M1 Cosmology - 3 - Relativistic Models

Nick Kaiser

Contents

1 Hubble's discovery of the expanding universe

- Hubble observed Cepheid variable stars in "spiral nebulae" (late 20's)
  - period-luminosity relation: period $\rightarrow L \rightarrow$ distance D
- large distances firmly established them to be well outside of the MW
  - external 'galaxies'
- he combined these with Slipher's spectra $\rightarrow$ redshift $\rightarrow$ "recession velocity"
- he found that *the local universe is expanding*
  - recession velocity roughly proportional to distance
  - Lemaitre had earlier shown the same thing (published in French)
- Later studies extended the 'Hubble diagram' (magnitude vs log-z) to larger distances and improved precision and accuracy
  - $cz = H_0 r$ with (current) *expansion rate* $H_0 \simeq 70\text{km/s/Mpc} \simeq (1.4 \times 10^{10}\text{yr})^{-1}$
  - departures on small-scales – growth of structure $\rightarrow$ 'peculiar' velocities
  - and departures from linearity on very large scales $D \sim c/H$

Hubble's observations – apparent distances as a function of recession velocity – are (and were at the time; 1929) interpreted in terms of *homogeneous and isotropic world-models* (Friedmann, 1922) based on Einstein's then newly created *general theory of relativity* (GR, 1915). Let's now introduce the salient features of GR:

2 The essential features of Einstein's theory of gravity

2.1 The Galilean equivalence principle

In GR there is no 'force' of gravity; the fact that all objects fall the same way under the influence of gravity (the Galilean principle of equivalence) allowed Einstein to propose that gravity is curvature of space-time and that particles follow straight lines – or 'geodesics' – in space-time.

- *The curvature of space-time tells matter how to move.* (John Wheeler)

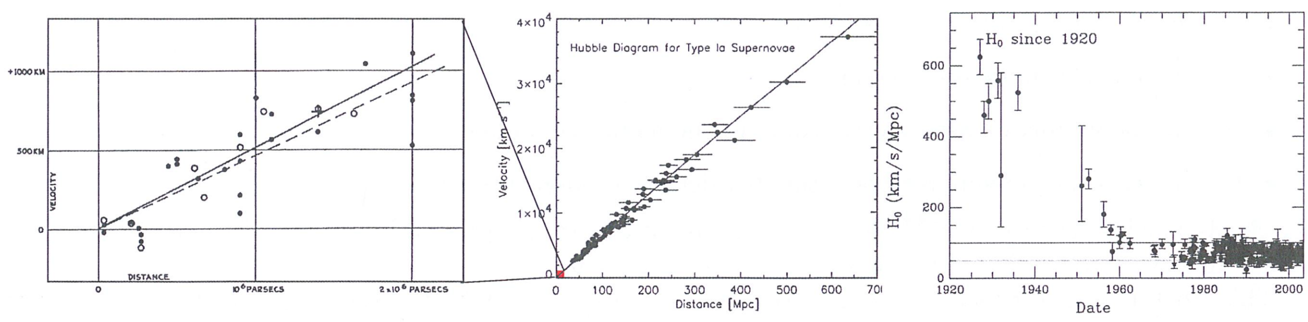

Figure 1: The Hubble diagram over the ages. On the left is Hubble's original plot from 1929. In the centre is a modern determination using supernovae. The red box at the lower left indicates the range of data that Hubble used. On the right is shown a compilation by John Huchra of historical estimates of the present-day expansion rate H_0. A common feature of such measurements is an underestimate of systematic errors that plague measurements of cosmological distances.

2.2 Space-time is everywhere locally the same as in special relativity

Just as the surface of a smooth object like an apple is locally flat and indistinguishable from Euclidean space, the so-called *'manifold'* of space-time is assumed to be everywhere locally indistinguishable from Minkowskian space.

Now, living on Earth, we don't see that; we see the effect of gravity everywhere. But to see the true, locally Minkowskian, nature of the manifold of space-time, one need only jump out of the window of a tall building. Until you hit the ground, there is no gravity. According to Einstein, going into free-fall locally *'nulls-out'* the effects of gravity.

That means that there is a *light-cone structure* built into the manifold of space-time. It is an *absolute property of space-time*. The light cones allow one to categorise 4-vectors – the separation between neighbouring two events being the prototypical vector – into 3 classes:

1. **time-like:** a pair of events that a massive particle – or *'observer'* – can travel between

2. **space-like:** a pair of events that no observer can travel between

3. **null:** a pair of events that can be linked by a photon – whose path lies in the light cone

and all freely falling observers, no matter how they are moving, agree on this categorisation.

Another absolute property of space-time is that, while there are a family of possible freely falling observers at any point in space-time – depending on how they are moving and how they are oriented – *they can sense if they are rotating*. There is, again locally, a *non-rotating frame* that is somehow intrinsic to the manifold.

In special relativity, all physical laws are expressed in terms of 4-vectors and tensors. An example is given by Maxwell's equations, which can be expressed as ${F^\nu}_{\mu,\nu} = \mu_0^{-1} j_\mu$ where ${F^\nu}_\mu$ is the *Faraday 4-tensor* containing the components of the **E** and **B** and where $\vec{j} \longrightarrow j^\mu$ is the 4-current, and this obeys the continuity equation $\nabla \vec{j} \longrightarrow {j^\mu}_{,\mu} = 0$. Einstein proposed that all such laws remain valid (locally, and in terms of quantities measured by freely falling observers) even in the presence of gravitating matter.

2.3 Matter controls the curvature of space-time via its stress-energy tensor

In Newton's theory, the gravitational potential ϕ is generated by the density of mass ρ via *Poisson's equation:*

$$\boxed{\nabla^2 \phi = 4\pi G \rho} \tag{1}$$

and the effect of the tidal field – second spatial derivative of the potential – on separations of freely falling particles is

$$\boxed{\Delta \ddot{x}_i = -\frac{\partial^2 \phi}{\partial x_i \partial x_j} \Delta x_j.} \tag{2}$$

Einstein argued that the generalisation of this is the *geodesic deviation equation*, in which the tidal field is replaced by the curvature tensor.

In special relativity there is a rank-two symmetric tensor **T**, whose 10 components $T^{\mu\nu}$ are

- T^{00} the energy density (equal to ρc^2)
- T^{0i} the momentum density vector
- T^{i0} the energy flux density vector (equal to the momentum density)
- T^{ij} the stress or pressure 3-tensor (the flux density of momentum)

For slowly moving masses only T^{00} is important (the others are suppressed by powers of v/c or v^2/c^2) and – aside from the factor c^2 – is the same as the mass density ρ.

This led Einstein to propose that, in a relativistic theory of gravity, the source of gravity must be the stress-energy tensor. So he set about determining an analogue of Poisson's equation (1) with, on the right hand side, $\mathbf{T}$.

And, just as $\nabla^2 \phi = \partial^2 \phi / \partial x_i \partial x_i$ is a contraction (the trace) of the tidal field 3-tensor, he reasoned that the left-hand side must contain a rank-two tensor constructed from the geometric curvature tensor.

In determining this, he was guided by the fact that $\mathbf{T}$ obeys the *continuity equation* $\nabla \mathbf{T} = 0$. He found that there is an essentially unique rank-2 tensor $\mathbf{G}$ determined from the curvature of space-time that obeys the same law. We call $\mathbf{G}$ the *Einstein tensor* in his honour, and it is this that is driven (or 'sourced') by the stress-energy tensor in *Einstein's equations:*

$$\boxed{\mathbf{G} = 8\pi\kappa\mathbf{T}} \tag{3}$$

This equation expressed the way that, in the second part of John Wheeler's beautiful aphorism, *"matter tells space-time how to curve"*. It contains a single free parameter κ, and correspondence with Newtonian theory requires $\kappa = G/c^4$.

Concerns about the implications of (3) for cosmology led Einstein to propose to add an extra term $\Lambda\mathbf{g}$ to the left-hand side, where Λ is a constant, known as the *cosmological constant.*

3 Brief review of the machinery of general relativity

Special relativity is all about how the world appears to observers in relative motion with respect to each other; hence the word 'relativity'. General relativity is sometimes said to be founded on Einstein's equivalence principle, stated in the form that *gravity and acceleration are equivalent.* We saw in the last lecture how the metric of space-time with coordinates tied to rulers and clocks in an accelerated rocket appears 'warped'; clocks run faster in the nose of a rocket as compared to the tail.

That, together with the EEP as stated above, might lead one to think that this warping is a manifestation of the gravitational field. But that would be entirely misleading; the gravitational field in GR is the *curvature*. And in an accelerating rocket the curvature vanishes. There is no gravitational field. Similarly, most phenomena we observe sitting on the Earth – balls falling, photons being redshifted in the Pound & Rebka (1959) experiment etc. – are not directly measuring the gravitational field *qua* curvature; that appears only when one looks at the *relative acceleration* of objects at widely different places in the field.

Motions of particles and the form of e.g. the equations of electromagnetism in an accelerated frame are an application of *generalised covariance*. The manner in which such equations can be expressed in terms of arbitrary coordinates was a major part of Einstein's massive contribution to science in 1915. But it still 'just' special relativity, and should not be confused with gravitation.

As argued by Synge in his textbook, the very terminology general *relativity* is rather bizarre, if not repellant. There is really nothing 'relative' about the gravitational field. If the 256 components of the curvature tensor vanish – as they do for the metric we developed to describe physics in an accelerating frame – then there is no gravitational field.[1] Whereas if there is curvature, then it is, as a geometric entity, absolute (though the *components* of the curvature tensor depend on the relative frame of motion or orientation of the measurer).

Misner, Thorne and Wheeler, in their textbook, argue that one should draw a clear distinction between the fact that special relativistic laws of physics can be expressed in an arbitrary coordinate system and the effects of gravitational curvature – including the fact that there is no *a priori geometry* in GR. We will follow their advice here, and first, in §3.1, develop the mathematics of SR – i.e. flat space-time – in arbitrary coordinates and then, in §3.2, generalise this to curved space-time.

[1] Wolfgang Rindler takes a very different view from Synge on this. Rindler believes that the phenomena in an accelerating frame can be thought of as gravitational field arising from the rest of the matter in the Universe – which, from the perspective of the rocketeer, is suffering a 'reflex' acceleration.

MTW: 1) Generalized covariance $F^{\mu\nu}{}_{,\mu} = \frac{1}{\mu_0} j^{\nu}$ $x^{\alpha}(x')$
AE 1915
2) Curvature = gravity, curved manifold, no a priority geometry.

3.1 Generalised covariance (MTW)

The mathematics of tensor (and vector and 1-form) calculus in curvilinear coordinates is well demonstrated by the case of 2-dimensional Euclidean space. The extra complexity of the fact that space-time is 4-dimensional and is non-Euclidean – the metric being reducible to diagonal form with one negative eigenvalue – does not greatly change things.

3.1.1 Curvilinear coordinates in 2-dimensions

This is illustrated in figure 2 which shows the 2D Euclidean plane with usual rectilinear coordinates $\{x, y\}$ – here these will, mostly, be denoted as primed coordinates – and overlaid contours of curvilinear coordinates $\xi(x, y)$ and $\eta(x, y)$.

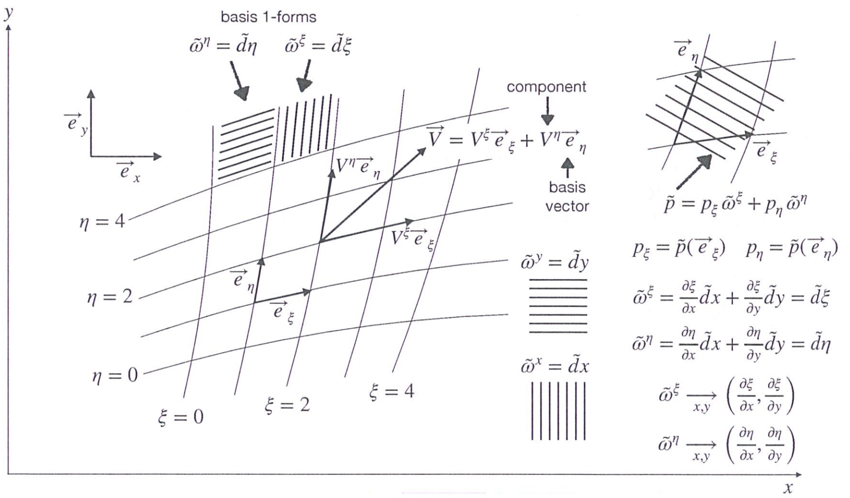

Figure 2: The long curved lines show surfaces of constant curvilinear coordinates $\xi(x, y)$ and $\eta(x, y)$, which are functions of the rectilinear coordinates $\{x, y\}$. The vectors $\vec{e}_{\xi}$ and $\vec{e}_{\eta}$ are so-called 'coordinate basis vectors'; $\vec{e}_{\xi}$ points along the η = constant surface from one iso-ξ surface to the next. As shown, an arbitary vector $\vec{V}$ can be expressed as a sum of basis-vectors times components. Also indicated are the basis 1-forms $\tilde{\omega}^{\xi}$ and $\tilde{\omega}^{\eta}$.

