Redshift: Observation and Geometry

H. Karcher (Bonn)

Abstract. The observation of redshift is one of the most dominant experimental facts in astronomy. In this lecture I want to explain the differential geometric nature of redshift. Jacobi fields are a principal tool.

Redshift Observation. In this section I use historical language, in particular the word "distance" is used without relativistic conotation. A light beam from the sun can be spread with a prism into a rainbow band of coloured light. A more powerful instrument shows a very large number of dark lines in such a spectrum of the sun; these Fraunhofer lines represent frequencies that are missing (or are at least weakened) in a continuous band of frequencies. In a laboratory one can obtain a spectrum with few dark lines if one sends light with a continuous spectrum (e.g. emitted by a very hot object) through vapour of one chemical element and then spreads the light with a prism: The vapour absorbs (and therefore weakens) frequencies that are specific to the element from which the vapour is made. A vapour which consists of many elements can still be analyzed via its spectrum: There are enough characteristic frequencies in the absorption spectrum of each single element that one can identify elements in the vapour by matching their families of characteristic frequencies to a subset of absorption lines of the vapour. The Fraunhofer lines therefore tell us what elements are in the outer atmosphere of the sun. Finer details of the spectrum (such as width and strength of the absorption lines) show the temperature of the sun's atmosphere and also the relative proportions of the elements in the atmosphere. We can also obtain the spectra of other stars and one should imagine that their spectra are so complicated that it is fair to assume:

Two stars with the same spectrum are physically very similar.

This is the crucial hypothesis for distance comparisons: If two stars with the same spectrum are observed and one is a quarter as bright as the other, then the brighter one is half as far away. The nearest fixed stars for which we also have trigonometric distance measurements support this fundamental hypothesis. – For cosmological observations it is more important to identify "physically very similar galaxies" or "physically very similar clusters of galaxies", but this is more difficult to define than for stars.

We have the following fundamental observational fact: *In all the very many cases in which a spectrum was obtained from some astronomical object one found, without exception, that all the frequency* RATIOS *of the Fraunhofer lines are as expected but the absolute frequency* VALUES *are shifted.* For nearer objects both higher frequencies (blueshift) and lower frequencies (redshift) are observed, but for distant objects, remarkably, only redshift is observed. The simplest explanation interprets this frequency shift as a "Doppler effect": if a light source and an observer move towards each other, then the observed frequencies are higher than the emitted ones, while, if they move away from each other, then the observed frequencies are lower than the emitted ones (see below for a quantitative statement).

Natural Clocks. Fundamental discussions of relativity theory are concerned with the comparison of time measurements of different observers. For such discussions one has to give a meaning to the phrase: "Different observers use the same type of clock." Suppose we agree that the frequency of a specified transition of a specified element defines the units of time. Then we have a very pleasant surprise: The above fundamental fact about the *joint frequency shift* of all the Fraunhofer lines from the same astronomical object means that all the atomic clocks there agree among each other. Transitions of quantummechanical systems are also at the heart of radioactive decay: We may, therefore, also take as units of time the half life time of a specified radioactive material. Such clocks then convey very strongly the impression that time passes whether we measure it or not, and the amount of material left is always a record of how much time passed. We cannot as directly as with the Fraunhofer lines observe that all such half life time clocks (on the same astronomical object) agree, but since our models of stellar evolution depend heavily on half life times of various nuclei we do have indirect confirmation. For the discussion of time in relativity it is essential to realize that time measurements are not artificial conventions: Radioactive elements naturally decay, thus keeping track of the passing of time; moreover this is in agreement also with time units specified by transition frequencies. This natural passing of time is part of our world, everywhere. The time measured by natural clocks is called *proper time.*

