Omega_0 z + 1}} \quad \text{(B26)}$$

Of course, for the physical, as opposed to comoving, density, an *additional* factor of $(1+z)^3$ must be added to the denominator.

B.4.4. $\lambda_0 = 0$ and $\eta = \frac{2}{3}$

For $\eta = \frac{2}{3}$ and $\lambda_0 = 0$ there is also an analytic solution (see SEF):

$$D_{xy} = \frac{c}{H_0}\frac{2}{3\Omega_0^2}(1+z_x) \times \left(R_1(z_x)R_2(z_y) - R_2(z_x)R_1(z_y)\right), \quad \text{(B27)}$$

with

$$R_1(z) = \frac{1}{(1+z)^2} \quad \text{(B28)}$$

and

$$R_2(z) = \frac{\sqrt{\Omega_0 z + 1}(\Omega_0 z + 3\Omega_0 - 2)}{(1+z)^2}. \quad \text{(B29)}$$

B.5. Other cases

We can offer no proof that no other easier solutions, either reducing Eq. (33) to a more easily (numerically) integrated form or even to an analytic solution, exist. This is left as an exercise to the interested reader. The authors are of course interested in such solutions and are willing to verify them. As far as we know, Eq. (33) must be used except in the special cases mentioned in this appendix.

B.6. Light travel time

Feige (1992) not only gives the distance but also the light travel time by means of elliptic integrals. As for the distance, and for the same reasons, simple analytic formulae can and must be used for the special cases $\Omega_0 = 0$ and $\lambda_0 = 0$. For $k = 0$, an analytic expression for the light travel time exists, although the elliptic integrals can also be used in this case. For completeness, we give these special cases here for the light travel time $t_{xy} = t_x - t_y$.

For $\Omega_0 = 0$ we have:

$$t(z) = \frac{1}{H_0\sqrt{|\lambda_0|}} \begin{cases} \arcsin(\psi) & \lambda_0 < 0 \\ \frac{\sqrt{|\lambda_0|}}{(1+z)} & \lambda_0 = 0 \\ arcsinh(\psi) & 0 < \lambda_0 < 1 \\ -\sqrt{|\lambda_0|}\ln(1+z) & \lambda_0 = 1 \\ arccosh(\psi) & \lambda_0 > 1 \end{cases}, \quad \text{(B30)}$$

where $\psi := \frac{1}{(1+z)}\sqrt{\frac{|\lambda_0|}{|1-\lambda_0|}}$

For $\lambda_0 = 0$ we have:

$$t(z) = A \times \begin{cases} \frac{\sqrt{(\Omega_0 z+1)(\Omega_0-1)}}{\Omega_0(1+z)} - \arcsin\left(\sqrt{\frac{\Omega_0-1}{\Omega_0(1+z)}}\right) & \Omega_0 > 1 \\ -\frac{\sqrt{\Omega_0-1}^3}{\Omega_0}\frac{2}{3}\left(\sqrt{\frac{1}{1+z}}\right)^3 & \Omega_0 = 1 \\ arcsinh\left(\sqrt{\frac{1-\Omega_0}{\Omega_0(1+z)}}\right) - \frac{\sqrt{(\Omega_0 z+1)(1-\Omega_0)}}{\Omega_0(1+z)} & 0 < \Omega_0 < 1 \end{cases}, \quad \text{(B31)}$$

where

$$A = -\frac{\Omega_0}{H_0(\sqrt{|\Omega_0 - 1|})^3}$$

(For $\Omega_0 = 0$ the appropriate case from Eq. (B30) must be used.)

For $k = 0$ we have:

$$t(z) = \frac{2}{3H_0} \times \begin{cases} \frac{1}{\sqrt{\Omega_0 - 1}} \arcsin(\psi) & \Omega_0 > 1 \\ \left(\sqrt{\frac{1}{1+z}}\right)^3 & \Omega_0 = 1 \\ \frac{1}{\sqrt{1-\Omega_0}} arcsinh(\psi) & 0 < \Omega_0 < 1 \end{cases} , \quad \text{(B32)}$$

where

$$\psi = \begin{cases} \sqrt{\frac{\Omega_0 - 1}{\Omega_0 (1+z)^3}} & \Omega_0 > 1 \\ \sqrt{\frac{1-\Omega_0}{\Omega_0 (1+z)^3}} & 0 < \Omega_0 < 1 \end{cases} . \quad \text{(B33)}$$

(For $\Omega_0 = 0$ the appropriate case from Eq. (B30) must be used.)

C. Volume element

Sometimes the distance is only a means of calculating the volume element at a given redshift. In the static Euclidean case the volume element is of course

$$dV = 4\pi r^2 dr. \quad \text{(C1)}$$

In the cosmological case, the volume element is, with $r = R_0\sigma$,

$$dV = 4\pi r^2 dD^P = 4\pi r^2 \frac{c}{H_0} \frac{dz}{\sqrt{Q(z)}}. \quad \text{(C2)}$$

$R_0\sigma_y$ is, for $\eta = 1$, simply $D_{y0} = (1+y)D_{0y}$; see Sect B.3. Thus, the distance D_{y0} is all that is needed to calculate the volume; this first can be calculated by Eq. (33) with $\eta = 1$ (This applies even if one would calculate *distances* with another value of η since the volume element is a quantity related to the global geometry of the universe—alternatively, one can use elliptic integrals, as in Sect. B.3 and Feige (1992).) If one has an expression for $\sigma(z)$, then, since

$$dD^P = R_0 \frac{d\sigma}{\sqrt{1 - k\sigma^2}}, \quad \text{(C3)}$$

which follows directly from Eq. (1), one can write

$$dV(\sigma) = 4\pi R_0^3 \int_0^\sigma \frac{\sigma'^2 d\sigma'}{\sqrt{1 - k\sigma'^2}}. \quad \text{(C4)}$$

where R_0 is given by Eq. (7) for the present values:

$$R_0 = \frac{c}{H_0} \frac{1}{\sqrt{|\Omega_0 + \lambda_0 - 1|}}. \quad \text{(C5)}$$

and

$$\sigma = \frac{D_{y0}(1 + z_y)}{R_0} \quad \text{(C6)}$$

Integration gives

$$V(\sigma) = \begin{cases} 2\pi r^3 \left(\frac{\sqrt{1+\sigma^2}}{\sigma^2} - \frac{arcsinh\,\sigma}{\sigma^3} \right) & k = -1 \\ \frac{4}{3}\pi r^3 & k = 0 \\ 2\pi r^3 \left(\frac{\arcsin\sigma}{\sigma^3} - \frac{\sqrt{1-\sigma^2}}{\sigma^2} \right) & k = +1 \end{cases} \quad \text{(C7)}$$

Thus, for $k = +1$, the total volume of the universe is $2\pi^2 R_0^3$. (See, e.g., Sandage (1995), Sect. 1.6.1; Sandage's d is our D^P and his r is our σ.) Since

$$d\chi = \frac{d\sigma}{\sqrt{1 - k\sigma^2}}$$

Eq. (C7) can also be written as (cf. Feige (1992), Eq. (116); Feige's r is our σ)

$$V(\chi) = 2\pi R_0^3 \begin{cases} \sinh(\chi)\cosh(\chi) - \chi & k = -1 \\ \frac{2}{3}\chi^3 & k = 0 \\ \chi - \sin(\chi)\cos(\chi) & k = +1 \end{cases} \quad \text{(C8)}$$

Of course, all this refers to volumes *now* at the distance corresponding to $z = y$. If the volume at another time is important, say at the time of emission of the light we see now—for instance if one is concerned with the space density of some comoving objects—then the volume element must be divided by $(1 + z)^3$.

References

Dyer C. C., Roeder R. C., 1972, ApJ 174, L115

Dyer C. C., Roeder R. C., 1973, ApJ 180, L31

Feige B., 1992, Astr. Nachr. 313, 139

Kayser R., 1985, doctoral thesis, University of Hamburg

Linder E. V., 1988, A&A 206, 190

Mattig W., 1958, Astr. Nachr. 284, 109

Sandage A., 1995, Practical Cosmology: Inventing the Past. In: Binggeli B., Buser R. (eds.) The Deep Universe. Springer, Berlin

Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer-Verlag, Heidelberg

This article was processed by the author using Springer-Verlag LaTeX A&A style file *L-AA* version 3.

3.3 Follow-up

This paper is my fourth-highest-cited cosmology paper[3]; of the three with more citations, two are major collaboration papers (Chae *et al.*, 2002; Browne *et al.*, 2003, with 23 and 22 authors respectively). Most citations to our paper are because the citing authors had used the code and/or as a reference for definitions of cosmological distances. In particular, the Supernova Cosmology Project (Perlmutter *et al.*, 1999) made use of my code in their landmark paper for which Saul Perlmutter, as leader of the Supernova Cosmology Project, was awarded the 2011 Nobel Prize in Physics. Other citations are from people refining the theory, discussing (semi-)numerical implementations, presenting approximations, or testing the ZKDR distance against numerical simulations or other schemes for representing the mass distribution of a universe.

[3] My fourth-highest-cited paper overall deals with paleoclimatology, making this paper my fifth-highest-cited overall.

Chapter 4

ANGSIZ User's Guide

4.1 Context

While Kayser, Helbig and Schramm (1997) discussed the theory behind and some aspects of the implementation of the code, the *User's Guide* is intended for those who wish to actually use the publicly available code. It also includes a λ_0–Ω_0 diagram to explain the classification of cosmological models which corresponds to the σ_0–q_0 diagram in Stabell and Refsdal (1966), although as far as the code itself is concerned it is necessary only to know if the cosmological model has a maximum redshift in order to issue the appropriate error message should someone try to calculate the distance for $z > z_{\mathrm{max}}$.

ANGSIZ User's Guide

Phillip Helbig

February 1996

Note: For more information, see the article 'A general and practical method for calculating cosmological distances' by Rainer Kayser, Phillip Helbig and Thomas Schramm in *Astronomy and Astrophysics*, 1996.

a. General description

The **ANGSIZ** routine calculates the angular size distance between two objects as a function of their redshifts. (In world models which have no big bang, but contract from infinity to a finite size before expanding, there are two or four independent distances. In these cases, the first distance is returned by **ANGSIZ** and the rest by the routine **BNGSIZ**.) The distance also depends on the cosmological parameters λ_0, Ω_0 and $\eta(z)$. These are not passed to **ANGSIZ** but rather to an auxiliary routine, **INICOS**, which calculates all z-independent information for a given world model.

The result is in units of cH_0^{-1}. Multiply the result by 3×10^3 (actually 2.99792458×10^3) to get the distance in Mpc for $H_0 = 100\,\mathrm{km \cdot s^{-1} \cdot Mpc^{-1}}$. One can define h to be H_0 in units of $H_0 = 100\,\mathrm{km \cdot s^{-1} Mpc^{-1}}$. Thus, multiplying by 3×10^9 corresponds to an h of 1. For other values of h, simply *divide* the result by h, for example by 0.5 for $H_0 = 50\,\mathrm{km \cdot s^{-1} Mpc^{-1}}$. (Be aware that h can be and sometimes is defined as 1 for other values of H_0, usually 50. *Sometimes* this is indicated by a subscript, *i.e.* h_{50}, h_{100} and so on.)

The calling sequence is described below in section b.

All routines are in absolutely standard **FORTRAN77** and have been tested on a number of different combinations of compiler/hardware/OS so essentially the same numerical results should be obtained everywhere.

i. INICOS

The routine **INICOS** takes the cosmological parameters and 'user wishes' as input and calculates quantities needed by **ANGSIZ**—which are relayed internally to **ANGSIZ**—and also returns these to the calling routine. In addition, as in **ANGSIZ** itself, an error message is returned. **INICOS** can be used independently of **ANGSIZ** of course, but not the other way around.

INICOS returns the **INTEGER** variable **WMTYPE**, world model type, which gives a qualitative classification of the cosmological model. The table shows the correspondence between the value of **WMTYPE** and some standard classifications from the literature, and also gives some brief information on the temporal and spatial properties of the corresponding world model. The figure shows the location of the various world model types in the λ_0-Ω_0-plane.

`WMTYPE`	$t=\infty$	$R=\infty$	$\exists z_{\max}$	bounce	name	λ_0	k	Ω_0	σ_0	q_0	SR	HB
1	no	yes	no	no		<0	-1	0	0	>0	$O(0)$	1(iii)
2	no	yes	no	no		<0	-1	>0	>0	>0	$O(1)$	1(iii)
3	no	yes	no	no		<0	0	>0	>0	$>\frac{1}{2}$	$O(2)$	1(ii)
4	no	no	no	no		<0	$+1$	>0	>0	$>\frac{1}{2}$	$O(3)$	3(iv)
5	no	no	no	no	MTW	0	$+1$	>1	>0	$>\frac{1}{2}$	$O(4)$	3(iv)
6	no	no	no	no		>0	$+1$	$>\Omega_{0,\mathrm{c}}$	$>\sigma_{0,\mathrm{c}}$	$>\frac{1}{2}$	$O(5)$	3(iii)(a)
7	yes	yes	no	no	Milne	0	-1	0	0	0	$M_1(1)$	1(ii)
8	yes	yes	no	no	'LCDM'	0	-1	>0	>0	$0<q_0<\frac{1}{2}$	$M_1(2)$	1(ii)
9	(yes)	yes	no	no	Einstein-de Sitter	0	0	1	$\frac{1}{2}$	$\frac{1}{2}$	$M_1(3)$	2(ii)
10	yes	yes	no	no		>0	-1	0	0	$-1<q_0<0$	$M_1(0)$	1(i)
11	yes	yes	no	no		>0	-1	>0	>0	$-1<q_0<\frac{1}{2}$	$M_1(4)$	1(i)
12	yes	yes	no	no	de Sitter	1	0	0	0	-1	S	2(i)
13	yes	yes	no	no	'ΛCDM'	>0	0	>0	>0	$-1<q_0<\frac{1}{2}$	$M_1(5)$	2(i)
14	yes	no	no	no	Lemaître	>0	$+1$	note	note	note	$M_1(6)$	3(1)
15	(yes)	no	no	no		>0	$+1$	$\Omega_{0,\mathrm{c}}$	$\sigma_{0,\mathrm{c}}$	$>\frac{1}{2}$	A_1	3(ii)(b)
16	(yes)	no	(yes)	(no)	Einstein	>0	$+1$	$\Omega_{0,\mathrm{c}}=\infty$	$\Omega_{0,\mathrm{c}}=\infty$	∞	E	3(ii)(a)
17	yes	no	yes	no	Eddington	>0	$+1$	$\Omega_{0,\mathrm{c}}$	$\sigma_{0,\mathrm{c}}$	<-1	A_2	3(ii)(c)
18	yes	no	yes	yes	bounce	>0	$+1$	$0<\Omega_0<\Omega_{0,\mathrm{c}}$	$0<\sigma_0<\sigma_{0,\mathrm{c}}$	<-1	$M_2(1)$	3(iii)(b)
19	yes	no	yes	yes	Lanczos	>0	$+1$	0	0	<-1	$M_2(0)$	3(iii)(b)

Table 1. Classification of cosmological models as done by `INICOS`. Further information is provided in the text, particularly concerning parenthetical expressions and when 'note' appears instead of the expression in question.

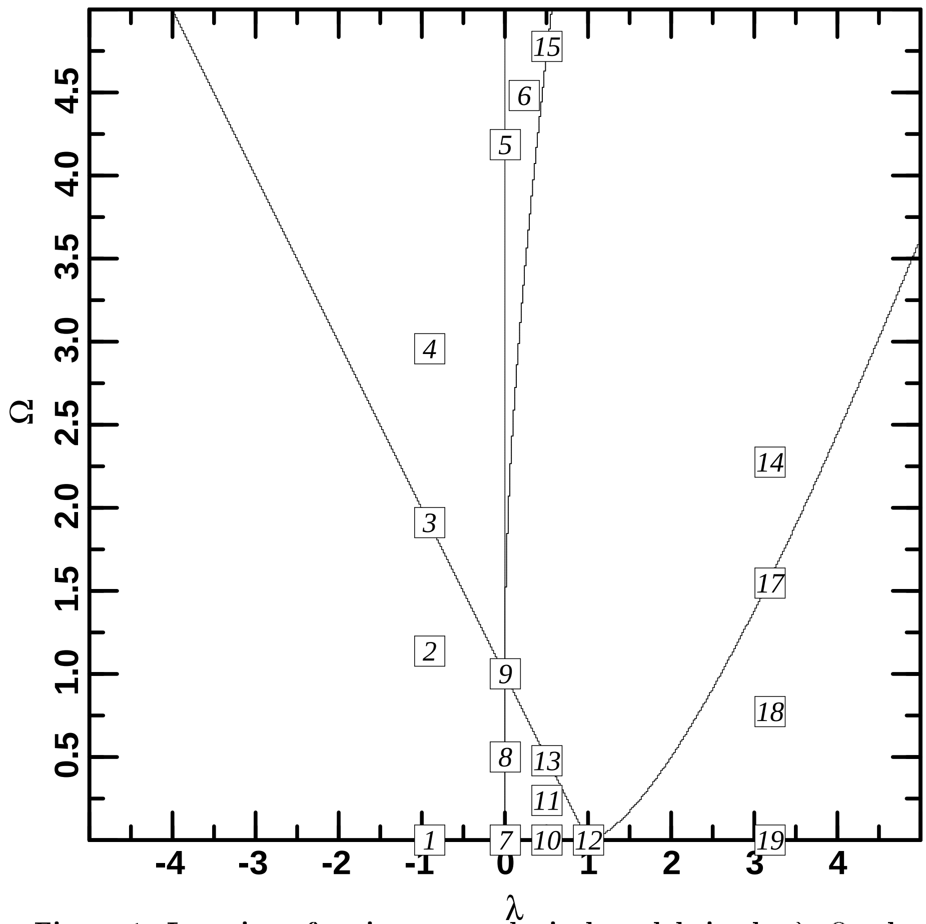

Figure 1. Location of various cosmological models in the λ_0-Ω_0-plane. A given world model is either on a line or curve segment (bounded by at least one vertex), on a vertex or in the space between the various lines and curves. The diagonal line corresponds to $k = 0$, the vertical line to $\lambda_0 = 0$, the curve immediately to the right of this line divides models which will eventually collapse (to the left) to those which will not (to the right) and the curve at lower right divides world models with a big bang (to the left) from those with no big bang (to the right).

`WMTYPE` refers to the classification returned by `INICOS` as the value of the variable `WMTYPE`. The following cosmological models are *definitively* ruled out by observations: 1, 7, 10, 12, 16, 17, 18, 19. 16 is ruled out because it has no expansion; the rest which aren't ruled out by the fact that they are empty are ruled out by the fact that a z_{max} which is at least as large as the largest observed redshift implies a value for Ω_0 which is so low as to be definitively ruled out.

$t = \infty$ answers the question whether the universe will expand forever or, more precisely, is the opposite of the answer to the question 'will the universe collapse to $R = 0$ in the future'. The subtlety arises because of the fact that `WMTYPE`s 9, 15 and 16, while not collapsing, don't necessarily 'expand forever' (perhaps depending on how this is precisely defined) although they might 'exist forever'. In the Einstein-de Sitter model `WMTYPE 9`, the expansion rate asymtotically approaches 0; the value of the scale factor R at $t = \infty$ is, however, $R = \infty$. Thus, the universe expands forever, although at $t = \infty$ the value of $\dot{R}$ is 0. The Eddington universe, `WMTYPE 15`, expands forever, as well, but at $t = \infty$ not only is the value of $\dot{R}$ 0 but R reaches a finite maximum as well. The Einstein universe, `WMTYPE 16`, is static. It exists forever, but doesn't really expand forever, since it doesn't expand at all. R is fixed at a finite value and $\dot{R}$ is 0 at all times.

$R = \infty$ answers the question whether the 3-dimensional space is infinite in extent. It is infinite for $k = -1$ or $k = 0$ and finite for $k = +1$ (assuming a simple topology).

$\exists z_{\max}$ If a $z_{\max}$ exists, then there was no big bang. (In the de Sitter model the big bang occurs at $t = -\infty$.) Rather, the universe expands from a minimum value of R. The static Einstein universe (`WMTYPE 16`) has a redshift of 0 for all objects, thus distance cannot be determined from redshift in this case.

bounce If there has been a bounce, the universe is expanding after contracting from $R = \infty$ to $R = R_{\min}$. This means that there are four independent distances for a given redshift, unless the dependence of η on z is the same during the collapsing and expanding phases, in which case there are only two independent distances. (In the Einstein static universe, `WMTYPE` 16, there are an infinite number of distances for the only possible redshift of 0.)

name refers to a commonly used term for the given world model. MTW refers to the fact that this world model is discussed in *Gravitation* by Misner *et al.* [1973] in §27.10 and Box 27.4 as an example of the type of cosmological model preferred by Einstein ($\lambda = 0$ because Einstein preferred a 'simpler' universe without the cosmological constant after it became clear that a) his static model cannot describe the real universe, not even as a 0^{th} approximation and b) non-static solutions with a cosmological constant exist; $k = +1$ in order to have no problems with boundary conditions at ∞). This world model is noted as having been 'investigated by Einstein' in Stabell & Refsdal [1966]; the name 'Einstein universe', although perhaps more appropriate for this cosmological model, is historically irrevocably associated with the static universe (`WMTYPE 16`). Milne's cosmological model is not really our `WMTYPE 7`, although it is equivalent; Milne had $G = 0$ in his kinematic relativity, and either $G = 0$ or $\rho = 0$ has the effect of making the expansion independent of gravitation. 'LCDM' refers to the fact that these cosmological parameters are used in the typical 'low density standard cosmological model'; here CDM is an abbreviation for 'cold dark matter', which includes details on scales small enough where homogeneity cannot be used as an approximation and doesn't concern us here. Naturally, global properties of cosmological models are independent of such local structure. Of course, the parameters are the same for LHDM (hot) and LHCDM (or LMDM: mixed). 'ΛCDM' refers to the fact that these cosmological parameters are used in the typical 'low density flat standard cosmological model with a cosmological constant'. Note: our 'Eddington model' (`WMTYPE` 17) is referred to as the Lemaître model in Stabell & Refsdal [1966] and as the Eddington-Lemaître model in Bondi [1961]. Harrison [1981] also uses our term Eddington model for this case. Similarly. our Lemaître model (`WMTYPE` 14) is called the Eddington-Lemaître in

Berry [1986]. Bondi [1961] and Harrison [1981] also use our term Lemaître model for this case.

λ_0 the range of λ_0.

k the sign of k.

Ω_0 the range of Ω_0. $\Omega_{0,\mathrm{c}} = 2\sigma_{0,\mathrm{c}}$. In `WMTYPE` 14, Ω_0, σ_0 and q_0 can take on all possible values, but not all combinations are possible. Specifically, for $-1 < q_0 < \frac{1}{2}$, $\sigma_{0,\mathrm{c}}$ ($\Omega_{0,\mathrm{c}}$) has no meaning, and the range $0 < \sigma_{0,\mathrm{c}}$ is allowed. Otherwise, we must have $\Omega_0 > \Omega_{0,\mathrm{c}}$ ($\sigma_0 > \sigma_{0,\mathrm{c}}$). (Of course, $\Omega_0 = 2\sigma_0$.)

σ_0 the range of σ_0; $\sigma = 0.5 \times \Omega_0$. $\sigma_{0,\mathrm{c}}$ is defined as

$$\sigma_{0,\mathrm{c}} = \frac{1}{6}(q_0+1)\left(q_0+1 \pm \sqrt{(q_0+1)\left(1_0 - \frac{1}{3}\right)}\right) \text{ (SR Eq. 12)}$$

In `WMTYPE` 14, Ω_0, σ_0 and q_0 can take on all possible values, but not all combinations are possible. Specifically, for $-1 < q_0 < \frac{1}{2}$, $\sigma_{0,\mathrm{c}}$ ($\Omega_{0,\mathrm{c}}$) has no meaning, and the range $0 < \sigma_{0,\mathrm{c}}$ is allowed. Otherwise, we must have $\sigma_0 > \sigma_{0,\mathrm{c}}$ ($\Omega_0 > \Omega_{0,\mathrm{c}}$).

q_0 the sign of q_0; $q_0 = \sigma_0 - \lambda_0$. In `WMTYPE` 14, Ω_0, σ_0 and q_0 can take on all possible values, but not all combinations are possible. Specifically, for $-1 < q_0 < \frac{1}{2}$, $\sigma_{0,\mathrm{c}}$ ($\Omega_{0,\mathrm{c}}$) has no meaning, and the range $0 < \sigma_{0,\mathrm{c}}$ is allowed. Otherwise, we must have $\sigma_0 > \sigma_{0,\mathrm{c}}$ ($\Omega_0 > \Omega_{0,\mathrm{c}}$).

SR refers to the description of cosmological models in Stabell & Refsdal [1966, p. 383] where models are classified with a scheme similar to that used by `INICOS`.

HB Refers to the classification in chapter IX of Bondi [1961]. Since Bondi implicitly assumes $\Omega_0 > 0$, this is not entirely correct in the cases where there is a qualitative difference for $\Omega_0 = 0$, as in `WMTYPE`s 7, 10 and 12. In this case, q_0 is never positive, so the qualitative curves given by Bondi should be extrapolated back to $R = 0$ without the negative curvature part at the beginning.

ii. ANGSIZ

This routine calculates the angular size distance between two redshifts (just one of the possible distances in the bounce models). This is done by the numerical integration of a second order differential equation. (For details see Kayser, Helbig & Schramm [1996].) As far as we know, no analytic solution for the general case exists. An advantage of using `ANGSIZ` even when an analytic solution for a special case exists is that `ANGSIZ` *always* gives a valid result, *i.e.*, the equation holds for *all* cosmological models and no cases must be distinguished. (`ANGSIZ` is not recommended for use in the Eddington model, `WMTYPE` 17. If the user specifically wants to use the routines for this cosmological model, it is possible to allow this by declaring the `COMMON` block `ARTHUR` in the calling routine, containing a logical variable, which should be set to `.TRUE.` before calling `ANGSIZ`. This will override the default behaviour.)

iii. ETAZ

The variable η gives the fraction of homogeneously distributed matter in the cosmological model. On large scales, homogeneity is assumed, in accordance with the Robertson-Walker metric, the Cosmological Principle, and so on. However, on smaller scales, matter can of course be clumped. The fraction $1-\eta$ of matter is in these clumps, which means that they are outside (and sufficiently far from) the cone of light rays between source and observer. (If a clump of matter is near or in the beam itself, one must take account of this explicitly as a gravitational lens effect.)

It is important to remember that the value of η—for a fixed real situation—depends on the angular scale involved. For example, a halo of compact MACHO type objects around a galaxy in a distant cluster would be counted among the homogeneously distributed matter if one were concerned with the angular size distance to background galaxies further away, but would be considered clumped on scales such as those important when considering microlensing by the compact objects themselves. Thus, the clumps must have a scale comparable to the separation of light rays from an extragalactic object or larger.

Since we don't know exactly how dark matter is distributed, different η values can be examined to get an ideas as to how this uncertainty affects whatever it is one is interested in.

Of course, η can be a function of z. If the user wishes η to be constant for all z, then the variable `VARETA` should be set to `.FALSE.` when calling `INICOS`. In this case, the value of `ETA` passed to `INICOS` determines η. If `VARETA` ist `.TRUE.`, then the value of `ETA` is irrelevant and is given by the `REAL FUNCTION ETA(Z)`, which the user can modify as needed. If a constant value of `ETA` is desired, this can of course be achieved through defining the `FUNCTION ETA(Z)` appropriately, but this will typically make the computing time 2–3 times longer than setting `VARETA` to `.FALSE.` and using `ETA` to determine the value of η.

iv. BNGSIZ

This `SUBROUTINE BNGSIZ` is essentially the same as `ANGSIZ` except that it calculates the other three distances in bounce models. As long as `ETAZ` doesn't make the value of η depend on whether the universe is expanding or contracting, there is only one additional independent distance, `D14`; `D34` is the same as `D12` and `D32` is the same is `D14`. `D12` is the distance returned by `ANGSIZ`, where both the starting point and the end point of the integration are in the expansion phase. `D14` has its starting point `Z1` in the expansion phase and its end point `Z2` in the contraction phase, thus the angular size distance is found by integrating from `Z1` to `ZMAX` and back to `Z2`. `D34` has both boundaries in the contraction phase, and `D32` its starting point in the contraction phase and its end point in the expansion phase.

If one is not interested in the bounce models, this routine isn't needed. If one is only interested in the 'primary' distance in bounce models, this is returned by `ANGSIZ`, so also in this case `BNGSIZ` isn't needed.

`BNGSIZ` must be called after calling `ANGSIZ`: there are no input parameters; these and other necessary information are obtained by `BNGSIZ` from `ANGSIZ`.

b. Use as a black box: what the user needs to know

In the descriptions of the input variables in the parameter lists to the routines called by the user we also indicate a range of suggested values in square brackets. Of course, others are possible, but the ones we suggest correspond to cosmological models which are not firmly ruled out by other arguments (and some which are) and have been tested. Physically meaningless values should result in an error message. We didn't think it meaningful to check the routines for correctness for values outside these ranges. For output variables, the range indicates possible values. The mathematical and physical meaning of the variables can be found in Kayser *et al.* [1996].

i. INICOS

The interface to `INICOS` is

```
      SUBROUTINE INICOS(LAMBDA,OMEGA,ETA,VARETA,DEBUG,
C                       ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
C                                    in
C
     $                  WMTYPE,MAXZ,ZMAX,BOUNCE,ERROR)
C                       ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
C                                    out
```

The input variables are:

`REAL LAMBDA` is the normalised cosmological constant. [−10,...,+10]

`REAL OMEGA` is the density parameter. [0,...,10]

`REAL ETA` is the homogeneity parameter. [0,...,1]

`LOGICAL VARETA` if `.TRUE.` means that η is not given by `ETA` but by the value returned by the `REAL FUNCTION ETAZ`.

`LOGICAL DEBUG` if `.TRUE.` means that error messages should be written to standard output.

The output variables are:

`INTEGER WMTYPE` gives some information about the cosmological model (see the table). [1,...,19]

`LOGICAL MAXZ` if `.TRUE.` indicates that a maximum redshift exists; only in this case is the variable `ZMAX` correctly defined and only in this case should the variable `ZMAX` be used.

`REAL ZMAX` gives the maximum redshift in the cosmological model, if `LOGICAL MAXZ` is `.TRUE.`; otherwise, the maximum redshift is infinite, and for convenience set to zero.

`LOGICAL BOUNCE` if `.TRUE.` indicates that we have a 'bounce' model (see the table); only in this case can the other distances be computed by `BNGSIZ`.

`INTEGER ERROR` indicates if an error has occurred and what this means (see section vi.). [0,...,13]

ii. ANGSIZ

The interface to **ANGSIZ** is:

```
      SUBROUTINE ANGSIZ(Z1,Z2, D12,ERROR)
C                       ^^^^^  ^^^^^^^^^
C                        in       out
```

The input variables are:

REAL Z1 is a redshift; set Z1 to 0 for the common case of the 'distance from a normal observer'. [0,...,5]

REAL Z2 is a redshift; set Z2 to the redshift of the object if one is interested in the angular size distance from a 'normal observer' to the object and correspondingly Z1 has been set to 0. [0,...,5]

The output variables are:

REAL D12 Is the angular size distance between Z1 and Z2. In bounce models (and of course in big bang models) this corresponds to the distance such that the universe has always been in a state of expansion between the times of light emission and reception.

INTEGER ERROR indicates if an error has occurred and what this means (see section vi.). [0,...,13]

ANGSIZ is not recommended for **WMTYPE** 17, since distances can become arbitrarily large, and the necessary overhead would complicate the routines to such a degree that 'normal' performance would be significantly hampered. If the user specifically wants to use the routines for this cosmological model, it is possible to allow this by declaring the **COMMON** block **ARTHUR** with a logical variable, which should be set to **.TRUE.** in the calling routine. This will override the default behaviour.

iii. BNGSIZ

The interface to **BNGSIZ** is:

```
      SUBROUTINE BNGSIZ(D14,D34,D32,ERROR)
C                       ^^^^^^^^^^^^^^^^^
C                              out
```

REAL D14 is the angular size distance between Z1 and Z2. This corresponds to the distance such that the universe is *not* in a state of *contraction* at Z1 and *not* in state of *expansion* at Z2.