3.1.2 Basis vectors and 1-forms

A vector $\vec{V}$ can be expressed as

$$\vec{V} = V^{\alpha}\vec{e}_{\alpha} \tag{4}$$

where $\{x^{\alpha}\} = (\xi, \eta)$ and where, as usual, we are using the summation convention. Note that the subscript α here is not an *index*; it is a *label* indicating which basis vector we are referring to. Each of the basis vectors here has two components, which would be indicated by a superscript.

A 1-form $\tilde{p}$, perhaps $\tilde{p} = \tilde{d}\phi \longrightarrow \partial_{\alpha}\phi$, can similarly be expressed as

$$\tilde{p} = p_{\alpha}\tilde{\omega}^{\alpha} \tag{5}$$

where $\{\tilde{\omega}^{\alpha}\}$ are a set of basis 1-forms.

The basis vectors are not generally orthogonal to one another, nor are the basis 1-forms, but the former are orthogonal to the latter in the sense that the number of iso-x^{α} surfaces pierced by $\vec{e}_{\beta}$ vanishes if $\alpha \neq \beta$, and is otherwise unity:

$$\tilde{\omega}^{\alpha}(\vec{e}_{\beta}) = \delta^{\alpha}_{\beta}. \tag{6}$$

This means that we can extract the α^{th} component of a vector $\vec{V}$ by letting it act on $\tilde{\omega}^{\alpha}$, since

$$\vec{V}(\tilde{\omega}^{\alpha}) = V^{\beta}\vec{e}_{\beta}(\tilde{\omega}^{\alpha}) = V^{\beta}\delta^{\alpha}_{\beta} = V^{\alpha}. \tag{7}$$

3.1.3 Bases for tensors

We can use outer products of basis vectors and 1-forms to generate bases for tensors of arbitrary kind. For example, the stress-energy tensor

$$\mathbf{T} = \int \frac{d^3p}{p^0} f(\mathbf{p})\vec{p} \otimes \vec{p} \tag{8}$$

can be written as

$$\mathbf{T} = T^{\alpha\beta}\vec{e}_{\alpha} \otimes \vec{e}_{\beta} \tag{9}$$

where the $T^{\alpha\beta}$ are the contravariant components of this particular tensor and the $\vec{e}_\alpha \otimes \vec{e}_\beta$ form a basis for rank-2 tensors in general. The outer product $\vec{e}_\alpha \otimes \vec{e}_\beta$ here is that thing which, acting on an ordered pair of basis 1-forms $\tilde{\omega}^\mu, \tilde{\omega}^\nu$, gives

$$(\vec{e}_\alpha \otimes \vec{e}_\beta)(\tilde{\omega}^\mu, \tilde{\omega}^\nu) = \vec{e}_\alpha(\tilde{\omega}^\mu)\vec{e}_\beta(\tilde{\omega}^\nu) = \delta^\mu_\alpha \delta^\nu_\beta \tag{10}$$

so we can extract the $(\mu, \nu)^{\text{th}}$ component of $\mathbf{T}$ by letting it act on the two 1-form bases: $T^{\mu\nu} = \mathbf{T}(\tilde{\omega}^\mu, \tilde{\omega}^\nu)$.

3.1.4 Transformation of the bases

The rectilinear (primed) frame components of the curvilinear (un-primed) basis vectors $\vec{e}_\alpha$ and 1-forms $\tilde{\omega}^\alpha$ are seen from figure 2 to be

$$\begin{aligned} (\vec{e}_\alpha)^{\beta'} &= \Lambda^{\beta'}{}_\alpha \equiv \frac{\partial x^{\beta'}}{\partial x_\alpha} \\ (\tilde{\omega}^\alpha)_{\beta'} &= \Lambda^\alpha{}_{\beta'} \equiv \frac{\partial x^\alpha}{\partial x_{\beta'}} \end{aligned} \tag{11}$$

where we have defined the transformation matrices for transforming between the two 'frames'. These are analogous to the Lorentz transformation matrices for boosts and rotations.

It follows that the curvilinear bases can be expressed as linear combinations of the rectilinear ones as

$$\begin{aligned} \vec{e}_\alpha &= (\vec{e}_\alpha)^{\beta'} \vec{e}_{\beta'} = \Lambda^{\beta'}{}_\alpha \vec{e}_{\beta'} \\ \tilde{\omega}^\alpha &= (\tilde{\omega}^\alpha)_{\beta'} \tilde{\omega}^{\beta'} = \Lambda^\alpha{}_{\beta'} \tilde{\omega}^{\beta'} \end{aligned} \tag{12}$$

while the inverse transformation is

$$\begin{aligned} \vec{e}_{\beta'} &= \Lambda^\alpha{}_{\beta'} \vec{e}_\alpha \\ \tilde{\omega}^{\beta'} &= \Lambda^{\beta'}{}_\alpha \tilde{\omega}^\alpha \end{aligned} \tag{13}$$

since the matrix inverse of e.g. $\Lambda^{\beta'}{}_\alpha \equiv \partial x^{\beta'}/\partial x^\alpha$ is $\Lambda^\alpha{}_{\beta'} \equiv \partial x^\alpha/\partial x^{\beta'}$.

3.1.5 The derivatives of the basis vectors – the connection

The rectilinear-frame basis vectors $\vec{e}_{\alpha'}$ are independent of position. It is for that reason that we did not use them explicitly previously. The variation of their curvilinear counterparts $\vec{e}_\alpha$, however, play an essential role when we work with derivatives of vectors (or other geometric entities) since these are (sums of) products of components and bases.

Since $\vec{e}_{\beta'}$ is independent of position, it follows from the first of (12) that

$$\vec{e}_{\alpha,\beta} = \frac{\partial \vec{e}_\alpha}{\partial x^\beta} = \Lambda^{\beta'}{}_{\alpha,\beta} \vec{e}_{\beta'} \tag{14}$$

and, using the first of equations (13) we have

$$\begin{aligned} \vec{e}_{\alpha,\beta} &= \Lambda^\gamma{}_{\beta'} \Lambda^{\beta'}{}_{\alpha,\beta} \vec{e}_\gamma \\ &= \frac{\partial x^\gamma}{\partial x^{\beta'}} \frac{\partial^2 x^{\beta'}}{\partial x^\alpha \partial x^\beta} \vec{e}_\gamma. \end{aligned} \tag{15}$$

Thus the derivative of a basis vector can be expressed as a linear combination of the basis vectors:

$$\boxed{\vec{e}_{\alpha,\beta} = \Gamma^\gamma{}_{\alpha\beta} \vec{e}_\gamma} \tag{16}$$

with coefficients, known as the *connection* or the *Christoffel symbols*,

$$\boxed{\Gamma^\gamma{}_{\alpha\beta} \equiv \Lambda^\gamma{}_{\beta'} \Lambda^{\beta'}{}_{\alpha,\beta} = \frac{\partial x^\gamma}{\partial x^{\beta'}} \frac{\partial^2 x^{\beta'}}{\partial x^\alpha \partial x^\beta}.} \tag{17}$$

One should note that, despite appearances, these are *not* the components of a tensor; they don't transform in the correct manner. It follows from the latter form above that the connection is invariant under exchange of the lower two indices.

3.1.6 The covariant derivative of a vector field

The derivative of a vector field $\vec{V}(\vec{x}) = V^\alpha(\vec{x})\vec{e}_\alpha(\vec{x})$ with respect to the β^{th} spatial coordinate is

$$\begin{aligned}\partial_\beta \vec{V} = \partial_\beta(V^\alpha \vec{e}_\alpha) &= V^\alpha{}_{,\beta}\vec{e}_\alpha + V^\alpha \vec{e}_{\alpha,\beta} \\ &= V^\alpha{}_{,\beta}\vec{e}_\alpha + V^\alpha \Gamma^\gamma{}_{\alpha\beta}\vec{e}_\gamma \\ &= (V^\gamma{}_{,\beta} + \Gamma^\gamma{}_{\alpha\beta}V^\alpha)\vec{e}_\gamma.\end{aligned} \tag{18}$$

If we contract this with the components of a displacement $\vec{dx}$ we get the change in the vector $d\vec{V} = dx^\beta \partial_\beta \vec{V}$. It follows that $V^\gamma{}_{,\beta} + \Gamma^\gamma{}_{\alpha\beta}V^\alpha$ are the components of a (mixed) rank-2 tensor which we can write symbolically as

$$\nabla \vec{V} = \underbrace{(V^\gamma{}_{,\beta} + \Gamma^\gamma{}_{\alpha\beta}V^\alpha)}_{V^\gamma{}_{;\beta}} \tilde{\omega}^\beta \otimes \vec{e}_\gamma. \tag{19}$$

with the understanding that it is the basis 1-form that gets fed the displacement $\vec{dx}$. If we feed $\nabla\vec{V}$ the 4-velocity of a particle $\vec{U} = d\vec{x}/d\tau$ we get the rate of change with respect to proper time of the vector field along the world-line of the particle $d\vec{V}/d\tau$. This is sometimes written as $d\vec{V}/d\tau = \nabla_{\vec{U}}\vec{V}$. Similarly for an arbitrarily parameterised path $\vec{x}(\lambda)$, if we feed $\nabla\vec{V}$ the tangent vector $d\vec{x}/d\lambda$, we get $d\vec{V}/d\lambda$.

The key take-away from the above is that (again despite appearances) $V^\gamma{}_{,\beta}$ are not the components of a mixed rank-2 tensor (something that we can dot with a displacement to get the change in the vector), but those of

$$\boxed{\nabla\vec{V} \longrightarrow V^\gamma{}_{;\beta} \equiv V^\gamma{}_{,\beta} + \Gamma^\gamma{}_{\alpha\beta}V^\alpha} \tag{20}$$

are. We call $\nabla\vec{V}$ the *covariant derivative* of $\vec{V}$.

3.1.7 Derivatives of 1-forms and tensors

One can do the same thing with 1-forms, scalars, tensors etc.. The derivative of a scalar is straightforward, No Christoffel symbols are involved; for a scalar field $\phi(\vec{x})$, $\phi_{;\alpha} = \phi_{,\alpha}$, just the ordinary partial derivative.

This gives a quick way to figure out the components $p_{\alpha;\beta}$ of the covariant derivative $\nabla\tilde{p}$ of a 1-form $\tilde{p}$. Both ordinary (partial) and covariant derivatives obey the usual rule when applied to a product, so, taking the ordinary derivative of the scalar $p_\mu V^\mu$ we have

$$(p_\mu V^\mu)_{,\nu} = p_\mu V^\mu{}_{,\nu} + p_{\mu,\nu}V^\mu \tag{21}$$

but this is the same as

$$\begin{aligned}(p_\mu V^\mu)_{;\nu} &= \overbrace{p_\mu V^\mu{}_{;\nu}} + p_{\mu;\nu}V^\mu \\ &= p_\mu(V^\mu{}_{,\nu} + \Gamma^\mu{}_{\alpha\nu}V^\alpha) + p_{\mu;\nu}V^\mu\end{aligned} \tag{22}$$

and subtracting (21) from (22) gives

$$(p_{\mu;\nu} - p_{\mu,\nu})V^\mu = -p_\mu \Gamma^\mu{}_{\alpha\nu}V^\alpha = -p_\alpha \Gamma^\alpha{}_{\mu\nu}V^\mu \tag{23}$$

where, in the last step, we have swapped the dummy indices $\mu \Leftrightarrow \alpha$. For this to be true for any $\vec{V}$ implies

$$\boxed{\nabla\tilde{p} \longrightarrow p_{\mu;\nu} = p_{\mu,\nu} - \Gamma^\alpha{}_{\mu\nu}p_\alpha.} \tag{24}$$

The same argument can be generalised to give derivatives of tensors of any rank; the covariant derivative is the ordinary partial derivative plus a series of terms, one per index, involving a Christoffel symbol times the tensor in question. These enter with a sign that is positive (negative) for upstairs (downstairs) indices. Thus, for example,

$$T^\alpha{}_{\beta;\nu} = T^\alpha{}_{\beta,\nu} + \Gamma^\alpha{}_{\gamma\nu}T^\gamma{}_\beta - \Gamma^\gamma{}_{\beta\nu}T^\alpha{}_\gamma \tag{25}$$

in which, when we deal with the upstairs index (in the second term) we contract with a downstairs index in the Christoffel symbol, and when we deal with the downstairs index (in the last term) we contract with the upstairs index in $\Gamma^\alpha{}_{\mu\nu}$. As always, the ordering of downstairs indices in Christoffel symbols is arbitrary.