Doppler redshift and Minkowski metric. Mathematicians are probably inclined to begin with $\mathbb{R}^4$ together with Minkowski's bilinear form as "the geometry" of special relativity. The discussion of redshift allows one to deduce this geometry from fewer assumptions. Assume we have at first one distinguished observer A (representing us) and another observer B moving with constant velocity v away from A. A uses $\mathbb{R}^3 \times \mathbb{R}$ to plot his observations, and the goal is to find the relevant geometry for $\mathbb{R}^3 \times \mathbb{R}$. (Then all observers who move without acceleration will become equivalent.) Observer A calls $\{0\} \times \mathbb{R}$ his own world line and $\{(t \cdot v, t) \in \mathbb{R}^3 \times \mathbb{R},\ v \in \mathbb{R}^3 \ \text{ constant}\}$ the worldline of B. A has chosen a natural clock so that $\{0\} \times \mathbb{R}$ has a unit of time. The problem is: What is the unit of time on the worldline of B? We will now show that the units of time on all the other world lines form a level surface of an indefinite quadratic form, the Minkowski metric. We look only at the plane spanned by the two world lines of A and B. A sends two light signals one unit of time apart to follow B. They arrive with the unknown time difference $T \cdot 1$ at B. Knowing T would fix the time units on the world line of B. B returns each of the two light signals immediately. Since A moves away from B with the *same* speed as vice versa the relativity principle says that the time difference between the returned light signals, when they arrive at A, is observed as changed by the *same* factor T again, i.e. they arrive with time difference $T \cdot T$ at A. Now A can compute T^2 as a function of v as follows (see figure):

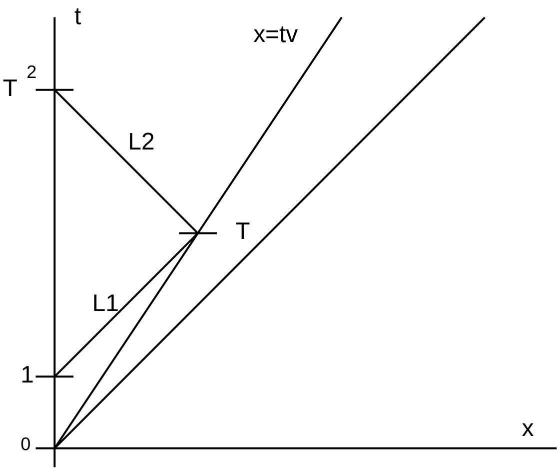
Observer A uses space units so that the velocity of light (known to be finite) is 1. The first light signal is emitted and returned while A and B meet; the figure shows only the light signal which is emitted one unit of time later.

The second light signal, as described by the outgoing light ray $L_1 : s \to (s, 1+s)$, reaches the worldline of B: $t \to (t \cdot v, t)$ at parameter value t such that $s = t \cdot v$ and $1 + s = t$, i.e. at $t = 1/(1-v)$, $s = t - 1 = v/(1-v)$.

(Note that so far t is an artificial parameter, proportional to proper time, on the world line of B, it has not yet a specific connection with the physical units of time of the natural clocks of A or B.)

The returning light ray $L_2 : s \to \left(\frac{v}{1-v}, \frac{1}{1-v}\right) + (-s, +s)$ reaches the world line of A at $s = v/(1-v)$, i.e. at the point $\left(0, \frac{1+v}{1-v}\right)$. Since this world line has already a time unit we find for the arrival time at A of the second signal

$$T^2 = (1+v)/(1-v).$$



Redshift in Special Relativity: $T = \big((1+v)/(1-v)\big)^{1/2}$.

The time unit point on the worldline of B is therefore the point

$$\frac{1}{T}\cdot\left(\frac{v}{1-v},\frac{1}{1-v}\right) = \left(\frac{v}{\sqrt{1-v^2}},\frac{1}{\sqrt{1-v^2}}\right).$$

In the coordinates of A these unit points lie on the quadratic surface $t^2 - |x|^2 = 1$. So indeed, we arrived at the Minkowski Metric, and we used only the relativity principle and also that the world lines of light rays are independent of the speed (or the world line) of the emitter. I repeat, that this Minkowski geometry is physically highly relevant: Let us choose as units of time the half life time of radium and let us consider the straight world lines of several chunks of radium flying apart with various constant speeds from $(0,0)$. Each of these these worldlines meets the unit time surface $t^2 - |x|^2 = 1$ of the Minkowski geometry at a physically relevant point: exactly one half of each of the radium chunks has vanished by decay.

If a world line is not a straight line, then, using the Minkowski metric, one may still parametrize the curve such that each tangent vector is a unit timelike vector. The model interpretation is then extended as follows: *The geometrically distinguished parameter on the world line gives the time of a natural clock on this world line.* This extended interpretation is in excellent agreement with accelerator experiments. Note that this eliminates vagueness from the "Twin discussions": The geometry of the world line determines how much radium decays; if two world lines start from a common point and meet again, but the distinguished parameter on each gives them a *different* age, then this different age is indeed the intended interpretation, namely, smaller age means less radium has decayed. If one suspects a contradiction to the principle of relativity, then one has not accepted the geometric nature of natural (or proper) time on a straight or curved world line as part of the model interpretation in special relativity.