REAL D34 is the angular size distance between Z1 and Z2. This corresponds to the distance such that the universe is in a state of *expansion neither* at Z1 *nor* at Z2.

REAL D32 is the angular size distance between Z1 and Z2. This corresponds to the distance such that the universe is *not* in a state of *expansion* at Z1 and *not* in a state of *contraction* at Z2.

INTEGER ERROR indicates if an error has occurred and what this means (see section vi.). [0,...,13]

iv. ETAZ

The interface to ETAZ is:

```
REAL FUNCTION ETAZ(Z)
```

which, being a function with no side effects, has no output variables. The one input variable is:

REAL Z this is the redshift; η is a function of z and is given by this function and *not* by the value of ETA supplied to INICOS if VARETA is .TRUE.; only in this case will ETAZ be used.

The COMMON block WHICH contains the INTEGER variable CHOICE. If this COMMON block is also in the calling routine, then a decision can be made there as to which dependency of η on z is to be used; see the use of CHOICE in our example ETAZ. This is initialised in our BLOCK DATA COSANG and so should not be initialised by the user.

For bounce models, it is of course possible that the dependence of η on z depends on whether the universe is expanding or contracting. Since VALUE2 is set to .TRUE. during the phase of *contraction* and otherwise to .FALSE. by ANGSIZ and BNGSIZ, which can access VALUE2 through the COMMON block CNTRCT, this variable can be used in ETAZ order to have two different dependencies, as in our example. *Additionally, in this case, as in our example, the value of* CHOICE *must be* negative *and otherwise must be positive.* This distinction is needed by BNGSIZ in order to avoid calculating D34 and D32 if these are not independent of the other distances.

v. Calling sequence

Thus, one should call INICOS anew for each cosmological model tested, indicating the cosmological model, whether or not η is to be a constant or is to be calculated by ETAZ and whether or not error messages are to be displayed. (ERROR is of course always appropriately set.) If one wants to calculate distances, then ANGSIZ should be called with two redshifts, and the angular size distance is returned. One should use the output variables of INICOS to determine whether or not to call ANGSIZ. If desired, ETAZ can be used; it can also be modified by the user. The calling routine can also make use of the COMMON block WHICH to enable different z dependencies of η to be calculated without having to change the code, recompile and relink for each case. BNGSIZ can be called after calling ANGSIZ if one is interested in the additional distances in the case of the bounce models.

vi. Error messages

There are fourteen possible 'error states', namely the values [0,...,13] of the INTEGER variable ERROR. Errors 1–3 can occur in INICOS, 4–12 in ANGSIZ, 12–13 in BNGSIZ. (Thus, only one error variable is needed in the calling routine.) ERROR is *always* appropriately set; the variable DEBUG merely controls whether or not messages are written to standard output. With the exception of ERROR .EQ. 12, the message is just the information contained in the following overview.

0 No error was detected.

1 (OMEGA .LT. 0.0).

2 `((ETA .LT. 0.0) .OR. (ETA .GT. 1.0))` (disabled if `VARETA` is `.TRUE.`, in which case the value of `ETA` is irrelevant.) `ETAZ`, however, checks to see if `ETA` is within the allowed range.

3 The world model is so close to the A2 curve that the calculation of `ZMAX` is numerically unstable. This usually happens—when it does—very near the de Sitter model. However, it should be rare in practice, as noted below. Even if one were able to calculate a good value for `ZMAX`, this would probably lead to an error in `ANGSIZ` if `Z1` or `Z2` is near `ZMAX`.

4 The world model is `WMTYPE` 17. In this world model, the universe is infinitely old, and there is a finite `ZMAX`. This means that distances (especially for $\eta = 0$) can become arbitrarily large. The routines are not recommended for calculation in this cosmological model, since enabling the detection of overflow would unnecessarily complicate the routines. Also, the practical value of `ZMAX`, used internally, would have to be appreciable less than the real `ZMAX`. If the user specifically wants to use the routines for this cosmological model, it is possible to allow this by declaring the `COMMON` block `ARTHUR` with a logical variable, which should be set to `.TRUE.` in the calling routine. This will override the default behaviour.

5 `(Z1 .LT. 0.0)`

6 `(Z2 .LT. 0.0)`

7 `(Z1 .GT. ZMAX)`

8 The value of `ZMAX` returned by `INICOS` is the calculated value; however, internally a somewhat smaller `ZMAX` is sometimes necessary for numerical stability. This internal value is close enough to the real `ZMAX` for all practical purposes (at worst it has an error of about 10^{-3}). This message means that `Z1` is too large for numerical stability, being very near the real `ZMAX` but slightly smaller.

9 `(Z2 .GT. ZMAX)`

10 The value of `ZMAX` returned by `INICOS` is the calculated value; however, internally a somewhat smaller `ZMAX` is sometimes necessary for numerical stability. This internal value is close enough to the real `ZMAX` for all practical purposes (at worst it has an error of about 10^{-3}). This message means that `Z2` is too large for numerical stability, being very near the real `ZMAX` but slightly smaller.

11 Overflow error is possible. Calculation can be continued but the result might not be as exact as usual or subsequent calculation with the same cosmological parameters might lead to an 'unhandled exception'.

12 Very rare. This means an error has occurred deep down in the numerical integration. If `DEBUG` is `.TRUE.` then a message will be printed identifying where.

13 This means that `BNGSIZ` was called for a non-bounce world model.

When an error is returned, with the exception of 11, then *none* of the other output variables of the routine should be used and all distances are set to 0.

c. Inside the black box: description of all routines

Here is a short description of all of the `SUBROUTINE`s and `FUNCTION`s:

`BLOCK DATA COSANG` initialises `COMMON` block variables.

`SUBROUTINE INICOS` is described above.

`REAL FUNCTION QQ` calculates $Q^2(z)$.

`SUBROUTINE ANGSIZ` is described above.

`SUBROUTINE BNGSIZ` is described above.

`SUBROUTINE LOWLEV` Performs the low-level integration, mainly calling `ODEINT`.

`SUBROUTINE ASDRHS` calculates the right hand sides of the differential equation for the angular size distance

`SUBROUTINE ODEINT` is a 'driver' routine for the numerical integration.

`SUBROUTINE MMM` is for the modified midpoint method used by `ODEINT`.

`SUBROUTINE POLEX` performs polynomial extrapolation used by `ODEINT`.

`SUBROUTINE BSSTEP` performs one Bulirsch-Stoer integration step.

`REAL FUNCTION ETAZ` is described above.

i. Call tree

```
[COSANG]
[INICOS]
|  [QQ]
[ANGSIZ]
|  [LOWLEV]
|  |  [ODEINT]
|  |  |  [ASDRHS]
|  |  |  |  [ETAZ]
|  |  |  [BSSTEP]
|  |  |  |  [MMM]
|  |  |  |  |  [ASDRHS]
|  |  |  |  |  |  [ETAZ]
|  |  |  |  [POLEX]
|  [QQ]
[BNGSIZ]
|  [LOWLEV]
|  |  [ODEINT]
|  |  |  [ASDRHS]
|  |  |  |  [ETAZ]
|  |  |  [BSSTEP]
|  |  |  |  [MMM]
|  |  |  |  |  [ASDRHS]
|  |  |  |  |  |  [ETAZ]
|  |  |  |  [POLEX]
|  [QQ]
```

d. Caveats

The following `COMMON` block names are used and, since these are global names, should not be used in a conflicting way in any other routines: `COSMOL`, `ANGINI`, `WHICH`, `BNCSIZ`, `LOWSIZ`, `CNTRCT`, `ARTHUR`, `MERROR` and `PATH`. With the exception of the optional `WHICH`, all of these are used internally and need not further concern the user. The same is true of the `SUBROUTINE` and `FUNCTION` names listed above (except for `INICOS`, `ANGSIZ`, `BNGSIZ` and `ETAZ`, of course.)

The rest of this section should be of absolutely no concern to almost all users, since 'incorrect' results due to the reasons discussed below should only occur due to roundoff error corresponding to a ridiculous accuracy in the cosmological parameters; in most of these cases this would only be in world models which are uninteresting because they are ruled out conclusively (see the table).

The value of `WMTYPE` can differ from the 'real' value as calculated analytically from λ_0 and Ω_0 but only if one is so close to the boundary between two regions of parameter space that this happens due to ordinary inexactness in the internal representation of 'real' numbers. The same goes for the values of `MAXZ` and `BOUNCE`, since these are trivially related to `WMTYPE`. The A1 and A2 curves (see Stabell & Refsdal [1966]), corresponding to `WMTYPE`s 15 and 17, have been 'drawn with a numerically thick pencil' but this should only be noticeable for `LAMBDA` and `OMEGA` values precise to 10^{-5}. This is also true of the endpoints of the curves, `WMTYPE`s 9 and 12. `WMTYPE` 16 cannot be returned by `INICOS`, since this corresponds to the static Einstein cosmological model, in which λ_0 and Ω_0 are both $= \infty$. The value of `ZMAX` is probably exact in the third figure after the decimal point for world models *extremely near* the A1 or A2 curve; otherwise, it is as good as any numerically calculated quantity.

The values of `D12` and `D34` returned should be correct in the third digit after the decimal point and shouldn't be off by more than one digit in the fourth. The values of `D14` and `D32` are probably somewhat less precise, correct in the second decimal digit.

e. Development and tests

The development and most testing of the routines was done on a Digital VAXStation 3100 Model 76 running VMS 5.5-2 and with DEC FORTRAN. To make sure that numerical accuracy is monitored in a robust way, comparisons were made to the following systems:

- DEC ALPHA 4000/710, VMS, DEC FORTRAN (`FORTRAN 77`)
- DEC ALPHA 4000/710, VMS, DEC FORTRAN 90
- Cray C916, UNICOS 8, native `FORTRAN 77` compiler
- Convex C3840 and C220, Convex OS 11.0, native `FORTRAN 77`
- DSM Infinity 8000, SCO-UNIX, Green Hills Fortran Version 1.8.5
- Fujitsu/Siemens S100, UXP/M, UXP/M Fortran 77 EX/VP (V12)
- Hewlett-Packard 9000/735, HPUX, native `FORTRAN 77` and Fortran90 compilers
- IBM RS/6000, AIX, xlf90
- IBM RS/6000, AIX, xlf90, optimisation -qarch=pwr

- IBM RS/6000, AIX, xlf90, optimisation -qarch=ppc
- IBM RS/6000, AIX, xlf90, optimisation -qarch=pwr2
- 'generic InTel PC', Linux, f2c+gcc
- Intel Pentium, MS-DOS, Lahey Fortran90
- Apple PowerMac 6100/60, Macintosh OS D 7.5, Language Systems Fortran PPC
- IBM 9121-440, MVS, VS FORTRAN

f. Updates

Corrections, updates and so on will be posted under the subject 'ANGSIZ' in the newsgroup sci.astro.research. Comments, bug reports, *etc.* should be sent by email to `phelbig@hs.uni-hamburg.de`. Please put 'ANGSIZ' in the subject line.

g. Disclaimer

We've tested the routines to a greater degree than usual for serious scientific work, though of course not as extensively as for (serious) commercial software. To the best of our knowledge, they work as described, but of course we can take no responsibility for the consequences of any errors in the code. As far as possible we will correct any errors; in any case we will make them known in the newsgroup posting.

References

[1986] Berry, M. V.: *Cosmology and Gravitation*
Bristol: Adam Hilger, 1986

[1961] Bondi, H.: *Cosmology*
Cambridge: Cambridge University Press, 1961

[1981] Harrison, E. R.: *Cosmology, the science of the universe*
Cambridge: Cambridge University Press, 1981

[1996] Kayser, R., P. Helbig, T. Schramm: 'A general and practical method for calculating cosmological distances'
A&A (accepted)

[1973] Misner, C. W. , K. S. Thorne, J. A. Wheeler: *Gravitation*
New York: Freeman, 1973

[1966] Stabell, R., S. Refsdal: 'Classification of general relativistic world models'
MNRAS, **132**, 3, 379 (1966)

4.3 Follow-up

The Supernova Cosmology Project (Perlmutter *et al.*, 1999) made use of my code in their landmark paper for which Saul Perlmutter, as leader of the Supernova Cosmology Project, was awarded the 2011 Nobel Prize in Physics.

The original code presented by Kayser, Helbig and Schramm (1997) was written in FORTRAN77. I have trivially converted it to fixed-form FORTRAN95 and am in the process of converting it to free-form FORTRAN95 (or, if meaningful, some later version) as well as replacing FORTRAN77-style constructs for which there are now better alternatives. Plans are to include this in a larger software package which will also include elliptic-integral calculations for special values of η as well as the calculation of non-distance cosmological quantities.

Part III

General applications

Chapter 5

Are the clumps at a redshift of 2.39 really sub-galactic?

5.1 Context

Pascarelle *et al.* (1996) claimed to have detected 'Sub-Galactic Clumps at a Redshift of 2.39'; we pointed out that this implicitly assumes the standard distance, and if $\eta = 1$ isn't assumed, the evidence for the claim is substantially weakened. We noted that the *linear* size could have been a factor of 4 larger than claimed, which would correspond to a factor of 64 in *volume*.

Are the clumps at a redshift of 2.39 really sub-galactic?

Phillip Helbig **Rainer Kayser**

Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg, Germany

Recently, Pascarelle *et al.* reported the discovery of 'sub-galactic clumps' at a redshift of 2.39. The physical size of the clumps was inferred from the angular size assuming a 'standard' cosmological model with $H_0 = 80$ km/s/Mpc and $q_0 = 0.5$.[1] (This value for q_0 corresponds to $\Omega_0 = 1$ with the assumption of $\lambda_0 = 0$ or $k = 0$ which the authors do not state but which is apparent from their numerical values.) However, the influence of the adopted cosmological model on the result is so large at this redshift that their conclusion does not necessarily follow from their data, and thus this evidence for a particular scenario of galaxy evolution is premature.

For constant angular size, the physical size is proportional to the angular size distance, by definition. The angular size distance is a function of the Hubble constant H_0, the density parameter Ω_0, the cosmological constant λ_0 and the degree of homogeneity of the matter distribution (for instance as described by the smoothness parameter η).[2] We note that it is not unlikely that H_0 is smaller than 80 km/s/Mpc,[3] that Ω_0 is less than one,[4] that λ_0 is larger than zero[5] and/or that η (implicitly assumed to be 1 by Pascarelle *et al.*) is less than one (corresponding to a distribution of matter in the universe clumped on scales approximately equivalent to the angular size of a galaxy, which is more probable if Ω_0 is appreciably less than 1).[2] It is interesting to note that *each* of these possibilities *increases* (for $z \approx 2.39$) the angular size distance and thus the inferred physical size. (The angular size distance is inversely proportional to H_0; the dependence on the other parameters is more complicated.)

We have calculated the physical size in kpc for an angular size of $0.11''$ (corresponding to the mean value of the continuum scale length of the objects in Pascarelle *et al.*) for 6 representative cosmological models, all of which are compatible with the current observational situation. The results are presented in Table 1. It is clear that (the linear size of) the 'sub-galactic clumps' could be four times larger than Pascarelle *et al.* claim; until the observational status of the cosmological parameters is clearer, we feel it is too early to extract detailed conclusions about galaxy formation on the basis of the angular size of objects at large redshift.

---

1. Pascarelle, S. M., Windhorst, R. A., Keel, W. C. & Odewahn, S. C. *Nature* **383**, 45–50 (1996).
2. Kayser, R., Helbig, P., Schramm, T. *A&A* (in press).
3. Sandage, A. *ApJ*, **402**, 3–14 (1993).
4. Coles, P., Ellis, R. *Nature*, **370**, 609–615 (1994).
5. Carroll, S., Press, W. H., Turner, E. L. *ARAA*, **30**, 499–542 (1992).
6. Ostriker, J. P., Steinhardt, P. J. *Nature*, **377**, 600-602 (1995).

H_0 [km/s/Mpc]	Ω_0	λ_0	η	r [kpc]	t_0 [10^9a]
80	1	0	1	0.54	8.15
50	1	0	1	0.86	13.04
50	0.1	0	1	1.34	17.56
50	0.1	1.1	1	1.70	29.53
50	0.1	1.1	0	2.14	29.53
70	0.35	0.65	1	0.86	15.71

Table 1. The physical size r of an object of angular size $0.11''$ for some representative cosmological models
The first row corresponds to the model of Pascarelle *et al.*; the next four rows successively change the value of one parameter in what is the probable direction according to the current observational situation. The last row is a currently popular 'standard model' which is something of a 'best fit' to the current observational situation[6] *among models which are flat or without a cosmological constant*. **Also shown is the age of the universe t_0 for each cosmological model.**

5.3 Follow-up

This article was a direct response to Pascarelle *et al.* (1996) and submitted, as was that paper, as a letter to *Nature*. In retrospect, it is not surprising that *Nature* didn't accept it; not only is the rejection rate high, but *Nature* also tends to publish claims which are especially interesting or unexpected. While there is nothing wrong with that (as long as there are other outlets for more bread-and-butter work, which there are), *Nature* does not seem to be interested in debunking especially interesting or unexpected claims, especially those which have been published in *Nature*. As the point of our paper was obvious to us, and should be, at least qualitatively, to everyone, it doesn't make sense to publish it elsewhere, as the simple point is worth making in this form only in direct response to a contrary (though implicit) claim. (Nevertheless, we did submit it as a *letter* to *Astronomy & Astrophysics*, but it was rejected.[1]). However, it does serve as an easy-to-understand introduction to the effects of η on the angular-size distance, so is included here.

The idea grew out of conversations with Rainer Kayser; I performed the calculations and physically wrote the paper.

[1]I have had only two other papers rejected. In both cases I was the last of three authors on the paper, the first author being a student in Hamburg who then moved on, in one case leaving astronomy altogether. As we all had more important things to do, I didn't follow those up, though I think that both could have been published after some (not necessarily meaningful) revision requested by the referees. One of those was the generalization to three dimensions of the idea of using Poisson statistics in the context of large-source microlensing as developed by Refsdal and Stabell (1991) for the one-lens-plane case. The other was an examination of the claim by Liebscher *et al.* (1992) that the Lyman-α forest, *under the assumption that there is no evolution in the absorbing matter*, supports a low-density Universe with a large cosmological constant. Their paper was heavily criticized, with some critics claiming that their conclusions didn't follow from the data used and their assumptions. We showed that that criticism is false, by repeating the entire analysis (including re-measuring original spectra), and that the same conclusion is obtained using a different statistical method developed for another cosmological test (Kayser, 1995). However, the corresponding cosmological model was already ruled out on other grounds when their paper was written, but one could turn the argument around and use the data to calculate the evolution in the Lyman-α forest for an assumed cosmological model, which we did. The paper appeared on arXiv (`astro-ph/9607117`; note that my co-author Liebscher is unrelated to the other Liebscher). It was even cited in a paper (Overduin and Priester, 2001) on which one of the authors (Wolfgang Priester) was also on the paper by Liebscher *et al.* (1992). While I have written several papers which debunk claims from the literature (Helbig and Kayser, 1996a; Jackson *et al.*, 1998; Helbig, 1998a; Zackrisson *et al.*, 2003; Helbig, 2012, 2020b,d, 2021), it is rare that they are acknowledged, much less that the debunked author admits that I could be right.

Part IV

Gravitational lensing

Chapter 6

Measuring the Hubble Constant with lens time delays in an inhomogeneous universe

6.1 Context

In a gravitational-lens system with more than one image of the source, the light-travel time can be different for each image. Since all other observables (flux ratios, angles, *etc.*) are dimensionless[1], measuring the time delay will thus set a physical scale. Since distances are, to first order, inversely proportional to H_0, one can thus measure H_0 by measuring the time delay (Refsdal, 1964), it being proportional to the combination of angular-size distances: $(D_{\rm d}D_{\rm s})/D_{\rm ds}$. At higher order, distances depend on λ_0 and Ω_0, hence if one has more than one such system and/or an independent measurement of H_0, one can measure those additional parameters (Refsdal, 1966). At even higher order, η plays a role as well, so, especially with higher-redshift systems, one might hope to use measurements of time delays to determine η (Kayser and Refsdal, 1983).

Kayser and Refsdal (1983) investigated the influence of η for several world models with $\lambda_0 = 0$ by comparing the $\eta = 1$ and $\eta = 0$ cases. For the double quasar 0957+561 (Walsh *et al.*, 1979), the cosmological correction factor (which gives the influence of the cosmological model compared to the limiting low-redshift case) was calculated for σ_0 values[2] ranging from 0 to 2 (corresponding to $0 \leq \Omega_0 \leq 4$) with q_0 values of 1.0, 0.5, 0.0, and -1 ($\lambda_0 = \sigma_0 - q_0$). I repeated the exercise for arbitrary combinations of λ_0, Ω_0, and η, again showing the importance of η, which has become even more important now that the values of λ_0 and Ω_0 are so well known. Those results were presented at a workshop at Jodrell Bank which I helped organize while I was working there (Helbig and Jackson, 1997).

[1] While the observables needed to model a gravitational-lens system are dimensionless, one can of course obtain additional observations, such as the velocity dispersion of the lens, which will also set an absolute scale.

[2] $\sigma := \Omega/2$, not to be confused with σ as defined in Chap. 1.

Measuring the Hubble constant with lens time delays in an inhomogeneous universe

Phillip Helbig

NRAL, University of Manchester, Jodrell Bank, Macclesfield, UK-Cheshire SK11 9DL, England

Abstract

The effects of a locally inhomogeneous universe on the uncertainty of the Hubble constant as determined from measured time delays in gravitational lens systems is discussed. The effect has been described adequately in the literature, but it is usually not taken into account when discussing measurements of H_0 using gravitational lens time delays. Depending on the cosmological model and the redshifts of the particular lens system considered, the effect of local inhomogeneity can significantly increase the uncertainty in the determination of H_0, and in 'probable' cosmological models can be the dominant uncertainty.

a. Introduction

The idea of measuring the Hubble constant H_0 using the time delay between images of a source which is multiply imaged due to the gravitational lens effect was introduced by Refsdal (1964), who also discussed the higher-order dependence on the other main cosmological parameters, in modern notation the cosmological constant λ_0 and the density parameter Ω_0 (Refsdal 1966). In particular, Refsdal (1966) introduced the 'cosmological correction function' T which describes these higher-order effects. Kayser & Refsdal (1983) showed that this same formalism also applies in the case of an arbitrary lens mass distribution and in the extreme case of a locally inhomogeneous universe, the so-called empty-cone approximation. Since the cosmological correction function depends only on the redshifts of the lens and source and on the distances involved, it is straightforward to generalise even further, using the formalism and methods set out in Kayser *et al.* (1997), to cases intermediate between the traditional approach (which assumes an idealised universe consisting of a perfect fluid) and the empty-cone approximation.

Recently, not only has the general idea of measuring H_0 by lens time delays become more acceptable, but (partly the cause of this) other uncertainties, such as measuring (and interpreting!) the time delay itself (see Pelt *et al.* (1996) and references therein) and modelling the lens mass distribution have become better understood, so that now the dominant uncertainties are cosmological—the values of λ_0 and Ω_0 and the parameter η discussed below, which describes local inhomogeneity.

b. Basic theory

i. Time delay

One can write an expression for the time delay (*cf.* Kayser & Refsdal (1983))

$$H_0 = (\Delta t)^{-1} T f \qquad (1)$$

where H_0 is the Hubble constant, Δt the time delay, T the cosmological correction function and f is a function of observational quantities and the mass distribution of the lens and will not be further discussed here. The cosmological correction function

$$T = \frac{H_0}{c}\frac{D_{\rm d} D_{\rm s}}{D_{\rm ds}}(1 + z_D)\frac{z_{\rm s} - z_{\rm d}}{z_{\rm d} z_{\rm s}} \qquad (2)$$

is defined so that $T \rightarrow 0$ for $z_{\rm s} \rightarrow 0$.

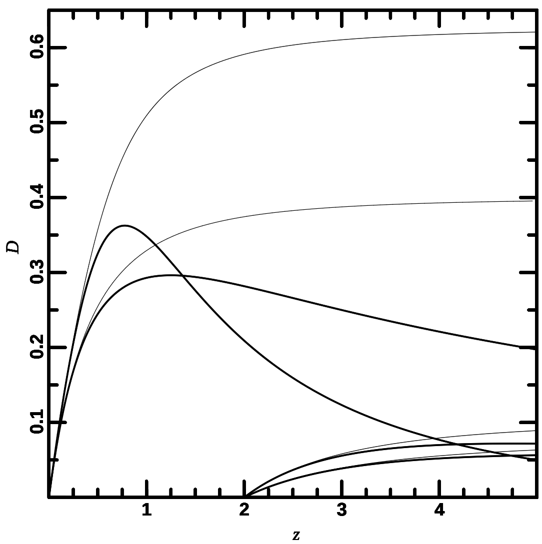

Figure 1 Dependence of the angular size distance D on λ_0 and η
The angular size distance from the observer and from an object at $z = 2$ to another at higher redshift as a function of the redshift z for different cosmological models. Thin curves are for $\eta = 0$, thick for $\eta = 1$. The upper curves near $z = 0$ ($z = 2$ at lower right) are for $\lambda_0 = 2$, the lower for $\lambda_0 = 0$. $\Omega_0 = 1$ for all curves. The distances are given in units of c/H_0.

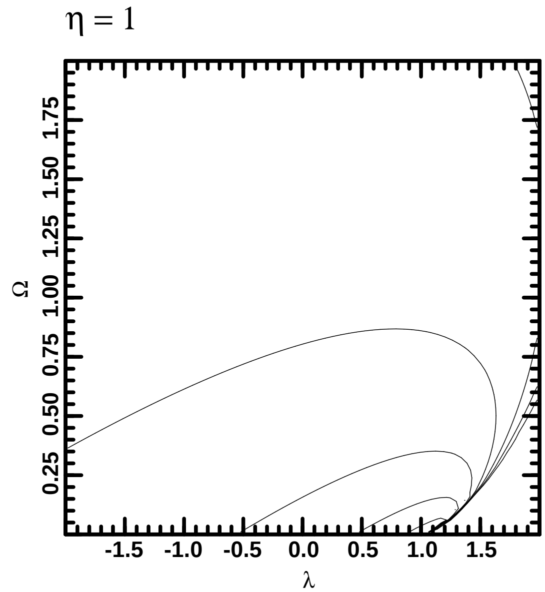

Figure 2 Dependence of D on λ_0 and Ω_0
For $\eta = 1$ $D(\lambda_0, \Omega_0)$ is plotted. The source redshift is $z = 2$. Starting from $(\lambda_0, \Omega_0) = (1, 0)$ and spiraling clockwise, contours are at $0.6, 0.5, 0.4, 0.3, 0.2, 0.1, b$ where b separates the cosmological models with and without a big bang (in the latter the distance is not defined for $z = 2$).

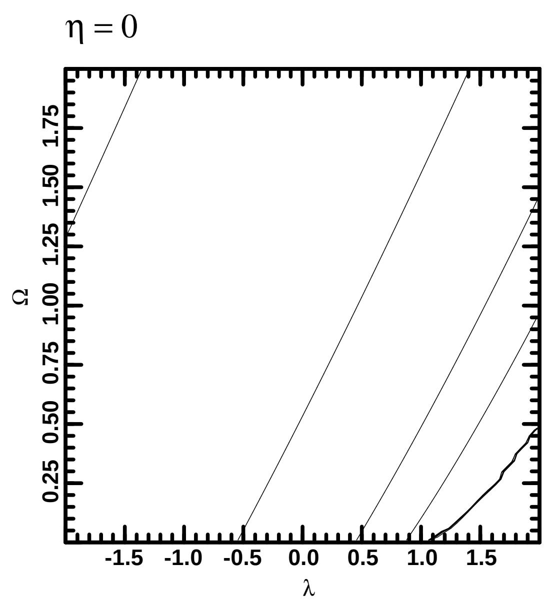

Figure 3 Dependence of D on λ_0 and Ω_0
The same as Fig. 2 but for $\eta = 0$. From upper left to lower right, contours are at $0.3, 0.4, 0.5, 0.6, b$.

ii. Cosmological distances and the effects of a locally inhomogeneous universe

See, *e. g.*, Kayser *et al.* (1997) for an overview of cosmological distances and for a method of taking inhomogeneities into account when calculating cosmological distances. Figures 1–3 show the dependence of the angular size distance (the relevant distance for gravitational lensing) on the cosmological model and on the inhomogeneity parameter η, which is the fraction of smoothly distributed matter within the light cone which determines the distance; $1 - \eta$ is the fraction of the matter distributed clumpily. Here it is assumed that all clumps are outside the light cone ('clumps' inside having been taken into account explicitly as a gravitational lens effect) and far enough away so that the effects of shear can be ignored.

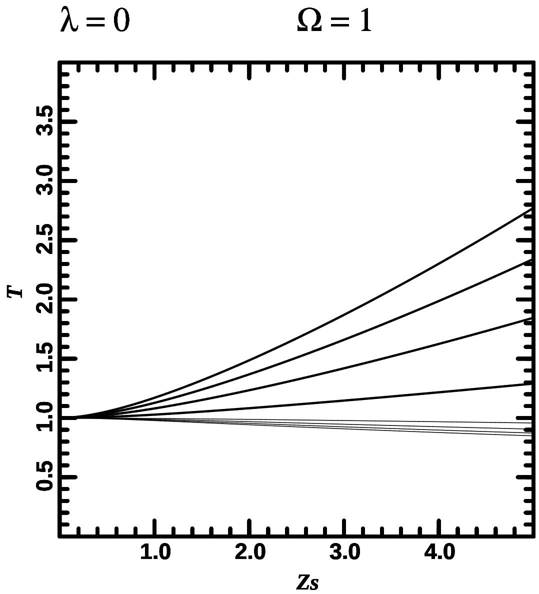

Figure 4 Dependence of T on $D_{\rm s}$, $D_{\rm ds}$ and η
For a fixed cosmological model ($\lambda_0 = 0$ and $\Omega_0 = 1$, as indicated) $T(z_{\rm s})$ is plotted. Thin curves correspond to $\eta = 1$, thick to $\eta = 0$. From top to bottom, $z_{\rm d}/z_{\rm s} = 0.7, 0.5, 0.3, 0.1, 0.1, 0.3, 0.5, 0.7$.

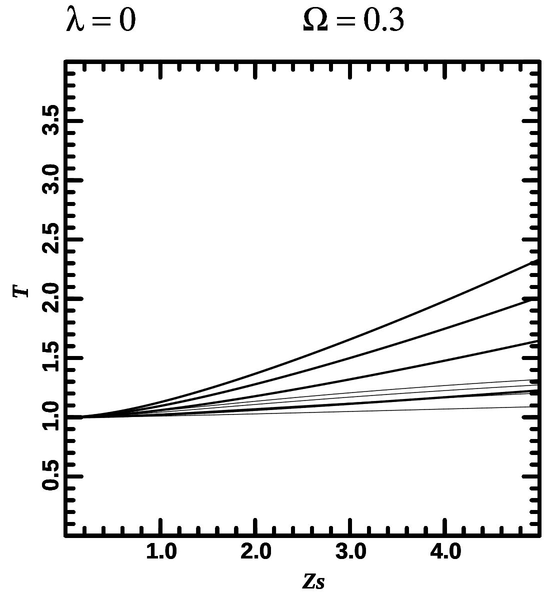

Figure 5 Dependence of T on $D_{\rm s}$, $D_{\rm ds}$ and η
The same as Fig. 4 but for a different values of λ_0 and Ω_0.

c. The cosmological correction function

Figures 4–15 show the dependence of T on the cosmological model. The parameter space examined *roughly* corresponds to cosmological models which cannot be ruled out observationally. Thus, the spread of T gives an idea of the uncertainty in H_0 when determined from a measured time delay, in addition to any uncertainties in (the interpretation of) the measurement itself and the lens model. Alternatively, if H_0 and the lens models are well-constrained by other means, each lens system with a measured time delay provides an independent constraint on λ_0 and Ω_0. The dependence of T on the cosmological parameters comes solely from the influence of the latter on the angular size distances. Since $\eta = 0$ is an extreme case, one could then rule out world models above a contour line such as in Fig. 11; this is interesting since the direction of these contours is such that a degeneracy present in many other cosmological tests—lensing statistics, m-z relation, age of the universe—can be broken.