$\nabla P = -a[\rho + P/c^2]$ Why is the enthalpy there.

3.1.8 Determining the connection from the metric

Equation (17) above gives the connection in terms of the transformation matrix $\Lambda^{\alpha'}{}_{\beta}$ and its partial derivative. The same transformation matrix determines the components of the curvilinear-frame metric:

$$g_{\alpha\beta} = \Lambda^{\alpha'}{}_{\alpha}\Lambda^{\beta'}{}_{\beta} g_{\alpha'\beta'} \tag{26}$$

which follows from invariance of $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$, and where, in Minkowski space, the rectilinear-frame metric is $g_{\alpha'\beta'} = \eta_{\alpha'\beta'}$.

It proves to be very useful to have an explicit expression for the connection in terms of the components of the metric (and their derivatives). This can be obtained using the fact that the metric $\mathbf{g}$ is independent of position: $\nabla \mathbf{g} = 0$. Thus

$$g_{\alpha\beta;\nu} = g_{\alpha\beta,\nu} - \Gamma^{\gamma}{}_{\alpha\nu} g_{\gamma\beta} - \Gamma^{\gamma}{}_{\beta\nu} g_{\alpha\gamma} = 0, \tag{27}$$

which tells us how the curvilinear-frame components of the metric need to vary in order that $\mathbf{g} = g_{\alpha\beta}\,\tilde{\omega}^\alpha \otimes \tilde{\omega}^\beta$ remains constant.

The trick needed in order to obtain an explicit expression for the connection from this is to write down three equivalent forms of this by permuting the indices (i.e. letting the third index be ν, α and β in turn) and take the sum of the first two minus the third. This gives $2\Gamma^{\alpha}{}_{\mu\nu} g_{\alpha\beta} = g_{\beta\mu,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta}$ from which we find the explicit expression

$$\boxed{\Gamma^{\alpha}{}_{\mu\nu} = \frac{1}{2} \underbrace{g^{\alpha\beta}}_{\text{inverse of } g_{\alpha\beta}} (g_{\beta\mu,\nu} + g_{\beta\nu,\mu} - g_{\mu\nu,\beta}).} \tag{28}$$

This should be familiar; it is what appeared in the geodesic equation obtained in the last chapter, and which, in terms of the Christoffel symbols, is

$$\frac{dp^\gamma}{d\lambda} + \Gamma^{\gamma}{}_{\mu\nu} p^\mu p^\nu = 0 \tag{29}$$

This is no coincidence; recalling that $\vec{p} = d\vec{x}/d\lambda$ (where $d\lambda = d\tau/m$ or $\lim_{m\to 0} d\tau/m$ for a massless particle) and with $\vec{p} = p^\gamma \vec{e}_\gamma$ we have

$$\begin{aligned} \frac{d\vec{p}}{d\lambda} &= \frac{dp^\gamma}{d\lambda}\vec{e}_\gamma + p^\gamma \frac{d\vec{e}_\gamma}{d\lambda} = \frac{dp^\gamma}{d\lambda}\vec{e}_\gamma + p^\gamma \frac{dx^\mu}{d\lambda}\frac{\partial \vec{e}_\gamma}{\partial x^\mu} \\ &= \frac{dp^\gamma}{d\lambda}\vec{e}_\gamma + p^\gamma p^\mu \Gamma^{\nu}{}_{\gamma\mu}\vec{e}_\nu = \left(\frac{dp^\gamma}{d\lambda} + \Gamma^{\gamma}{}_{\mu\nu} p^\mu p^\nu\right)\vec{e}_\gamma \end{aligned} \tag{30}$$

where, in the last step, we swapped $\gamma \Leftrightarrow \nu$. So (29) is simply telling us that, for paths that extremise τ, the components p^γ vary along the path in such a manner that $\vec{p} = p^\gamma \vec{e}_\gamma$ remains constant.

3.2 Differential geometry on a curved manifold

A section of a curved 2-dimensional manifold – it might be the surface of an apple, and it is assumed to be smooth – is shown in figure 3. We can imagine blind ants living on this surface who are equipped with tape measures with which they can measure distances – along extremal lines in the surface – between points on the surface.

It is obvious that, with a little ingenuity, the ants can determine that they are on curved surface. They could, for instance, lay out a set of points that are equidistant from some central point, and then measure the circumference and compare it with the radius. For the surface shown in the figure, they would find that the circumference is generally smaller than 2π times the radius, and they would find that the fractional deficit grows quadratically with radius. But, if they were on a saddle-shaped surface, they would find that there is a quadratically growing fractional excess to the circumference. Writing the circumference as $l = 2\pi r(1 - r^2/R^2 + \ldots)$ allows them to determine a local squared radius of curvature R^2, which would be negative on a saddle.

What else can they determine about the geometry of the surface? Clearly, if they were not blind and could measure distances off the surface they would be able to measure the deviation of the surface from a flat reference surface – a so-called *'tangent plane'*. And this would be described by 3 numbers; which one could take to be the squared curvature radii for the principle axes and an angle giving their orientation. But that would be a measurement of the *extrinsic curvature*. Armed only with measurements in the manifold, what can they learn? If they are on an egg, for instance, can they tell, from local measurements, that they are not on a sphere?

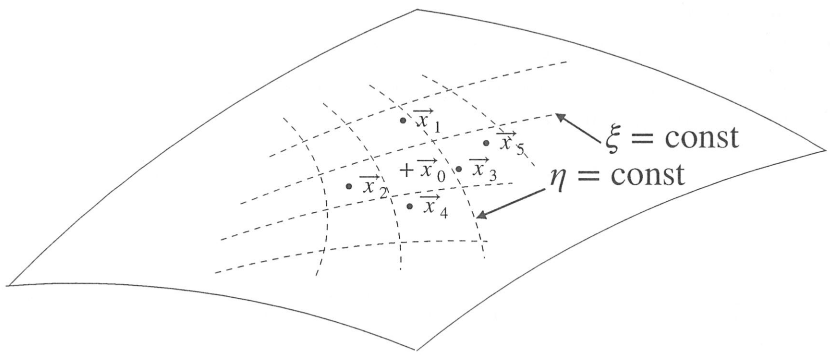

Figure 3: A portion of a 2-dimensional manifold on which lines of (unprimed) coordinates $x^\alpha = (\xi, \eta)$ are shown along with a set of points (events in 4D). Distances – in the manifold itself, not taking a 'short-cut' through the fictitious embedding space – can be measured using pieces of string or rulers. This allows the determination of the metric; which we can think of as a distillation of such distance measurements.

3.2.1 The metric

On the section of the 2D manifold shown in figure 3 someone has drawn lines of constant coordinates ξ, η. These are also assumed to be smooth and to foliate the surface. Scattered within some relatively limits regions are glued a set of little pucks with labels (an index i) and with their (ξ, η) coordinates written on them in braille (so our ants can read off their coordinates).

Our ants can measure physical distances (in the surface) dl_{ij} between pairs of points and can also read-off the coordinate displacements $(d\xi, d\eta)_{ij}$. If they square the former, they find, empirically, that the results vary, in the limit of small separations, bi-linearly with the displacements in $x^\alpha = (\xi, \eta)$ coordinates:

$$dl^2 = g_{\alpha\beta} dx^\alpha dx^\beta \tag{31}$$

where we use, as always, the summation convention.

This furnishes the ants with a 2×2 symmetric matrix at some point $\vec{x}$. They can repeat this exercise at different positions on the manifold. In this way, they can determine the components of the metric $g_{\alpha\beta}(\vec{x})$ as a function of position $\vec{x}$. And from this, they can obtain the partial derivatives $g_{\alpha\beta,\gamma}$ that appear in the formula above for the connection. They will find, in general that these are non-vanishing.

It is important to realise that the metric is something *measurable*. It is a distillation of measurements of physical distances. As we will elaborate, it allows one to determine the curvature of the manifold. The components of the metric depend also on the arbitrary coordinate system that was imprinted on the manifold.

3.2.2 Local flatness theorem and locally inertial coordinates

At any point on the surface, the ants can erect a locally Euclidean coordinate system. They can, for example, stretch out a measuring tape across the surface to define the x-axis, and mark off intervals of constant physical distance. Taking a pair of points on the x-axis they can find a set of points that are equidistant, which they can take to be the y-axis and which is orthogonal to the x-axis. Mathematically, this is described by saying that the displacements $dx^{\alpha'} = (dx, dy)$ are a linear transformation of the displacements $dx^\alpha = (d\xi, d\eta)$:

$$dx^{\alpha'} = \Lambda^{\alpha'}{}_{\alpha} dx^\alpha. \tag{32}$$

With $dx^\alpha = \Lambda^\alpha{}_{\alpha'} dx^{\alpha'}$, where $\Lambda^\alpha{}_{\alpha'}$ is the inverse of $\Lambda^{\alpha'}{}_{\alpha}$, the line element is

$$dl^2 = g_{\alpha\beta} dx^\alpha dx^\beta = \underbrace{\Lambda^\alpha{}_{\alpha'} \Lambda^\beta{}_{\beta'} g_{\alpha\beta}}_{g_{\alpha'\beta'}} dx^{\alpha'} dx^{\beta'} \tag{33}$$

where, by construction, $g_{\alpha'\beta'} = \text{diag}\{1, 1\}$.

There is clearly some freedom in how they can do this (the direction of the x-axis). Mathematically, this reflects the fact that the metric $g_{\alpha\beta}$, being symmetric, has only three degrees of freedom, while the transformation matrix has four.

This is at a point $\vec{x}_0$, which we can take to be the origin. The *local flatness theorem* states that, given some manifold with a set of coordinates defined on it and with a metric g distilled from measurements of distances as described above, one can always find a coordinate system with which, at any given point, the metric is as above – i.e. locally Euclidean in the 2-D case illustrated above – and, moreover, that *deviations appear only at 2nd order in distance from the point.*

The details are given in appendix B, where it is shown that while (32) derives from the linearly transformed coordinate system $x^{\alpha'}(\vec{x}) = \Lambda^{\alpha'}{}_{\alpha} x^{\alpha}$, if we make a 2nd order Taylor series expansion $x^{\alpha'}(\vec{x}) = \Lambda^{\alpha'}{}_{\alpha} x^{\alpha} + \frac{1}{2}\Lambda^{\alpha'}{}_{\alpha\beta} x^{\alpha} x^{\beta}$ (so $\Lambda^{\alpha'}{}_{\alpha\beta} = \partial^2 x^{\alpha'}/\partial x^{\alpha}\partial x^{\beta}$ at $\vec{x}_0$) for which

$$dx^{\alpha'} = (\Lambda^{\alpha'}{}_{\alpha} + \Lambda^{\alpha'}{}_{\alpha\beta} x^{\beta}) dx^{\alpha} \tag{34}$$

then we find

$$ds^2 = \Lambda^{\mu}{}_{\mu'} \Lambda^{\nu}{}_{\nu'} dx^{\mu'} dx^{\nu'} \left[g_{\mu\nu} + x^{\gamma} \left\{ g_{\mu\nu,\gamma} - g_{\beta\nu} \Lambda^{\beta}{}_{\alpha'} \Lambda^{\alpha'}{}_{\mu\gamma} - g_{\mu\beta} \Lambda^{\beta}{}_{\alpha'} \Lambda^{\alpha'}{}_{\nu\gamma} \right\} + \ldots \right]. \tag{35}$$

where $g_{\mu\nu}$ and $g_{\mu\nu,\gamma}$ are the components of the metric and their first derivatives at $\vec{x}_0$ and $\ldots$ denotes terms of 2nd or higher order in displacement from $\vec{x}_0$.

Thus if, given the curvilinear frame metric components $g_{\mu\nu}$ and their derivatives $g_{\mu\nu,\delta}$ at $\vec{x}_0$, we can find a set of transformation coefficients $\Lambda^{\alpha'}{}_{\mu\nu}$ that make the quantity in parentheses $\{\ldots\}$ vanish then we will not only have $g_{\mu'\nu'} = \delta_{\mu'\nu'}$ at $\vec{x}_0$ but its derivatives $g_{\mu'\nu',\gamma'}$ will vanish there also.