Redshift in Lorentz Manifolds. Let M be a 4-dimensional manifold and g a Lorentz metric on M. We need some preliminary facts about light cones which we derive before we come back to the discussion of redshift. The (forward) light cone at $p \in M$ is made up of the (forward) null geodesics starting at p. At conjugate points these light cones may not be hypersurfaces, but I do not go into such details. Any curve $\varepsilon \to c(\varepsilon)$ on a light cone defines a variation $(s,\varepsilon) \to c(s,\varepsilon)$ of null geodesics $s \to c(s,\varepsilon)$ connecting p and $c(\varepsilon)$. Hence

$$\frac{\nabla}{\partial s}\frac{\partial}{\partial s}c(s,\varepsilon) = 0, \quad g\left(\frac{\partial}{\partial s}c(s,\varepsilon), \frac{\partial}{\partial s}c(s,\varepsilon)\right) = 0, \quad c(0,\varepsilon) = p,$$

$$\frac{d}{ds}g\left(\frac{\partial}{\partial \varepsilon}c(s,\varepsilon), \frac{\partial}{\partial s}c\right) = g\left(\frac{\nabla}{\partial s}\frac{\partial}{\partial \epsilon}c(s,\varepsilon), \frac{\partial}{\partial s}c(s,\varepsilon)\right) = 0, \quad \frac{\partial}{\partial \varepsilon}c(0,\varepsilon) = 0.$$

This proves: Since each tangent vector X to a light cone at the point $c(s)$ of a generating null geodesic c can be represented as $X = \frac{\partial}{\partial \varepsilon}c(s,\varepsilon)$ we have

$$g\left(X, \frac{\partial}{\partial s}c\right) = 0.$$

Except at conjugate points, this equation defines the tangent hyperplane of the null cone at $c(s)$ (at $s = 0$ in the direction $\frac{\partial}{\partial s}c(0)$). If we extend X to a parallel vectorfield along $c(s)$, then $g(X(s), \frac{\partial}{\partial s}c(s)) = 0$ along c. This says that the tangent hyperplanes to the geodesic null cone along the null ray c are *Levi-Civita parallel*. Finally, if $c(s,\varepsilon)$ is a variation of null geodesics not necessarily on the same light cone (i.e. without the assumption $c(0,\varepsilon) = p$ in the above computation), then the Jacobi field $\frac{\partial}{\partial \varepsilon}c(s,\varepsilon)|_{\varepsilon=0}$ can be split as $u(s) + T(s)$, where $u(s)$ is *parallel* along the null ray $c(s,0)$ and $T(s)$ is tangent to the light cone along $c(s,0)$. (In general $u(s)$ and $T(s)$ are not Jacobi fields and not even unique – they may be altered by a parallel vector field tangent to the light cone along c.)

Now consider a world line $a(t)$, parametrized by proper time, i.e. $g(a', a') = -1$; we call $a()$ the world line of a light source. Consider also the family of light cones from $a(t)$. Light signals from this source are represented by null rays on these light cones. In addition consider another world line $b(\tau)$, $g(b', b') = -1$, called the world line of an observer. In general the observer world line will meet the light cones of the source transversally and therefore we have a family of null rays joining the two world lines. Infinitesimally we have a Jacobi field J along a joining null ray; it is transversal to the light cone: $J(0) = a'(t_0)$, $J(1) = \lambda \cdot b'(\tau_1)$. The interpretation of this is: Light signals which leave the source at unit time intervals of the clock of $a(\ )$ arrive at λ times unit time intervals of the clock of $b(\ )$. The convention in physics is to take the ratio of the emitted frequency $\omega_{emitted}$ (measured by a) divided by the observed frequency $\omega_{observed}$ (measured by b) as the observed frequency shift. This ratio is λ in our Jacobi field description.

Finally we apply the above preliminary facts. Consider a light signal which leaves the source (which has time unit vector a') with frequency $\omega_{emitted}$, travels along the affinely parametrized null ray $c(s)$ and arrives at the observer (time unit vector b') with frequency $\omega_{observed}$. Then, since the scalar product $g(J(s), c'(s))$ is *constant* along the joining null ray, i.e. $g(a', c') = g(\lambda \cdot b', c')$, we have derived the

Observed Frequency Shift:

$$\lambda = \frac{g(a', c')}{g(b', c')} = \frac{\omega_{emitted}}{\omega_{observed}}$$

for observers a', b' (i.e. $g(a', a') = -1 = g(b', b')$) which are joined by a null ray c with affine parametrization, i.e $\frac{\nabla}{\partial s}c' = 0$.