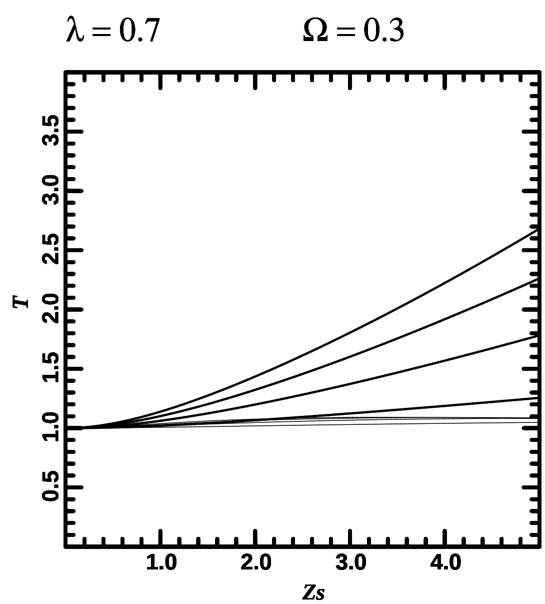

Figure 6 Dependence of T on $D_{\rm s}$, $D_{\rm ds}$ and η
The same as Fig. 4 but for a different values of λ_0 and Ω_0.

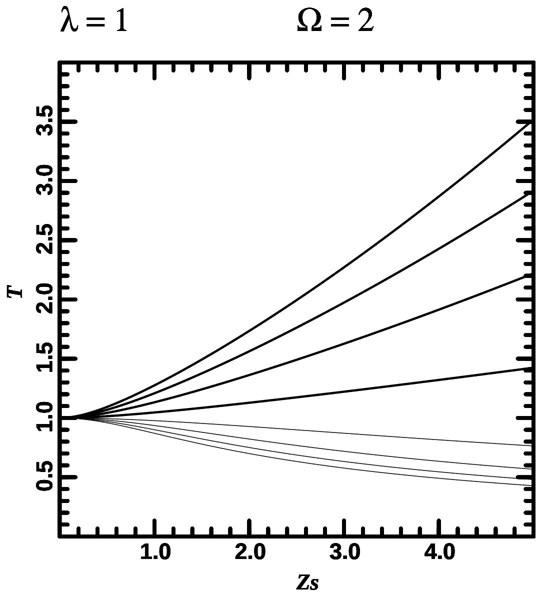

Figure 7 Dependence of T on $D_{\rm s}$, $D_{\rm ds}$ and η
The same as Fig. 4 but for a different values of λ_0 and Ω_0.

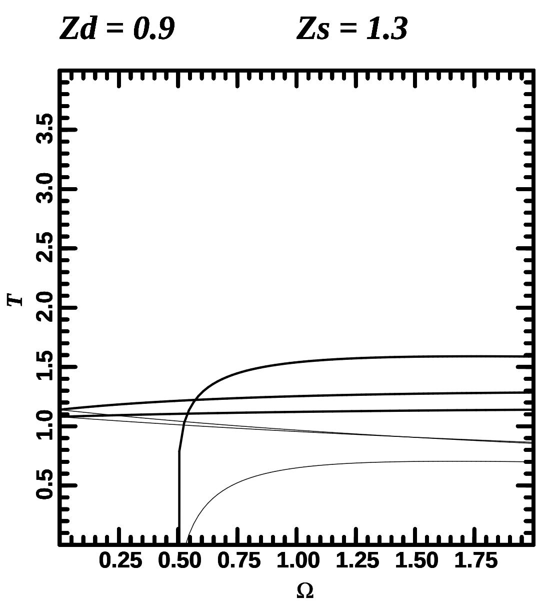

Figure 8 Dependence of T on λ_0, Ω_0 and η
For fixed source and lens redshifts ($z_{\rm d} = 0.9$ and $z_{\rm s} = 1.3$) $T(\Omega_0)$ is plotted. As in Fig. 4, thin curves correspond to $\eta = 1$, thick to $\eta = 0$. The curves for which $T < 0$ for $\Omega_0 < 0.5$ are for $\lambda_0 = 2$; in this case lower values of Ω_0 correspond to the so-called bounce models (see, *e.g.*, Kayser *et al.* (1997)). For the other curves, from top to bottom $\lambda_0 = 1.0, 0.0, 0.0, 1.0$.

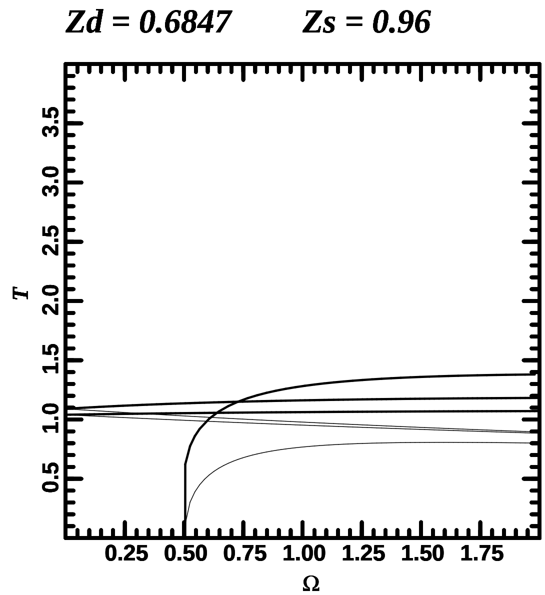

Figure 9 Dependence of T on λ_0, Ω_0 and η
The same as Fig. 8 but with $z_{\rm d}$ and $z_{\rm s}$ equal to the values in the gravitational lens system $0218 + 357$.

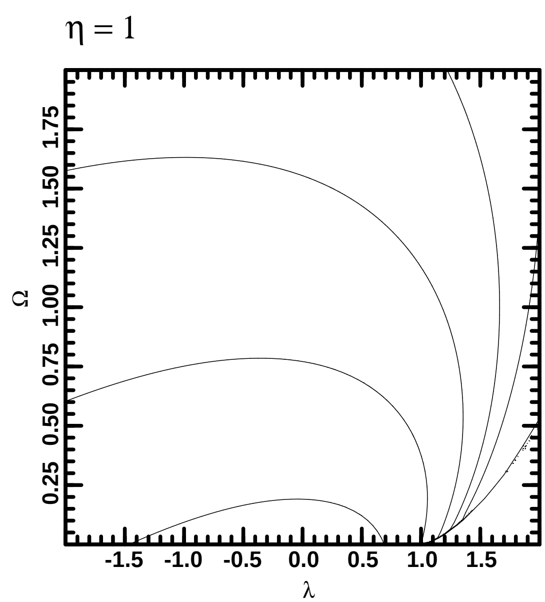

Figure 10 Dependence of T on λ_0 and Ω_0
For fixed source and lens redshifts ($z_{\rm s} = 1.3$ and $z_{\rm d} = 0.9$) $T(\lambda_0, \Omega_0)$ is plotted, here for the case of $\eta = 1$. From $(\lambda_0, \Omega_0) = (0, 0)$ spiraling clockwise, contours are at $1.1, 1.0, 0.9, 0.8, 0.7, b$.

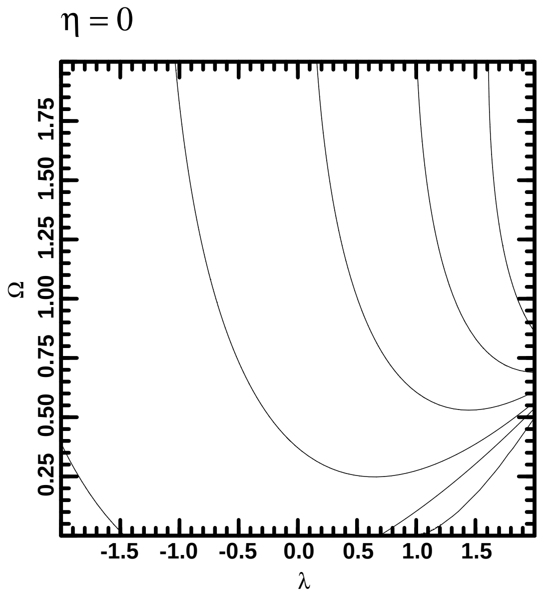

Figure 11 Dependence of T on λ_0 and Ω_0 The same as Fig. 10 but for $\eta = 0$. From lower left to upper right, contours are at $1.1, 1.2, 1.3, 1.4, 1.5$. The contour at lower right is b, the one next to it 1.1.

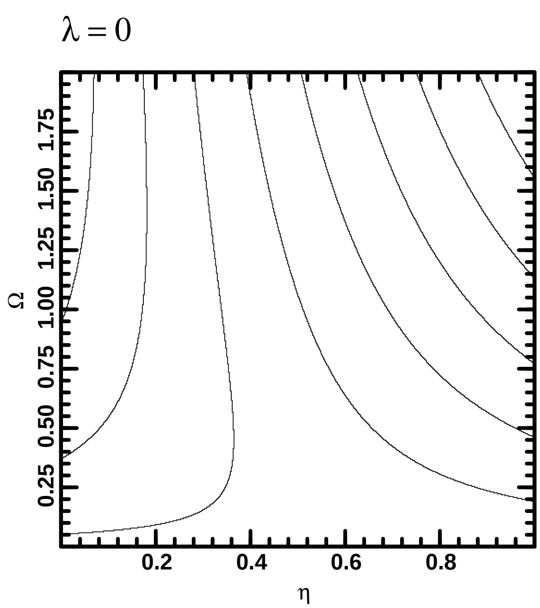

Figure 12 Dependence of T on η and Ω_0 For fixed source and lens redshifts ($z_s = 1.3$ and $z_d = 0.9$) $T(\eta, \Omega_0)$ is plotted, here for the case of $\lambda_0 = 0$. From left to right, contours are at $1.25, 1.20, 1.15, 1.10, 1.05, 1.00, 0.95, 0.90$.

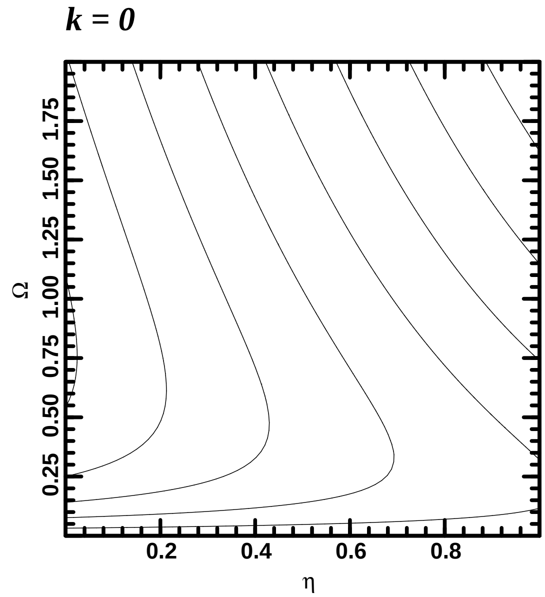

Figure 13 Dependence of T on η and Ω_0 The same as Fig. 12 but for $k = 0$. Contours as in Fig. 12.

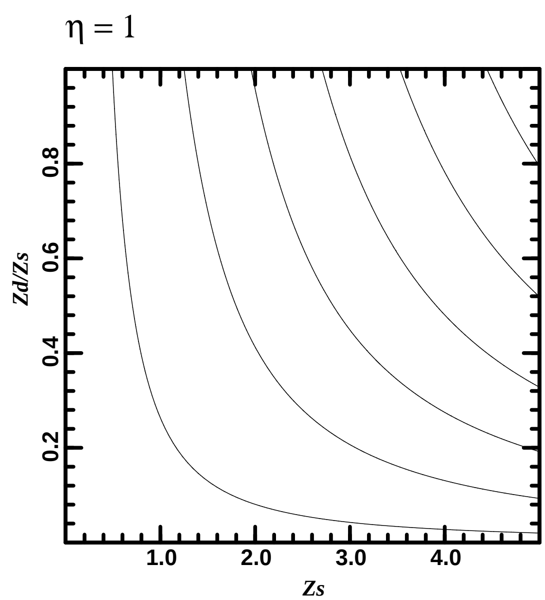

Figure 14 Dependence of T on z_s and z_d For a fixed cosmological model, ($\lambda_0 = 0$ and $\Omega_0 = 1$) $T(z_d, z_s)$ is plotted, here for $\eta = 1$. From lower left to upper right, contours are at $0.99, 0.96, 0.93, 0.90, 0.87, 0.84$.

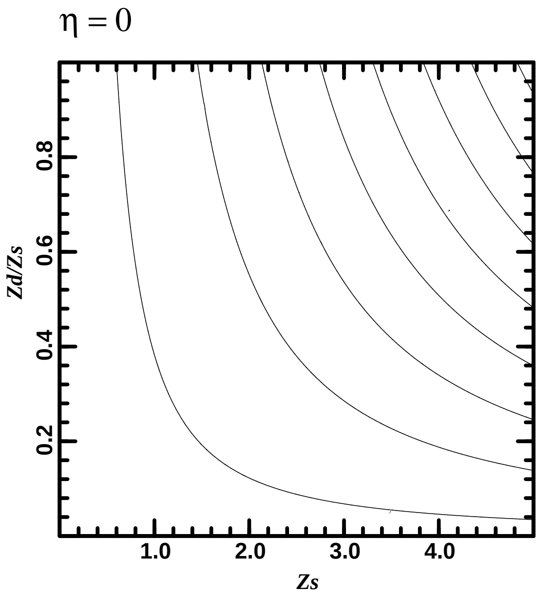

Figure 15 Dependence of T on z_s and z_d **The same as Fig. 12 but for $\eta = 0$. From lower left to upper right, contours are at** $1.1, 1.4, 1.7, 2.0, 2.3, 2.6, 2.9, 3.2$**.**

d. Summary and conclusions

The uncertainty due to cosmological considerations, parametrised by the cosmological correction function T, in the value of H_0 as derived from a measured time delay generally behaves as follows when the other parameters are held constant:[1]

- $|T|$ increases with increasing z_d
- $|T|$ increases with decreasing η
- T increases with z_s for $\eta = 1$ and decreases for $\eta = 0$
- $|T|$ increases with increasing Ω_0
- $|T|$ increases with increasing λ_0 except when Ω_0 is small

in order of generally decreasing importance. Thus, if one is interested in minimising this uncertainty, one should measure the time delay preferentially in systems where z_d/z_s is relatively low and, less important, where z_s itself is small. Should η prove to be ≈ 1 then the need for small source and (relatively) small lens redshifts is less urgent, and the dependence on λ_0 would be made even smaller than it already generally is. Similarly, a small value for Ω_0 would decrease the uncertainties due to η and λ_0. Of course, if one knows H_0 already, then the criteria for desirable source and lens redshifts and for desirable values of the other cosmological parameters are reversed, since then one could use the observations to constrain λ_0 and Ω_0.

It is interesting to contrast the dependency of T on the cosmological parameters λ_0, Ω_0 and η with that of the statistics of multiply-imaged systems in surveys (see, *e. g.*, Fukugita *et al.* (1992)): the order of decreasing importance in the latter case is λ_0, Ω_0 and η, just the opposite as for the case of T considered here.

[1] See also Kayser & Refsdal (1983)

Acknowledgements

This research was supported by the European Commission, TMR Programme, Research Network Contract ERBFMRXCT96-0034 'CERES'.

References

Fukugita, M., K. Futamase, M. Kasai & E. L. Turner: 'Statistical properties of gravitational lenses with a nonzero cosmological constant'
APJ, **393**, 1, 3 (1992)

Kayser, R., P. Helbig, T. Schramm: 'A general and practical method for calculating cosmological distances'
A&A, **318**, 3, 680 (1997)

Kayser, R., S. Refsdal: 'The difference in light travel time between gravitational lens images. I. Generalization of the wavefront method to arbitrary deflectors and inhomogeneous universes'
A&A, **128**, 156 (1983)

Pelt, J., R. Kayser, S. Refsdal, T. Schramm: 'The light curve and the time delay of QSO 0957 + 561'
A&A, **305**, 1, 97 (1996)

Refsdal, S.: 'On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect'
MNRAS, **128**, 4, 307 (1964)

Refsdal, S.: 'On the possibility of testing cosmological theories from the gravitational lens effect'
MNRAS, **132**, 1, 101 (1966)

6.3 Follow-up

Just 30 years ago, H_0 was uncertain by a factor of two. Today, it is uncertain by $\approx 6\%$, with formal uncertainties being much smaller, thus creating 'tension' between various measurements. While the effect of η on H_0 is relatively small, because 'direct' measurements of H_0 necessarily take place at low redshift where the effect of η is small, one does not expect uncertainty in η to be an important contributor to uncertainties in such measurements of H_0 (*e.g.* Odderskov *et al.*, 2016). However, since many cosmological tests actually constrain combinations of cosmological parameters, η can play a role, and indeed Fleury *et al.* (2013) have suggested that taking η into account can explain the 'tension'.

Since H_0 can be determined by measuring time delays in gravitational-lens systems because the time delay depends on the combination of distances $(D_\mathrm{d}D_\mathrm{s})/D_\mathrm{ds}$ which in turn is inversely proportional to H_0, η can potentially have an influence (Kayser and Refsdal, 1983). While that is now largely appreciated, measurements have become so precise that using just the single parameter η to describe the inhomogeneity is too coarse an approximation, so that uncertainty is taken into account by explicitly modelling the observed mass distribution along the line of sight (*e.g.* Rusu *et al.*, 2017; Bonvin *et al.*, 2017). Nevertheless, investigating the influence of η on the derived value of H_0 gives one an estimate of the size of the effect.

At the time, that was a very hot topic. I was one of the main organizers of a workshop at Jodrell Bank (Helbig and Jackson, 1997) to address some of those issues.

Chapter 7

The image-separation–source-redshift relation as a cosmological test

7.1 Context

One-third of my refereed-journal papers point out mistakes in other papers; this is one of them. Park and Gott (1997) claimed that there is an inverse correlation between the image separation $\Delta\theta$ and the source redshift z_{s} for gravitational-lens systems and noted that in a flat universe one expects no correlation at all, *e.g.* the average image separation is independent of z_{s}, as long as a singular isothermal sphere is used to model the lens (while that is mathematically simple, it is also observationally motivated). That the flat universe exhibits this property is not obvious and was probably new information to most readers. The paper suggested two questions to me: Can one use the $\Delta\theta$–z_{s} relation as a cosmological test? What happens to the $\Delta\theta$–z_{s} relation if one does not use only the standard distance, but allows for a universe with small-scale inhomogeneities?

The paper by Park and Gott (1997) was typical of many at the time: complicated formalism which was new to many readers, but sparse data of perhaps low quality. In my paper for *MNRAS* (Helbig, 1998a), I concentrated on the effect of an inhomogeneous universe. It turns out that a negative correlation is expected in the extreme $\eta = 0$ case, regardless of the values of the cosmological parameters. That is even less obvious than the correlation itself, and I was the first to see it. (With the standard distance, a negative correlation is expected for a $k = -1$ universe and a positive one for $k = +1$.) However, at least with the data available at the time, no interesting constraints on the cosmological parameters could be derived. Also, the negative correlation observed by Park and Gott (1997) goes away if one additional gravitational-lens system is included, showing that the test is not very robust. (I included one system which Park and Gott (1997) had missed, 0218+357 (which is also the subject of a paper by Biggs *et al.* (1999)); it had a published source redshift, though only in a proceedings contribution.)

Helbig (1998a) plots the relationship for various cosmological models for the extreme $\eta = 1$ and $\eta = 0$ cases, as well as all observational data in one plot (to save space). For four samples (that used by Park and Gott (1997), that used by Park and Gott (1997) with the addition of 0218+357, the JVAS/CLASS sample (which includes 0218+357), and the union of all samples), relative probabilities in the λ–Ω plane were calculated, but to save space the plots were only discussed by Helbig (1998a), not displayed. In a poster (Helbig, 1998b) for a cosmology workshop in Potsdam (Müller *et al.*, 1998), more space was available; in particular, the individual samples were plotted separately as well as the relative probabilities in the λ–Ω plane for all four samples. Those are shown after the paper. (Since the proceedings contribution (Helbig, 1998b) otherwise has considerable overlap with the paper Helbig (1998a), I am not including the former as a separate chapter, but rather including supplementary material from it in this chapter.)

Mon. Not. R. Astron. Soc. **298,** 395–398 (1998)

The $\Delta\theta$–z_s relation for gravitational lenses as a cosmological test

Phillip Helbig★
University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL

Accepted 1998 March 9. Received 1998 March 6; in original form 1997 November 27

ABSTRACT

Recently, Park & Gott claimed that there is a statistically significant, strong, negative correlation between the image separation $\Delta\theta$ and source redshift z_s for gravitational lenses. This is somewhat puzzling if one believes in a flat ($k = 0$) universe, since in this case the typical image separation is expected to be independent of the source redshift, while one expects a negative correlation in a $k = -1$ universe and a positive one in a $k = +1$ universe. Park & Gott explored several effects that could cause the observed correlation, but no combination of these can explain the observations with a realistic scenario. Here, I explore this test further in three ways. First, I show that in an inhomogeneous universe a negative correlation is expected regardless of the value of k. Secondly, I test whether the $\Delta\theta$–z_s relation can be used as a test to determine λ_0 and Ω_0, rather than just the sign of k. Thirdly, I compare the results of the test from the Park & Gott sample with those using other samples of gravitational lenses, which can illuminate (unknown) selection effects and probe the usefulness of the $\Delta\theta$–z_s relation as a cosmological test.

Key words: cosmology: observations – cosmology: theory – gravitational lensing.

1 INTRODUCTION

Historically, there has been little interest in the $\Delta\theta$–z_s relation compared with other cosmological tests based on gravitational lensing statistics, perhaps because the inflationary paradigm (e.g. Guth 1981), which began about the same time as the discovery of the first gravitational lens (Walsh, Carswell & Weyman 1979), has become so influential. Since a flat ($k = 0$) universe is a robust prediction of inflation, many researchers assume this and consider only flat universes (or, at most, $k = -1$ cosmological models with $\lambda_0 = 0$). Owing to the fact that for the popular singular isothermal sphere model for a single-galaxy lens the average image separation $\Delta\theta$, integrated over the lens redshift z_d from $z_d = 0$ to $z_d = z_s$, is *completely independent* of the source redshift z_s in a flat universe, there is little point in pursuing the $\Delta\theta$–z_s relation if one is interested primarily in flat cosmological models. If one is not committed to a flat universe, then of course one should not assume $k = 0$, but even if one believes that the universe must be flat, it is still important to test this belief observationally. The situation is somewhat worsened by the fact that most ‘standard’ cosmological tests such as the m–z (magnitude–redshift or ‘standard candle’) and θ–z (angular size–redshift or ‘standard rod’) relations, ‘conventional’ gravitational lensing statistics and the age of the universe are relatively insensitive to the radius of curvature of the universe [$R_0 \sim (|\Omega_0 + \lambda_0 - 1|)^{-1/2}$], being degenerate in combinations of λ_0 and Ω_0 in directions roughly perpendicular to lines of constant R_0 in the λ_0–Ω_0 plane. A notable exception is the constraints derived from CMB anisotropies (e.g. Scott, White & Silk 1995; Hu, Sugiyama & Silk 1997).

2 THEORY

For a singular isothermal sphere lens, the angular image separation is given by (e.g. Turner, Ostriker & Gott 1984)

$$\Delta\theta = 8\pi\left(\frac{v}{c}\right)^2\frac{D_{ds}}{D_s}, \tag{1}$$

where v is the velocity dispersion and D is the angular size distance (see below). Even if the singular isothermal sphere is not a perfect model for the gravitational lens systems considered, it is still a good approximation when one is concerned only with the image separation. For a given v, by combining equations (5) and (6) of Gott, Park & Lee (1989) and using the more appropriate and more general angular size distances, one obtains an expression for the average image separation $\Delta\theta$, by integrating over the lens redshift z_d from $z_d = 0$ to $z_d = z_s$:

$$\frac{\Delta\theta(z_s)}{\Delta\theta(0)} = \frac{\left(\int_0^{z_s} dz_d \, \frac{D_{ds}^3 D_d^2 (1+z_d)^2}{D_s^3 Q}\right)}{\left(\int_0^{z_s} dz_d \, \frac{D_{ds}^2 D_d^2 (1+z_d)^2}{D_s^2 Q}\right)}, \tag{2}$$

where

$$Q = \sqrt{\Omega_0(1+z_d)^3 - (\Omega_0 + \lambda_0 - 1)(1+z_d)^2 + \lambda_0}\,. \tag{3}$$

The D_{ij} (with $D_k := D_{0k}$) in equations (1) and (2) are angular size distances, which are functions of the lens and source redshifts z_d and

★E-mail: p.helbig@jb.man.ac.uk

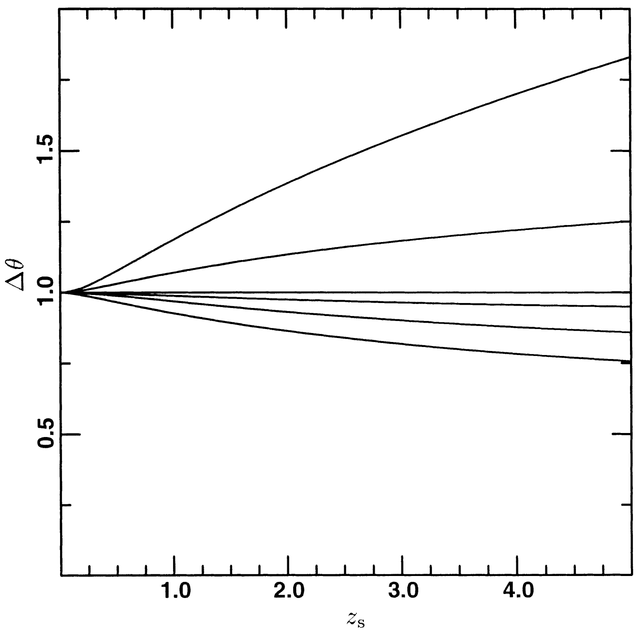

Figure 1. Normalized image separation as a function of source redshift. From the top, the (λ_0, Ω_0) values are (2, 4), (0, 4), $k = 0$, (0, 0.7), (0, 0.3) and (−5, 1). For $k = 0$ the result is valid for all (λ_0, Ω_0) values the sum of which is 1. $\eta = 1$.

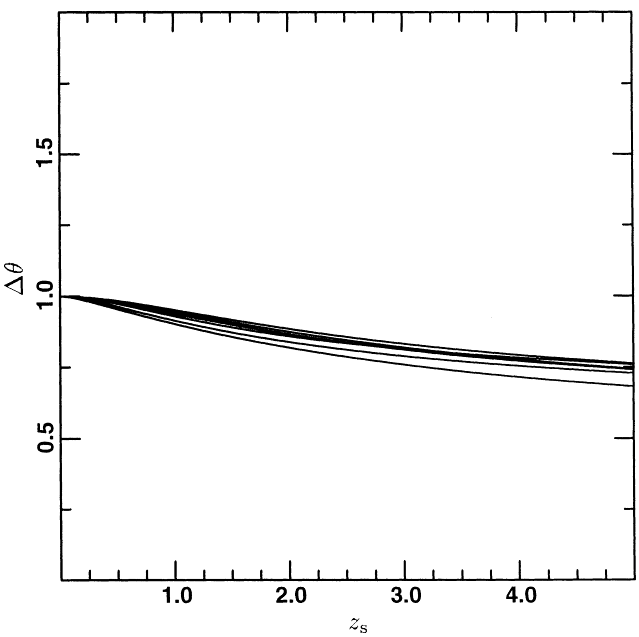

Figure 2. The same as Fig. 1 except that here $\eta = 0$.

z_s, the cosmological parameters λ_0 and Ω_0 as well as the 'homogeneity parameter' η, which gives the fraction of smoothly, as opposed to clumpily, distributed matter along the line of sight. Note that equation (2) is valid for all combinations of λ_0, Ω_0 and η. The angular size distances can be computed for arbitrary combinations of these parameters by the method outlined in Kayser, Helbig & Schramm (1997).

Figs 1 and 2 show $\Delta\theta$ as a function of z_s for various cosmological models, for $\eta = 1$ (the traditional case assuming a completely homogeneous universe) and $\eta = 0$ as extreme cases. Note in Fig. 1 that the curve is a horizontal line for $k = 0$, has positive slope for $k = +1$ and negative slope for $k = -1$, where $k := \mathrm{sign}(\Omega_0 + \lambda_0 - 1)$. In Fig. 2, for $\eta = 0$, the slope is negative regardless of the value of k. Thus, at first sight it appears that an inhomogeneous universe, a possibility not investigated by Park & Gott (1997, hereafter PG), might be able to explain the puzzling negative correlation between $\Delta\theta$ and z_s. However, it is shown in Section 5 that even the extreme $\eta = 0$ scenario produces an anti-correlation which is much weaker than that found by PG. This effect can be qualitatively understood by realizing how equation (2) is affected by decreasing η: inspection shows that this might be estimated by examining D_{ds}/D_s. All other things being equal, the angular size distance increases with decreasing η. Also, the effect of η is more noticeable at large redshift differences. Since $z_s \geq z_s - z_d$, the denominator is the more important term, and so decreasing η increases D_s and so decreases D_{ds}/D_s and thus $\Delta\theta(z_s)/\Delta\theta(0)$.

3 DATA

PG used an inhomogeneous sample of gravitational lenses from the literature. While this seems problematic at first sight, PG noted that there is no reason to believe that this should influence the analysis. Nevertheless, it is worth comparing the PG results with those obtained from a better defined sample.

The observational data provided by the JVAS and CLASS surveys offer an independent sample of gravitational lenses. JVAS is the Jodrell Bank VLA Astrometric Survey (Patnaik et al. 1992); CLASS is the Cosmic Lens All-Sky Survey (Myers et al., in preparation). Even though the observational tasks are not yet complete, the JVAS and CLASS surveys which constitute the data base have already yielded sufficient gravitational lenses to enable one to make an independent analysis. Table 1 shows the current state of knowledge about the JVAS/CLASS gravitational lenses. Note that the questionable source redshift for 2114 + 022 is probably the redshift of an additional lensing galaxy (this interpretation is supported by several independent lines of evidence).

Although not all source redshifts in the JVAS/CLASS sample are known, 8 out of 11 are, and based on our survey, discovery and follow-up strategies there is no reason to suspect the unknown source redshifts to be statistically different from those already known. Fig. 3 shows the source redshifts and image separations of the gravitational lens systems used in this paper: the PG sample and the JVAS/CLASS sample.

4 CALCULATIONS

All calculations here implement the method of PG, which uses the Spearman rank correlation test to generate a relative probability for a given cosmological model. PG noted the fact that they always obtained a low probability with their sample, even when allowing for non-flat cosmological models (albeit in a limited area of parameter space), galaxy evolution or departure from the singular isothermal sphere model. As PG noted, allowing for these effects increases the probability, since they all tend to create a negative correlation in a flat universe, but the magnitude of the effect is not large enough to explain the observations. Again as noted by PG, if the lenses are parts of clusters, then this will work in the opposite direction, making the observed negative correlation even more puzzling.

Calculations were performed for four samples:

(i) the PG sample,

Table 1. The JVAS/CLASS gravitational lenses.

Name	# images	$\Delta\theta$ [arcsec]	lens galaxy type	z_d	z_s
0218 + 357	ring +2	0.33	spiral	0.6847	0.96
0414 + 0534	4	2.0	elliptical	?	2.62
0712 + 472	4	1.2	?	0.406	1.339
1030 + 074	2	1.6	peculiar	0.599	1.535
1422 + 231	4	1.2	?	0.65	3.62
1600 + 434	2	1.4	spiral	0.4144	1.589
1608 + 656	4	2.2	spiral?	0.64	1.39
1933 + 503	4 + 4 + 2	0.9	?	0.755	?
1938 + 666	4 + 2	0.9	?	?	?
2045 + 265	4 + 1?	2.0	?	0.87	1.28
2114 + 022	2 + 2?	2.4	?	0.316	0.588?