But we know how to solve $\{\ldots\} = 0$; we simply need to choose $\Lambda^{\alpha'}{}_{\nu\gamma}$ so that $\Lambda^{\beta}{}_{\alpha'} \Lambda^{\alpha'}{}_{\nu\gamma} = \Gamma^{\beta}{}_{\nu\gamma}$ where the Christoffel symbols are calculated from $g_{\mu\nu}$ and $g_{\mu\nu,\delta}$ using (28). It doesn't matter that (28) was obtained in the context of a flat space-time whereas now we are working on a curved manifold; the algebra is valid regardless.

Thus the required 2nd derivatives in the transformation above are

$$\frac{\partial^2 x^{\alpha'}}{\partial x^{\alpha} \partial x^{\beta}} = \Lambda^{\alpha'}{}_{\alpha\beta} = \Lambda^{\alpha'}{}_{\gamma} \Gamma^{\gamma}{}_{\alpha\beta} \tag{36}$$

where we may note that the symmetry of $\Gamma^{\gamma}{}_{\alpha\beta}$ under $\alpha \Leftrightarrow \beta$ assures that we have just the right number of degrees of freedom in the connection to determine $\partial^2 x^{\alpha'}/\partial x^{\alpha}\partial x^{\beta}$.

This is valid for arbitrary manifolds, either locally Euclidean or locally Minkowskian, subject to the constraints of smoothness and also that the inverse of the metric in (28) exist; which is essentially the requirement that the coordinates we have laid down are such that the determinant of the metric be non-vanishing.

With the transformation above, we obtain a prime-frame coordinate system in which the 1st derivatives of the metric components vanish, and consequently the connection also vanishes in the primed-frame:

$$\Gamma^{\gamma'}{}_{\alpha'\beta'} = 0. \tag{37}$$

And this means that the geodesic equation is

$$d^2 x^{\alpha'}/d\lambda^2 = 0 \tag{38}$$

so, for the case of interest – a locally Minkowskian manifold – particles following geodesics have world-lines for which the primed coordinates increase linearly with proper time, and their paths are unaccelerated. These so-called *inertial coordinate systems* can be realised by freely falling observers carrying rulers and clocks, as illustrated in figure 4.

3.2.3 Beyond local flatness

As an alternative, we might have argued, somewhat loosely, but correctly, that the number of coefficients in the second-order term involving $\partial^2 x^{\alpha'}/\partial x^{\alpha}\partial x^{\beta}$ in N dimensions is $N^2(N+1)/2$, there being N choices for α' and $N(N+1)/2$ choices for α and β (N where $\alpha = \beta$, plus $N(N-1)/2$ where $\alpha \neq \beta$), which is the same as the number of independent components in $g_{\alpha'\beta',\gamma'}$ since $g_{\alpha'\beta'}$ is symmetric, so the transformation above should have enough freedom to render $g_{\alpha'\beta',\gamma'} = 0$.

What if we try to extend this and consider a 3rd order Taylor expansion of $x^{\alpha'}(\vec{x})$ with an additional set of coefficients $\partial^3 x^{\alpha'}/\partial x^{\alpha}\partial x^{\beta}\partial x^{\gamma}$. Can we use this to find a coordinate system in which the *second derivatives* of the metric $g_{\alpha'\beta',\gamma'\delta'}$ also vanish?

Let's do this in 2-dimensions. The metric derivatives – being symmetric under $\alpha' \Leftrightarrow \beta'$ and $\gamma' \Leftrightarrow \delta'$ – has $3 \times 3 = 9$ independent components – while in $x^{\alpha'}{}_{,\beta\gamma\delta}$ the index α' can take values $0, 1$, while the distinct combinations of the other indices are $(\beta\gamma\delta) = (000), (001), (011)$ & (111), for a total of $2 \times 4 = 8$ independent

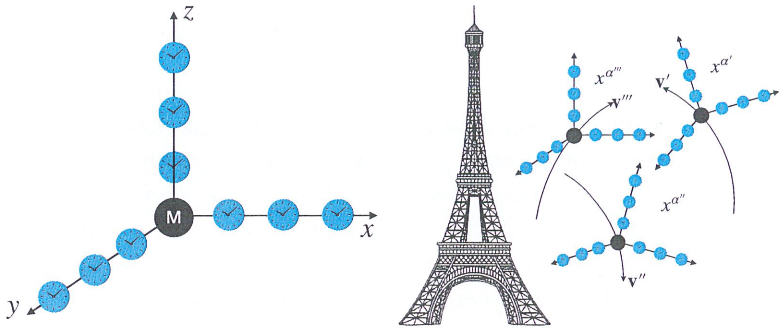

Figure 4: Inertial coordinates $x^{\alpha'}$ can be realised by having a massive freely falling non-spinning observer carry a set of rigid rulers on which there are other (massless) observers with clocks. This is shown at the left. The clocks are assumed to have been synchronised by exchanging light signals. These observers can record the $x^{\alpha'}$ coordinates of events that are labelled with the general coordinates x^α. From this they can establish the transformation matrix $\Lambda^\alpha{}_{\alpha'} = \partial x^\alpha / \partial x^{\alpha'}$ of the mapping $x^\alpha = x^\alpha(x^{\alpha'})$. As illustrated at the right – with the tower there to remind us that we are in the presence of a gravitational field – we can have multiple inertial frames that are moving and/or rotated with respect to each other. They form a 6-parameter family.

components. So we do not have enough freedom at our disposal to make $g_{\alpha'\beta',\gamma'\delta'}$ vanish. But we are only short by one; this says that the curvature of a 2-dimensional manifold is described by a single number at each point.

Coming back to the question posed at the outset, evidently the blind ants cannot sense the orientation of the egg! They can send out a pair of lines from a point with a certain opening angle and measure the deficit in the length of the arc joining their ends as compared to the Euclidean expectation. Or they can send a pair of ants marching along initially parallel paths and sense the change of separation. But the result of any such experiment is independent of the direction they do this measurement. Only if they were to do *non-local* experiments such as circumnavigating the egg can they learn, for example, that they are not living on a sphere.

More generally, the number of distinct combinations of β, γ & δ indices in $x^{\alpha'}{}_{,\beta\gamma\delta}$ is, in general, N where all indices are equal plus $N(N-1)$ where two are the same and the other is different and $N!/(N-3)!3!$ where they are all different (that is if $N > 2$; for $N = 2$, as we saw, there is no way to have 3 indices all different). Multiplying by N, for the possible values of α', gives $N(N^2 + N!/(N-3)!3!)$ (or 80 for $N = 4$) as the number of independent combinations of indices of $x^{\alpha'}{}_{,\beta\gamma\delta}$.

On the other hand, $g_{\alpha\beta,\gamma\delta}$ has, in general, $(N(N+1)/2)^2$ independent components, so for $N = 4$ there are 100 independent components.

Thus the intrinsic curvature in 4D is characterised by 100 - 80 = 20 numbers, and, in general, the curvature in N dimensions is characterised by $(N(N+1)/2)^2 - N(N^2 + N!/(N-3)!3!) = N^2(N^2-1)/12$ numbers (which works for $N = 2$ if we define $(-1)! = 0$).

In 4D then, the curvature – which we will see shortly is a tensor – has 20 independent components. This is more than the 6 independent components of the Newtonian tidal field tensor $\phi_{\mathrm{N},ij}$ that appears in the Newtonian geodesic deviation equation $d^2\Delta x_i/dt^2 = -\phi_{\mathrm{N},ij}\Delta x_j$. That's not unreasonable as we would expect the relativistic version of this to involve a 4-vector rather that a 3-vector displacement.

3.2.4 Parallel transport

In flat space (or space-time) we can compare vectors at different points simply by comparing their rectilinear-frame components. A vector field $\vec{V}(\vec{x})$ is constant if $\partial_{\beta'} V^{\alpha'} = V^{\alpha'}{}_{,\beta'} = 0$. The curvilinear components of such a field will vary with position, but the covariant derivative will vanish: $\nabla\vec{V} \longrightarrow V^\alpha{}_{;\beta} = 0$

Given a vector $\vec{V}$ at a point in a curved manifold one can *parallel transport* it along a line. This can be realised physically on a 2-dimensional curved surface as follows: if you drive over a line of wet paint, the vectors between the splotches of paint are parallel transported. Another way to construct a sequence of vectors that are parallel-transported copies of one another is by means of *'Schild's ladder'*, as illustrated in figure 5.

Mathematically, a vector that is being transported along a curve $\vec{x}(\lambda)$, with tangent vector $\vec{U} = d\vec{x}/\lambda$

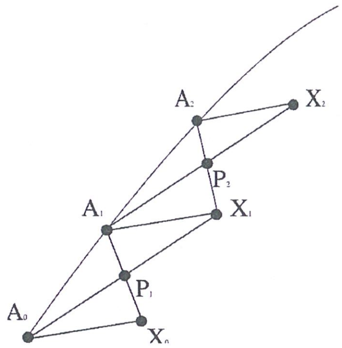

Figure 5: Left: 'Schild's ladder'. If we have a vector (here $\vec{X}_0$) at some point $\vec{x} = \vec{A}_0$, and a path $\vec{x}(\lambda)$ through that point, we can make a parallel transported copy $\vec{X}_1$ of $\vec{X}_0$ at $\vec{A}_1$ by erecting a vector $\vec{A}_1 - (\vec{A}_0 + \vec{X}_0)$. We then construct a vector from $\vec{A}_0$ to the mid point of that vector and extend it the same distance. This gives the end point of the vector $\vec{X}_1$. If we make the steps of the ladder smaller we obtain in the limit the (scaled down) parallel transported vector along the path.

has components that locally obey

$$d\vec{V}/d\lambda = \nabla_{\vec{U}}\vec{V} \longrightarrow U^{\beta}V^{\alpha}{}_{;\beta} = 0 \tag{39}$$

just as in flat space(-time). The reason this remains valid is that it only involves first derivatives (of components and basis vectors); the curvature appears, as we will see, only when we consider second derivatives.

But there is an important distinction: On a flat manifold, the result of transporting a vector is independent of the path and we can unambiguously construct a constant field. On a curved manifold, the result of parallel transport depends on the path taken, as illustrated in figure 6.

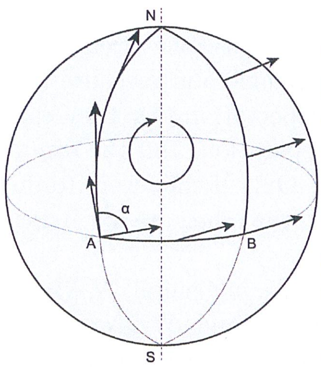

Figure 6: Parallel transport on a sphere. If we have a vector (here $\vec{X}_0$) at some point $\vec{x} = \vec{A}_0$, The result of parallel transporting a vector on a curved manifold depends on the path taken. If we transport a vector around a closed loop, the end result will not agree with what we started with. For a small loop – one for which the dimensions are small compared with the radius of curvature of the sphere in the case illustrated – difference is proportional to the area area of the loop. This property serves to define the *curvature tensor*. It can be thought of as a machine, or sub-routine, that takes an initial vector $\vec{V}$, 2 other vectors $\vec{a}$ and $\vec{b}$ that can be the edges of a parallelogram loop, and returns the change $\Delta\vec{V}$. It is a rank-4 tensor.

noncordinate treatment.

3.2.5 Covariant differentiation of fields

If we have a vector *field* $\vec{V}(\vec{x})$ defined on a manifold, the *covariant derivative* $\nabla\vec{V}$ is the answer to the question: *how does the field change with respect to a parallel transported copy of itself?* This is illustrated in figure 7. It is well defined as a limit; but the change for a finite path is path dependent.

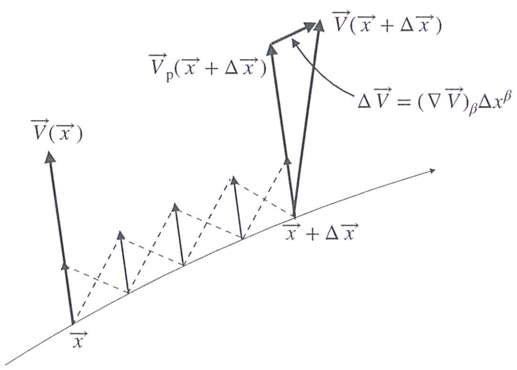

Figure 7: We define the *covariant derivative* $\nabla\vec{V}$ of a vector field $\vec{V}(\vec{x})$ to be the rate at which the field is changing with respect to a parallel transported version of itself. Starting with the value of the field at $\vec{x}$ we make a parallel transported copy $\vec{V}_{\text{p}}$ at a nearby position $\vec{x}+\Delta\vec{x}$ which we subtract from the value of the field there to give $\Delta\vec{V}$. The components of the difference vector are given by $(\Delta\vec{V})^{\alpha} = V^{\alpha}{}_{;\beta}\Delta x^{\beta}$ where the components $V^{\alpha}{}_{;\beta}$ of the rank-2 tensor $\nabla\vec{V}$ are given by the same formula as in curvilinear coordinates in flat space-time.