This formula generalizes what we know from special relativity (see Redshift and Minkowski Metric). The special situation was: $a' = (0,1)$, $b' = \left(v/\sqrt{1-v^2},\ 1/\sqrt{1-v^2}\right)$, $c' = (1,1)$, and the quotient of scalar products gives the special relativistic Doppler redshift

$$\frac{g(a',c')}{g(b',c')} = \sqrt{\frac{1+v}{1-v}}$$

i.e., the redshift which we have already discussed.

Gravitational redshift in the Schwarzschild geometry. Arguably the physically most important Lorentz manifold was found by Schwarzschild [Sc]. On the manifold $(2m,\infty) \times S^2 \times I\!R$ he constructed the Ricci flat Lorentz metric:

$$\frac{dr^2}{1-2m/r} + r^2 \cdot d\sigma^2 - (1-2m/r) \cdot dt^2.$$

This geometry models the empty space outside a static and rotationally symmetric star. We can apply it to describe the gravitational field of the sun and find many correctly predicted measurable deviations from the Newtonian theory; the most famous effects are the correct bending of light and the perihelion advance of Mercury. (The treatment in [Be] does not use almost Newtonian coordinates, it relies on Jacobi fields.) We can also apply this geometry to model the vicinity of the earth, a neccessity if one wants to operate global arrays of radio telescopes. – In the region where $r/(2m) >> 1$ one may compare the motion of test particles in this geometry with Kepler orbits around a star. These predictions are, for increasing r, asymptotically the same. Therefore we can use this comparison to interpret m as the mass of the star and (always for $r/(2m) >> 1$!) r as the Newtonian distance from the star. As r decreases the deviations from the Newtonian predictions increase. On the submanifold $r = 3m$ we have photons, or null geodesics, circling the sun, an extreme case of bending of light. The Schwarzschild geometry has time translations as isometries; let X be the corresponding Killing field. The timelike unit vector field $u = X/\sqrt{1-2m/r}$ we will call the Killing observer. Schwarzschild's coordinates are adapted to this Killing observer and become singular at $r = 2m$. I like the treatment in [St] which explains how adaption to a different family of observers leads to coordinates which are valid beyond the so called horizon at $r = 2m$. – On the earth, for example, u represents observers who stand on the surface of the planet, or in some high tower. Consider a light signal that goes from the bottom to the top of this tower. What is the observed redshift? Since X is a Killing field, we have $g(X(c(s)), c'(s)) = const$ along the joining null ray and therefore find for the observed red shift

$$\lambda = \frac{g\left(X/\sqrt{1-2m/r_{emitter}}\ ,\ c'\right)}{g\left(X/\sqrt{1-2m/r_{observer}}\ ,\ c'\right)} = \sqrt{\frac{1-2m/r_{observer}}{1-2m/r_{emitter}}} > 1.$$

This redshift measurement in a 40 meter tower became possible by using the Mösbauer effect and in fact was one of its first spectacular applications. The redshift between Killing observers on the photon sphere $r = 3m$ and at infinity is only $\sqrt{3}$; this redshift goes to infinity as r approaches $2m$, an explanation for calling the Schwarzschild geometry a black hole. Note that the above redshift computation also means that we cannot synchronize natural clocks on the earth: Radioactive material decays faster at the top of a tower than at the bottom. I repeat that this is a “geometry of world line phenomenon”: If one starts with two equal chunks of radium, transports one to the top of some tower, leaves it there for a while and then gets it back for comparison with the one which was stored at the bottom, then the chunk from the top contains less radium in spite of the fact that the decay of each defines a natural clock, i.e measures the physically relevant proper time. Or, put in other words: Radio waves from the stars to the earth have a higher observed frequency at the bottom of the tower than at the top. This phenomenon is called gravitational redshift. Obviously it cannot be explained by relative motion. (If we interprete, following Planck, $h \cdot \nu$ as the kinetic energy of a photon and assume that this energy increases as usual as it looses potential energy while falling towards the earth, then we would predict the same redshift as above, but we would not predict that radium decays faster higher up in this potential field.)