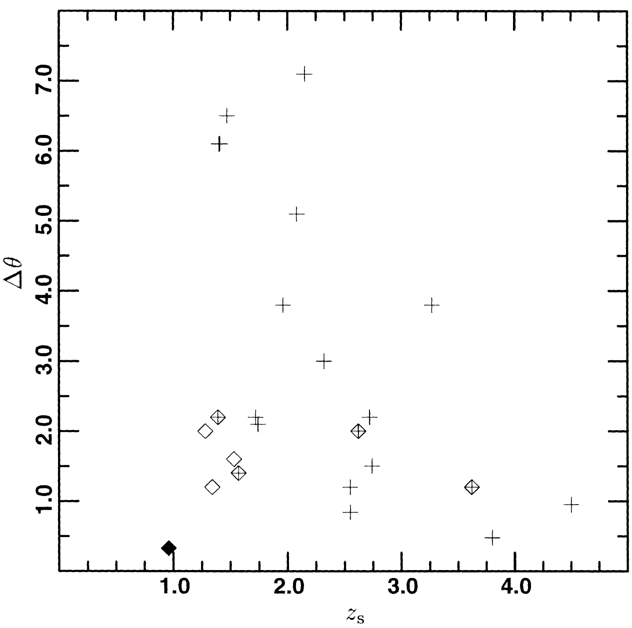

Figure 3. Source redshifts z_s and image separations $\Delta\theta$ (in arcsec) for the gravitational lens systems studied in this paper. Crosses represent the PG sample (20 systems; note that two data points with $\Delta\theta \approx 6$ arcsec almost coincide); diamonds represent the JVAS/CLASS sample (eight systems; of course only those with known source redshifts are included). Note that there is an overlap of four data points. The filled diamond represents the system 0218 + 357, which was not used by PG although its source redshift had been published before the PG analysis was carried out (Lawrence 1996).

(ii) the PG sample with the addition of the system 0218 + 357,
(iii) the JVAS/CLASS sample, and
(iv) the union of all samples.

Note that the source redshift of 0218 + 357 had been published before the PG analysis was performed (Lawrence 1996). Since 0218 + 357 lies below and to the left of all other data points, it is clear that including it will weaken the puzzling negative correlation found by PG; this is discussed more quantitatively in Section 5.

5 RESULTS AND DISCUSSION

Since the PG test assigns a low probability to a $k = 0$ universe, the question arises as to whether it can be used as a general cosmological test to determine the values of λ_0 and Ω_0. This is not the case. For all four samples I have calculated the Spearman rank correlation probability as a function of λ_0 and Ω_0 in a range of parameter space ($-8 < \lambda_0 < 2$ and $0 < \Omega_0 < 10$) much larger than that allowed even by a generous interpretation of observations. This was done with a resolution of 0.1 in both λ_0 and Ω_0 for both $\eta = 1$ and $\eta = 0$. The Spearman rank correlation probability is essentially constant over a wide range of parameter space; basically, either all cosmological models are probable, or all are improbable, depending on the sample used.

The probability is a weak function of the cosmological model, with the sharpest transition occurring when crossing the $k = 0$ line in the λ_0–Ω_0 plane. For all samples except the PG sample, the probability is > 5 per cent in almost the entire parameter space;[1] those cosmological models with a lower probability are among those ruled out by current observations. Thus, the Spearman rank correlation probability does not allow one to reject any otherwise viable cosmological models, which shows both that there is no reason to expect unknown effects in the gravitational lens samples and that this probability is not very useful as a cosmological test. For the PG sample, the 1 per cent contour corresponds almost exactly to the $k = 0$ line, with higher values for a negatively curved universe. Thus, the PG sample is *marginally* compatible with a $k = -1$ cosmological model, although the probability values are low throughout the λ_0–Ω_0 plane, with values near the maximum of 0.025 being attained only for small (but realistic) Ω_0 values and large (in absolute value) negative values of λ_0. Since there are no known selection effects that can account for the differences between the PG sample and other samples, either the test is not very useful and/or it is pointing to unknown selection effects in the literature sample used by PG. The fact that the PG result changes dramatically (probability ≈10–20 per cent in most of the λ_0–Ω_0 plane) by the inclusion of just one additional data point, which could have been included in their analysis, argues in favour of the former possibility.

The above discussion was for $\eta = 1$. For $\eta = 0$ the situation is qualitatively the same and quantitatively involves only slightly different values of probabilities derived from the Spearman rank correlation test.

It is interesting to compare the probabilities from the Spearman rank correlation test for the PG sample using the actual values of z_s and $\Delta\theta$ as used by PG to those obtained using more up-to-date data for the *same* lens systems. If two values are very near each other, rounding them off to the same values produces a different result for the rank correlation test than if they differed by even a small

[1] For the JVAS/CLASS sample, the maximal probability is 0.955 and is realized in almost the entire $k = +1$ area of the parameter space.

amount. Using more up-to-date data, an even lower probability is obtained for the PG sample, for $\eta = 1$ and $\eta = 0$, for a wide variety of cosmological models.

6 CONCLUSIONS

Park & Gott (1997) pointed out that the image separations in gravitational lens systems show a strong significant negative correlation with the source redshift, while in a flat universe one would expect no correlation (a negative correlation would be expected in a universe with negative curvature and a positive one in a universe of positive curvature). None of the possibilities they examined was strong enough to explain the effect. A possibility not examined by them, namely an inhomogeneous universe, produces a negative correlation regardless of the sign of the curvature, but it too is not strong enough to account for the effect. As a general test for the values of λ_0 and Ω_0 the test is of no use; all cosmological models are assigned roughly the same probability, but *which* value they are assigned depends on the sample used.

The strong dependence of the result on the sample used seems to indicate that the result of Park & Gott (1997) is due not to some physical cause but rather to unidentified selection effects in the sample of gravitational lenses taken from the literature. The large number of JVAS and CLASS lenses gives us an independent comparison sample, thus demonstrating the need for discovering a large number of lenses in a well-defined sample. As Park & Gott (1997) point out, since many conclusions based on ‘conventional’ gravitational lensing statistics are based on essentially the same lenses as in their literature sample, if this sample is for some unknown reason atypical then conclusions drawn from statistical analyses of it must be examined with care. It will thus be interesting to see what conclusions can be drawn from a statistical analysis of the JVAS/CLASS sample after the observational tasks have been completed. (We expect to find more lenses, but have no qualms about using the present incomplete sample in this analysis since there is no reason to believe that a larger sample would show a different $\Delta\theta$–z_s relation.)

ACKNOWLEDGMENTS

I thank Asantha Cooray for comments on the manuscript and my collaborators in the CERES project for helpful discussions. This research was supported by the European Commission, TMR Programme, Research Network Contract ERBFMRXCT96-0034 ‘CERES’.

NOTE ADDED IN PRESS

Since this work was completed, two other responses to Park & Gott (1997) (apart from Helbig 1998) have appeared. The first (Williams 1997) is complementary to this work in that it assumes the effect is real and explores the astrophysical consequences, while the second (Cooray 1998) is more similar to this analysis, arriving at essentially the same conclusions though using different observational data (and exploring neither the question of usefulness as a general test for λ_0 and Ω_0 nor the effects of a locally inhomogeneous universe).

REFERENCES

Cooray A., 1998, AJ, submitted (astro-ph/9711179)
Gott J. R., III, Park M. G., Lee H. M., 1989, ApJ, 338, 1
Guth A. H., 1981, Phys. Rev. D, 23, 347
Helbig P., 1998, in Müller V., Gottlöber S., Mücket J. P., Wambsganß J., eds., Large Scale Structure: Tracks and Traces. World Scientific, Singapore, p. 319
Hu W., Sugiyama N., Silk J., 1997, Nat, 386, 37
Kayser R., Helbig P., Schramm T., 1997, A&A, 318, 680
Lawrence C. R., 1996, in Kochanek C. S., Hewitt J. N., eds, Astrophysical Applications of Gravitational Lensing. Kluwer Academic Publishers, Dordrecht, p. 299
Park M. G., Gott J. R., III, 1997, ApJ, 489, 476 (PG)
Patnaik A. R., Browne I. W. A., Wilkinson P. N., Wrobel J. M., 1992, MNRAS, 254, 655
Scott D., White M., Silk J., 1995, Sci, 268, 829
Turner E. L., Ostriker J. P., Gott J. R., III, 1984, ApJ, 284, 1
Walsh D., Carswell R. F., Weymann R. J., 1979, Nat, 279, 381
Williams L. L. R., 1997, MNRAS, 292, L27

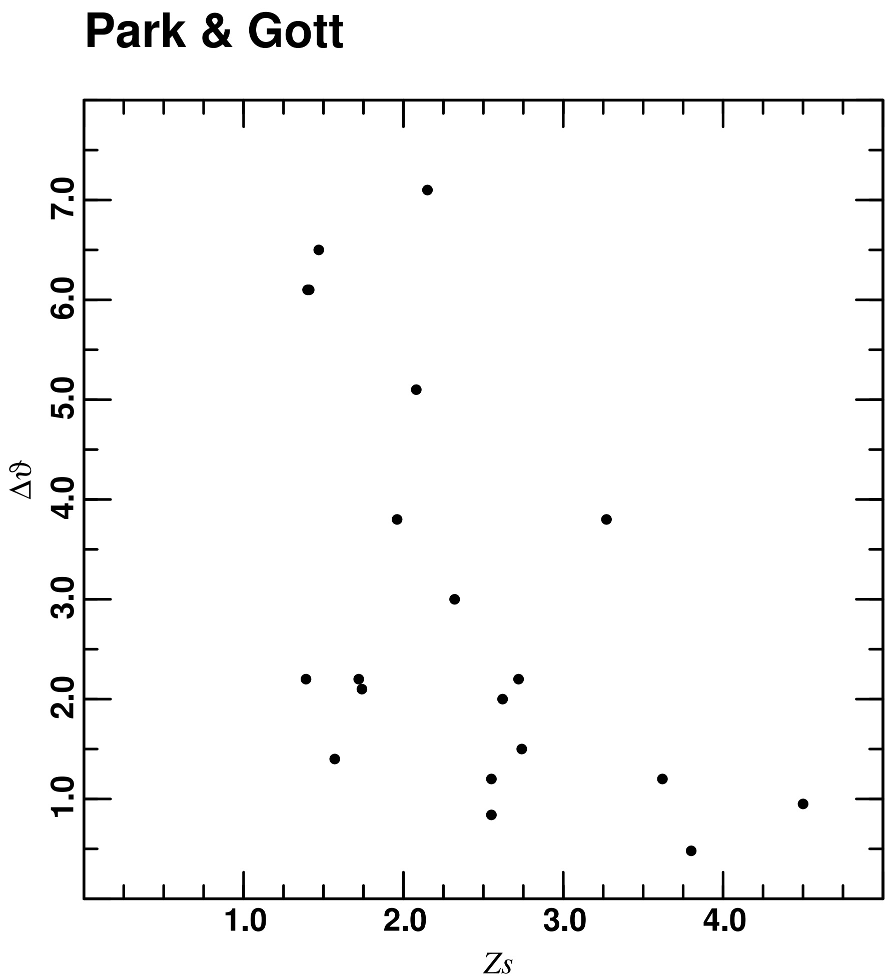

Figure 7.1: The sample used by Park & Gott.

7.3 Further details and follow-up

I show each of the four samples used and the resulting relative probability in the λ–Ω plane. As discussed by Helbig (1998a), the probability is more or less the same for a wide range of cosmological parameters and thus not useful as a cosmological test. (Although included in Helbig (1998a) (though with different symbols for different samples), for clarity Fig. 7.7 is included below as well.)

Park & Gott

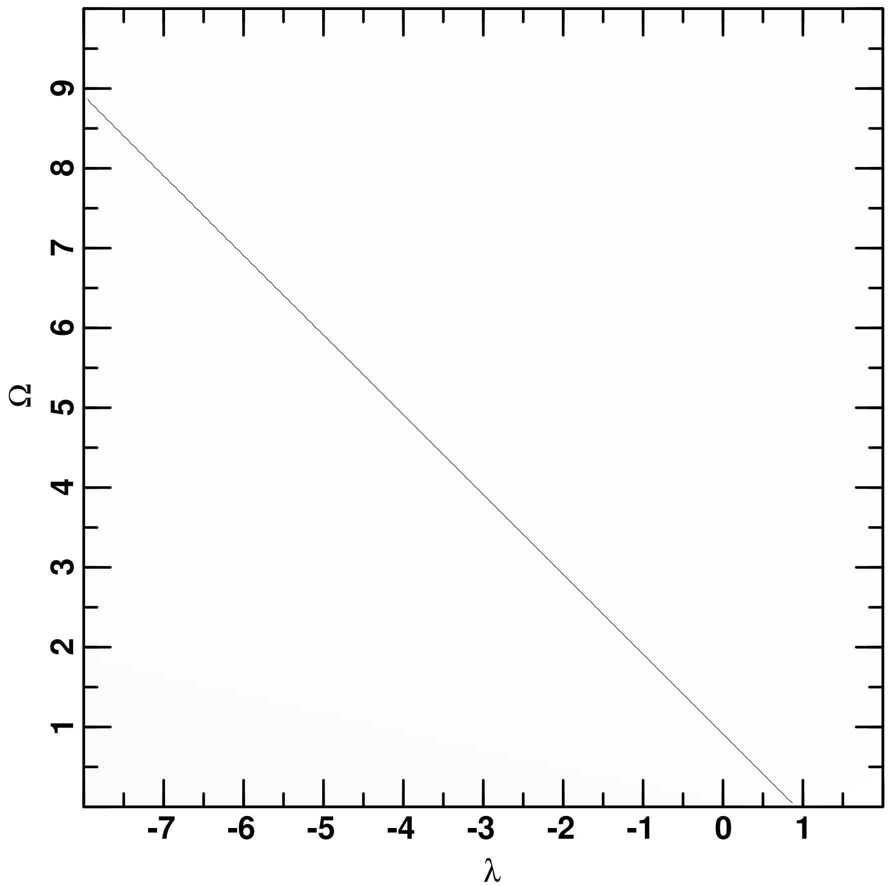

Figure 7.2: Spearman rank-correlation probabilities as a function of λ_0 and Ω_0 for the Park & Gott sample. The figures plotting the probability all utilize the same linear grey scale with 0 being white and 1 black. The highest value on this plot is 0.025, barely visible in the lower left corner. The diagonal line is the 1% contour level. The area above contains smaller values; larger ones are below.

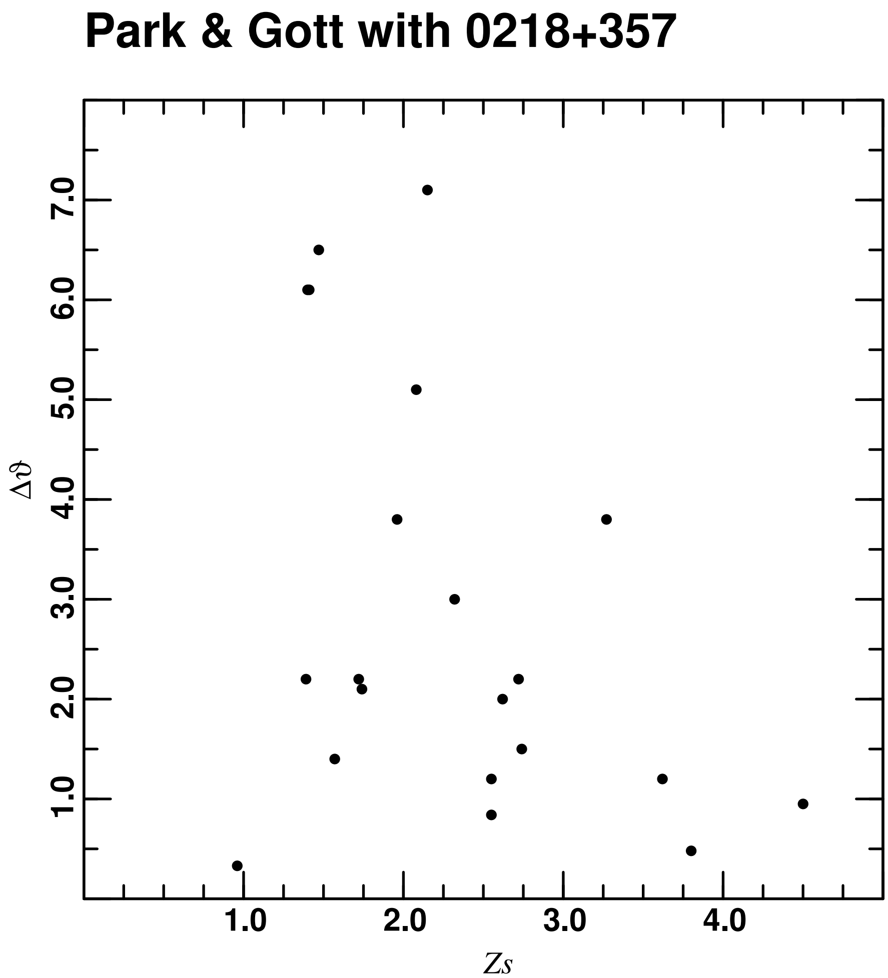

Figure 7.3: The sample used by Park & Gott with the addition of the gravitational lens system $0218 + 357$.

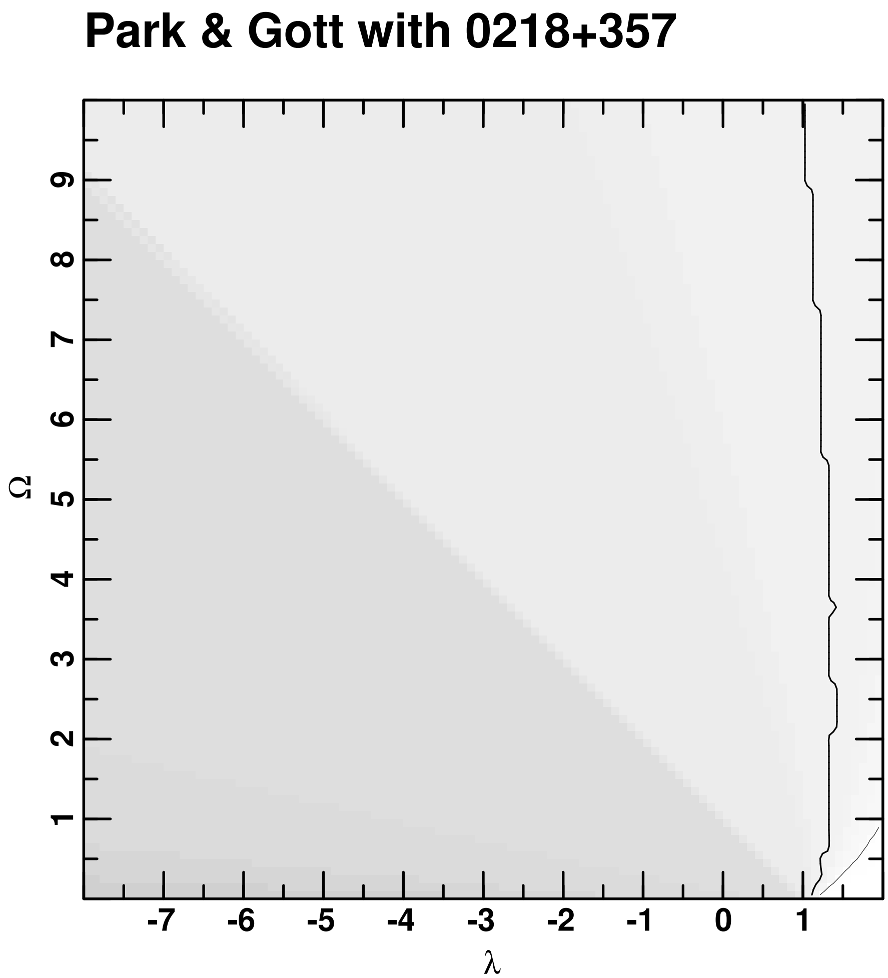

Figure 7.4: The same as Fig. 7.2 but for the Park & Gott sample with the addition of $0218 + 357$. The maximum is 0.184. The thin curve is the 1% contour, the thick curve is the 5% contour.

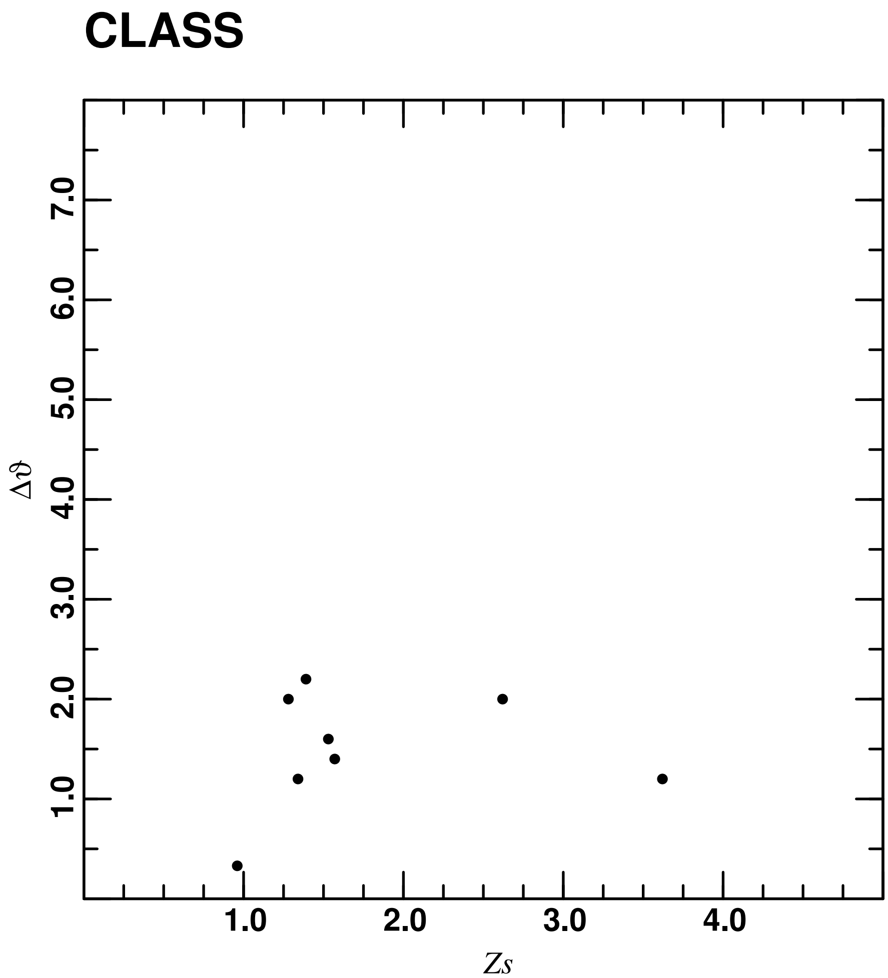

Figure 7.5: The CLASS sample of gravitational lenses.

Helbig (1998a) discussed the sensitivity of the Spearman rank probability to roundoff error; in particular, if two values are very near each other, rounding them off to the same values produces a different result for the rank-correlation test than if they differ by even a small amount. Another aspect of round-off error is seen in computing the Spearman rank probability for $k = 0$ models. Park and Gott (1997) give a probability of 0.012 for a flat Universe. I can reproduce that value by using $\lambda_0 = 0.5$ and $\Omega_0 = 0.5$. Other values of λ_0 and Ω_0 (with the sum of 1, corresponding to $k = 0$) result in values between 0.008 and 0.017, while inserting $k = 0$ 'by hand' instead of doing the computations for explicit λ_0 and Ω_0 values (that is, using the fact that the $\Delta\theta$–z_{s} relation is flat in this case rather than computing it) results in a value of 0.011. That and the problem with inexact observations mentioned above suggest that the probabilities computed using the Spearman rank-correlation test should be taken with a grain of salt.

While the $\Delta\theta$–z_{s} relation alone is not useful as a cosmological test, the image separation is of course observational data which can be used, along with other observational data, in 'full-scale' lens-statistics analyses (*e.g.* Quast and Helbig, 1999; Helbig *et al.*, 1999; Helbig, 1999; Macias Perez *et al.*, 2000; Chae *et al.*, 2002). The same is true of the lens redshift: not useful as a cosmological test if used alone (Helbig and Kayser, 1996a), but such data are used in 'full-scale'

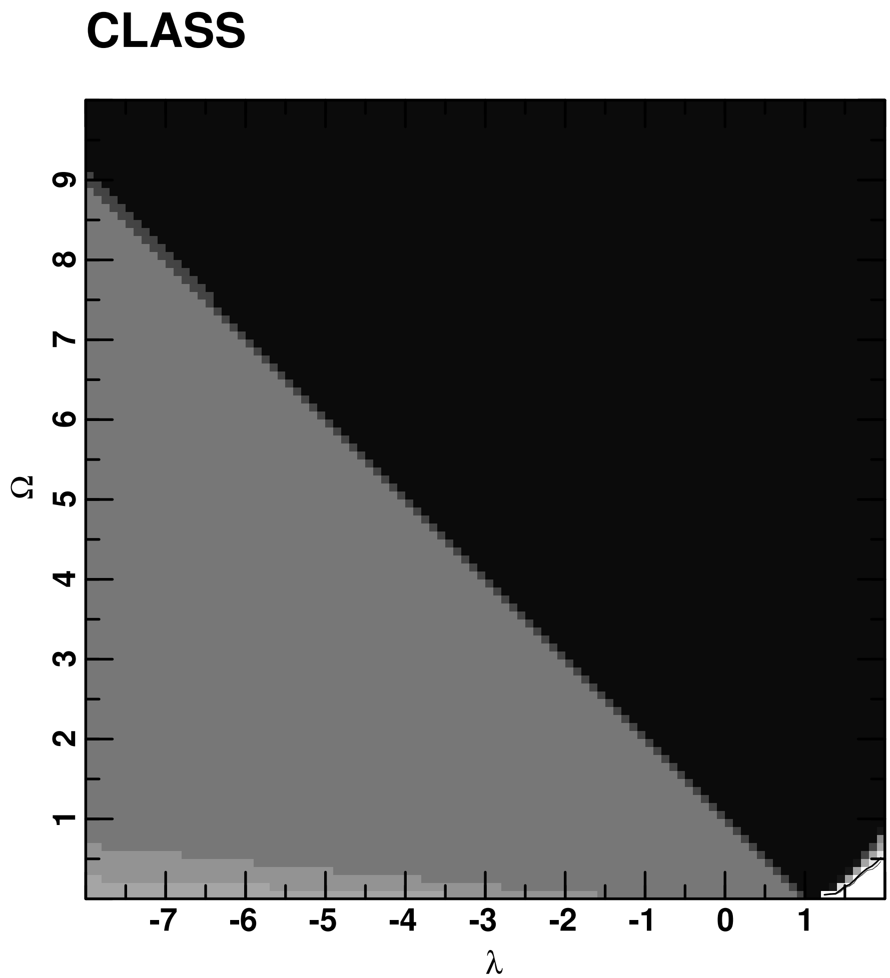

Figure 7.6: The same as Fig. 7.2 but for the CLASS sample. The maximum is 0.955.

lens-statistics analyses.

This paper (Helbig, 1998a) is one of my favourites among my own papers: it corrects an erroneous claim, presents interesting new information, and is short and to the point. Like many of my papers, it is theoretical though closely connected with observations and is concerned with the intersection of cosmology, gravitational lensing, and statistics.

As is often the case with papers which mainly point out errors in other papers, Helbig (1998a) has received only a handful of citations (Cooray, 1999; Zhu, 2000; Dev *et al.*, 2004; McKean *et al.*, 2004; Han and Park, 2015; Rana *et al.*, 2017a,b). Cooray (1999) cites it only for the incidental information which lens systems have spiral galaxies as lenses. Zhu (2000) notes that (not stated, but assuming $\eta = 1$) one can use the effect "to test directly the curvature of the universe". Dev *et al.* (2004) mention it as one example of using gravitational lensing to study cosmology, and go on to look at other aspects of the image separation in that context. McKean *et al.* (2004) report results from Keck spectroscopy of CLASS lens systems, noting that redshift information is necessary for the test discussed by Helbig (1998a). Han and Park (2015) discuss my work in a bit more detail, including the effect of η and the statistical problems with a small sample, hence their work applying the test to the Sloan Digital Sky

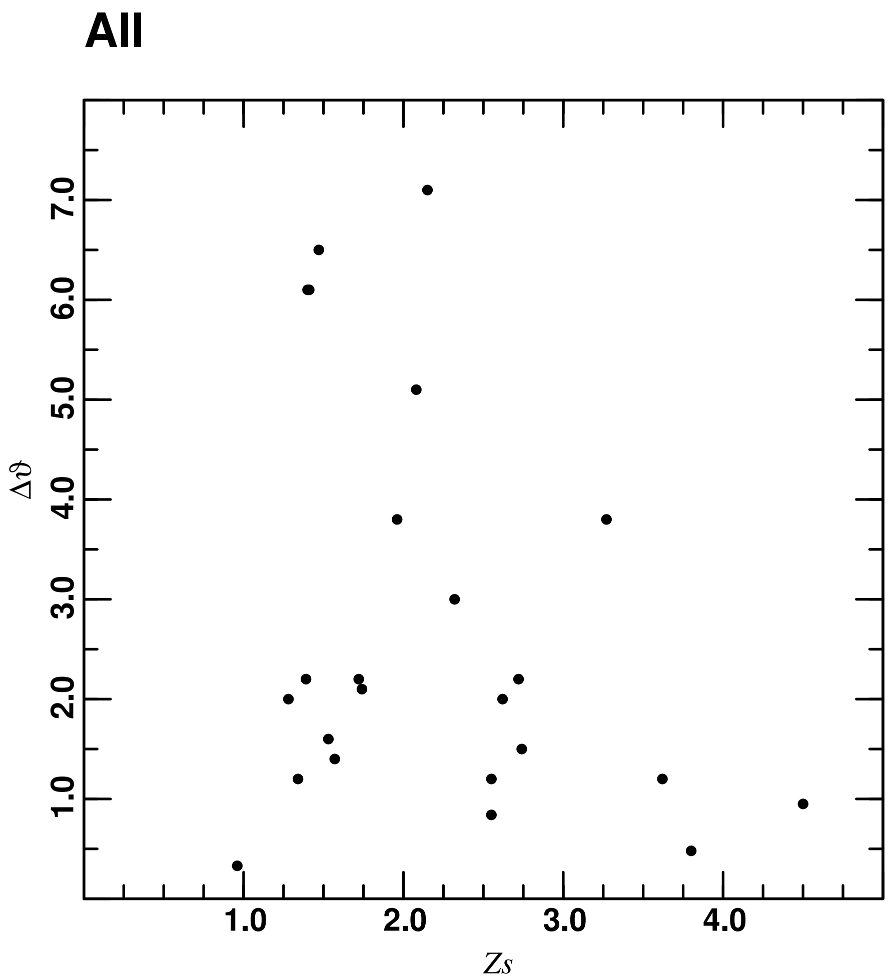

Figure 7.7: Combined sample: union of Park & Gott and CLASS samples.

Survey. Rana *et al.* (2017a), among other things, repeat the test with more data, and also cite Kayser, Helbig and Schramm (1997) for mathematical expressions concerning cosmological distances. Rana *et al.* (2017b) mention my work merely as an example of using gravitational lensing to study cosmology, along with another of my papers (Helbig *et al.*, 1999).