3.2.6 Curvature

The curvature tensor $\vec{R}$ can be defined as the change $\Delta\vec{V}$ obtained after transporting a vector $\vec{V}$ around a parallelogram defined by two vectors $\vec{a}$ and $\vec{b}$ (see figure 8).

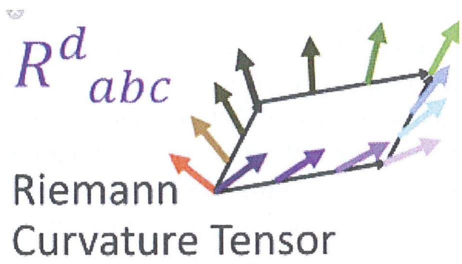

Figure 8: Curvature may be defined in terms of parallel transport of a vector around a loop, which can be taken to be a small parallelogram. The change in the vector depends on the original vector and the 2 vectors defining the path in a linear manner: $\Delta\vec{V} = \mathbf{R}(\ ,\vec{V},\vec{a},\vec{b})$, which is the equivalent, in GR, of the Newtonian (tidal) gravitational field. It is computable from the connection, which in turn derives from the metric, and when contracted gives the Einstein tensor on the LHS of his field equations.

We can write – in geometric notation –

$$\Delta\vec{V} = \mathbf{R}(\ ,\vec{V},\vec{a},\vec{b}) \tag{40}$$

or – in component notation –

$$\Delta V^\alpha = R^\alpha{}_{\beta\mu\nu} V^\beta a^\mu b^\nu . \tag{41}$$

One way to calculate the curvature, which must be expressible in terms of the connection, is illustrated in figure 9.

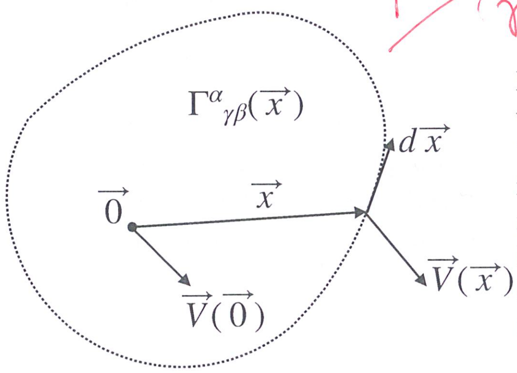

Figure 9: To calculate the curvature – the change of a vector $\vec{V}$ when transported around a loop – we first construct a vector *field* $V^\alpha(\vec{x}) = V^\alpha(0) - \Gamma^\alpha{}_{\gamma\beta}(\vec{0})V^\gamma(\vec{0})x^\beta$ by parallel transporting $\vec{V}(\vec{0})$ from $\vec{0}$ to points $\vec{x}$ on the loop. We then use that as an approximation to evaluate the integral in $\Delta V^\alpha = -\oint dx^\beta \Gamma^\alpha{}_{\gamma\beta}(\vec{x})V^\gamma(\vec{x})$. Note that we need here only 1st order precision for both $\Gamma^\alpha{}_{\gamma\beta}(\vec{x})$ and $V^\gamma(\vec{x})$ in order to obtain an approximation to the integral that is valid at second order in the loop size.

The result of this is the *Riemann curvature tensor*, expressed in terms of the connection by

$$\boxed{R^\alpha{}_{\beta\mu\nu} \equiv \Gamma^\alpha{}_{\beta\nu,\mu} - \Gamma^\alpha{}_{\gamma\nu}\Gamma^\gamma{}_{\beta\mu} - \{\nu \Leftrightarrow \mu\}} \tag{42}$$

which evidently contains both derivatives and products of the Christoffel symbols.

It is a bit tedious, but one can show that the symmetries of the Riemann tensor are such that it does indeed have 20 independent components in 4-dimensions.

3.2.7 Geodesic deviation

Curvature plays a key role in the *equation of geodesic deviation.* (see figure 10):

$$\left(\frac{d^2\vec{\xi}}{d\lambda^2}\right)^\alpha = \underbrace{(R^\alpha{}_{\mu\beta\nu}p^\mu p^\nu)}_{\text{tidal field tensor}} \xi^\beta \tag{43}$$

whereby the curvature controls the focussing – or defocussing – of neighbouring geodesics. In Newtonian gravity the same role is played by the gravitational *tidal field*:

$$\frac{d^2\boldsymbol{\xi}}{dt^2} = -\boldsymbol{\xi} \cdot \underbrace{\nabla\nabla\phi}_{\text{tidal field}} . \tag{44}$$

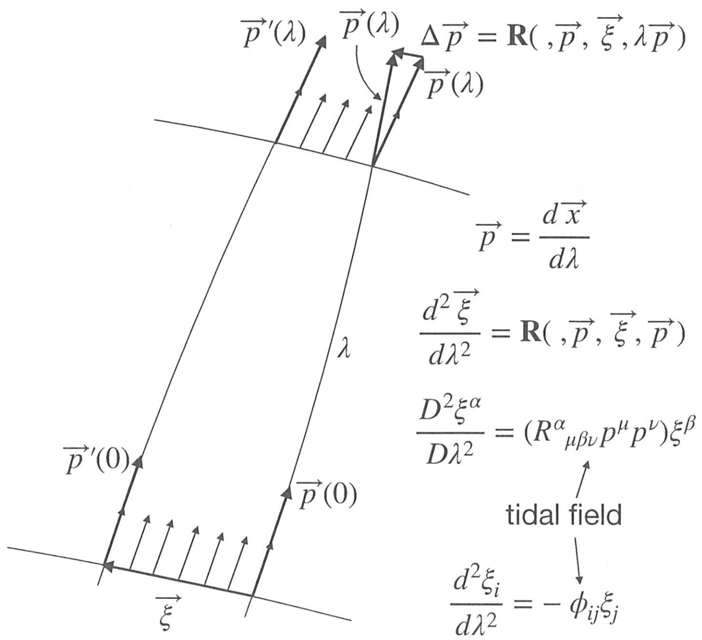

Figure 10: Geodesic deviation. This is a space-time diagram; time increases vertically. We start with a particle with 4-momentum $\vec{p}(0)$ at the bottom right. We can choose a parameter λ along the path – proportional to the proper time on a clock carried by the particle such that $\vec{p} = d\vec{x}/d\lambda$. Clone the particle and parallel transport it's momentum along the *separation vector* $\vec{\xi}$ to make $\vec{p}'(0)$. We have two particles with parallel momenta. Advance them forward in time (or λ) to make $\vec{p}(\lambda)$ and $\vec{p}'(\lambda)$ which will no longer be parallel (unless there is no curvature of space-time). Transport $\vec{p}'(\lambda)$ back to the location of the original particle and subtract the momenta. The result $\Delta\vec{p}$ is, from the definition of curvature, as indicated. But since $\vec{p} = d\vec{x}/d\lambda$, the rate of change of $\Delta\vec{p}$ is the same as $d^2\vec{\xi}/d\lambda^2$. This gives the *geodesic deviation equation.*

3.2.8 The Ricci and Einstein tensors

The *Ricci tensor* is defined to be the contraction of the Curvature tensor on the 1st and 3rd indices: $R_{\mu\nu} = R^\gamma{}_{\mu\gamma\nu}$. This looks like a candidate for the left hand side of the field equations. But it does not obey the continuity law $R^{\mu\nu}{}_{;\mu} = 0$. But, it turns out, if one subtracts from it half of $\mathbf{g}$ times the *Ricci scalar* $R \equiv R^\mu{}_\mu$ the so-called *Einstein tensor*

$$G_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R \tag{45}$$

does.

3.2.9 Einstein's equations

The geodesic deviation equation allows one to identify certain components of the Riemann tensor with the corresponding components of $\partial^2\phi/\partial x_i\partial x_j$. This ties down the single free parameter κ in Einstein's gravity to be $\kappa = G_\mathrm{N}/c^4$ so Einstein's equation becomes

$$\mathbf{G} = 8\pi(G_\mathrm{N}/c^4)\mathbf{T} \tag{46}$$

where, just as it is a *contraction* of the tidal field tensor – its trace ϕ_{ii} – that appears in Poisson's equation

$$\nabla^2\phi = 4\pi G_\mathrm{N}\rho \tag{47}$$

the *Einstein tensor* $\mathbf{G}$ is a certain contraction of the rank-4 curvature tensor.

3.2.10 Raychaudhuri's equation

A related useful result is *Raychaudhuri's equation.* While the relative acceleration of a pair of particles depends on both the tidal focussing caused by the local matter and the tidal field coming from distant matter. But the latter does not change the volume occupied by a collection of particles. If one takes the inverse cube root of the volume: $a = \sqrt[3]{V}$, and, if one assumes an irrotational flow – a sensible approximation in an expanding volume element as momentum conservation implies that any rotational motions decay – one finds that this obeys the equation

$$\ddot{a} = -\frac{4\pi}{3}G(\rho + 3P/c^2)a \tag{48}$$

This is exactly as in Newtonian gravity except that pressure appears along with the density. This is one of the fundamental equations of cosmology.

3.2.11 The cosmological constant

Einstein obtained the field equations $\mathbf{G} = 8\pi\kappa\mathbf{T}$, which are the simplest compatible with Newtonian gravity with ρ replaced by $\mathbf{T}$ as the source, in 1915. In 1917, in order to allow static (non-expanding) solutions with vanishing pressure, he proposed the modification

$$\mathbf{G} + \Lambda\mathbf{g} = 8\pi\kappa\mathbf{T} \tag{49}$$

which introduces a new constant of nature Λ with units of inverse length squared.

If we move this over to the right-hand side of the field equations, we see that a positive Λ would correspond to a matter source with $T^{\mu\nu} = (\Lambda/8\pi\kappa)\mathrm{diag}(1, -1, -1, -1)$, or, equivalently, to matter with positive density $\rho = T^{00}/c^2$ and strong, but negative, pressure $P = -\rho c^2$. I.e. matter with strong *tension*.

This means that a positive Λ causes a pair of test particles to accelerate away from one another – because the pressure terms in $\rho + 3P/c^2$ outweigh the effect of the density – rather as would a negative mass density without pressure or tension.

With the discovery of the expansion of the universe, the cosmological constant was discarded. But it has re-emerged recently with vigour, and in two different situations.

First, in the theory of *inflation* it is assumed that the dynamics of the universe was at early times dominated by the effect of a nearly spatially constant and slowly time-varying relativistic scalar field dubbed the 'inflaton' (a cousin of the Higgs field, if you like).

Second, in 1999, it was shown convincingly[2] that the universe is now entering an accelerating phase. This is naturally interpreted as the effect of the cosmological constant or, perhaps, the influence of another scalar field analogous to the inflaton.

3.3 What is *the* gravitational field?

In Newtonian gravity we have the potential ϕ, the gravitational acceleration $\mathbf{g} = -\boldsymbol{\nabla}\phi$ and the tide $\phi_{ij} = \partial^2\phi/\partial x_i \partial x_j$.

In Einstein's gravity we have, analogously, the metric $g_{\mu\nu}$, the connection $\Gamma^\alpha{}_{\mu\nu}$ and the curvature $R^\alpha{}_{\beta\gamma\delta}$. Which of these deserves to be called *the* gravitational field?

The answer is the *curvature*. In many situations, one can have non-trivial metric tensors – i.e. coordinate systems in which $g_{\mu\nu} \neq \mathrm{diag}(-1, 1, 1, 1)$ – and non-vanishing connection also. These include simply working in curvilinear coordinate systems, such as spherical coordinates, but can also be used include affects of acceleration, as in a rocket, or on a rotating roundabout. But they have nothing to do with gravity *per se*.

The curvature, on the other hand, vanishes in such situations, and it would be zero everywhere and always in a world where Einstein's constant κ (or Newton's G_{N}) were zero. The curvature, or tidal field, considered as a geometric entity $\mathbf{R}$ is generated by the presence of matter $\mathbf{T}$ (along with boundary conditions; the field equations, like Poisson's equations, not providing explicitly all the components of the curvature). If it vanishes – and being a tensor, its vanishing or not-vanishing is an absolute fact – then there is no gravity. In this sense, there is nothing 'relative' about general relativity.