Cosmological redshift in Friedmann universes. The standard cosmological models are based on the product manifold $M^+_\kappa \times I\!R_+$ with conformally flat (non product) metrics (where M^+_κ denotes a space of constant curvature κ). All the physics is in the conformal factor, the simpler product geometry has no physical interpretation; nevertheless, such a representation is computationally helpful since redshift computations are trivial for the product geometry. As a preliminary we therefore discuss the change of redshift under a conformal change of the metric.

Let $\tilde g(X,Y) = \mu^2 \cdot g(X,Y)$. We have for the Christoffel symbols

$$\Gamma(X,Y) := \tilde D_X Y - D_X Y = \frac{d_Y \mu}{\mu} \cdot X + \frac{d_X \mu}{\mu} \cdot Y - g(X,Y) \ \text{ grad } \mu.$$

Then we get

$$\frac{D}{ds} c' = 0, g(c',c') = 0 \Longrightarrow \frac{\tilde D}{ds} c' = \Gamma(c',c') = 2 \frac{d_{c'} \mu}{\mu} \cdot c'.$$

So: g-null geodesics remain $\tilde g$-null geodesics, but not with an affine parametrization. We reparametrize

$$\begin{aligned} \tilde c(s) =& c(\psi(s)) \ , \ \tilde c'(s) = c'(\psi(s)) \cdot \psi'(s) \\ \frac{\tilde D}{ds} \tilde c'(s) =& 0 \Leftrightarrow \psi'' + \psi'^2 \cdot \frac{2 d_{c'} \mu}{\mu} = 0 \Leftarrow \psi'(s) = \frac{1}{\mu^2}(c(\psi(s))). \end{aligned}$$

Although the reparametrization is given via the solution of a differential equation we still get an *explicit* formula for the change of the redshift since we only need ψ' as a function of μ. If u, v are the g-observers, then $\tilde u = u/\mu$, $\tilde v = v/\mu$ are the $\tilde g$-observers and $\tilde c' = c' \cdot \psi' = c'/\mu^2$. Thus, if we denote by $\mu_{emitter}$ the function μ evaluated at the source, then

$$\begin{aligned} \tilde\lambda = \frac{\tilde\omega_{emitter}}{\tilde\omega_{observer}} &= \frac{\tilde g(\tilde u, \tilde c')}{\tilde g(\tilde v, \tilde c')} = \frac{g(u,c')/\mu_{emitter}}{g(v,c')/\mu_{observer}} = \frac{\omega_{emitter}/\mu_{emitter}}{\omega_{observer}/\mu_{observer}} \\ &= \frac{\mu_{observer}}{\mu_{emitter}} \cdot \lambda. \end{aligned}$$

For emphasis we repeat: λ is the observed redshift between the observers u, v for the Lorentz metric g, and $\tilde{\lambda}$ is the observed redshift for the conformally changed situation, i.e. $\tilde{g} = \mu^2 \cdot g$, $\tilde{u} = u/\mu$, $\tilde{v} = v/\mu$. Therefore the above computation proves the
Conformal Redshift Transformation:

$$\tilde{\lambda} = \frac{\mu_{observer}}{\mu_{emitter}} \cdot \lambda.$$

We return to the simplest cosmological models, which are given by Lorentz metrics on $M^4 = M^3_\kappa \times I\!R_+$. The physicists prefer since Friedman [Fr] to write the desired metric in the form

$$\overline{g} = a^2(\tau) \cdot g_\kappa - d\tau^2,$$

while I will get slightly more explicit final formulas by working with the conformal change of the product metric [Ka]:

$$\tilde{g} = \mu^2(t) \cdot (g_\kappa - dt^2) \quad \left(\text{with } dt = \frac{d\tau}{a(\tau)} \right).$$

The divergence free Einstein tensor

$$G = Ric - \frac{1}{2}(\text{trace } Ric) \cdot id$$

has the tangent spaces to the factors $M^3_\kappa, I\!R_+$ as eigenspaces. The eigenvalues for the simple metric $g_\kappa(.,.) - dt^2$ are $-\kappa, -\kappa, -\kappa, -3\kappa$. The eigenvalues of the Einstein tensor for the physically relevant metric $\tilde{g}$ are

$$\lambda_M = \mu^{-2} \cdot (\kappa - 2\frac{\mu''}{\mu} + \frac{\mu'^2}{\mu^2}), \quad \lambda_{I\!R} = \mu^{-2} \cdot (-3\kappa - 3\frac{\mu'^2}{\mu^2}).$$