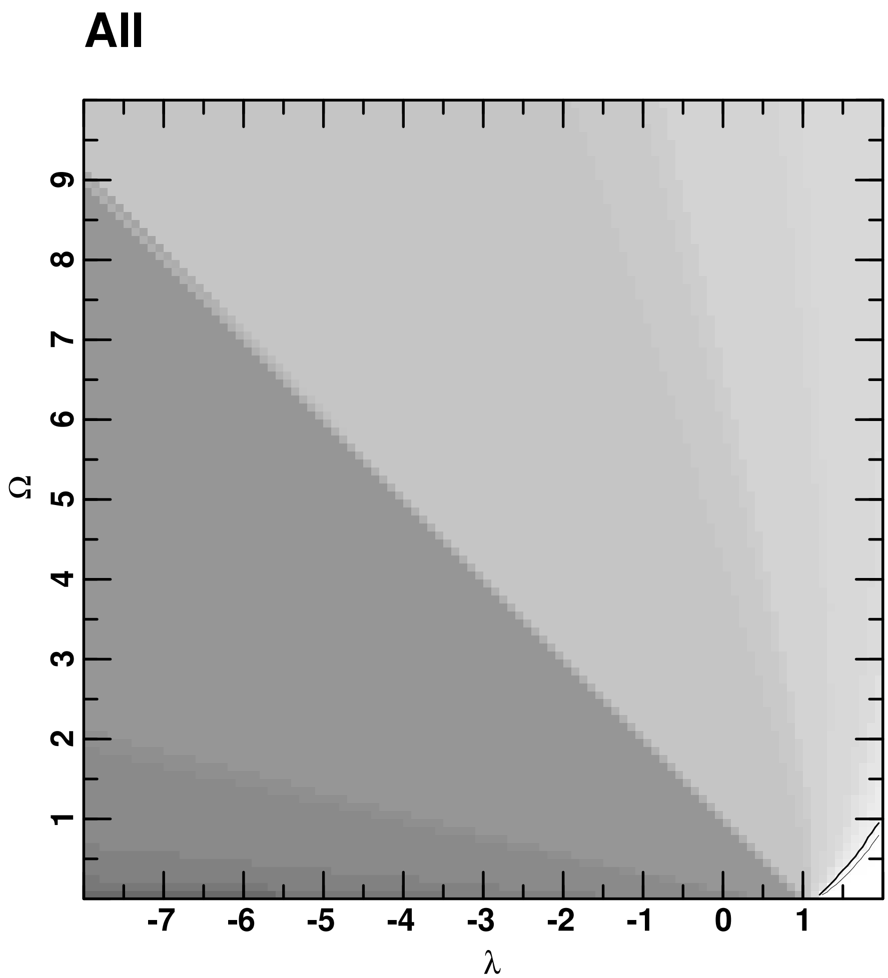

Figure 7.8: The same as Fig. 7.2 but for the combined sample. The maximum is 0.583.

Part V

Type-Ia supernovae

Chapter 8

The magnitude–redshift relation for Type Ia supernovae, locally inhomogeneous cosmological models, and the nature of dark matter

8.1 Context

Although Perlmutter *et al.* (1999) had investigated the influence of η on the values of λ_0 and Ω_0 obtained from the m–z relation for Type Ia supernovae, using the code of Kayser, Helbig and Schramm (1997), in later studies that was conspicuous by its absence. As a result, it had been in the back of my mind to investigate that, because more and higher-redshift data had become available, but the concrete motivation was something else. I had written a FORTRAN program to reduce n-dimensional data cubes; given data on an n-dimensional grid, one can reduce them to a grid of one dimension fewer in various ways: marginalizing (averaging) over the dimension to be eliminated, maximizing it (*i.e.* using the highest value rather than the average), or taking a cut through the higher-dimensional cube at a specified value of the dimension to be eliminated. The goal is often to obtain constraints in a two- or one-dimensional parameter space, which can then be plotted. Another approach is to calculate contours in the higher-dimensional space and project them onto one with one dimension fewer. To test those procedures, I wanted to use real data and also have a feel for the influence of the parameters on the calculated quantity, so I chose the likelihood derived from the m–z relation for Type Ia supernovae as a function of λ_0, Ω_0, and η, using publicly available data to calculate a three-dimensional data cube as a starting point. Since the results were scientifically interesting, I wrote them up (Helbig, 2015a).

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **451,** 2097–2107 (2015) doi:10.1093/mnras/stv1074

The *m*–*z* relation for Type Ia supernovae, locally inhomogeneous cosmological models, and the nature of dark matter

Phillip Helbig★
Thomas-Mann-Str. 9, D-63477 Maintal, Germany

Accepted 2015 May 11. Received 2015 May 9; in original form 2015 February 15

ABSTRACT

The *m*–*z* relation for Type Ia supernovae is one of the key pieces of evidence supporting the cosmological 'concordance model' with $\lambda_0 \approx 0.7$ and $\Omega_0 \approx 0.3$. However, it is well known that the *m*–*z* relation depends not only on λ_0 and Ω_0 (with H_0 as a scale factor) but also on the density of matter along the line of sight, which is not necessarily the same as the large-scale density. I investigate to what extent the measurement of λ_0 and Ω_0 depends on this density when it is characterized by the parameter η ($0 \le \eta \le 1$), which describes the ratio of density along the line of sight to the overall density. I also discuss what constraints can be placed on η, both with and without constraints on λ_0 and Ω_0 in addition to those from the *m*–*z* relation for Type Ia supernovae.

Key words: supernovae: general – cosmological parameters – cosmology: theory – dark energy – dark matter.

1 INTRODUCTION

In the last 15 years or so, cosmological observations have improved greatly and it also appears that the values are converging on their true values.[1] This allows us to answer such questions (provided, of course, that 'standard assumptions' hold) as whether the Universe will expand for ever (yes), how old it is, whether it is accelerating now (yes), when it started accelerating, etc. (With the assumption of a simple topology, the Universe is finite if $\lambda_0 + \Omega_0 > 1$, but since observations indicate that this value is very close to 1, we cannot yet answer this question.) Among the most important of these observations are those by the Supernova Cosmology Project (e.g. Goobar & Perlmutter 1995; Perlmutter et al. 1995, 1998, 1999; Amanullah et al. 2010; Suzuki et al. 2012) and the High-*z* Supernova Search team (e.g. Garnavich et al. 1998; Riess et al. 1998; Riess et al. 2000) (see also the reviews by Riess 2000, Leibundgut 2001, 2008, and Goobar & Leibundgut 2001) which provide joint constraints on λ_0 and Ω_0. Combined with other observations (e.g. Komatsu et al. 2011; Planck Collaboration XVI, 2014), these lead to quite well constrained values for the cosmological parameters (e.g. fig. 5 in Suzuki et al. 2012). Although the supernova data alone allow a relatively wide range of significantly different other models, it is interesting that the best-fitting values obtained from these observations using the current data are quite close to the much better constrained values using combinations of several observations without the supernova data, at least under the assumptions with which the former were calculated. However, since the *m*–*z* relation depends not only on λ_0 and Ω_0 (with H_0 as a scale factor) but also on the distribution of matter along and near the line of sight, the dependence of conclusions drawn from the *m*–*z* relation for Type Ia supernovae on this matter distribution should be investigated. Alternatively, these observations can perhaps tell us something about this distribution.

The plan of this paper is as follows. Section 2 sketches the basic theory used in this paper. In Section 3, I briefly review previous investigations of the influence of a locally inhomogeneous universe on the *m*–*z* relation. Section 4 describes the calculations done and discusses the results. Summary, conclusions and outlook are presented in Section 5.

2 BASIC THEORY

Kayser, Helbig & Schramm (1997, hereafter KHS) developed a general and practical method for calculating cosmological distances in the case of a locally inhomogeneous universe. See KHS for details (and for a description of the notation, which is followed here); here I repeat only the most important points for the purpose of this paper.

If the Universe is homogeneous, then the fact that light propagates along null geodesics provides sufficient information to calculate distances from redshift. If the Universe is locally inhomogeneous, then distances which depend on angular observables related to the

★E-mail: helbig@astro.multivax.de

[1] The case for the 'concordance model' with $\lambda_0 \approx 0.7$ and $\Omega_0 \approx 0.3$ was already made by Ostriker & Steinhardt (1995); the values of the concordance model thus do not need the supernova data, though of course adding more data improves the constraints. While the corresponding uncertainties have dramatically decreased (e.g. Komatsu et al. 2011; Planck Collaboration XVI, 2014), the values themselves have remained constant over the last twenty years.

propagation of radiation will differ from the homogeneous case because more or less convergence will change the angle involved (the angle at the observer in the case of the angular-size distance, that at the source in the case of the luminosity distance). The basic idea is that one considers a universe which is homogeneous and isotropic on large scales, this determining the global dynamics via the Friedmann–Lemaître equation. Local inhomogeneities are modelled as clumps, where the extra matter in the clumps is taken from the surrounding matter. Thus, a beam which propagates between clumps will have only this thinned-out matter inside the beam, while outside the beam the average density (taking both the thinned-out background matter and the clumps into account) is approximately equal to the global density (precisely so in the limit of an infinitesimal beam). Zeldovich (1964), Dashevskii & Zeldovich (1965) and Dashevskii & Slysh (1966) developed a general differential equation for the distance l between two light rays on the boundary of a small light cone (the beam) propagating far away from all clumps of matter in a locally inhomogeneous universe:

$$\ddot{l} = -4\pi G \eta \rho \, l + \frac{\dot{R}}{R}\, \dot{l}, \tag{1}$$

where G is the gravitational constant, R the scale factor, and η (defined below) and the density ρ are functions of time (a dot indicates differentiation with respect to time). The first term can be interpreted as Ricci focusing due to the matter inside the beam, and the second term is due to the expansion of space during the light propagation. The key assumption here is that while the density ρ within the beam can differ from the overall density, the overall dynamics of the universe is still described by the Friedmann–Lemaître equation. The assumption that the light propagates far from all clumps means that Weyl focusing (shear) can be neglected. In the case that the densities inside the beam and outside the beam are the same, one of course recovers the homogeneous case.

Since the angular-size distance is defined as $D = l/\theta$, where θ is the angle at the apex of the beam (at the observer, not at the observed object), D follows the same differential equation as l. Making use of this, one can derive a general equation for the angular-size distance, valid for all (perturbed, in the sense described above) Friedmann–Lemaître cosmological models and all reasonable (see below) values of η. KHS described the inhomogeneity via the parameter $0 \le \eta \le 1$, where η is ratio of the density inside the beam to the global density or, alternatively, the fraction of matter which is homogeneously distributed, as opposed to being clumped.[2] This leads to a second-order differential equation for the angular-size distance (equation (33) in KHS) which can be efficiently integrated numerically:

$$Q\,D'' + \left(\frac{2Q}{1+z} + \frac{1}{2}\,Q'\right)\,D' + \frac{3}{2}\,\eta\,\Omega_0(1+z)\,D = 0, \tag{2}$$

where

$$Q(z) = \Omega_0(1+z)^3 - (\Omega_0 + \lambda_0 - 1)(1+z)^2 + \lambda_0. \tag{3}$$

In the locally inhomogeneous case as well the luminosity distance, which is needed in this paper, is larger than the angular-size distance by a factor of $(1+z)^2$.

This change, compared to the perfectly homogeneous case, is essentially a negative gravitational-lensing effect. In a conventional gravitational-lensing scenario, if the density at a given redshift between two light rays is higher than the overall density (the corresponding overdensity being 'the lens'), then there will be more convergence than in the case where the two densities are the same. In the case of light propagating between clumps, as described above, the situation is reversed, and the density between the light rays defining the distance-related angle is less than the overall density. This means that there is (negative) Ricci focusing (and no Weyl focusing), making objects appear fainter than they would be in the completely homogeneous case. Of course, this is only a rough model, but can be expected to be more realistic than the completely homogeneous case and to determine not just the sign of the difference but also give at least an estimate of its strength.

Obviously, one cannot have $\eta < 0$. However, it does not make sense to have $\eta > 1$ either. While it is certainly possible that the average density inside the beam could be greater than the global density, such cases are either unrealistic or not useful. The limiting case where the density in the beam is greater than the global density by a constant factor at every redshift is unrealistic because this would imply the existence of overdense regions with an extreme length-to-width ratio which are aligned between us and the source, which is incompatible with homogeneity and isotropy on large scales and would also put us in a special position. The other limiting case where a single compact object increases the density in the beam to above the global density is certainly possible, but observationally would show up as a gravitational-lens effect and should be analysed as such (perhaps by adopting $\eta \approx 0$ for the distance calculation and explicitly calculating the amplification). Of course, cases between these two extremes are possible, but it is clear that η must be between 0 and 1 if it is used as an additional parameter in the manner described by KHS; lines of sight which, due to fluctuations, are slightly denser than the overall density are certainly possible, but are not usefully parametrized by η in the style of KHS. (But see Lima, Busti & Santos 2014 for a toy model with an interesting extension of the η concept.) Note that η does not have to be constant as a function of redshift, and the code described in KHS supports an arbitrary dependence of η on z. Also, it could be different for different lines of sight. It was pointed out by Weinberg (1976) that η must be 1 when averaged over all lines of sight (allowing for the moment higher-than-global densities to be parametrized by $\eta > 1$), which follows from flux conservation. However, in practice lines of sight will probably avoid concentrations of matter, due to selection effects or design: distant objects will be more difficult to observed if there is luminous matter along the line of sight or if there is absorbing matter along the line of sight.[3] If these selection effects do not exist, and if the sample is large enough, then the 'Safety in Numbers' effect (Holz & Linder 2005) allows one to effectively assume $\eta = 1$, with inhomogeneity merely increasing

[2] This is sometimes denoted by α. I, and some others, use η because locally inhomogeneous cosmological models are often used in gravitational lensing (which per se implies local inhomogeneities), where α is almost always used to denote the deflection angle.

[3] Matter along the line of sight can increase the apparent brightness and thus make objects visible which otherwise would not be. This phenomenon, known as 'amplification bias' in gravitational lensing, is relevant only if the luminosity function is steep enough (since otherwise the magnification of the area of sky observed, which reduces the number of objects per observed area, will dominate, resulting in fewer objects in a flux-limited sample). However, the whole point of the m–z relation for Type Ia supernovae is that they are standard candles, or can be adjusted to behave as standard candles with the help of other observations, which means that the differential (integral) luminosity function is essentially a delta (Heaviside) function, so the amplification bias plays no role here. Also, since objects much fainter than supernovae can be detected in the corresponding observations, no realistic amplification would make an otherwise undetectable object visible.

the dispersion, roughly linearly with redshift. However, Clarkson et al. (2012) point out that most narrow-beam lines of sight are significantly underdense, even for beams much thicker than those considered in this paper. (On the other hand, they also point out that this does not necessarily lead to a reduction in brightness if one drops the assumption that inhomogeneities can be modelled as perturbations on a uniformly expanding background, a point also emphasized by Bolejko & Ferreira 2012; see also Bagheri & Schwarz 2014.)

The situation discussed above corresponds to the situation where the beam contains a density η times the global density at a given redshift and outside the beam the density is equal to the global density. In practice, this means that a fraction η of the mass in the universe is smoothly distributed and a fraction $1 - \eta$ is contained in clumps outside the beam. Of course, 'smoothly' depends on the size of the beam; for example, small objects are part of the 'smooth' component, not only in the limiting case where the smooth component consists of free elementary particles. The important point is that their angular size is small compared to that of the beam. η is thus also a function of angle: the larger the angle, the more representative is the matter within the beam, so that η approaches 1 for large enough angles. Since the beams of supernovae at cosmological distances are extremely thin objects (the thinnest objects ever studied by science), evidence for $\eta < 1$ should be most obvious in the m–z relation for Type Ia supernovae.

A given value of η along a given line of sight does not imply that this value does not change along the line of sight, although that is of course a possibility, but rather that the influence on angle-dependent distances can be described by an effective value of η which is some appropriate average of a value which varies along the line of sight. This means that it is possible for the density along the line of sight to be larger than the global density at some points, but this is not in contrast with the claim above that $\eta > 1$ is not useful as long as the *effective* value $\eta_{\text{eff}} \le 1$. Another complication is that essentially all lines of sight to supernovae will have a density higher than the globally average cosmological density due to the overdensities associated with the Milky Way and with the supernova host galaxy (and corresponding clusters).[4] However, since the absolute magnitudes of supernovae are not known from first principles, but rather calibrated from observations, this effect is, to a first approximation, unobservable, since it is essentially a renormalization of the absolute magnitude. Even if this extra matter associated with the galaxies at the ends of the beam would increase the density inside the beam to larger than the global density, it is not useful to think of this as $\eta > 1$, since I want to compare the standard assumption (completely homogeneous Universe, at least as far as light propagation is concerned) with that of a more realistic distribution. The point of comparison, the m–z-relation for a homogeneous Universe, also contains extra matter at each end of the beam, and hence extra convergence. As far as I know, no-one has ever taken this into account and it is not necessary if one is interested only in the differences. (This would have to be taken into account, though, if the absolute magnitude of objects at cosmological distances were known independently of observation.)

Although the term 'dark matter' suggests something opaque, the defining characteristic is lack of interaction with electromagnetic radiation. Thus, not only does dark matter not glow, it is also transparent. It is thus irrelevant whether dark-matter objects within the beam significantly cover a source as seen by an observer. (Of course, when comparing observed to calculated brightness, one must correct for extinction due to 'conventional' matter – it can also be dark in the sense that it does not radiate, but it is not transparent.) Here, I am using the term 'dark matter' to refer to the 'missing matter', i.e. that responsible for the difference between the density due to baryonic matter (other non-baryonic but known particles (neutrinos) do not increase this significantly) and the global density of the Universe, as measured on large scales. Of course, non-radiating baryonic matter does exist, but we know from constraints from big-bang nucleosynthesis that this cannot be a significant fraction of the missing matter. This reflects current usage, e.g. the 'DM' in 'ΛCDM', and is more convenient than 'not yet identified non-baryonic matter'.

Since we know that the Universe is not exactly homogeneous and isotropic, $\eta \neq 1$ is the most obvious departure from the simplest cosmological model (the Einstein–de Sitter model with $\lambda_0 = 0$, $\Omega_0 = 1$, and $\eta = 1$, although the last item is often not stated explicitly), but there is not much literature on this topic. (There are, though, several recent papers investigating whether 'dark energy' could be something other than the traditional cosmological constant, e.g. whether the equation of state w differs from -1, whether it changes with time etc, even though there are no observations which indicate this. Of course, that does not mean that one should not look.)

If η is allowed to vary from one line of sight to another, one could regard this as an additional contribution to the uncertainty in the distance modulus, much the same as the uncertainty in the absolute magnitude. Theoretically, fitting the observations for a constant value of η would result in a worse fit for such cases if this additional uncertainty is ignored or in a larger allowed region of parameter space if it is included in the error budget. With some assumptions, one could try to take this additional dispersion into account and/or correct for it; see e.g. Amanullah, Mörtsell & Goobar (2003), Gunnarsson et al. (2006), Jönsson et al. (2006), Jönsson, Mörtsell & Sollerman (2009). In practice, with a large number of objects and only a few variables, the difference in goodness of fit is well within the expected range of values for the case in which η is the same along all lines of sight. Alternatively, with current data it is also a relatively small contribution to the error budget. Thus, if observations suggest $0 < \eta < 1$, it would be unclear if this is evidence for the corresponding value of the global value of η or whether this is a compromise between lines of sight with lower and higher values. However, if observations indicate $\eta = 0$ or $\eta = 1$, then this would be evidence for the corresponding global value, because these are the extreme values of η and cannot result from averaging.

Of course, more complicated models are possible. In this paper, I consider only models in which η is a constant function of redshift and the same along all lines of sight[5]; also, in all cases but one, it is independent of the other cosmological parameters. The variation between these models, however, is certainly larger than the realistic range of the possible influence of η on the m–z relation for Type Ia supernovae.

3 BRIEF HISTORY

The effects of a locally inhomogeneous universe on quantities important for observational cosmology were first investigated in a

[4] I thank Philip Bull for first pointing this out to me.

[5] See Gunnarsson et al. (2006) for a discussion of a z-dependent η in the context of the m–z relation for Type Ia supernovae.

series of papers by Zeldovich (1964), Dashevskii & Zeldovich (1965), and Dashevskii & Slysh (1966). Dyer & Roeder (1972) discussed the special case of $\lambda_0 = 0$ but with Ω_0 as a free parameter for $\eta = 0$ (where there is an analytic solution) and for general η values (Dyer & Roeder 1973). As a result, the distance for $\eta = 0$ is sometimes referred to as the Dyer–Roeder distance. KHS presented a second-order differential equation and numerical implementation valid for the general case ($-\infty < \lambda_0 < \infty, 0 \le \Omega_0 \le \infty, 0 \le \eta \le 1$). Kantowski and collaborators (Kantowski 1969, 1998, 2003; Kantowski, Vaughan & Branch 1995; Kantowski, Kao & Thomas 2000; Kantowski & Thomas 2001) have stressed the importance of η for the interpretation of the m–z relation for Type Ia supernovae and have provided numerical implementations using elliptic integrals for the special values of η of 0, $\frac{2}{3}$, and 1. Perlmutter et al. (1999) considered the effect of $\eta \neq 1$ on their results (see their fig. 8) and concluded that, at least in the 'interesting' region of the λ_0–Ω_0 parameter space, it had a negligible effect (see also Jönsson et al. 2006). The reason for the current paper is that, with the larger number of supernovae now available, this is no longer the case. Further investigation has often been motivated by the m–z relation for Type Ia supernovae (e.g. Goliath & Mörtsell 2000; Mörtsell, Goobar & Bergström 2001). It has also been investigated, via comparison with explicit ray-tracing through mass distributions derived from simulations or observations, whether η is a useful parametrization for local inhomogeneity (e.g. Bergström et al. 2000; Mörtsell 2002) (and the conclusion is that it is a useful approximation, at least for cosmological models which are otherwise realistic).

There seem to be three schools with respect to the attitude taken to the possible influence of inhomogeneities on cosmological parameters derived from the m–z relation for Type Ia supernovae. One school ignores it completely, assuming a completely homogeneous Universe as far as the calculation of the luminosity distance is concerned (e.g. Riess et al. 1998), or provides some limited justification for not considering it further (e.g. Betoule et al. 2014). Another school emphasizes that the problem is not completely understood, the amount of uncertainty is unknown, and even the sign of some effects is unclear (e.g. Clarkson et al. 2012). A third school uses some approximation to at least get an idea of the size of possible effects (e.g. Mörtsell et al. 2001). (While Perlmutter et al. 1999 did consider the possible influence of η, hence belonging to the third school, at least at that time, with their data then it was not a significant source of uncertainty in their main result. One purpose of this paper is to show that this is no longer the case.)

4 CALCULATIONS, RESULTS AND DISCUSSION

I have used the publicly available 'Union2.1' sample of supernova data (Suzuki et al. 2012) and calculated χ^2 and the associated probability following Amanullah et al. (2010) on regularly-spaced grids of various extents and resolutions in the λ_0–Ω_0–η parameter space. This *assumes*, of course, that η is a free parameter on the same footing as λ_0 and Ω_0. My goal is not to obtain the 'best' cosmological parameters, not even the 'best' ones from the supernova data alone. Rather, it is to investigate the influence of $\eta \neq 1$ on the interpretation of the m–z relation for Type Ia supernovae. I have thus intentionally made the supernova data as precise as possible, by using only the statistical uncertainties (i.e. column 4 in the publicly available data file) and fixing H_0 at 70 km s^{-1} Mpc^{-1}, which implies $M = -19.318\,276\,1161$. Thus, all increase in the allowed region of parameter space (at a given confidence level) is due only to the influence of η.[6]

I have calculated χ^2 and the corresponding probability on two three-dimensional grids: a larger, lower-resolution grid

$-5 < \lambda_0 < 5$	$\Delta\lambda_0 = 0.02$	(500 values)
$0 < \Omega_0 < 10$	$\Delta\Omega_0 = 0.02$	(500 values)
$0 < \eta < 1$	$\Delta\eta = 0.01$	(100 values)

and a smaller, higher-resolution grid

$0 < \lambda_0 < 1.5$	$\Delta\lambda_0 = 0.003\,125$	(480 values)
$0 < \Omega_0 < 1.0$	$\Delta\Omega_0 = 0.003\,125$	(320 values)
$0 < \eta < 1.0$	$\Delta\eta = 0.01$	(100 values).

(This paper contains no plots based on the larger, lower-resolution grid; the corresponding calculations were done to make sure that there is no appreciable probability outside of the range of the smaller, higher-resolution grid.) Since three-dimensional contours cannot be fully represented in two dimensions, I present various two- and one-dimensional visualizations in order to illustrate the influence of η.

All contours in two (three) dimensions have been calculated as the smallest-area closed curve (smallest-volume closed surface) which encloses the corresponding fraction of the probability. I have used the standard values 0.683, 0.954, and 0.997; these correspond to 1σ, 2σ, and 3σ in the Gaussian case. However, I have made no assumption about Gaussianity, since I have calculated the contours explicitly, rather than plotting them at the corresponding fraction of the peak likelihood under the Gaussian assumption. For all plots, the area outside of the plot has been assigned a probability of zero. Otherwise, no priors other than those explicitly stated have been used. In particular, no prior information on the values of the cosmological parameters from other tests have been used; what I show, depends on the supernova data only. Figs 1, 2, and 3 show *projections* of the three-dimensional contours along one axis on to the plane spanned by the other two axes for the smaller, higher-resolution grid. It can be seen that the combination of λ_0 and Ω_0 is well constrained, as are both individually, while η is hardly constrained at all. Note also that λ_0 and Ω_0 are less constrained for lower values of η. (The contours at 0.954 and 0.997 cannot be distinguished in these plots.) The relatively sharp bend in the lower-right contours in Figs 1 and 2 is due to the fact that I have assigned a probability of 0 to models which have no big bang (see the discussion of figs 1 and 2 in Helbig 2012 and references therein for an explanation).

Another way of visualizing these three-dimensional contours is to make *cuts* through them for a fixed value of one of the parameters. Figs 4, 5, and 6 show cuts for $\eta = 0.005, 0.455, 0.955$. The contours become smaller and move to lower values of λ_0 and Ω_0 as η becomes larger. (Again, the contours at 0.954 and 0.997 cannot be distinguished in these plots.) Fig. 6 is quite similar to standard presentations of the supernova constraints (e.g. Suzuki et al. 2012),

[6] Since the goal is not to obtain the best constraints on λ_0 and Ω_0, but rather to investigate the influence of η on the constraints, I have retained the Union 2.1 sample with which I began this investigation, rather than updating it to use, e.g., that used by Betoule et al. (2014). Those with better access to such data will always have a better sample than that which is publicly available. Since even Betoule et al. (2014) do not consider η at all, it is perhaps important at the moment for a theorist to take a step back for a more general view in order to contrast with continual updates using somewhat better samples. It is important, though, that the Union 2.1 sample is significantly larger than those used in the early works discussed in Section 1.

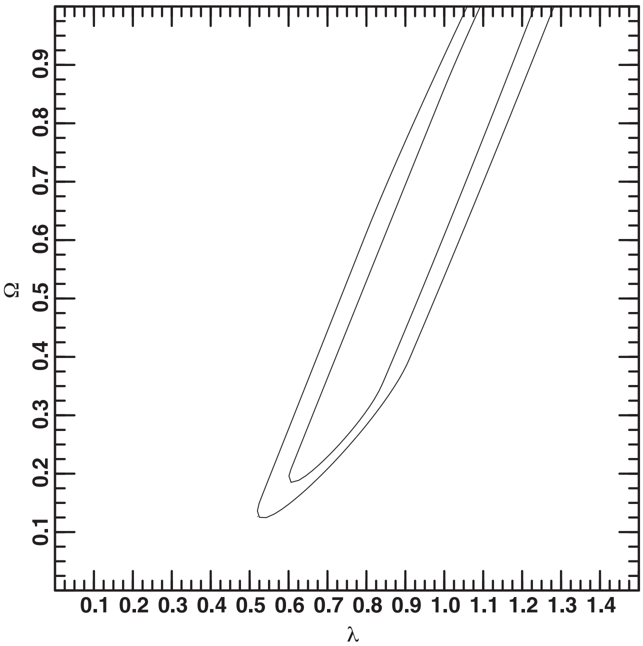

Figure 1. Projection of three-dimensional probability distribution along the η-axis.

but keep in mind that these contours are a cut through the three-dimensional contours for a fixed value of η, not two-dimensional contours. If η is substantially less than 1, then not only is the allowed region much larger, but the 'concordance model' with $\lambda_0 \approx 0.7$ and $\Omega_0 \approx 0.3$ is ruled out. Qualitatively, this behaviour is easy to understand: there is some degeneracy between η and $\lambda_0 + \Omega_0$ since both increase the amount of focusing in the beam, the former because there is more matter in the beam and the latter because of the increase in the global curvature, which is essentially $\lambda_0 + \Omega_0$. When there is essentially no matter in the beam, then the value of Ω_0 is

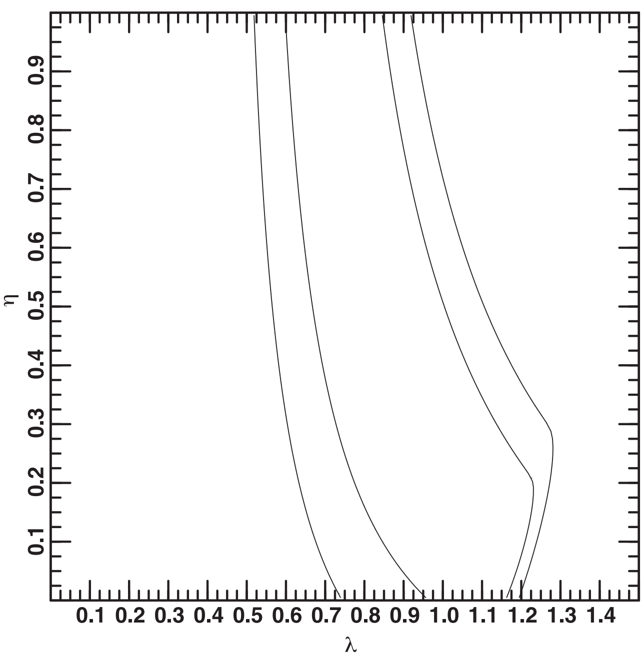

Figure 2. Projection of three-dimensional probability distribution along the Ω_0-axis.

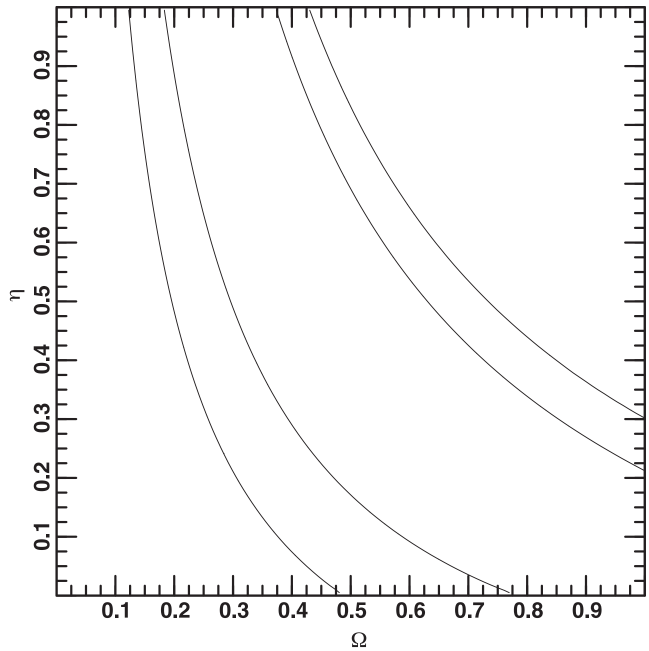

Figure 3. Projection of three-dimensional probability distribution along the λ_0-axis.

less important and hence not as well constrained. This means that $\lambda_0 + \Omega_0$ can be realized via a larger range of each parameter, making the allowed region larger. The middle value of η is that of the global maximum probability. (Since λ_0 and Ω_0 are better constrained, the corresponding plots for fixed values of these parameters, not shown here, are less interesting.)