This makes nonsense of the oft stated form of 'Einstein's principle of equivalence' (EEP): *that gravity and acceleration are indistinguishable.*. If you are being accelerated in a rocket out in empty space then there is no curvature; the gravitational field vanishes. Adopting this view, we would have to say that Pound and Rebka did not measure the *gravitational redshift* in their celebrated experiment in 1959. What they measured was merely the affect of their apparatus being accelerated.

A better way to state the EEP is to say that, if you are freely falling, you will not see any effect of gravity (locally). Note that this is very different from the situation with e.g. electromagnetism, where one can measure e.g. the electric field $\mathbf{E}$ locally. Fundamentally that is because, in electromagnetism, particles generally have different charge-to-mass ratios, whereas in gravity, the Galilean equivalence principle – that all particles fall the same way under the influence of a gravitating body – means that all objects have the same effective charge-to-mass ratio.

[2] A critical addition to the Hubble diagram for supernovae was the evidence from the CMB that the spatial curvature of the universe must be very close to being spatially flat, so this required something to augment the rather low measured matter density.

4 Homogeneous Expanding Universe Models

4.1 The Friedmann, Lemaitre, Robertson & Walker (FLRW) world model

4.1.1 The cosmological principle

Observations – particularly of the *cosmic microwave background* (CMB) – show our universe to be highly *isotropic* on large scales.

Establishing that the Universe is *spatially homogeneous* is more difficult, but there is good evidence for that as well.

At the time of Hubble there was very little firm evidence for either. But, perhaps presciently, or perhaps because the only solutions of Einstein's equations were those of very high symmetry, cosmologists adopted the *cosmological principle* that the Universe is spatially homogeneous (i.e. that, aside from local irregularities, it looks the same from all points in space – at a given time).

A subset of cosmologists went further than this and argued for what is called the *perfect cosmological principle:* that not only does the Universe appear the same at all points in space but also at all times. This was not widely adopted[3], but, interestingly it is now believed by most cosmologists that both during *inflation* in the very early universe and in the distant future as well, the Universe will become time invariant and obey the perfect cosmological principle.

4.1.2 Cosmic-time, comoving coordinates and fundamental observers

In a homogeneous universe that started with a big-bang, a useful time coordinate is the proper-time τ since the explosion. This is called *cosmic time.*

We now need to set up three spatial coordinates. We choose these to be tied to so-called *'fundamental observers'* – which we can think of practically as galaxies – to whom the universe looks identical at equal cosmic time.

We can take one of these observers to define the origin of the spatial coordinates. We will choose it to be the Milky Way, but it could equally have been any other galaxy. And, since the universe looks isotropic around us, we set up angular coordinates θ, ϕ such that solid angle on the sky is

$$d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2. \tag{50}$$

Light comes to us from other galaxies along lines of constant θ and ϕ.

Local flatness implies that, in our vicinity, space-time looks Minkowskian, so $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$ where t is coordinate time. But for galaxies close to us, with small recession velocities, γ is close to unity and $d\tau = dt/\gamma$ is close to dt with only $\mathcal{O}(v^2/c^2)$ corrections, which we'll ignore for now. Making the transformation of spatial coordinates $(x, y, z) = a(\tau)r(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$

$$ds^2 = -c^2 d\tau^2 + a^2(\tau)(dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)). \tag{51}$$

Here $a(\tau)$ is the universal *scale factor*, with units of length, and is chosen so that other local fundamental observers have constant r, θ, ϕ, which are called *comoving coordinates* and which are all dimensionless.

This metric – or *world model* – has the symmetries required by the cosmological principle, but is only valid in our local neighbourhood. This form of the metric is said to be *spatially flat*, because the spatial part of the metric is Euclidean. It turns out, in fact, that our Universe appears to be spatially flat, to a very good approximation, but it didn't have to be like that.

The situation here is analogous to 2D geometry on a planar surface or on a sphere, as illustrated in figure 11. Both of these are homogeneous spaces – all points on a plane or sphere being equivalent – but the latter can be thought of as a generalisation of the planar metric $dl^2 = dr^2 + r^2 d\phi^2$ with $r^2 \Rightarrow f(r)$, with $f(r)$ chosen to be $R = \sin(r/R)$.

In fact, there is another generalisation of the plane that is a space of constant curvature, which is the saddle, for which $f(r) = R\sinh(r/R)$, as illustrated in figure 12.

Similarly, a generalisation of the metric above that respects the symmetries of isotropy and homogeneity is to replace $r^2 \Rightarrow f^2(r)$. But, as in 2D, the requirement that the curvature be independent of position

[3] In 1953, Herbert Dingle, who was president of the Royal Astronomical Society said 'Since it causes me considerable discomfort to use names that are clearly misleading, I shall refer to the "cosmological principle" as the *cosmological assumption* and to the "perfect cosmological principle" as the *cosmological presumption*'

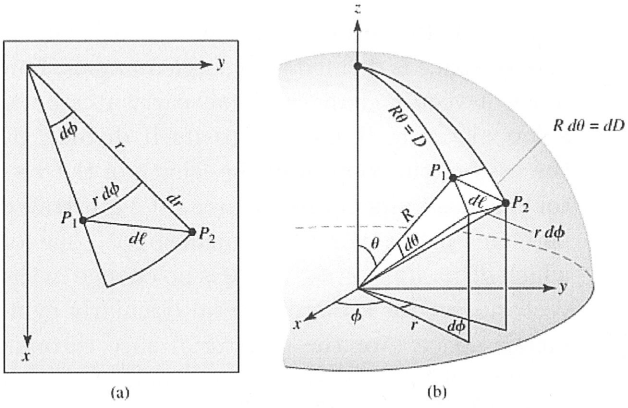

Figure 3.2: (Reproduced from Carroll & Ostlie's *Modern Astrophysics*).

Figure 11: Illustration of the line element $dl^2 = g_{\alpha\beta}dx^\alpha dx^\beta$ in polar coordinates on a plane (left) where $dl^2 = dr^2 + r^2 d\phi^2$ and on a sphere embedded in 3-D (right) where $dl^2 = R^2(d\theta^2 + \sin^2(\theta)d\phi^2)$. Both spaces are homogeneous and isotropic and locally flat. The metric on a sphere can be written as $dl^2 = dr^2 + f^2(r)d\phi^2$, where $r = R\theta$, with particular choice for the function $f(r) = R\sin(r/R)$, for which choice the curvature is independent of position. For small displacements from any point – which can be taken to be the pole – the two metrics are equivalent.

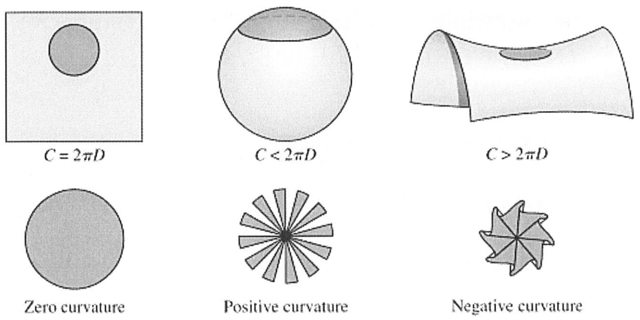

Figure 12: In 2-dimensions there are 3 choices of geometry that are homogeneous and isotropic: the sphere, the plane and the saddle. They are distinguished by whether the circumference of a circle is less than, equal to, or greater than 2π times the radius. An alternative way to determine the curvature locally is to take a vector and 'parallel transport' it around a small closed loop and differencing the transported copy from the original (see figure 8).

restricts the form of the function. To obtain $f(r)$ it is sufficient to consider the 'equatorial plane' $\theta = \pi/2$ on which 2D surface (at some chosen cosmic time) we know that the curvature is characterised by a single number. We can take this to be $(|\Delta \mathbf{V}_2| - |\Delta \mathbf{V}_1|)/|\mathbf{V}|A$ where $\mathbf{V}$ is a radial vector lying in this surface and $\Delta \mathbf{V}_2$ and $\Delta \mathbf{V}_1$ are the changes in the vector when transported either first radially and then tangentially or in the opposite order around a loop of area A, as illustrated in figure 13. For this to be independent of r requires that

$$f''/f = \text{constant} \tag{52}$$

where the prime denotes differentiation of $f(r)$ with respect to r. The possible solutions are, unsurprisingly,

$$f(r) = \begin{cases} R\sin(r/R) \\ r \\ R\sinh(r/R) \end{cases} \tag{53}$$

where R is the curvature radius in comoving coordinates.

4.1.3 The FLRW line element

If we define $\chi \equiv r/R$, which is often called *conformal distance* and absorb R into the scale factor, we obtain one of the conventional forms for the *FLRW line element*

$$\boxed{ds^2 = -c^2 d\tau^2 + a^2(\tau)(d\chi^2 + S_k(\chi)^2(d\theta^2 + \sin^2\theta d\phi^2))} \tag{54}$$

where the *spatial coordinates* (χ, θ, ϕ) are dimensionless fixed labels attached to *fundamental observers* (FOs) who carry clocks measuring proper time τ, and where

$$S_k(\chi) = \begin{cases} \sin\chi \\ \chi \\ \sinh\chi \end{cases} \quad \text{for} \quad k = \begin{cases} +1 \\ 0 \\ -1 \end{cases} \tag{55}$$

where k is the *curvature constant* (sometimes called the *curvature eigenvalue*).

The *spatial geometry* – i.e. the geometry of *hypersurfaces* of constant time τ – may be analogous to a sphere (called a *hypersphere*), flat (Euclidean) or hyperbolic (saddle-like), depending on the curvature constant k.

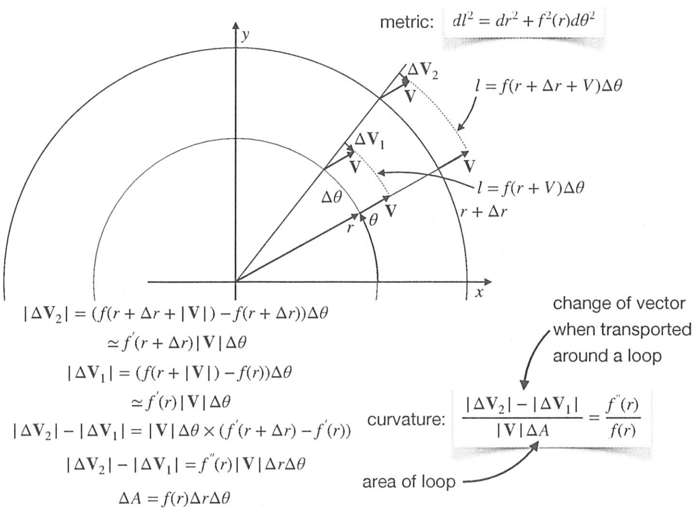

Figure 13: In 2-dimensions, the intrinsic curvature of space is defined by a single number; how much a vector changes if you carry it around a loop keeping it parallel to itself divided by the area of the loop (and the length of the vector). It has units of inverse area. It's illustrated here on a plane. For a Euclidean space – one for which $dl^2 = dr^2 + r^2 d\theta^2$ there is no change in the vector. But for a more general circularly symmetric space – or the equatorial slice through a 3-dimensional spherically symmetric space – with $dl^2 = dr^2 + f^2(r) d\theta^2$ – the curvature is non-zero and is equal to f''/f. Homogeneous spaces must have f''/f = constant, the possible solutions of which are $f(r) = R\sin(r/R)$, $f(r) = r$ or $f(r) = R\sinh(r/R)$.

We can read off from (54) the components of the metric (in τ, χ, θ, ϕ) coordinates

$$\mathbf{g} \longrightarrow \mathrm{diag}\{-c^2, a^2, a^2 S_k(\chi)^2, a^2 S_k(\chi)^2 \sin^2\theta\}. \tag{56}$$

As in *special relativity* (Minkowski space-time), the (squared) interval of proper separation between neighbouring 'events'

$$ds^2 \text{ may be} \begin{cases} \text{negative} \\ \text{zero} \\ \text{positive} \end{cases} \text{for} \begin{cases} \text{timelike} \\ \text{null} \\ \text{spacelike} \end{cases} \text{intervals} \tag{57}$$

The scale factor $a(\tau)$ is analogous to the radius of curvature of a sphere. For $k = +1$, so $S_k(\chi) = \sin(\chi)$, the range of χ is finite, and if we take a radially directed spatial geodesic (extremal line on surface of τ = constant) from one pole ($\chi = 0$) to the other ($\chi = \pi$), it has length $\pi a(\tau)$, and if you extend it twice as far you get back to the starting point. Such models are said to be (spatially) *closed*. If composed of ordinary matter, they are also closed in time as they end in a 'big-crunch'. For $k = -1$, the universe is infinite in extent, and is said to be spatially *open*. In all cases, if we consider a region of the universe of physical size much less than $a(\tau)$, the effects of spatial curvature are small.