Now we use the additional physical assumption that the cosmological "matter" is a dust (with the galaxies the dust grains). The stress energy tensor T of such a simple matter is diagonal in the rest frame of the dust with eigenvalues $0, 0, 0, \rho$. The Einstein equation $G = 8\pi T$ therefore reduce to the ODE $\lambda_M = 0$ for the conformal factor μ. (Vice versa, if a Lorentz manifold has an Einstein tensor with eigenvalues $0, 0, 0, \rho$, then it is a model of a dust world.) We get an explicit solution in terms of the function s_κ defined by $s''_\kappa + \kappa s_\kappa = 0$, $s_\kappa(0) = 0$, $s'_\kappa(0) = 1$ (note ${s'}^2_\kappa + \kappa \cdot s^2_\kappa = 1$), which, in the case of vanishing cosmological constant, is:

$$\mu(t) = \frac{s^2_\kappa(\frac{t}{2})}{s^2_\kappa(\frac{td}{2})}, \quad (td \text{ 'time today' and curvature } \kappa \text{ are the model parameters}).$$

The world lines of the dust grains are the geodesics $t \to (m, t) \in M_\kappa \times I\!R_+$. The time like unit vector field $(0, 1/\mu(t)) \in T_{(m,t)}(M_\kappa \times I\!R_+)$ represents the observers which are at rest relative to the dust particles. This model is so simple that the orthogonal distribution of the distinguished family of (dust) observers is integrable; therefore one cannot resist to call the integral hypersurfaces $M_\kappa \times \{const\}$ "space" at time $t = const$.

At $t = 0$ this metric becomes singular ("big bang"). If we interpret "today" as $t = td$, then κ is the curvature of the space like hypersurface $t = td$. The main physical property of the dust, its mass density ρ, is, for this family of models, given by

$$\rho(t) = 3 \cdot \frac{s_\kappa^4(\frac{td}{2})}{s_\kappa^6(\frac{t}{2})} = \mu^{-3}(t) \cdot \rho(td) \ , \ \rho(td) = \text{ mass density today.}$$

The simple metric $g_\kappa - dt^2$ has time translation as an isometry, the corresponding Killing field $(0, 1)$ has constant length and therefore there is no frequency shift between these Killing observers ($\lambda = 1$). In the physically relevant metric $\mu^2(t) \cdot (g_\kappa - dt^2)$ we assume that light observed today ($t = td$) is emitted at $t = t$. By our formula for the conformal redshift transformation we obtain:

$$\tilde{\lambda} = \frac{\mu(td)}{\mu(t)} = \left(\frac{\rho(t)}{\rho(td)}\right)^{\frac{1}{3}} .$$

This says that the observed redshift between dust particle observers depends on the ratio of the mass densities at emission and observation. Note that the redshift goes to infinity as $t \to 0$. If we choose to interpret $\rho(t)^{-1/3}$ as a measure of the size of "space" at time t, then the observed redshift is the quotient of the sizes of "space" at observation and at emission, i.e. it measures this kind of "expansion" of "space".

Note that the mass density at a point of a Lorentz manifold is on the one hand a measurable physical quantity and on the other hand computable from the Ricci tensor of the modelling Lorentz manifold, while, in contrast to this, the definition of the "spaces" as $t = const$ for distinguished observers – more precisely as the *integral hypersurfaces of the rest spaces* of some distinguished vectorfield of observers – depends on the simplicity of the model. The orthogonal complements of families of observers are in general *not* integrable distributions.

References

[Be] Besse, A. L.: Einstein Manifolds, Springer-Verlag Berlin, Heidelberg 1987.

[Fr] Friedmann, A.: Über die Krümmung des Raumes. Z. Phys. 10, 377-386 (1922)

[Ka] Karcher, H.: Infinitesimale Charakterisierung von Friedmann-Universen. Archiv d. Math. 38 (1982), 58-64.

[Sc] Schwarzschild, K.: Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. Sitzber. Deut. Akad. Wiss. Berlin, Kl. Math.-Phys. Tech., 189-196 (1916).

[St] Stephani, H.: Allgemeine Relativitätstheorie, VEB Deutscher Verlag der Wissenschaften, Berlin 1980.

Hermann Karcher UNM416@uni-bonn.de
Mathematisches Institut, Beringstr. 1
D-53115 Bonn