The 'standard procedure' for reducing the number of parameters shown in a plot is to *marginalize* over the less interesting or 'nuisance' parameters. This is shown in Figs 7, 8, and 9. Here, and in

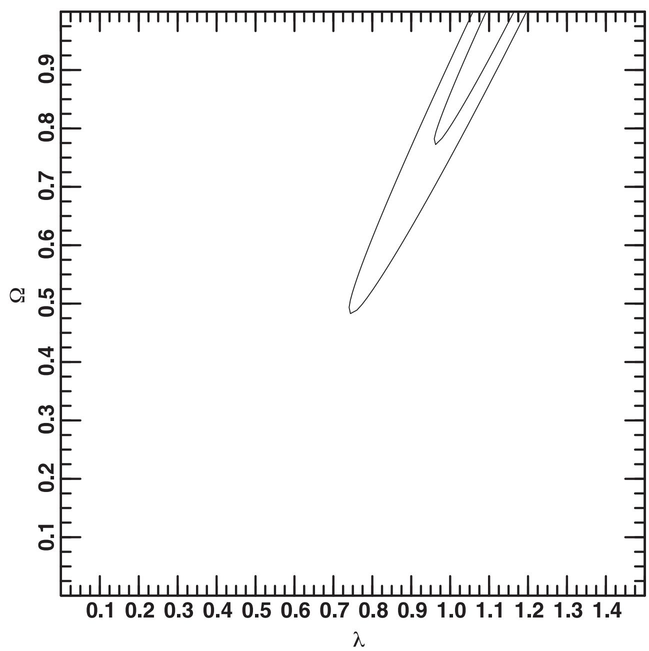

Figure 4. Cut through the three-dimensional probability distribution perpendicular to the η-axis for $\eta = 0.005$.

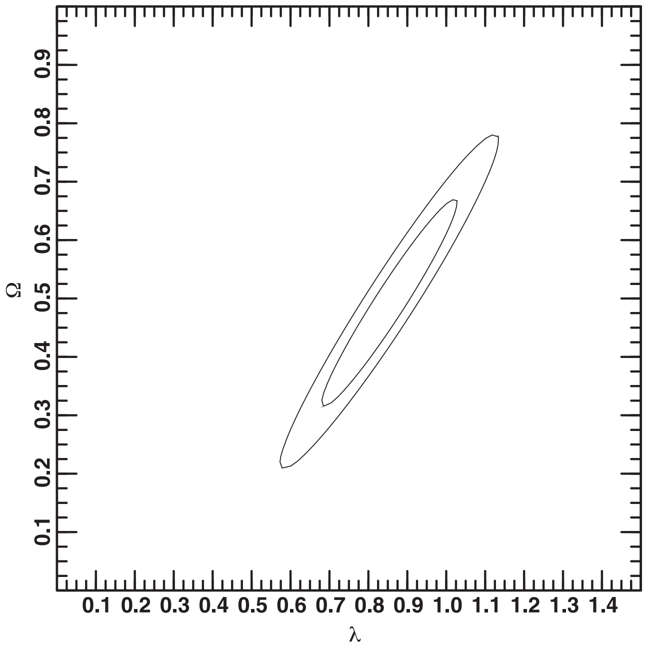

Figure 5. Cut through the three-dimensional probability distribution perpendicular to the η-axis for $\eta = 0.455$.

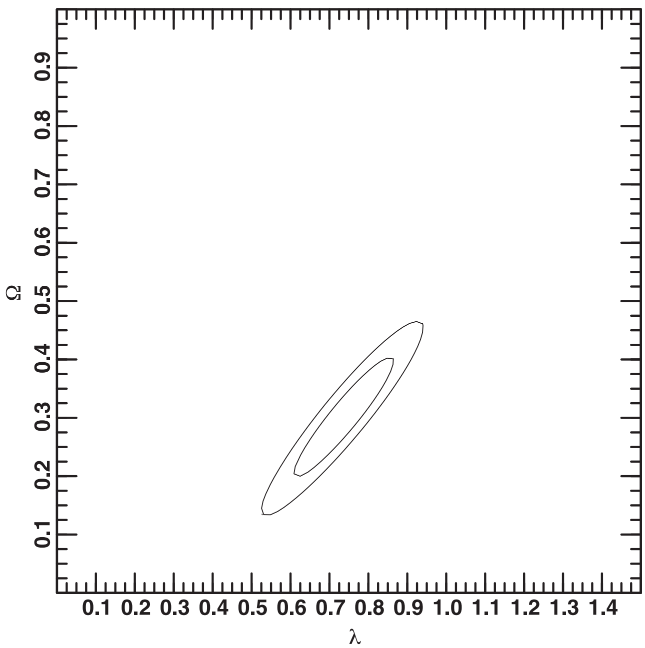

Figure 6. Cut through the three-dimensional probability distribution perpendicular to the η-axis for $\eta = 0.955$.

similar figures below, the grey-scale corresponds to the probability.[7] These are qualitatively similar to the projections. Fig. 7 also contains two straight lines corresponding to a flat universe with $\lambda_0 + \Omega_0 = 1$

[7] It has become fashionable to plot contours and have the regions between the contours filled with a certain colour (or perhaps shade of grey). This conveys no information in addition to the contours themselves. Of course, the probability between two contours, or within the smallest contour, is not everywhere the same, as is obvious from Fig. 7. I have chosen to display this potentially important information in addition to the contour curves.

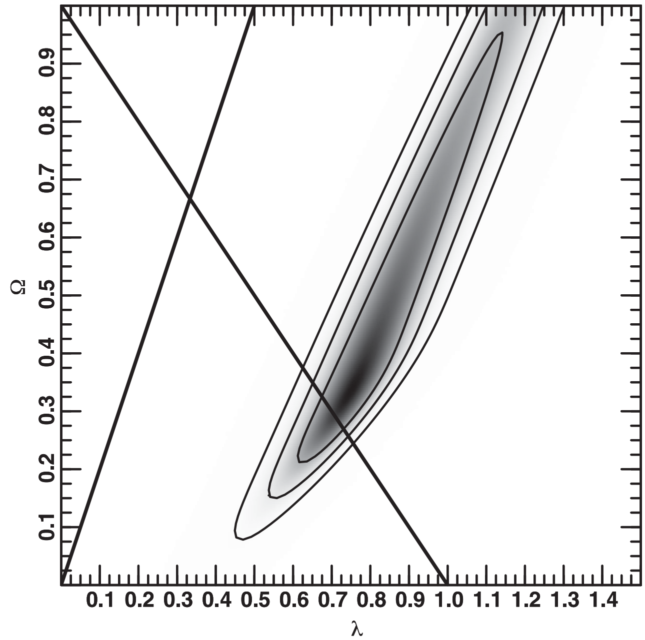

Figure 7. Two-dimensional probability distribution obtained by marginalizing over η.

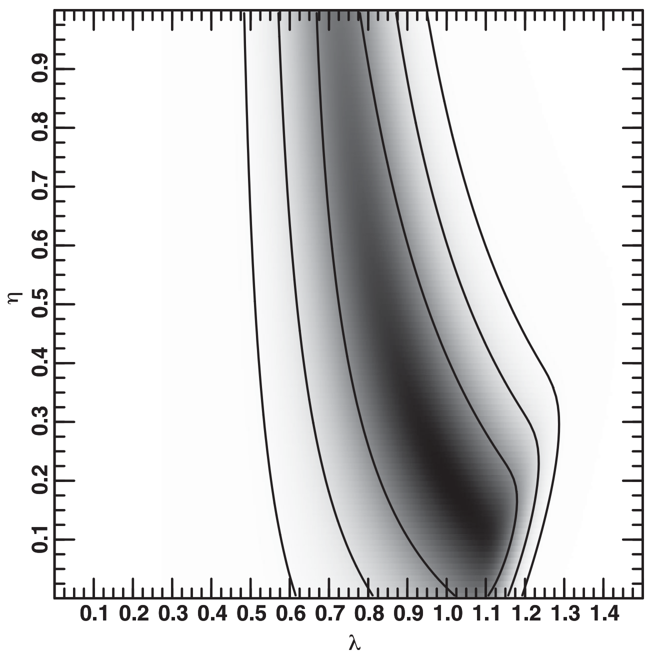

Figure 8. Two-dimensional probability distribution obtained by marginalizing over Ω_0.

(negative slope) and zero acceleration ($q_0 = \frac{\Omega_0}{2} - \lambda_0 = 0$) (positive slope). Note that a flat universe is compatible with the data but not required by them; in fact, the degeneracy in the constraints is almost perpendicular to the flat-universe line. In this particular plot, the degeneracy corresponds roughly to $q_0 \approx -0.6$; in many of the other plots, the degeneracy in the λ_0–Ω_0 plane is closer to a constant value of $\Omega_0 - \lambda_0$ than to a constant value of $q_0 = \frac{\Omega_0}{2} - \lambda_0$. ($q_0$ was important historically since the departure from the linearity of the m–z relation at low redshift is proportional to q_0; nowadays quoting a value for q_0 derived from the m–z relation for

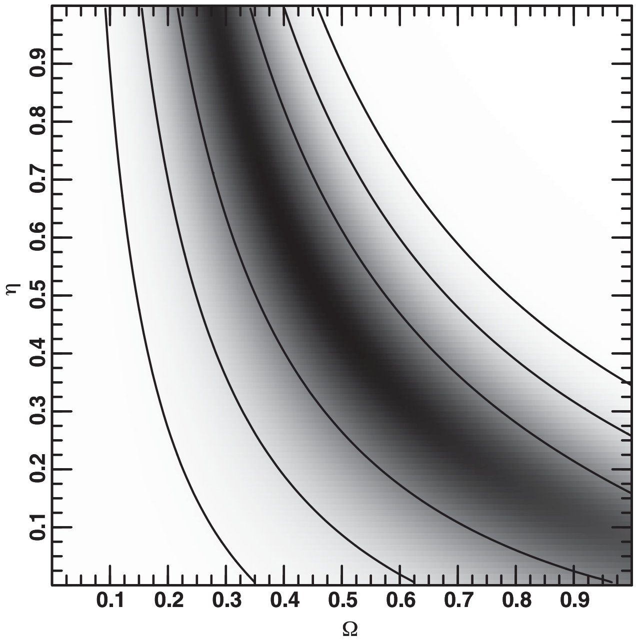

Figure 9. Two-dimensional probability distribution obtained by marginalizing over λ_0.

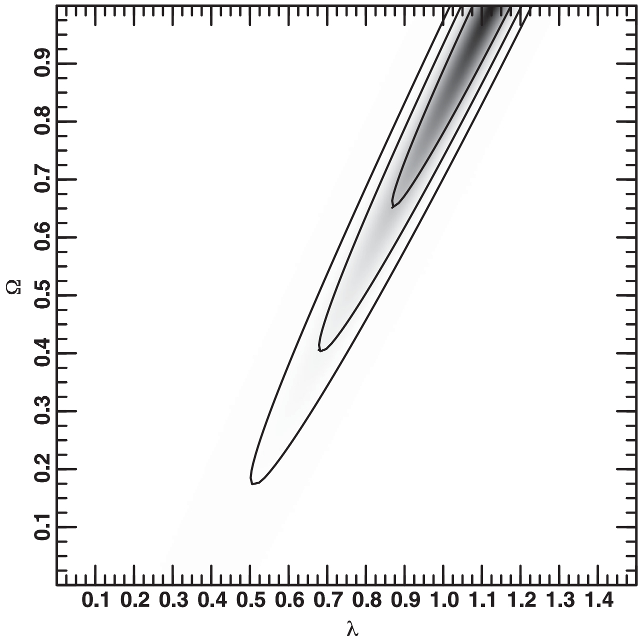

Figure 11. Two-dimensional probability distribution for $\eta = 0$.

higher-redshift objects is neither necessary nor sufficient nor, in general, meaningful.)

Another approach is to *maximize* the 'nuisance' parameter, i.e. for a given point in the plane of the plot, find the value of the third parameter which maximizes the probability. This is shown in Fig. 10. (For these data, such plots are very similar to those where the third parameter has been marginalized over, so only this one example is shown.)

Most discussion of the *m*–*z* relation for Type Ia supernovae has concentrated not on contours of more than two dimensions, nor on some reduction (projection, cut, marginalization, maximization) of these higher-dimensional contours to two dimensions, but rather on two-dimensional contours, i.e. with a δ-function prior on the nuisance parameters. Almost always, of course, the (often implicitly assumed) prior is $\eta = 1$. For comparison, in Figs 11, 12, and 13 I show constraints in the λ_0–Ω_0 plane for fixed values of η, namely 0, 0.455 (the value at the maximum of the three-dimensional probability distribution) and 1. The last should be compared with e.g. fig. 11 in Kowalski et al. (2008), but keep in mind that, as mentioned above, I have fixed H_0 and use only the statistical uncertainties. (See also figs 1a and 5a in Amanullah et al. 2003.) Thus, Fig. 13 has slightly smaller contours than similar plots elsewhere in the literature. Again, this is intentional so that any deviations from this

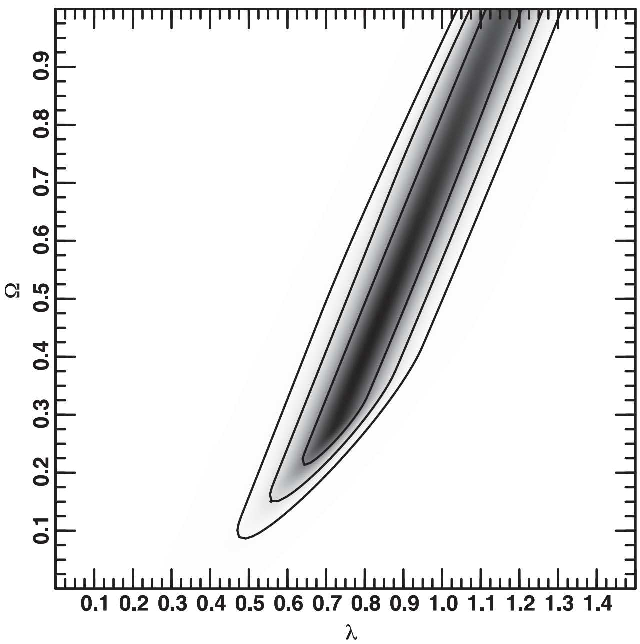

Figure 10. Two-dimensional probability distribution obtained by maximizing η.

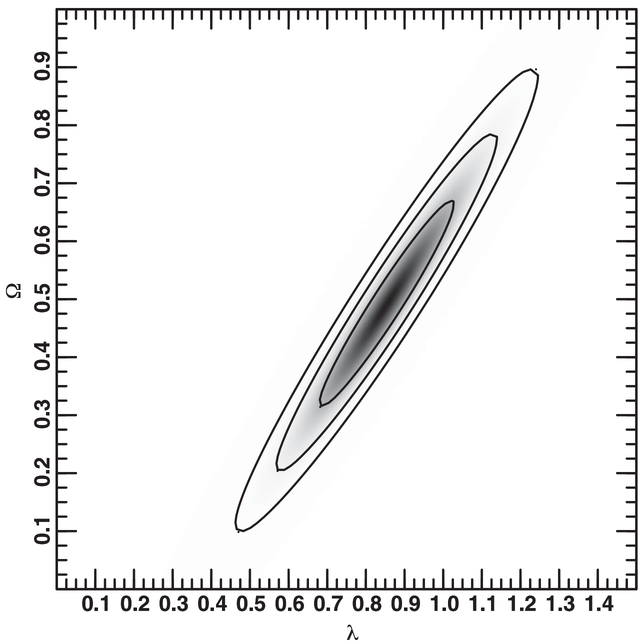

Figure 12. Two-dimensional probability distribution for $\eta = 0.455$.

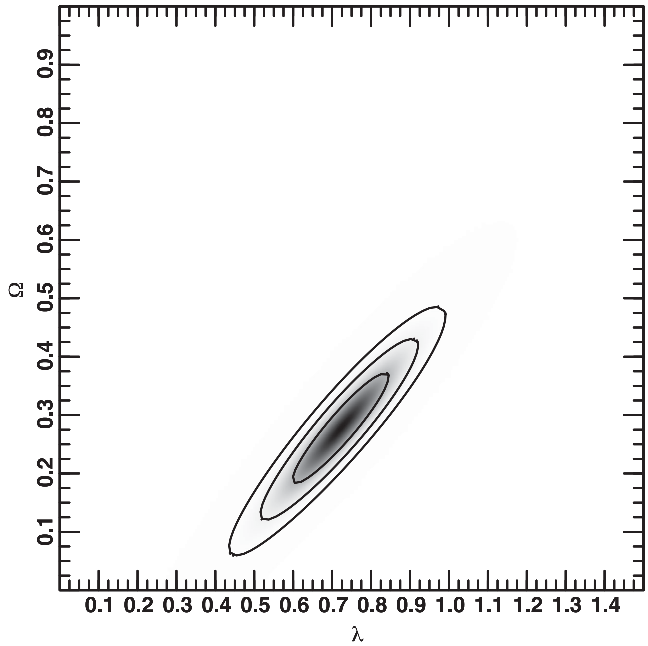

Figure 13. Two-dimensional probability distribution for $\eta = 1$.

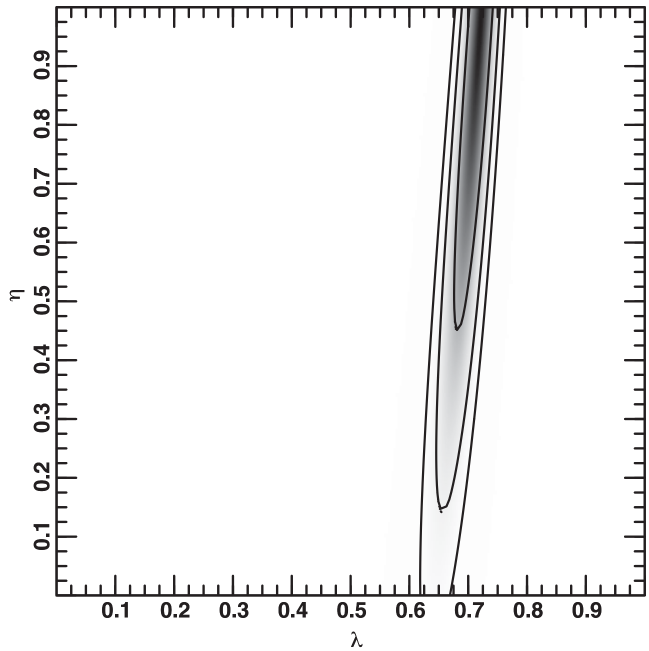

Figure 14. Two-dimensional probability distribution for $k = 0$.

fiducial plot (larger and/or shifted contours) are due solely to the influence of η.

Of course, little significance should be placed on variations in the probability within the innermost contour, since the probability that the point representing the true values of λ_0 and Ω_0 is only about twice as likely to lie inside this contour than outside it. Nevertheless, it is remarkable that the maximum of the probability in Fig. 13 is at $\lambda_0 = 0.721\,0938$ and $\Omega_0 = 0.277\,3438$, i.e. at the values of the concordance model (within the small uncertainties; these are much smaller than even the 68.3 per cent contour in Fig. 13).[8] Note that when fewer supernova data were available, the best-fitting value was at much higher values of λ_0 and Ω_0; see e.g. fig. 1 in Helbig (1999). (As mentioned above, the best-fitting value is often not visible in modern versions of such plots, though of course it can be easily found in the data used to make the plots.) If the best-fitting value remains the same when significantly more supernova data are available, then very probably the true value will have been converged upon, even though the range of values allowed, even at the 68.3 per cent level, would include values well outside what is acceptable when other cosmological constraints are considered (i.e. joint constraints from several cosmological tests). Normally, when more data are available one expects the new best-fitting value to be consistent with, but different from, the old best-fitting value, as has been the case with the supernova data up until now. However, looking towards the future, I don't expect the best-fitting values for λ_0 and Ω_0 to change significantly, but do expect the constraints from the supernova data to improve, which appears somewhat puzzling. A possible explanation for this is that the statistical errors in the supernova data have been overestimated. Note, however, that the best-fitting values for the supernova data correspond to the concordance model only if one assumes $\eta \approx 1$. For $\eta = 0.455$, the concordance model lies very near the 95.4 per cent contour, and for $\eta = 0$ it is even outside the 99.7 per cent contour. The plots above illustrate that it is not possible to appreciably constrain η from the supernova data alone. However, the fact that the supernova data suggest the concordance model only for high values of η could be seen as evidence that $\eta \approx 1$.

A similar result is shown in Fig. 14, where a flat universe ($\lambda_0 + \Omega_0 = 1$) has been assumed. As in the other plots, λ_0 is reasonably well constrained, while η is quite weakly constrained. (In this case, since $\Omega_0 = 1 - \lambda_0$, Ω_0 is just as well constrained; in general, Ω_0 is less well constrained than λ_0.) However, note that the best-fitting value is for $\eta = 1$ and $\lambda_0 \approx 0.72$; in other words, again the best fit is for the concordance model with $\eta = 1$. (This plot also shows the importance of plotting the probability and not just a few contours.)

To illustrate the change in the effect of η now that more supernova data are available, Fig. 15 shows the constraints where η is a function of Ω_0, namely $\eta = 0$ for $\Omega_0 \leq 0.25$ and $1 - 0.25$ for $\Omega_0 \geq 0.25$. This should be compared with fig. 8 in Perlmutter et al. (1999). In that figure, the red contours were calculated in the same way as those in Fig. 15. In the same figure, the green contours were calculated in the same way as in Fig. 11. The comparison illustrates vividly the fact that the effect of η can no longer be neglected. While Perlmutter et al. (1999) concluded that, at least in the interesting part of parameter space, the constraints on λ_0 and Ω_0 from the supernova data did not depend heavily on the assumed value of η, this is definitely no longer the case.

While the supernova data cannot usefully constrain η, as has been shown above, the fact that they result in the concordance model if one assumes $\eta \approx 1$ suggests that $\eta \approx 1$. Since there are many cosmological tests completely independent of the supernova data, and also independent of the value of η, which suggest the concordance model (this is of course why it is called the concordance model), one can assume the concordance values for λ_0 and Ω_0 and calculate the probability of η from the supernova data with these additional constraints; this is shown in Fig. 16. The best-fitting value is

[8] For completeness, I quote the exact position of the maximum as calculated on the grid; of course, this does not imply that the maximum is known to greater precision than the resolution of the grid.

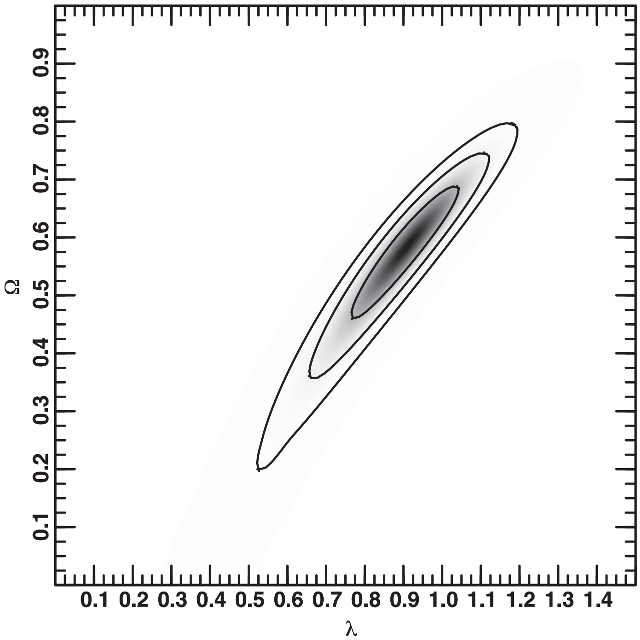

Figure 15. Two-dimensional probability distribution with $\eta = f(\Omega_0)$.

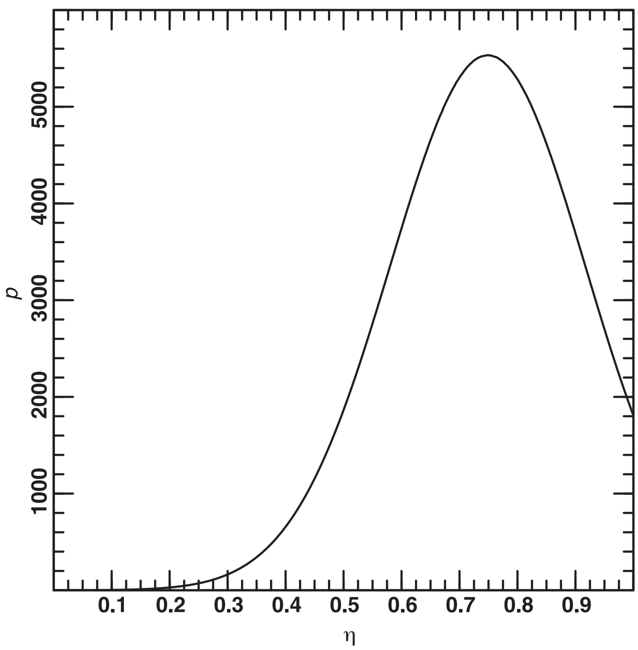

Figure 16. One-dimensional probability distribution for the concordance model.

$\eta = 0.7485$ while the formal statistical limits are

$0.60 < \eta < 0.90$ (68.3 per cent)

$0.46 < \eta < 1.00$ (95.4 per cent)

$0.28 < \eta < 1.00$ (99.7 per cent).

(This can be contrasted with Fig. 17 which shows the value of η preferred by the supernova data alone; λ_0 and Ω_0 have been marginalized over.) While $\eta = 1$ is not ruled out at high confidence, lower values of η are ruled out at a high level of statistical

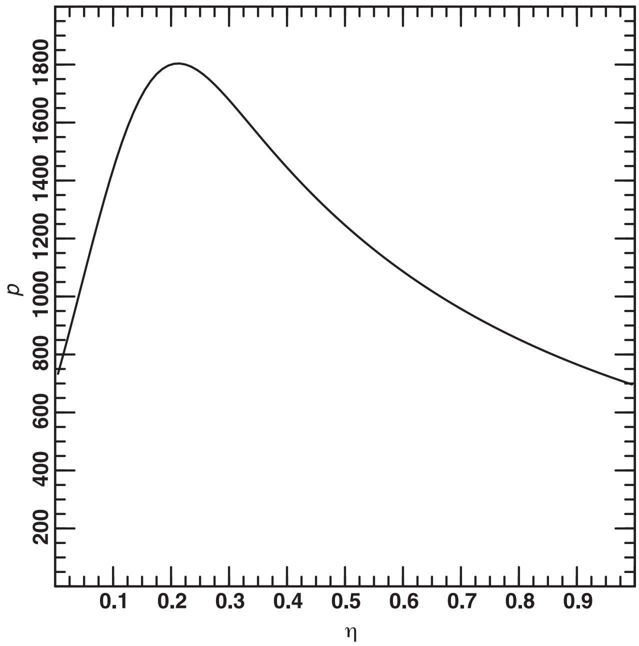

Figure 17. One-dimensional probability distribution after marginalizing over λ_0 and Ω_0.

significance.[9] This suggests that η is relatively large, even though the beams of supernovae at cosmological distances are extremely thin and this cosmological test should suggest a value of η lower than that of any other cosmological test known today. One would not expect to obtain $\eta = 1$ since some matter is associated with galaxies which are outside the beam; such mass contributes about 0.1 to Ω_0. Fig. 16 thus suggests that dark matter is distributed much more smoothly than galaxies. While the beam of a supernova at cosmological distance is almost a fair sample of the universe, it is an even fairer sample of dark matter. Dark matter is thus not significantly clumped at the scale of a supernova beam.

5 SUMMARY, CONCLUSIONS, AND OUTLOOK

The following conclusions were more or less expected.

(i) Constraints on λ_0 and Ω_0 are weaker if η is not constrained.
(ii) The concordance model is reasonably probable.
(iii) There is a degeneracy between η and the amount of spatial curvature $(\lambda_0 + \Omega_0)$.
(iv) λ_0 is constrained best, then Ω_0, then η.

The following conclusion was neither expected nor surprising.

(i) Even when η is allowed to be a free parameter, the m–z relation for Type Ia supernovae is not compatible with $q_0 = \frac{\Omega_0}{2} - \lambda_0 \geq 0$,

[9] This can be contrasted with the location of the maximum of the three-dimensional probability distribution, where the best-fitting values are $\lambda_0 = 0.860\,9375$, $\Omega_0 = 0.501\,5625$, and $\eta = 0.455$. While this lies outside the allowed region of parameter space as determined from cosmological tests other than the m–z relation for Type Ia supernovae, the allowed region is quite large and the concordance model with $\eta = 1$ is within the 68.3 per cent contour. Even though the constraints on λ_0 and Ω_0 are of course weaker if η is allowed to vary, a significant portion of the three-dimensional parameter space can be ruled out, and portions of the λ_0–Ω_0 plane are also ruled out, though no additional region is ruled out which is not already ruled out by other cosmological tests.

and thus implies that the universe is currently accelerating.[10] (Even though the *m*–*z* relation for Type Ia supernovae is one of the key pieces of evidence supporting the cosmological 'concordance model' with $\lambda_0 \approx 0.7$ and $\Omega_0 \approx 0.3$, it is not an essential piece in the sense that combinations of other tests still result in the same concordance model. Nevertheless, it is still an important piece of evidence in favour of the concordance model since it is the only single test which, without additional assumptions, implies $q_0 < 0$, i.e. a universe which is currently accelerating.)

The following conclusions are somewhat surprising.

(i) The overall (in the three-dimensional parameter space) best-fitting values for λ_0 and Ω_0 are ruled out by other cosmological tests. Probably, this best-fitting point is the result of overfitting: its probability is not significantly higher than elsewhere and the allowed region is quite large.

(ii) If one *assumes* $k = 0$, then the best fit is very close to the concordance model *and* has $\eta = 1$.

(iii) If one *assumes* $\eta = 1$, then the best fit is very close to the concordance model.

(iv) If one *assumes* the concordance model, then one can probably rule out low values of η, even though the relevant scale is extremely small, which implies that dark matter is much less clustered than galaxies are.

(v) We cannot rule out $\eta = 1$, and there is some tentative evidence for it.

To summarize, allowing η, which is otherwise only weakly constrained, as a free parameter significantly alters both the best fit in the λ_0–Ω_0 plane and the allowed region of this plane. The concordance model is, however, still allowed. There are hints that $\eta \approx 1$, though these are not statistically significant when examined in the three- or two-dimensional parameter space. On the other hand, if one assumes the concordance values for λ_0 and Ω_0, low values of η can probably be ruled out, which is not obvious considering the very small scales involved; this implies that dark matter is very homogeneously distributed.

One might have thought that the increase in the number of data points since Perlmutter et al. (1999) would allow some sort of useful constraint to be placed on η from the supernova data without further assumptions. This is not the case. Even worse, if η is allowed to vary, then the conclusions about the cosmological model derived from the *m*–*z* relation for Type Ia supernovae are not as robust. However, as discussed in Section 1, current constraints from combinations of cosmological tests without using the supernova data determine the 'concordance model' with $\lambda_0 \approx 0.7$ and $\Omega_0 \approx 0.3$ to rather high precision. It is thus perhaps more interesting to assume the concordance model and use the supernova data to constrain η, especially since η is otherwise difficult to measure. Indeed, as shown in Fig. 16, current data already provide interesting constraints. It is also extremely interesting that the supernova data have the best-fitting values for λ_0 und Ω_0 corresponding to those of the concordance model if and only if $\eta \approx 1$ is assumed. (Note that while the best-fitting value of η assuming the concordance model is $\approx$0.75, the best-fitting values of λ_0 and Ω_0 assuming $\eta \approx 0.75$ are different from those of the concordance model.) If this is not a statistical fluke, it could indicate that $\eta \approx 1$, which is somewhat surprising since the value of η as 'felt' by the supernova might be expected to be somewhat less, because the corresponding beams are extremely thin. The fact that even the supernova data 'want' $\eta \approx 1$ could indicate that dark matter is distributed extremely homogeneously. See Holz (1998) for a different expression of the same idea. Alternatively, this could be evidence that the 'Safety in Numbers' scenario mentioned in Section 2 is in fact a valid approximation.