The line element has many uses. For example, if we observe a distant spherical object of proper diameter dl that is lying in the surface $\theta = \pi/2$ then the coordinate displacement vector across the object is $(0, 0, 0, d\phi)$ for which the proper separation is $ds = a(\tau_{\rm em}) S_k(\chi) d\phi = dl$ where $\tau_{\rm em}$ is the cosmic time when the photons we observe left the object. This gives the *angular diameter distance* $D_{\rm a} = dl/d\phi = a(\tau_{\rm em}) S_k(\chi)$ which, if $k \neq 0$, is different from the Newtonian expression. The latter, however, coincides with the relativistic result if $k = 0$, which appears to be a good approximation for our universe.

Another useful form for the metric is obtained if we define a new dimensionless time coordinate η such that

$$c d\tau = a(\tau) d\eta \tag{58}$$

in terms of which the line element is

$$ds^2 = a^2(\eta)\left(-d\eta^2 + d\chi^2 + S_k^2(\chi)(d\theta^2 + \sin^2\theta d\phi^2)\right) \tag{59}$$

where $a(\eta) = a(\tau)$, and which is said to be a 'conformal transformation' of the simper metric in parentheses without the scale factor. This leads to the terminology of η being called *'conformal time'*.

One may note that if we define $r \equiv S_k(\chi)$ we have $dr^2 = (1 - kr^2) d\chi^2$ so an alternative way to write the metric is

$$ds^2 = -c^2 d\tau^2 + a^2(\tau)\left(\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right). \tag{60}$$

This form is what is used in the appendix to develop the Einstein equations.

The other function $a(\tau)$ appearing in the metric is called the *cosmological scale factor*. It is not determined by symmetry considerations; rather it is determined by Einstein's field equations.

4.2 The Friedmann equations

In 1922, Alexander Friedmann found the equations describing an expanding homogeneous and isotropic universe.

He assumed the metric above and that the stress-energy tensor for the matter was that of an 'ideal fluid'

$$T_{\alpha'\beta'} = \begin{bmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{bmatrix} \tag{61}$$

where the primes on the indices indicate that this is in physical coordinates erected in the vicinity of a fundamental observer

As already discussed, in this tensor:

- the upper left element is the *total energy density* $\mathcal{E} = \rho c^2$
- and the lower-right sub-matrix is the diagonal stress tensor containing the *isotropic pressure* P
- three zeros in the left column are the momentum density,
- the three zeros in the top row are the energy flux density
- all these are as would be measured in the frame of a fundamental observer
- sometimes called *'comoving observers'* as they are co-moving with the cosmological fluid in that they see vanishing momentum density and energy flux in their frame of reference

Dynamical equations for $a(\tau)$ and $\rho(\tau)$ can be obtained from a combination of the Einstein field equations (by computing the connection and hence the Einstein tensor $\mathbf{G}$ and equating it to $8\pi\kappa\mathbf{T}$) and the laws of continuity of energy and momentum. As shown in appendix A, the resulting *Friedmann equations* are the *acceleration equation:*

$$\boxed{\ddot{a} = -(4\pi/3)G(\rho + 3P/c^2)a} \tag{62}$$

where dot denotes derivative with respect to cosmic time τ, the *energy equation:*

$$\boxed{\dot{a}^2 = (8\pi/3)G\rho a^2 - c^2 k} \tag{63}$$

and the *matter continuity equation:*

$$\boxed{\dot{\rho} = -3(\dot{a}/a)(\rho + P/c^2).} \tag{64}$$

These all agree with the Newtonian cosmology equations if $P = 0$.

4.2.1 The Friedmann energy equation

Significant features of (63) are:

- it is (if multiplied by 1/2) identical in form to the Newtonian expression for conservation of energy (kinetic energy + potential energy = constant) for an expanding spherical shell of radius a containing mass $M = (4\pi/3)G\rho a^3$
- dividing by a^2 the left hand side is the observable H^2 so, augmented by a measurement of the density ρ, this allows one to determine the radius of curvature a_0 and the curvature eigenvalue k
- the sign of the spatial curvature is determined by whether the kinetic energy is greater than, less than, or equal to the gravitational energy
- the pressure P does not appear – so the spatial curvature is determined by ρ alone

4.2.2 The matter continuity equation

From continuity equation (64) can be obtained from the time component of $\nabla \cdot \mathbf{T} = 0$. With $P = 0$, this has a solution $\rho \propto a^{-3}$ and so (defining $V \equiv (4\pi/3)a^3$) this then simply expresses conservation of mass:

$$d(\rho V)/d\tau = 0 \tag{65}$$

While for $P \neq 0$, and since $dV = 3(da/a)V$, the extra term implies

$$d(\rho V)/d\tau = -(P/c^2)dV/d\tau \tag{66}$$

which, on multiplying by c^2, we recognise as the first law of thermodynamics

$$\boxed{dE = d(\mathcal{E}V) = -PdV.} \tag{67}$$

4.2.3 The acceleration equation

The energy and continuity equations are two 1st order equations. If we take the proper time derivative of the first we get

$$2\dot{a}\ddot{a} = (8\pi/3)G(\dot{\rho}a^2 + 2\rho a\dot{a}) \tag{68}$$

and if we use the second to eliminate $\dot{\rho}$ we get the 2nd order acceleration equation:

$$\ddot{a} = -(4\pi/3)G(\rho + 3P/c^2)a \tag{69}$$

- which, for $P = 0$, is what Newton would have written down for an expanding sphere of dust
- but which, in general, contains an additional deceleration from the pressure
- sometimes expressed by saying 'pressure gravitates in GR'

Note that the three equations for $\dot{a}^2$, $\dot{\rho}$ and $\ddot{a}$ above are not independent, as any one of them can be obtained from the other two.

It is important to realise that the presence of the pressure here is not the effect of e.g. the kinetic energy of motion of particles in the case of a gas. Such contributions to the energy do gravitate, but are already included in the density as ρc^2 is the total energy density.

4.2.4 The 'equation of state'

The energy and continuity equations (which, we have just seen, imply the acceleration equation) give us two 1st order equations for three unknown functions of time: $a(\tau)$, $\rho(\tau)$ and $P(\tau)$.

To obtain solutions we must augment these with an *'equation of state'* giving the pressure $P = P(\rho)$. This may be

- $\boxed{P_{\rm m} = 0}$ for pressureless matter ('dust')
  - an important component at present
- $\boxed{P_{\rm r} = \rho c^2/3}$ for radiation
  - dynamically negligible now, but dominant in the past since it has $\rho_{\rm r} \propto a^{-4}$ as compared to $\rho_{\rm m} \propto a^{-3}$

Observations that the expansion of the universe is speeding up suggest we need a third component: *'dark energy'* with negative pressure to give $\ddot{a} > 0$.

Our ignorance about its equation of state is encapsulated in an unknown function of time (or redshift)

$$\boxed{\omega \equiv P/\rho c^2} \tag{70}$$

though a note on terminology is in order. An 'equation of state' (EoS) in thermodynamics is a relation giving one thermodynamic variable in terms of 2 others (e.g. $P = P(\rho, T)$), and which may be adiabatic, isothermal etc.In cosmology people use EoS rather loosely for an expression for P in terms of density alone, and as just the ratio of pressure to density. This is either because the temperature is either assumed to be solely a function of the density – as for thermal radiation that is expanding adiabatically – or because one is not dealing with a thermodynamic system (e.g. quintessence or the inflaton) so temperature is not defined.

4.2.5 Dark energy and the cosmological constant

The Friedmann equations were obtained without any cosmological constant term $\Lambda \mathbf{g}$ on the left-hand side of the Einstein equations. In non-expanding (primed) coordinates, such a term would have components $\Lambda \mathbf{g} \longrightarrow \Lambda\, \text{diag}\{-1, 1, 1, 1\}$. That would be equivalent to having a source term $\mathbf{T} \longrightarrow (\Lambda/8\pi\kappa)\, \text{diag}\{1, -1, -1, -1\}$ on the right-hand side. I.e. a positive energy density but a strong, and *negative,* pressure.

$$\boxed{P_\Lambda = -\rho_\Lambda c^2} \tag{71}$$

The corresponds to an equation of state parameter $\omega = -1$. This equation of state arises in inflation in the early universe and also if the dark energy is provided by a scalar field (*'quintessence'*) during late-time inflation.

From the continuity equation, dark energy with $\omega = -1$ has $\rho = $ constant and so would have become negligible in the recent past, but will dominate in the future.

4.2.6 Why does pressure increase the deceleration of the universe?

The energy equation appears very Newtonian. The continuity equation can be understood as local conservation of energy with $E = Mc^2$ from special relativity thrown in. The most surprising feature of the Friedmann equations from this perspective is the appearance of pressure in the acceleration equation. Why is it there?

One answer is that it pops out of Einstein's equations, but that is not very satisfactory. Another view is simply that it would be unreasonable for the energy equation to contain pressure, which forces pressure to appear in the acceleration equation. It is important to understand that the pressure is really something we can control at will. Imagine we have an expanding universe full of 'dust'; gravitating matter with $P = 0$, but the dust is really H-bombs, all primed to explode at some cosmic time. Thereafter we have $P \neq 0$.

While there is now pressure, there are no *pressure gradients*, thus there is no force acting on the matter. If this were a finite sphere, there would be pressure gradients at the edge, but, at least initially, before any shock waves from the edge can propagate to the interior, there are no forces acting, so the the velocity of matter, and also the expansion rate, should be continuous. Hence the constant term in the energy should not change. And the energy equation, as we have seen, gives the curvature radius of the universe in terms of the energy density and the expansion rate. The energy density is not changed when the bombs go off. It would not make any sense for the energy equation to contain a pressure term, as this would then require an instantaneous change in the global curvature, and even the topology, of the universe.

So, in this thought experiment, nothing much changes instantaneously when the pressure is switched on, but, thereafter, the energy density evolves differently because of the presence of P in the continuity equation. With the curvature constant k being fixed, that means there must be a change in the expansion rate as compared to what would have been the case if the pressure had remained zero.

Another way to see how pressure gravitates in GR, without resorting to computing the Einstein equations for the FLRW metric, is given in the appendix where we show, using weak-field gravity, how the pressure enters into the geodesic deviation equation. This explains the presence of pressure in the acceleration equation and, which amounts to essentially the same thing, in Raychaudhuri's equation.

4.2.7 The expansion rate, critical density and the density parameters

The *expansion rate* is defined to be

$$\boxed{H \equiv \dot{a}/a} \tag{72}$$

with units of inverse time, and whose current value is measured to a few percent precision to be

$$H_0 \simeq 70\, \text{km/sec/Mpc}. \tag{73}$$

The Friedmann equation can be expressed as

$$H^2 = (8\pi/3)G\rho - c^2 k/a^2 \tag{74}$$

from which we can infer that

$$k = \text{sign}(\rho_0 - \rho_{\text{crit}}) \tag{75}$$

where the *'critical density'* is

$$\boxed{\rho_{\rm crit} \equiv 3H_0^2/8\pi G} \tag{76}$$

which is the density the universe would have to have for the potential and kinetic terms in the Friedmann equations to balance.

We often express densities in units of the present day critical density. So, for example, the *density parameter* for the matter is defined to be

$$\boxed{\Omega_{\rm m0} \equiv \rho_{\rm m0}/\rho_{\rm crit}.} \tag{77}$$

Observations indicate that $\Omega_{\rm m0} \simeq 0.3$ (see below), so the amount of matter is about 30% of that required to 'close the universe'. If that were all there was we would conclude that the universe must have negative (hyperbolic or saddle-like) spatial curvature since $\rho_0 < \rho_{\rm crit}$, and that the current proper distance to an object at a conformal distance equal to the curvature scale ($\chi = 1$) is

$$a_0 = cH_0^{-1}|1 - \Omega_{\rm m0}|^{-1/2} \tag{78}$$

i.e. somewhat greater than the *Hubble distance* $L_{\rm H} \equiv c/H_0 \simeq 4000{\rm Mpc}$.

However, as already mentioned, there is good reason to think that there is also a non-negligible dark energy component with density parameter $\Omega_{\Lambda 0}$, most probably very close to $1 - \Omega_{\rm m0}$ and, in the past one had to include radiation though its current density parameter is very small $\Omega_{\rm r0} \sim 10^{-4}$.