In contrast to the first useful determinations of λ_0 and Ω_0 from the *m*–*z* relation for Type Ia supernovae (e.g. Garnavich et al. 1998; Riess et al. 1998; Perlmutter et al. 1999), where the effect of $\eta \neq 1$ had a negligible effect on the constraints derived, at least in the 'interesting' region of the λ_0–Ω_0 parameter space, with the larger number of supernovae now available, this is no longer the case. At the same time, current supernova data alone cannot usefully constrain η (though this might be possible if other cosmological data are taken into consideration, as discussed in the previous paragraph). This should be taken into account in attempts to determine further parameters, such as w, the equation-of-state parameter for dark energy. When more supernova data become available, especially at higher redshift, it might be possible to usefully constrain η and/or discriminate between the effect of η and other parameters such as w. (The difference in apparent magnitude for different values of η increases with increasing redshift, while the difference due to different values of λ_0 and Ω_0 is stronger (than that due to variation in η) at lower redshift and, for some sets of models, decreases at higher redshift.) While allowing η to be a free parameter, but constant as a function of redshift and for different lines of sight, is certainly not the last word with respect to the influence of locally inhomogeneous cosmological models on the *m*–*z* relation for Type Ia supernovae, it does demonstrate that care is needed when interpreting conclusions derived from assuming $\eta = 1$. At least, the uncertainty in λ_0 and Ω_0 must be correspondingly increased. While it might be possible to decrease this with a more realistic model, it is no longer possible to assume $\eta = 1$ and have confidence in the parameters and their uncertainties resulting from an analysis of the *m*–*z* relation for Type Ia supernovae.

[10] Mörtsell & Clarkson (2009) have shown that this conclusion also holds for a much wider class of models than the Friedmann–Lemaître models considered here.

ACKNOWLEDGEMENTS

I thank Nils Bergvall, Phil Bull, Bruno Leibundgut, and an anonymous referee for helpful comments. Figures were produced with the GRAL software package written by Rainer Kayser.

REFERENCES

Amanullah R., Mörtsell E., Goobar A., 2003, A&A, 397, 819
Amanullah R. et al., 2010, ApJ, 716, 712
Bagheri S., Schwarz D. J., 2014, J. Cosmol. Astropart. Phys., 10, 073
Bergström L., Goliath M., Goobar A., Mörtsell E., 2000, A&A, 358, 13
Betoule M. et al., 2014, A&A, 568, A22
Bolejko K., Ferreira P. G., 2012, J. Cosmol. Astropart. Phys., 5, 003
Clarkson C., Ellis G. F. R., Faltenbacher A., Maartens R., Umeh O., Uzan J.-P., 2012, MNRAS, 426, 1121
Dashevskii V. M., Slysh V. J., 1966, SvA, 9, 671
Dashevskii V. M., Zeldovich Y. B., 1965, SvA, 8, 854
Dyer C. C., Roeder R. C., 1972, ApJ, 174, L115
Dyer C. C., Roeder R. C., 1973, ApJ, 180, L31
Garnavich P. M. et al., 1998, ApJ, 493, L53
Goliath M., Mörtsell E., 2000, Phys. Lett. B, 486, 249
Goobar A., Leibundgut B., 2001, Annu. Rev. Nucl. Part. Sci., 61, 251
Goobar A., Perlmutter S., 1995, ApJ, 450, 14
Gunnarsson C., Dahlén T., Goobar A., Jönsson J., Mörtsell E., 2006, ApJ, 640, 417

Helbig P., 1999, A&A, 350, 1
Helbig P., 2012, MNRAS, 421, 561
Holz D. E., 1998, ApJ, 506, L1
Holz D. E., Linder E. V., 2005, ApJ, 631, 678
Jönsson J., Dahlén T., Goobar A., Gunnarsson C., Mörtsell E., Lee K., 2006, ApJ, 639, 991
Jönsson J., Mörtsell E., Sollerman J., 2009, A&A, 493, 331
Kantowski R., 1969, ApJ, 155, 89
Kantowski R., 1998, ApJ, 507, 283
Kantowski R., 2003, Phys. Rev. D, 68, 123516
Kantowski R., Thomas R. C., 2001, ApJ, 561, 491
Kantowski R., Vaughan T., Branch D., 1995, ApJ, 447, 35
Kantowski R., Kao J. K., Thomas R. C., 2000, ApJ, 545, 549
Kayser R., Helbig P., Schramm T., 1997, A&A, 318, 680
Komatsu E. et al., 2011, ApJS, 192, 18
Kowalski M. et al., 2008, ApJ, 686, 749
Leibundgut B., 2001, ARA&A, 39, 67
Leibundgut B., 2008, Gen. Relativ. Gravit., 40, 221
Lima J. A. S., Busti V. C., Santos R. C., 2014, Phys. Rev. D, 89, 067301
Mörtsell E., 2002, A&A, 382, 787
Mörtsell E., Clarkson C., 2009, J. Cosmol. Astropart. Phys., 1, 44
Mörtsell E., Goobar A., Bergström L., 2001, ApJ, 559, 53
Ostriker J. P., Steinhardt P. J., 1995, Nature, 377, 600
Perlmutter S. et al., 1995, ApJ, 440, L41
Perlmutter S. et al., 1998, Nature, 391, 51
Perlmutter S. et al., 1999, ApJ, 517, 565
Planck Collaboration XVI, 2014, A&A, 571, 16
Riess A. G., 2000, PASP, 112, 1284
Riess A. G. et al., 1998, AJ, 116, 1009
Riess A. G. et al., 2000, ApJ, 536, 62
Suzuki N. et al., 2012, ApJ, 746, 85
Weinberg S., 1976, ApJ, 208, L1
Zeldovich Y. B., 1964, SvA, 8, 13

This paper has been typeset from a TEX/LATEX file prepared by the author.

8.3 Follow-up

Dhawan *et al.* (2018) carried out a similar study, but adding the equation of state for dark energy, w, and a parameter describing its possible evolution, but restricted to spatially flat models, and taking into consideration constraints from the CMB and BAO[1]. They found a degeneracy between η and w, such that a higher value of η implies a lower (*i.e.* more negative) value of w. If dark energy is taken to be the cosmological constant, *i.e.* $w = -1$, then their result is $\eta = 0.81 \pm 0.33$ at the 68% confidence level. That can be compared to my result of $\eta = 0.75 \pm 15$ for λ_0 and Ω_0 fixed at the values of the concordance model ($\lambda_0 = 0.7$ and $\Omega_0 = 0.3$). Of course, fixing those parameters causes the uncertainty to be smaller.

Important is the fact that, as in many other works (Yu *et al.*, 2011; Busti and Santos, 2011; Yang *et al.*, 2013; Bréton and Montiel, 2013; Li *et al.*, 2015; Dhawan *et al.*, 2018), it seems to be a robust result that η has a relatively high value, and also that that conclusion weakens only slighly if w is allowed to differ from -1 or even change with time. Assuming $\eta = 1$ Dhawan *et al.* (2018) find $w = -0.961 \pm 0.055$. In other words, if my suspicions are correct that $w = -1$ exactly (*i.e.* dark energy is the cosmological constant[2]) and that, effectively, $\eta \approx 1$, as in the references above, then it is interesting that assuming one implies the other.

I presented this work in seminar talks at the University of Oslo in May 2014, the University of Uppsala in May 2014, and the University of Sussex in August 2015, as well as in a talk at the the 28th Texas Symposium on Relativistic Astrophysics in Geneva in December 2015 (Helbig, 2015c), and in a talk at the cosmology session at the Rencontres de Moriond in March 2016 (Helbig, 2016a).

---

[1] Baryon acoustic oscillations, along with the CMB and the m–z relation for Type Ia supernovae, are now one of the most important cosmological tests, providing a standard ruler at intermediate redshift, complementing that of the CMB at high redshift. In both cases, relatively large angles are involved, so one can assume $\eta \approx 1$.

[2] Despite many attempts, no-one has ever found a value of w inconsistent with -1.

Chapter 9

The magnitude–redshift relation for Type Ia supernovae: safety in numbers or safely without worry?

9.1 Context

Having established that $\eta \approx 1$ (Helbig, 2015a), I wanted to understand what that means: does the Universe appear homogeneous only when averaged over all lines of sight or over the celestial sphere—the two are not necessarily equivalent (Kaiser and Peacock, 2016)— or is each line of sight a fair sample of the Universe, in which case one might expect to observe $\eta \approx 1$ along all lines of sight? In the latter case, is the Universe really approximately homogeneous, or is it just the case that the distance calculated from redshift is approximately the same as that in a homogeneous universe?

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **453,** 3975–3979 (2015) doi:10.1093/mnras/stv1796

The *m*–*z* relation for Type Ia supernovae: safety in numbers or safely without worry?

Phillip Helbig*

Thomas-Mann-Str. 9, D-63477 Maintal, Germany

Accepted 2015 August 3. Received 2015 July 26; in original form 2015 June 30

ABSTRACT

The *m*–*z* relation for Type Ia supernovae is compatible with the cosmological concordance model if one assumes that the Universe is homogeneous, at least with respect to light propagation. This could be due to the density along each line of sight being equal to the overall cosmological density, or to 'safety in numbers', with variation in the density along all lines of sight averaging out if the sample is large enough. Statistical correlations (or lack thereof) between redshifts, residuals (differences between the observed distance moduli and those calculated from the best-fitting cosmological model), and observational uncertainties suggest that the former scenario is the better description, so that one can use the traditional formula for the luminosity distance safely without worry.

Key words: supernovae: general – cosmological parameters – cosmology: observations – cosmology: theory – dark energy – dark matter.

1 INTRODUCTION

I recently investigated the dependence of constraints on the cosmological parameters λ_0 and Ω_0 derived from the *m*–*z* relation for Type Ia supernovae on the degree of local homogeneity of the Universe (Helbig 2015). When deriving such constraints, it is often assumed that the Universe is completely homogeneous, at least with regard to light propagation. However, the constraints on the cosmological parameters derived depend on this assumption. If the degree of local inhomgeneity is parametrized by the parameter η giving the fraction of homogeneously distributed matter on the scale of the beam size such that the density at a given redshift is equal to the average cosmological density $\rho = \frac{3H^2\Omega}{8\pi G}$ outside the beam and $\eta\rho$ inside the beam (see Kayser, Helbig & Schramm 1997, for definitions and discussion), and assuming that η is independent of redshift and the same for all lines of sight, then only if $\eta \approx 1$ do the constraints on the cosmological parameters λ_0 and Ω_0 derived from the *m*–*z* relation for Type Ia supernova correspond to the 'concordance model' (e.g. Ostriker & Steinhardt 1995; Komatsu et al. 2011; Planck Collaboration 2014).

Two important conclusions of Helbig (2015) are thus that the values of λ_0 and Ω_0 derived from the *m*–*z* relation for Type Ia supernovae depend on assumptions made about η, substantially so for current data, and that only for $\eta \approx 1$ are these values consistent with other measurements of the cosmological parameters. Perlmutter et al. (1999) considered the effect of $\eta \neq 1$ on their results (see their fig. 8 and the discussion in their section 4.3) and concluded that, at least in the 'interesting' region of the λ_0–Ω_0 parameter space (i.e. $\Omega_0 < 1$; even at that time there was substantial evidence against $\Omega_0 > 1$), it had a negligible effect. Not only is this effect no longer negligible with newer data (both because there are more data points altogether and because there are more data points at higher redshifts), but, especially since we now have good estimates of λ_0 and Ω_0 from other tests, it allows one to use the supernova data to say something about η. With a strong indication from the supernova data that $\eta \approx 1$, it is important to consider the question whether this is true only when several lines of sight are averaged or is true for a typical individual line of sight. Perlmutter et al. (1999) investigated the influence of η on the values obtained for λ_0 and Ω_0 but could draw no conclusions about its value from the supernova data alone. Even though the 'concordance model' had already been postulated at the time (though of course there was much less evidence in favour of it than is the case today), assuming the corresponding values for λ_0 and Ω_0 could not allow any statement to be made about η since there was significant overlap in the allowed regions of parameter space for the various η scenarios. (Also, in contrast to the case with newer data, the best-fitting values of λ_0 and Ω_0 were far from the concordance values, though the concordance values were allowed even at 1σ.) This is consistent with their claim, based on simulations, that the conclusions drawn from their data should not depend heavily on η. Their robust conclusion that the *m*–*z* relation for Type Ia supernovae implies that $\lambda_0 > 0$, and the somewhat stronger claim that $q_0 < 0$ (i.e. the Universe is currently accelerating), regardless of assumptions made about η, are of course the most interesting results of Perlmutter et al. (1999) (and similar studies by the High-*z* Supernova Search Team and later papers by both groups). Interestingly, both of these are still robust with current data.

*E-mail: helbig@astro.multivax.de

2 TWO SCENARIOS

There are two ways in which $\eta \approx 1$ can be explained. One is that $\eta \approx 1$ holds for each individual line of sight. (This does not necessarily imply that η is actually constant along the beam, but only that the distance modulus calculated from the cosmological parameters λ_0, Ω_0, and H_0 and from the redshift z is the same as that calculated assuming $\eta \approx 1$. In other words, there could be density variations along the beam (apart from the decrease in density with decreasing redshift due to the expansion of the Universe) as long as they appropriately average out.) The other is that $\eta < 1$ for some lines of sight and $\eta > 1$ for others, such that $\eta \approx 1$ when averaged over all lines of sight, though of course density variations along the beam as in the other case could also be present.[1] This has been dubbed the 'safety in numbers' effect by Holz & Linder (2005).

In the first case, the residuals (the differences between the observed distance moduli and those calculated from the best-fitting cosmological parameters) should not depend on redshift per se, while in the second case they should increase with redshift: all else being equal, the lower η, the larger the distance modulus, and the difference between this and that calculated using the traditional $\eta = 1$ assumption is a monotonically increasing function of redshift; see e.g. fig. 1 in Kayser et al. (1997).[2] In the first case, the residuals are due only to uncertainties in the observed distance moduli, while in the second case they are due also to variations in the actual distance moduli as a result of different average densities along the line of sight. Of course, in the first case there could be a dependence of the residuals on redshift if the observational uncertainties depend on redshift, and in the second case the residuals are due both to variations in the actual distance moduli and to observational uncertainties in them. One could call the first case 'safely without worry', meaning that one can safely use the traditional formula for the luminosity distance (corresponding to $\eta = 1$) when calculating the distance modulus, without worry.

3 CALCULATIONS, RESULTS, AND DISCUSSION

For purposes of comparison and consistency, I work with the same data as in Helbig (2015), namely the publicly available 'Union2.1' sample of supernova data (Suzuki et al. 2012). Fig. 1 shows the residuals Δ (points) with respect to the best-fitting model assuming $\eta = 1$ in Helbig (2015) ($\lambda_0 = 0.721\,0938$ and $\Omega_0 = 0.277\,3438$) and the uncertainties σ in the distance moduli (lines). These are shown separately (both as points) in Figs 2 and 3. There appear to be a positive correlation between the uncertainties and redshifts and a

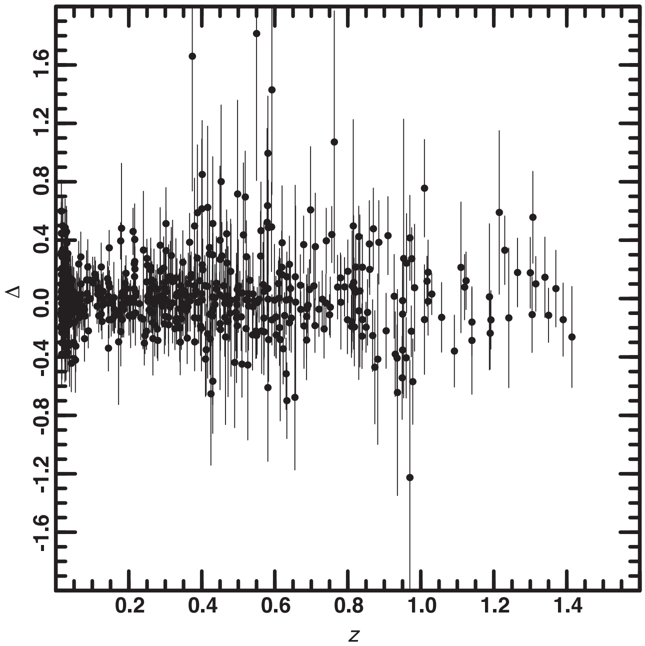

Figure 1. Residuals (differences between the observed distance moduli and those calculated from the best-fitting cosmological model) (points) and observational uncertainties (lines).

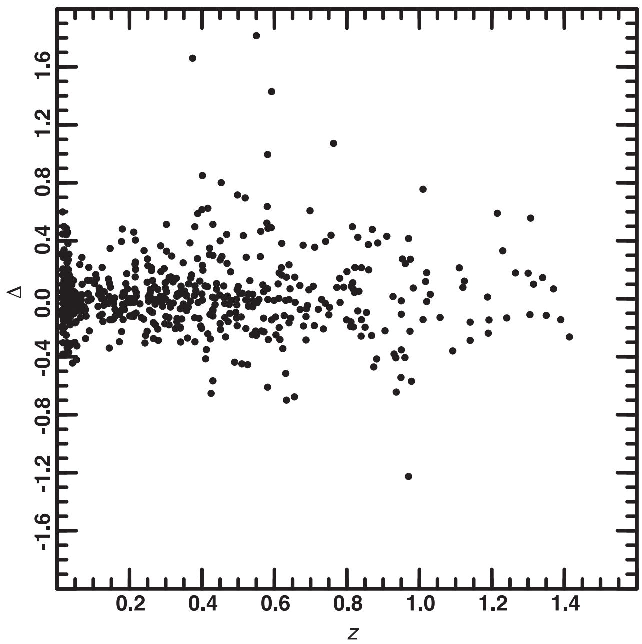

Figure 2. Residuals.

higher number of outliers at intermediate redshifts (though the fact that there are fewer at high redshifts might be due to the smaller number of objects there). If the first case discussed above holds, then (the absolute value of) the quotient $Q = \Delta/\sigma$ of the residuals and the uncertainties should show no trend with redshift, while if the second case holds there should be a positive correlation. Fig. 4 shows this quotient and, indeed, there appears to be no trend with redshift. Also, the width of the distribution seems to depend only on the number of points in the corresponding redshift range, i.e. there appear to be no outliers as such, or at least fewer.

[1] The second case requires a more general definition of η than that used in Kayser et al. (1997); see Lima, Busti & Santos (2014) and Helbig (2015) for discussion. Strictly speaking, as pointed out by Weinberg (1976), it is the magnification μ which averages to 1 over all lines of sight. Since $\eta \sim \kappa$, where κ is the convergence, and $\mu \sim (1-\kappa)^{-1}$, the relation is linear only in the limit of vanishing deviations, though approximately linear for the small deviations considered here. The actual situation is quite complicated. For example, the average angular-size distance $\langle D \rangle$, and hence the average luminosity distance $\langle D_L \rangle$, is biased even in the case of $\langle \mu \rangle = 1$. See Kaiser & Peacock (2015) for discussion of this and many other details in the still ongoing debate on this topic. I use the term 'average' here loosely; the important point is that the average of an observed quantity is the same as in the $\eta = 1$ case, not that η itself averages to 1.

[2] There is of course a similar effect with opposite sign for $\eta > 1$, as discussed in the previous footnote.

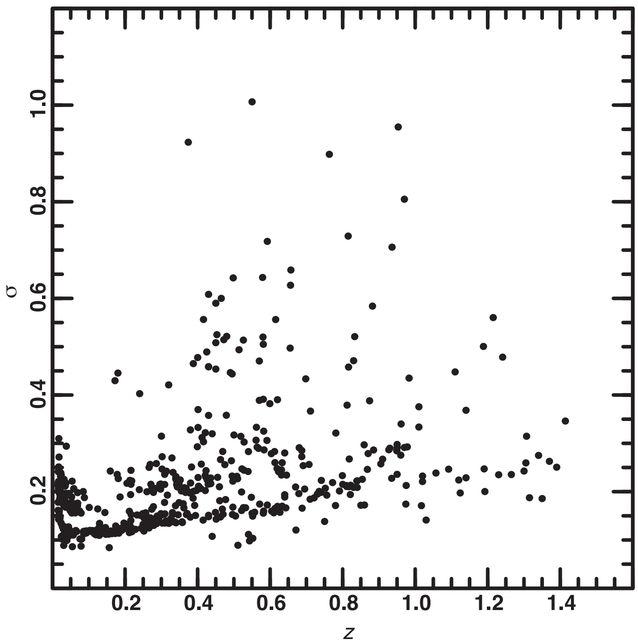

Figure 3. Observational uncertainties.

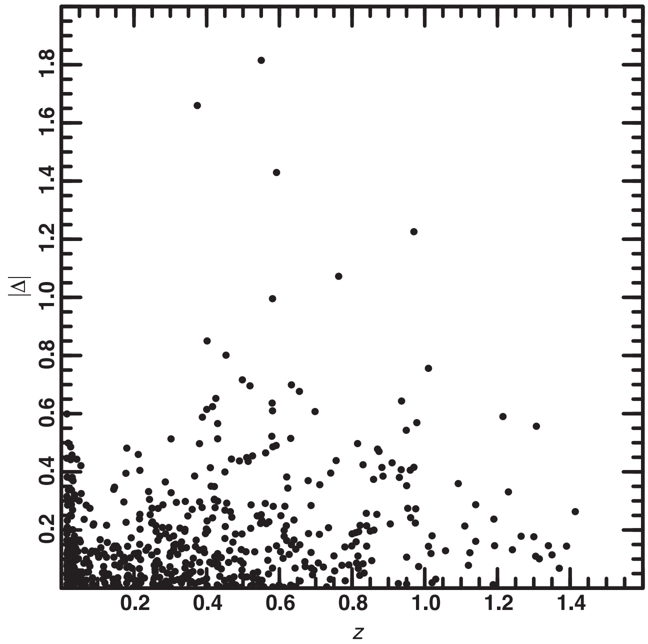

Figure 5. Absolute values of residuals.

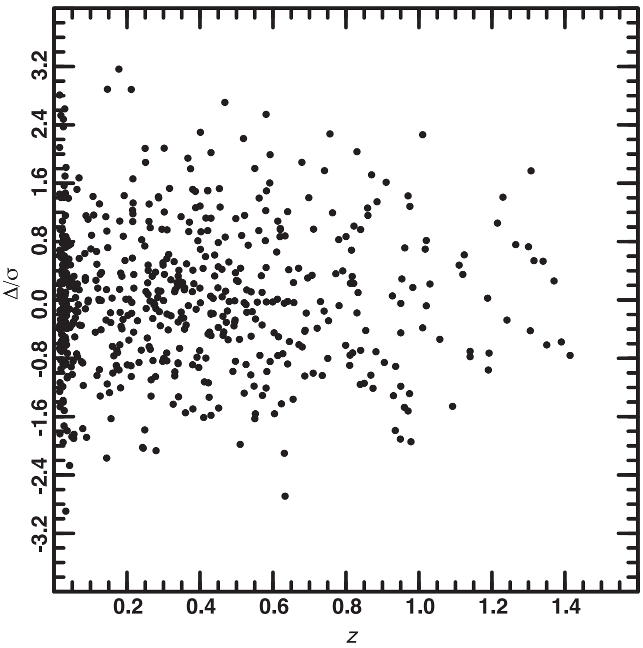

Figure 4. Quotients of residuals and observational uncertainties.

In order to quantify the dependence of the magnitude of the uncertainties on redshift, I have calculated various statistical measures, shown in Table 1, to investigate the existence of a correlation between the redshifts and the absolute values of the residuals $|\Delta|$ (plotted in Fig. 5), the observational uncertainties σ (Fig. 3), and the absolute value of the quotient of the residuals and the uncertainties, $|Q|$ (plotted in Fig. 6), as well as the corresponding statistical significance. Note that Fig. 5, like Fig. 3, appears to show a positive correlation between the absolute values of the residuals and redshifts and a higher number of outliers at intermediate redshifts. Also included in Table 1 are the corresponding quantities concerning the correlation between $|\Delta|$ and σ.

All three statistical tests agree about the sign of the correlation and whether or not it is significant. (The *values* of the correlations and the corresponding significance are not directly comparable.) Both the absolute values of the residuals, $|\Delta|$, and the observational uncertainties, σ, are positively correlated with redshift, but their quotient is not. This suggests that the first scenario described above, 'safely without worry', is the appropriate one, not the second scenario, 'safety in numbers'. If this is the case, then one would expect $|\Delta|$ and σ to be correlated, and indeed they are. (Note that this last test is not sufficient to rule out the 'safety in numbers' scenario, since even if the scatter in the actual distance moduli increased with

Table 1. Statistical quantities measuring the correlation between the redshifts z and the absolute values of the residuals $|\Delta|$, the observational uncertainties σ, and the quotient $|Q|$ of these, as well as between $|\Delta|$ and σ: r is Pearson's product-moment correlation coefficient, r_s is Spearman's rank-order correlation coefficient, and τ is Kendall's non-parametric rank-order correlation coefficient. The corresponding p values give the probability of getting a value as large as observed or larger in the case of the null hypothesis of no correlation. All values have been rounded to two significant figures.

data set	r	$p(r)$	r_s	$p(r_s)$	τ	$p(\tau)$
$z, \|\Delta\|$	0.23	1.3×10^{-8}	0.21	1.8×10^{-7}	0.14	7.8×10^{-7}
z, σ	0.41	1.5×10^{-25}	0.44	4.5×10^{-29}	0.28	2.6×10^{-23}
$z, \|Q\|$	3.6×10^{-2}	0.39	5.1×10^{-2}	0.22	3.3×10^{-2}	0.24
$\|\Delta\|, \sigma$	0.62	0.00	0.42	1.3×10^{-25}	0.29	1.7×10^{-25}

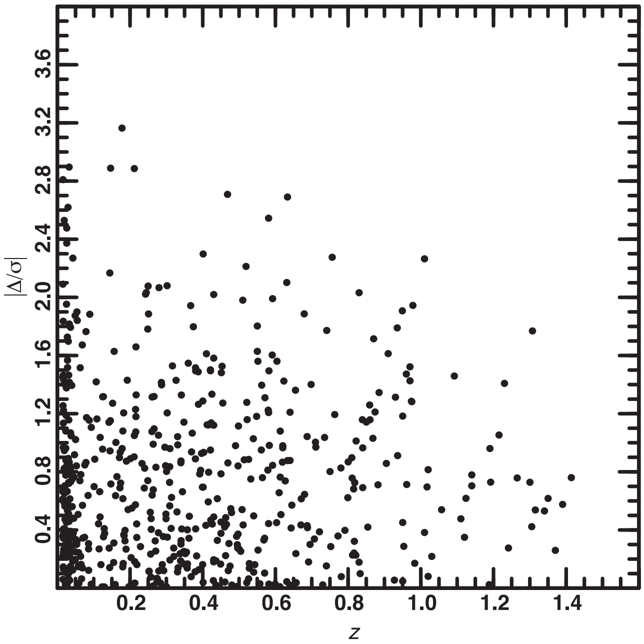

Figure 6. Quotients of absolute values of residuals and observational uncertainties.

redshift, there could still be a correlation between $|\Delta|$ and σ in addition to the one between $|\Delta|$ and z.)

Of course, this analysis takes the Union2.1 data set at face value, and relies on the assumption that the observational uncertainties have been correctly estimated. Also, $\eta \neq 1$ describes just a different amount of Ricci focusing due to more or less matter within the beam than in the standard case, as opposed to more general gravitational lensing. Note that Suzuki et al. (2012) explicitly correct for the amplification of supernovae known to be gravitationally lensed by galaxy clusters (see their section 2.1); in other words, the magnitudes used for the cosmology analysis are those which would have been observed in the absence of the corresponding galaxy clusters. To be sure, Suzuki et al. (2012), following the procedure described in section 7.3.5 of Amanullah et al. (2010), include as part of the error estimate $0.093z$ to take the statistical uncertainty due to gravitational lensing into account. If this were a significant part of the uncertainty, then it could explain the correlation between the uncertainties and redshifts, and thus favour the ‘safety in numbers’ scenario. This contribution to the error budget probably explains the slope of the lower envelope in Fig. 3. However, it is clear from Figs 3 and 4 that the main cause of the correlation is the absence of both large residuals and large uncertainties at low redshifts. The large residuals – much larger than $0.093z$ – also have large uncertainties, and occur mainly at intermediate redshifts. Finally, as described in section 7.2 of Amanullah et al. (2010) and section 4.4 of Suzuki et al. (2012), the Union2.1 data set was constructed by rejecting $3\,\sigma$ outliers, which would remove any strongly lensed supernovae from the sample. Both this use of median statistics and the $0.093z$ contribution contribute to the correlation at some level but, as explained above, cannot explain all, or even most, of it.

Note that Yu et al. (2011), using observational data other than the m–z relation for Type Ia supernovae, and assuming a flat Universe, arrive at essentially the same conclusion as Helbig (2015): $\eta \approx 1$ is favoured and low values of η can be ruled out.[3] Some of the assumptions in Yu et al. (2011) were questioned by Busti & Santos (2011), but even when these are corrected for, $\eta \approx 1$ is still favoured. As discussed in Helbig (2015), one expects to measure a larger value of η when larger angular scales, such as those investigated in Yu et al. (2011), are considered, so the result of Helbig (2015) remains interesting because of the small angular scales of supernova beams.

The Planck Collaboration (2015) measured the CMB lensing-deflection power spectrum at $40\,\sigma$, showing it to agree with the smooth ΛCDM amplitude (i.e. the $\eta = 1$ case) to within 2.5 per cent. Since all forms of gravitating clumps contribute to this, such a measurement of the power as a function of scale is fairly definitive about the smoothness of the energy-density distribution. This should be contrasted with the situation a few decades ago, when it was widely believed that there was no dark matter other than that required for flat rotation curves in spiral galaxies and for bound galaxy clusters; $\eta \approx 0$ was thought to be the best approximation even for objects as large as large galaxies (e.g. Gott et al. 1974; Roeder 1975). To be sure, most of the analysis done by the Planck Collaboration (2015) deals with $L \leq 400$, although $L < 2048$ is also investigated, where L is the multipole. $L = 400$ corresponds to an angular scale of somewhat less than a degree and $L = 2048$ to about 10 arcmin. This means that the corresponding physical size in the concordance model is about 5 Mpc at $z = 1$ (and about 40 kpc at the redshift of the CMB). Thus, the m–z relation for Type Ia supernovae probes much smaller scales, and indicates that even at these scales $\eta \approx 1$ is appropriate, i.e. that the Universe is homogeneous at even these very small scales.

[3] Yu et al. (2011) usually refer to (Ruth) Daly et al. (2008) as ‘Ruth et al.’.

4 CONCLUSIONS

There is a statistically significant correlation between the absolute value of the residuals, i.e. the difference between the observed distance moduli and those calculated from the best-fitting cosmological model, and the observational uncertainties in the Union2.1 sample of Type Ia supernova observations. Each of these quantities is also correlated with redshift but their quotients are not. This suggests that each individual line of sight to these supernovae is a fair sample of the Universe in the sense that the (average) density is approximately the same as the overall density; in other words, it is not necessary to average over several lines of sight in order to recover the overall density. Since most of the matter in the Universe is dark matter, it must be distributed smoothly enough so that most lines of sight contain the same density as the overall average density. When the resolution of cosmological numerical simulations becomes high enough to resolve the corresponding scale, this distribution must result. Rather than putting it in ‘by hand’, it would be more interesting if it emerged from other assumptions or theoretical considerations.

ACKNOWLEDGEMENTS

I thank an anonymous referee for helpful comments. Figures were produced with the GRAL software package written by Rainer Kayser.

REFERENCES

Amanullah R. et al., 2010, ApJ, 716, 712
Busti V. C., Santos R. C., 2011, Res. Astron. Astrophys., 11, 637
Daly R. A., Djorgovski S. G., Freeman K. A., Mory M. P., O’Dea C. P., Carb P., Baum S., 2008, ApJ, 677, 1
Gott J. R., III, Gunn J. E., Schramm D. N., Tinsley B., 1974, ApJ, 194, 543

Helbig P., 2015, MNRAS, 451, 2097
Holz D. E., Linder E. V., 2005, ApJ, 631, 678
Kaiser N., Peacock J. A., 2015, MNRAS, preprint (arXiv:1503.08506)
Kayser R., Helbig P., Schramm T., 1997, A&A, 318, 680
Komatsu E. et al., 2011, ApJS, 192, 18
Lima J. A. S., Busti V. C., Santos R. C., 2014, Phys. Rev. D, 89, 067301
Ostriker J. P., Steinhardt P. J., 1995, Nature, 377, 600
Perlmutter S. et al., 1999, ApJ, 517, 565
Planck Collaboration, 2014, A&A, 571, A16
Planck Collaboration, 2015, A&A, preprint (arXiv:1502.01591)
Roeder R. C., 1975, ApJ, 196, 671
Suzuki N. et al., 2012, ApJ, 746, 85
Weinberg S., 1976, ApJ, 208, L1
Yu H. R., Lan T., Zhang T. J., Wang B. Q., 2011, Res. Astron. Astrophys., 11, 125

This paper has been typeset from a TEX/LATEX file prepared by the author.

9.3 Follow-up

Of course, the concept of describing the small-scale inhomogeneity of the Universe by a single parameter, η, is an approximation. In principle, it can vary depending on line of sight and/or with redshift.[1] (One could of course allow for those effects in a more complex model, but one would need more data to usefully constrain all parameters.) Nevertheless, it seems to be a robust result that $\eta \approx 1$ (Yu *et al.*, 2011; Busti and Santos, 2011; Yang *et al.*, 2013; Bréton and Montiel, 2013; Li *et al.*, 2015; Helbig, 2015a; Dhawan *et al.*, 2018); the question is what that means. It could mean that the Universe is approximately homogeneous, which is presumably a possibility as long as we don't know what dark matter is, much less how it is distributed on the small scales relevant for the m–z relation for Type Ia supernovae, or it could mean that, *on average*, $\eta \approx 1$ when several lines of sight are considered (Weinberg, 1976).

It occurred to me, on my way back to a hotel in Stockholm after having given a talk in Uppsala about the work described in Helbig (2015a), that there is a way to test that: If the second mechanism were responsible, one would expect to see a dispersion, increasing with redshift, in the brightness of supernova at a given redshift, while that would not be the case if the first mechanism were responsible. One does indeed observe such an increasing dispersion with redshift. On the other hand, the observational uncertainties also increase with redshift. Interestingly, the quotient of the two (*i.e.* of the residuals compared to the best-fit model and the observational uncertainties) does not vary with redshift, which suggests that the first mechanism is responsible.

What does *that* mean? It could mean that the Universe actually is approximately homogeneous. On the other hand, it could also mean that the Universe is far from homogeneous, but nevertheless the angular-size distance (and the related luminosity distance) calculated as a function of redshift is approximately the same as in the homogeneous, $\eta = 1$, case. I suspect that the latter is the case. Both observations and numerical simulations suggest that matter in the Universe is distributed in a network of voids, filaments, sheets, and galaxy clusters. In other words, the assumptions on which the ZKDR distance is based are not valid in our Universe. Nevertheless, it could appear that $\eta \approx 1$: at low redshift, η makes little difference; at high redshift, a typical photon will have traversed several voids, filaments, sheets, and perhaps galaxy clusters, so that it has in fact traversed a fair sample of the Universe, resulting in the same focussing effect as if the Universe were indeed approximately homogeneous on small scales (Dyer and Roeder, 1976; Lima *et al.*, 2014).

Thus, despite all the work which has gone into the ZKDR distance, Swiss-cheese models, *etc.*, it appears that, in most cases, one can simply use the standard, $\eta = 1$ distance; Peel *et al.* (2014) arrived at a similar conclusion *via* rather different arguments.

I presented this work together with that described in the previous chapter in a seminar talk at the University of Sussex in August 2015, in a talk at the 28th Texas Symposium on Relativistic Astrophysics in Geneva in December 2015 (Helbig, 2015c), and in a talk at at the cosmology session at the Rencontres de Moriond in March 2016 (Helbig, 2016a). Despite the fact that I was extremely ill the entire week, I somehow managed to give my talk in the second session on Monday morning at the Moriond conference. A few days later, Cliff Burgess told me that he thought that my talk was the best one of the entire conference.

[1] One would expect η to increase with redshift for two reasons: less structure at high redshift, and, for a given angle at the observer, a larger volume sampled.

Part VI

Summary, conclusions, & supplementary material

Chapter 10

Summary and conclusions

As can be seen from Chap. 2, there is an extensive literature on the ZKDR distance. While there has been some debate from time to time, *e.g.* over the appropriateness of approximations or concerning related topics such as the difference between averaging over the sky or over lines of sight, by and large there is a consensus that the ZKDR distance is the appropriate distance measure to use in a universe with the corresponding mass distribution, at least since Fleury (2014) demonstrated with completely analytical arguments the equivalence of the ZKDR distance and that calculated from a certain class of Swiss-cheese models (which are exact solutions of the Einstein equations) at a well controlled level of approximation. Perhaps somewhat surprising historically is the fact that most cosmologists used the standard distances without even considering alternatives, despite the fact that important papers on this topic were published by well known people such as Zel'dovich (1964a,b) and Weinberg (1976). (One reason was perhaps the lack of efficient numerical implementation and of fast computers, which was also a reason why cosmogical models with special values of λ_0 and Ω_0 (and hence analytic distance formulae), such as the Einstein–de Sitter model, were used for the calculation of the standard distance.) While that didn't matter when observations were limited to low redshift, it can matter when computing the distance to higher-redshift sources. The first high-profile work in observational cosmology to take the ZKDR distance into account was that of the Supernova Cosmology Project (Perlmutter *et al.*, 1999), though most subsequent analyses used the standard distance; those that didn't usually reanalysed existing data, rather than taking the ZKDR distance into account in the analysis of new observations.

My own work in this field, prompted by work in gravitational lensing (where it is obvious that small-scale inhomogeneities, *i.e.* at least the gravitational lenses themselves, exist), started with developing an efficient numerical implementation (Helbig, 1996; Kayser, Helbig and Schramm, 1997, Chaps. 3 & 4) for the general form of the second-order differential equation for the ZKDR distance, *i.e.* for arbitrary values of λ_0, Ω_0, and η. I later applied that to extragalactic astronomy (Helbig and Kayser, 1996b, Chap. 5), to the calculation of the Hubble constant from gravitational-lens time delays (Helbig, 1997, Chap. 6), to the question of image separation in gravitational-lens systems (Helbig, 1998a,b, Chap. 7), and to the determination of cosmological parameters from the m–z relation for Type Ia supernovae (Helbig, 2015a,b, Chaps. 8 & 9). In particular, the last arrives at the conclusion, also found by others, that our Universe behaves *as if* $\eta \approx 1$. Although the nature of dark matter is unclear, it is probably not the case that (dark) matter is distributed such that $\eta \approx 1$ (and certainly not the case for luminous matter). Rather, the ZKDR model, with one component smoothly distributed and the other clumpily with light propagating far from all clumps, probably does not describe our Universe (though there is no debate that the ZKDR distance is the correct distance for the ZKDR mass distribution). At low redshift, η doesn't much matter, so the ZKDR distance is essentially the same as the standard distance. At higher redshift, light will have traversed many voids as well as sheets and filaments in the large-scale structure where the density is much higher than the average density of the Universe. Those under- and overdense regions 'average out' in a sense, so that the resulting distance is approximately the same as the standard distance, at least in most cases.

Does that mean that one can just use the standard distance in observational cosmology? For a rough approximation, probably. For detailed analyses, probably not. Even if most lines of sight traverse a fair sample of the Universe, that is not guaranteed. While the ZKDR distance for $\eta = 0$ gives the maximum possible distance for a given redshift, it is possible that the apparent distance is less than that given by the ZKDR distance for $\eta = 1$. In such cases, the assumptions on which the ZKDR distance are based are invalid, though for $\eta \gtrsim 1$ a meaningful calculation is still possible. The proper way to deal with such situations is to explicitly consider it as a gravitational-lens system, but that is difficult when the lens(es) might not be visible. At the very least, one should consider the appropriateness of all assumptions made when analysing data, not only with regard to distance measures.

Acknowledgements

Rainer Kayser introduced me to the topic of distance calculation in inhomogeneous cosmological models. As he was an assistant professor in Sjur Refsdal's group at the Hamburg Observatory at the time, I had heard some of his lectures. He had been one of the main organizers of the first international scientific conference I attended, the International Conference on Gravitational Lenses in Hamburg in September 1991 (Kayser *et al.*, 1992), which coincidentally had taken place just a few months before I started work on my master's thesis at the Hamburg Observatory. Despite having to work elsewhere for a living, I stayed on at the Hamburg Observatory after completing my master's thesis and obtaining the corresponding diploma in physics with a minor in astronomy. During that time I continued working with Rainer, both on my first paper (Helbig and Kayser, 1996a) and on the paper by Kayser, Helbig and Schramm (1997), while I was developing the ANGSIZ code. Years of lunchtime discussions with Rainer and Tom Schramm were an important part of the start of my work in cosmology.

That might have been the end of my career had Ian Browne not given me the chance to continue my work in cosmology at Jodrell Bank, as part of an EU network, CERES (Consortium for European Research on Extragalactic Surveys), which he coordinated. It was definitely the right place and the right time and I hope that I was the right person. When I was there, there were few theoreticians at Jodrell Bank. Although I remained a theoretician, it was an invaluable experience working at a major observatory. CERES also gave me a chance to work not only a bit more than two years at Jodrell Bank but also almost two years at the Kapteyn Institute in Groningen.

All of my colleagues at all three institutes have been an important part of my education. In many respects, the years spent there were the best years of my life.

Sjur Refsdal tended to work on one thing at a time. While I was in Hamburg, he was working mainly on large-source microlensing, somewhat removed from my interest in cosmology (though I later did a bit of work on microlensing (Zackrisson *et al.*, 2002, 2003)), thus I worked most closely with Rainer. Ironically, Sjur's work on cosmology with Rolf Stabell in the 1960s (Stabell and Refsdal, 1966; Refsdal, Stabell and de Lange, 1967) has been a major influence on my approach to cosmology. After leaving the Kapteyn Institute, I worked outside of academia for twenty years, always intending to obtain a doctorate and to return to research. Initially I had hoped to do that with the help of Sjur, but he died after a long illness just before I finally had the time to concentrate more on research. Nevertheless, I appreciate his confidence in me, along with that of Rolf.

In the last twenty years, I nevertheless managed to write a few papers, attend a few conferences (sometimes presenting a talk or poster, almost always as a single author), and give talks at institutes. I thank the people at the astronomical institutes in Uppsala, Oslo, and Brighton for their hospitality, and also the conference organizers and journal editors who did not reject me because of lack of institutional affiliation (some conference organizers and journal editors did reject me for that reason).

My second conference (after the one in Hamburg mentioned above, which I might not have attended had it not been in Hamburg and/or if I hadn't known the organizers) was the conference on gravitational lensing in Liège in 1993 (Surdej *et al.*, 1993), which 14 people from Sjur's group attended. Sjur was a collaborator of Jean Surdej, thus it is appropriate that things come full circle for me as a doctoral student in Liège. I'm particularly grateful for Jean's confidence in me, to Christian Barbier for agreeing to be my advisor, and to both for being part of the jury. I also thank the other jury members Ludovic Delchambre, Marc-Antoine Dupret, Rolf Stabell, and Michel Tytgat. I have always enjoyed my visits to Liège but unfortunately the COVID-19 pandemic caused me to spend much less time there than I would have liked.

I thank the writer and former astronomer Pippa Goldschmidt for helpful comments on my writing style.

Completing this thesis would not have been possible without the support of my family; I'm grateful that they allowed me to take the necessary time.

Figures were produced with the GRAL software package written by Rainer Kayser.

Bibliography

Barrow, J. D. 2012. *The Book of Universes* (London, Vintage)

Barrow, J. D. 2017. In The Philosophy of Cosmology, ed. K. Chamcham, J. Silk, J. D. Barrow, & S. Saunders (Cambridge, Cambridge Univ. Press), 83–108

Biggs, A., Browne, I. W. A., Helbig, P. *et al.* 1999. Time Delay for the Gravitational Lens System B0218+357. *MNRAS* **304:** 349–358

Bondi, H. 1960. *Cosmology* (Cambridge (UK), Cambridge Univ. Press)

Bonvin, V., Courbin, F., Suyu, S. H. *et al.* 2017. H0LiCOW V. New COSMOGRAIL time delays of HE 0435 − 1223: H_0 to 3.8 per cent precision from strong lensing in a flat ΛCDM model. *MNRAS* **465:** 4914–4930

Bréton, N. and Montiel, A. 2013. Observational constraints from supernovae Ia and gamma-ray bursts on a clumpy universe. *Phys. Rev. D* **87:** 063527

Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F. *et al.* 2003. The Cosmic Lens All-Sky Survey - II. Gravitational lens candidate selection and follow-up. *MNRAS* **341:** 13–32

Busti, V. C. and Santos, R. C. 2011. Comment on "Constraining the smoothness parameter and dark energy using observational $H(z)$ data". *Res. Astron. Astrophys.* **11:** 637–640

Chae, K.-H., Biggs, A. D., Blandford, R. D. *et al.* 2002. Constraints on Cosmological Parameters from the Analysis of the Cosmic Lens All Sky Survey Radio-Selected Gravitational Lens Statistics. *Phys. Rev. Lett.* **89:** 151301

Cooray, A. R. 1999. Cosmological Parameters from Statistics of Strongly Lensed Radio Sources. *A&A* **342:** 353-62

de Sitter, W. 1916a. On Einstein's Theory of Gravitation and its Astronomical Consequences. First Paper. *MNRAS* **76:** 699–728

de Sitter, W. 1916b. On Einstein's Theory of Gravitation and its Astronomical Consequences. Second Paper. *MNRAS* **77:** 155–184

de Sitter, W. 1916c. Space, Time, and Gravitation. *Obs.* **39:** 412–419

de Sitter, W. 1917a. On the Relativity of Inertia. Remarks concerning Einstein's latest Hypothesis. *Koning. Naturw. Akadm. Ned.* **19:** 1217–1225

de Sitter, W. 1917b. On the curvature of space. *Koning. Naturw. Akadm. Ned.* **20:** 229–243

de Sitter, W. 1917c. On Einstein's Theory of Gravitation and its Astronomical Consequences. Third Paper. *MNRAS* **78:** 3–28 de Sitter, W. 1934. On distance, magnitude, and related quantities in an expanding universe. *Bull. Astr. Inst. Neth.* **7:** 205

Dev, A., Jain, D. and Mahajan, S. 2004. Dark Energy and the Statistical Study of the Observed Image Separation of the Multiply-Imaged Systems in the Class Statistical Sample. *Int. J. Mod. Phys. D* **13:** 1005–1018

Dhawan, S., Goobar, A. and Mörtsell, E. 2018. The effect of inhomogeneities on dark energy constraints. *JCAP* **7:** 024

Dyer, C. C. and Roeder, R. C. 1976. Clusters of Galaxies as Gravitational Lenses. *Nat.* **260:** 764–765

Einstein, A. 1917. Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitzungsb. Kön. Pr. Akad. Wiss.* **VI:** 142–152

Einstein, A. and de Sitter, W. 1932. On the Relation between the Expansion and the Mean Density of the Universe. *Proc. Natl. Acad. Sci. USA* **18:** 213–214

Etherington, I. M. H. 1933. LX. On the Definition of Distance in General Relativity. *Philosophical Magazine* **15:** 761–733

Feige, B. 1992. Elliptic integrals for cosmological constant cosmologies. *AN* **313:** 139

Ferreira, P. G. 2014. *The Perfect Theory: A Century of Geniuses and the Battle over General Relativity* (Boston New York, Houghton Mifflin Harcourt)

Fleury, P. 2014. Swiss-cheese models and the Dyer-Roeder approximation. *JCAP* **6:** 054

Fleury, P., Dupuy, H. and Uzan, J.-P. 2013. Can All Cosmological Observations Be Accurately Interpreted with a Unique Geometry. *Phys. Rev. Lett.* **111:** 091302

Friedmann, A. A. 1922. Über die Krümmung des Raumes. *Zeitschr. Phys.* **1:** 377–386

Friedmann, A. A. 1924. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung. *Zeitschr. Phys.* **21:** 326–332

Gamow, G. 1956. The Evolutionary Universe. *Sci. Am.* **195:** 136–156

Gamow, G. 1970. *My World Line* (New York, Viking Press)

Han, D.-H. and Park, M.-G. 2015. Constraining Cosmological Parameters with Image Separation Statistics of Gravitationally Lensed SDSS Quasars: Mean Image Separation and Likelihood Incorporating Lens Galaxy Brightness. *J. Kor. Astr. Soc.* **48:** 83–92

Harrison, E. R. 1993. The Redshift-Distance and Velocity-Distance Laws. *ApJ* **403:** 28–31

Harrison, E. R. 2000. *Cosmology, the Science of the Universe, 2nd edn.* (Cambridge (UK), Cambridge Univ. Press)

Harwit, M. 2013. *In Search of the True Universe: The Tools, Shaping, and Cost of Cosmological Thought* (Cambridge (UK), Cambridge Univ. Press)

Heacox, W. D. 2015. *The Expanding Universe: A Primer on Relativistic Cosmology* (Cambridge (UK), Cambridge Univ. Press)

Helbig, P. 1996, `ANGSIZ` User's Guide, arXiv:astro-ph/9603028v3

Helbig, P. 1997. In Golden Lenses, ed. P. Helbig & N. Jackson, University of Manchester, NRAL, Jodrell Bank, `http://www.astro.multivax.de:8000/ceres/workshop1/proceedings.html`

Helbig, P. 1998a. The $\delta\theta$-$z_{\rm s}$ Relation as a Cosmological Test. *MNRAS* **298:** 395–398

Helbig, P. 1998b. In Large Scale Structure: Tracks and Traces. Proceedings of the 12th Potsdam Cosmology Workshop, held in Potsdam, September 15th to 19th, 1997, ed. V. Müller, S. Gottlöber, J.-P. Mücket, & J. Wambsganß (Singapore, World Scientific), 319–320

Helbig, P. 1999. Gravitational Lensing Statistics with Extragalactic Surveys. III. Joint Constraints on λ_0 and Ω_0 from Lensing Statistics and the m-z Relation for Type Ia Supernovae. *A&A* **350:** 1–8

Helbig, P. 2012. Is there a flatness problem in classical cosmology? *MNRAS* **421:** 561–569

Helbig, P. 2013. Review of *The Book of Universes* by John D. Barrow. *Obs.* **133:** 232–233

Helbig, P. 2014a. Review of *The Perfect Theory: A Century of Geniuses and the Battle over General Relativity* by Pedro G. Ferreira. *Obs.* **134:** 214–216

Helbig, P. 2014b. Review of *In Search of the True Universe: The Tools, Shaping, and Cost of Cosmological Thought* by Martin Harwit. *Obs.* **134:** 217–220

Helbig, P. 2015a. The m-z relation for type Ia supernovae, locally inhomogeneous cosmological models, and the nature of dark matter. *MNRAS* **451:** 2097–2107

Helbig, P. 2015b. The m-z relation for type Ia supernovae: safety in numbers or safely without worry? *MNRAS* **453:** 3975–3979

Helbig, P. 2015c. In Proceedings of the 28th Texas Symposium on Relativistic Astrophysics, `https://indico.cern.ch/event/336103/page/5620-proceedings`

Helbig, P. 2016a. In 2016 Cosmology: Proceedings of the 51st RENCONTRES DE MORIOND, ed. E. Augé, J. Dumarchez, & J. T. T. Vân, Cosmology, Rencontres de Moriond (ARISF), 171–174

Helbig, P. 2016b. Review of *The Expanding Universe: A Primer in Relativistic Cosmology* by William D. Heacox. *Obs.* **136:** 204–207

Helbig, P. 2017. A Formula for Confusion. *Obs.* **137:** 22–25

Helbig, P. 2019. Review of *The Oxford Handbook of the History of Modern Cosmology* editd by Helge Kragh and Malcolm S. Longair. *Obs.* **139:** 219–222

Helbig, P. 2020a. Calculation of distances in cosmological models with small-scale inhomogeneities and their use in observational cosmology: a review. *Open J. Astroph.* **3:** 1

Helbig, P. 2020b. The flatness problem and the age of the Universe. *MNRAS* **495:** 3571-3575

Helbig, P. 2020c. Whatever happened to the Dyer-Roeder distance? *Obs.* **140:** 128–144

Helbig, P. 2020d. Sonne und MOND, or, the good, the bad, and the ugly: comments on the debate between MOND and ΛCDM. *Obs.* **140:** 225–247

Helbig, P. 2021. Arguments against the flatness problem in classical cosmology: a review. *Eur. Phys. J. H* **46:** 10

Helbig, P. and Jackson, N., eds. 1997. Golden Lenses, University of Manchester, NRAL, Jodrell Bank, `http://www.astro.multivax.de:8000/ceres/workshop1/proceedings.html`

Helbig, P. and Kayser, R. 1996a. Cosmological Parameters and the Redshift Distribution of Gravitational Lenses. *A&A* **308:** 359–367

Helbig, P. and Kayser, R. 1996b. Are the clumps at a Redshift of 2.39 really sub-galactic? *rejected*, unpublished as individual paper

Helbig, P., Marlow, D. R., Quast, R. *et al.* 1999. Gravitational Lensing Statistics with Extragalactic Surveys. II. Analysis of the Jodrell Bank-VLA Astrometric Survey. *A&AS* **136:** 297–305

Jackson, N. J. F., Helbig, P., Browne, I. W. A. *et al.* 1998. Lensing Galaxies: Light or Dark? *A&A* **334:** L33–L36

Kaiser, N. and Peacock, J. A. 2016. On the bias of the distance-redshift relation from gravitational lensing. *MNRAS* **455:** 4518–4547

Kayser, R. 1985, doctoral dissertation, University of Hamburg

Kayser, R. 1995. A Cosmological Test with Compact Radio Sources. *A&A* **294:** L21–L23

Kayser, R., Helbig, P. and Schramm, T. 1997. A General and Practical Method for Calculating Cosmological Distances. *A&A* **318:** 680–686

Kayser, R. and Refsdal, S. 1983. The difference in light travel time between gravitational lens images. I. Generalization of the wavefront method to arbitrary deflectors and inhomogeneous universes. *A&A* **128:** 156–161

Kayser, R., Schramm, T. and Nieser, L., eds. 1992. Gravitational Lenses (Heidelberg, Springer-Verlag)

Kragh, H. and Longair, M. S., eds. 2019. *The Oxford Handbook of the History of Modern Cosmology* (Oxford, Oxford Univ. Press)

Lake, K. 1981. Comment on the time evolution of the cosmological redshift. *ApJ* **247:** 17–18

Lanczos, C. 1922. Bemerkung zur de Sitterschen Welt. *Phys. Zeitschr.* **23:** 539–543

Lemaître, G. 1927. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Ann. Soc. Sci. Brux.* **47:** 49–59

Lemaître, G. 1931a. A Homogeneous Universe of Constant Mass and Increasing Radius accounting for the Radial Velocity of Extra-galactic Nebulæ. *MNRAS* **91:** 483–490

Lemaître, G. 1931b. The Expanding Universe. *MNRAS* **91:** 490–501

Lemaître, G. 1931c. The beginning of the World from the point of view of quantum theory. *Nat.* **127:** 706

Lemaître, G. 1931d. Untitled contribution to the discussion 'The question of the relation of the physical universe to life and mind'. *Supplement to Nature* **127:** 704–706

Lemaître, G. 1931e. L'Expansion de l'Espace. *Publications du Laboratoire d'Astronomie et de Geodesie de l'Universit'e de Louvain* **8:** 101–120

Li, Z., Ding, X. and Zhu, Z.-H. 2015. Unbiased constraints on the clumpiness of the Universe from standard candles. *Phys. Rev. D* **91:** 083010

Liebscher, D.-E., Priester, W. and Hoell, J. 1992. Lyman α forests and the evolution of the universe. *AN* **313:** 265-273

Lima, J. A. S., Busti, V. C. and Santos, R. C. 2014. Studying light propagation in a locally homogeneous universe through an extended Dyer-Roeder approach. *Phys. Rev. D* **89:** 067301

Loeb, A. 1998. Direct Measurement of Cosmological Parameters from the Cosmic Deceleration of Extragalactic Objects. *ApJ* **499:** L111–L114

Luminet, J.-P. 2011. Editorial note to: Georges Lemaiîtrem, The beginning of the world from the point of view of quantum theory. *Gen. Rel. Grav.* **43:** 2911

Macias Perez, J. F., Helbig, P., Quast, R., Wilkinson, A. and Davies, R. 2000. Gravitational Lensing Statistics with Extragalactic Surveys. IV. Joint Constraints on λ_0 and Ω_0 from Gravitational Lensing Statistics and CMB Anisotropies. *A&A* **353:** 419–426

Mattig, W. 1958. Über den Zusammenhang zwischen Rotverschiebung und scheinbarer Helligkeit. *AN* **284:** 109-111

McKean, J. P., Koopmans, L. V. E., Browne, I. W. A. *et al.* 2004. Keck Spectroscopy of Cosmic Lens All-Sky Survey gravitational lenses. *MNRAS* **350:** 167–174

McVittie, G. C. 1962. Appendix to the change of redshift and apparent luminosity of galaxies due to the deceleration of selected expanding universes. *ApJ* **136:** 334–338

Milne, E. A. 1935. *Relativity, Gravitation, and world-structure* (Oxford, Clarendon Press)

Müller, V., Gottlöber, S., Mücket, J.-P. and Wambsganß, J., eds. 1998. Large Scale Structure: Tracks and Traces. Proceedings of the 12th Potsdam Cosmology Workshop, held in Potsdam, September 15th to 19th, 1997 (Singapore, World Scientific)

Nowakowski, M. 2001. The consistent Newtonian limit of Einstein's gravity with a cosmological constant. *Int. J. Mod. Phys. D* **10:** 649–662

Odderskov, I., Koksbang, S. M. and Hannestad, S. 2016. The local value of H_0 in an inhomogeneous universe. *JCAP* **2:** 001

O'Raifeartaigh, C., O'Keeffe, M., Nahm, W. and Mitton, S. 2018. One hundred years of the cosmological constant: from "superfluous stunt" to dark energy. *Eur. Phys. J. H* **43:** 73–117

Overduin, J. and Priester, W. 2001. Quasar absorption-line number density in a closed, Λ-dominated universe. *Astrophys. Space Sci.* **88:** 229–248

Park, M.-G. and Gott, III, J. R. 1997. Curvature of the Universe and Observed Gravitational Lens Image Separations versus Redshift. *ApJ* **489:** 476–484

Pascarelle, S. M., Windhorst, R. A., Keel, W. C. and Odewahn, S. C. 1996. Sub-Galactic Clumps at a Redshift of 2.39 and Implications for Galaxy Formation. *Nat.* **383:** 45–50

Peel, A., Troxel, M. A. and Ishak, M. 2014. Effect of inhomogeneities on high precision measurements of cosmological distances. *Phys. Rev. D* **90:** 123536

Perlmutter, S., Aldering, G., Goldhaber, G. *et al.* 1999. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ* **517:** 565–586

Quast, R. and Helbig, P. 1999. Gravitational Lensing Statistics with Extragalactic Surveys. I. A Lower Limit on the Cosmological Constant. *A&A* **344:** 721–734

Rana, A., Jain, D., Mahajan, S. and Mukherjee, A. 2017a. Constraining cosmic curvature by using age of galaxies and gravitational lenses. *JCAP* **3:** 028

Rana, A., Jain, D., Mahajan, S., Mukherjee, A. and Holanda, R. F. L. 2017b. Probing the cosmic distance duality relation using time delay lenses. *JCAP* **7:** 010

Refsdal, S. 1964. On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect. *MNRAS* **128:** 307–310

Refsdal, S. 1966. On the possibility of testing cosmological theories from the gravitational lens effect. *MNRAS* **132:** 101-111

Refsdal, S. and Stabell, R. 1991. Gravitational microlensing for large sources. *A&A* **250:** 62–65

Refsdal, S., Stabell, R. and de Lange, F. G. 1967. Numerical calculations on relativistic cosmological models. *Mem. R. Astron. Soc.* **71:** 143–248

Rindler, W. 1956. Visual horizons in world models. *MNRAS* **116:** 662–677

Robertson, H. P. 1935. Kinematics and World-Structure. *ApJ* **82:** 284–301

Robertson, H. P. 1936. Kinematics and World-Structure II.. *ApJ* **83:** 187–201

Rüdiger, R. 1980. Time dependence of the cosmological redshift in Friedmann universes. *ApJ* **240:** 384–386

Rusu, C. E., Fassnacht, C. D., Sluse, D. *et al.* 2017. H0LiCOW III. Quantifying the effect of mass along the line of sight to the gravitational lens HE 0435 – 1223 through weighted galaxy counts. *MNRAS* **467:** 4220–4242

Sandage, A. R. 1962. The Change of Redshift and Apparent Luminosity of Galaxies Due to the Deceleration of Selected Expanding Universes. *ApJ* **136:** 319–333

Sandage, A. R. 1970. Cosmology: a search for two numbers. *Physics Today* **23:** 34–41

Santos, R. C. and Lima, J. A. S. 2006, ZKDR Distance, Angular Size and Phantom Cosmology, arXiv:astro-ph/0609129

Smoot, G. F., Bennett, C. L., Kogut, A. *et al.* 1992. Structure in the COBE differential microwave radiometer first-year maps. *ApJ Lett.* **396:** L1–L5

Stabell, R. and Refsdal, S. 1966. Classification of general relativistic world models. *MNRAS* **132:** 379–388

Surdej, J., Fraipont-Caro, D., Gosset, E., Refsdal, S. and Remy, M., eds. 1993. Gravitational Lenses in the Universe (31^{st} Liège International Astrophysical Colloquium) (Liège, Université de Liège)

Tolman, R. C. 1934. Effect of inhomogeneity on Cosmological Models. *Proc. Natl. Acad. Sci. USA* **20:** 169–76

Walker, A. G. 1935. On Riemannian Spaces with Spherical Symmetry about a Line, and the Conditions of Isotropy in General Relativity. *Quart. J. Math.* **6:** 81–93

Walker, A. G. 1937. On Milne's Theory of World-Structure. *Proc. London Math. Soc.(Series 2)* **42:** 90–127

Walsh, D., Carswell, R. F. and Weymann, R. J. 1979. 0957 + 561 A, B: Twin Quasistellar Objects or Gravitational Lens? *Nat.* **279:** 381–384

Weinberg, S. 1976. Apparent luminosities in a locally inhomogeneous universe. *ApJ* **208:** L1–L3

Weinberg, S. 2005. Einstein's Mistakes. *Physics Today* **58:** 31–35

Yang, X., Yu, H.-R. and Zhang, T.-J. 2013. Constraining the smoothness parameter and the DD relation of Dyer-Roeder equation with supernovae. *JCAP* **6:** 007

Yu, H.-R., Lan, T., Wan, H.-Y., Zhang, T.-J. and Wang, B.-Q. 2011. Constraints on smoothness parameter and dark energy using observational $H(z)$ data. *Res. Astron. Astrophys.* **11:** 125–136

Zackrisson, E., Bergvall, N. and Helbig, P. 2002. In ESO Astrophysics Symposia, Vol. XIV, Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology. Proceedings of the MPA/ ESO/ MPE/ USM Joint Astronomy Conference held in Garching, Germany, 6-10 August 2001, ed. M. Gilfanov, R. Sunyaev, & E. Churazov (Springer), 550–552

Zackrisson, E., Bergvall, N., Marquart, T. and Helbig, P. 2003. Can microlensing explain the long-term optical variability of quasars? *A&A* **408:** 17–25

Zel'dovich, Y. B. 1964a. Nablyudeniya vo vselennoy odnorodnoy v srednem. *Astronomicheskii Zhurnal* **41:** 19-24

Zel'dovich, Y. B. 1964b. Observations in a Universe Homogeneous in the Mean. *SvA* **8:** 13-16

Zhu, Z.-H. 2000. Gravitational lensing statistical properties in general FRW cosmologies with dark energy component(s): analytic age of galaxies and gravitational results. *Int. J. Mod. Phys. D* **9:** 591–600