It is usual to define $\Omega_{k0} \equiv 1 - \sum\limits_{i={\rm m,r},\Lambda} \Omega_{i0}$,

- i.e. whatever would be needed to close the universe after accounting for all the matter content

using the fact that $\rho_{\rm m} \propto a^{-3}$, $\rho_{\rm r} \propto a^{-4}$, $\rho_\Lambda \propto a^0$ the Friedmann equation gives us the expansion rate when scale factor of the universe was a:

$$H = H_0[\Omega_{\rm m0}(a/a_0)^{-3} + \Omega_{\rm r0}(a/a_0)^{-4} + \Omega_\Lambda + \Omega_{k0}(a/a_0)^{-2}]^{1/2} \tag{79}$$

or, defining the *redshift* z by

$$\boxed{1 + z \equiv \frac{a_0}{a}} \tag{80}$$

we can compute the expansion rate as a function of redshift as

$$H(z) = H_0[\Omega_{\rm m0}(1+z)^3 + \Omega_{\rm r0}(1+z)^4 + \Omega_\Lambda + \Omega_{k0}(1+z)^2]^{1/2}. \tag{81}$$

4.2.8 Solutions of the Friedmann equations

If we specify type of constituents – and thus how their densities vary as a function of the size of the universe – then we can solve the Friedmann + continuity equation (or the acceleration equation) to get $a(\tau)$ and hence $\rho(\tau)$ and $P(\tau)$ also.

This requires, two boundary conditions: the present day density and expansion rate.

Unfortunately there are no analytic solutions with dark energy or pressure (except as limiting cases).

However, for a universe that contains only pressure free matter there is a parametric solution (the cycloid – for a closed universe – and hyper-cycloid for an open universe)

$$\begin{aligned} a(\eta) &= A(\cosh\eta - 1) \\ \tau(\eta) &= B(\sinh\eta - \eta) \end{aligned} \tag{82}$$

where A and B are constants and η is the conformal time, with the property that $d\eta \propto d\tau/a(\tau)$, and changes in conformal 'look-back time' and conformal distance are related by $d\eta = -d\chi$. A family of such solutions is shown in figure 14.

Unfortunately, if we include dark energy one needs to solve the equations numerically.

At early times things are much simpler, because at redshifts beyond a few we can neglect the dark energy and Ω_k terms. Then

- in the *matter dominated regime* we then have power law solution

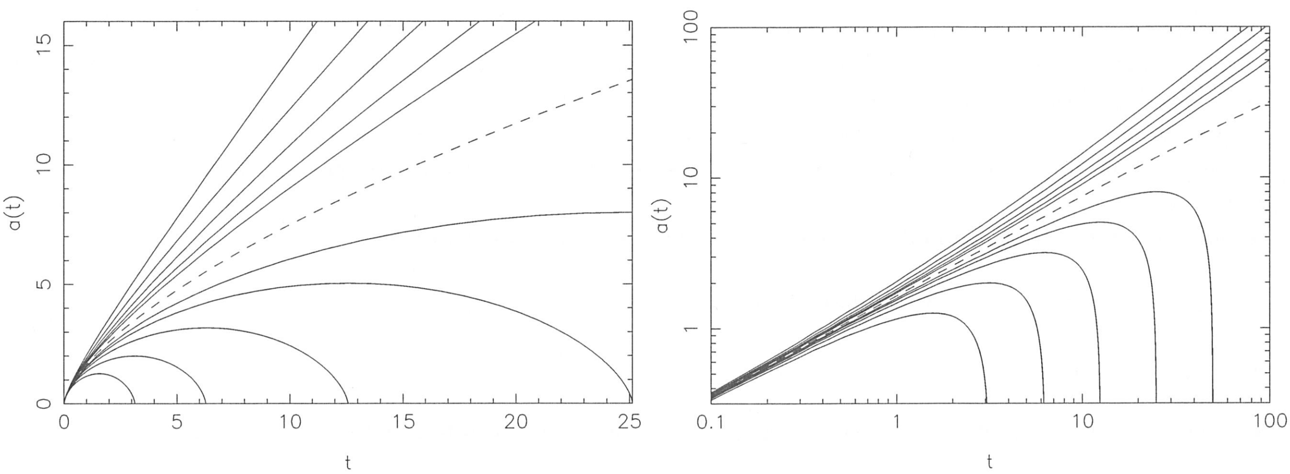

Figure 14: Cycloidal and hyper-cycloidal solutions of the Friedmann equation. These are the same as the solutions for a pebble fired upwards from a compact (Newtonian) mass. At very early times – when the kinetic energy and potential energy are both very large in relation to their difference – the solution is a power law $a \propto t^{2/3}$. The curves represent a sequence of increasing total energy.

  - $\boxed{a \propto \tau^{2/3}}$
  - which can be shown, either by taking the $\eta \ll 1$ limit of the hypercycloidal solution above,
  - or, more usefully, by noting that
    * for a power law expansion $a \propto \tau^\gamma$,
    * the expansion rate goes like $H = \dot{a}/a = \gamma/\tau \propto a^{-\gamma}$
    * but, since $H^2 \propto \rho \propto a^{-3}$, so $H \propto a^{-3/2}$, this requires $\gamma = 2/3$

- and in the *radiation dominated regime* applying the argument above, but with $\rho \propto a^{-4}$ we get $H \propto \sqrt{\rho} \propto a^{-2}$ so $\gamma = -1/2$ and therefore
  - $\boxed{a \propto \tau^{1/2}}$
- we will discuss these more below

Another interesting solution emerges when the universe is strongly Λ-*dominated*, or dominated by a field with $P \simeq -\rho c^2$, which, it seems, will happen in the not so distant future, and which, it is widely suspected, happened in the distant past in the *inflationary era* that preceded the hot big bang.

For a Λ-dominated universe the universe expands exponentially with

$$\boxed{a(\tau) \propto \exp(H\tau)} \tag{83}$$

- with H asymptotically constant
- while maintaining constant energy density
- and thus creating energy out of nothing, hence *"inflation is the ultimate free lunch"* (Guth)

4.3 Interpretation of observations in FRW models

The FLRW models preceded (just), and allowed interpretation of, Hubble's observations. He was measuring recession velocities inferred from the redshift and distances obtained from flux-densities of variable stars – which he was using as *'standard-candles'*.

These models allow prediction of, in addition to flux-density, angular sizes of objects of known size – or *'standard-rulers'* as a function of redshift.

Note that, to test the model, or determine the parameters of a model, we need to have at least two ways of determining the distance to an astronomical object.

Though it should be said that, at the modest distances he was observing, relativistic effects were negligible and a Newtonian interpretation would have been adequate. But, oddly enough, Newtonian cosmology was not well developed or understood at the time.

In FLRW models, light from distant objects is focused by the gravitational lensing effect of intervening matter. This depends on how much mass there is (and the redshift depends on how the universe is expanding).

This focussing is conventionally expressed in terms of *'apparent distances'*

- $\boxed{D_L(z)}$ for flux density (or luminosity L), and
- $\boxed{D_\mathrm{a}(z)}$ for angular size
- both of which, in any specific cosmological model, are computable functions of redshift

these are the answers to the questions:

- *'how for away would a object of known luminosity (size) have to be* <u>*in an empty universe*</u> *in order to have the flux density (angular size) computed in the model from the metric'*

while redshift depends on how much the universe has expanded since the light we see left the objects.

These apparent-distance vs. redshift relations $D_L(z)$ and $D_\mathrm{a}(z)$ can be used in two ways:

1. if we assume the cosmological density parameters are known we get the intrinsic properties from observed ones
2. if we assume the intrinsic properties are known then we determine the cosmological parameters

4.3.1 The cosmological redshift - measurement

We *defined* the redshift above such that

$$1 + z \equiv a_0/a. \tag{84}$$

The reason it is called the redshift is that it is directly observable as a shift in the wavelength of spectral lines

$$\boxed{\lambda_\mathrm{obs}/\lambda_\mathrm{em} = 1 + z} \tag{85}$$

and observed photon energies scale inversely with $1 + z$.

One way to prove this is to use the (covariant form of the) geodesic equation for the time component of the 4-momentum p_0, $dp_0/d\lambda = -\frac{1}{2}g_{\mu\nu,0}p^\mu p^\mu$. For a radial photon ($\theta$, ϕ = constant) this is

$$dp_0/d\lambda = -\tfrac{1}{2}g_{\chi\chi,0}p^\chi p^\chi \tag{86}$$

since g_{00} is time independent.

Now the normalisation condition for the (null) 4-momentum is $g_{\chi\chi}(p^\chi)^2 = -g_{00}(p^0)^2 = -g^{00}p_0^2$ so $p^\chi = -\sqrt{-g^{00}/g_{\chi\chi}}p_0 = -p_0/ac$ (negative since the photon is coming towards us). Since $g_{\chi\chi} = a^2$, its derivative is $g_{\chi\chi,0} = c^{-1}\partial_t a^2 = 2a\dot{a}/c$ and so

$$\frac{dp_0}{d\lambda} = -\frac{a\dot{a}}{c}\frac{p_0^2}{a^2c^2} = -p_0\frac{\dot{a}}{ac}p^0 = -p_0\frac{\dot{a}}{ac}\frac{cdt}{d\lambda} \tag{87}$$

so

$$\frac{dp_0}{p_0} = -\frac{da}{a} \tag{88}$$

which means that p_0, and hence also $p^0 = g^{00}p_0$, varies inversely with the scale factor.

Another way to understand this is by analogy with a standing wave in an expanding cavity (see left panel of figure 15).

One can also derive this by thinking about the Doppler shifts suffered by a photon bouncing repeatedly off the assumed to be steadily receding walls of an expanding cavity.

Weaknesses of this argument are:

- is radiation – e.g. that from a distant transient source – behave in the same way as a standing wave

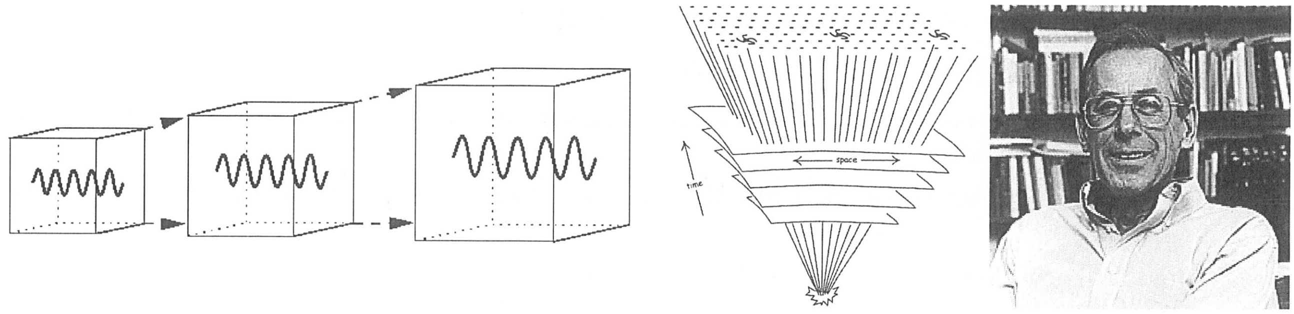

Figure 15: The cosmological redshift can be understood by analogy with standing waves in a cavity (left) for which the wavelength scales with the box size. A rigorous argument (due to Peebles - at right) is to consider the wavelength change as the product of a lot of infinitesimal shifts between a sequence of fundamental observers that the photon passes on its path.

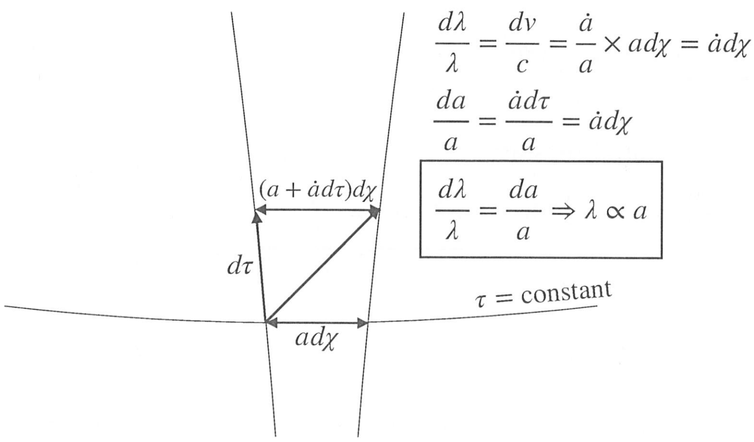

Figure 16: Peebles' argument for the cosmological redshift. We consider two neighbouring fundamental observers, one of whom sends a photon to the other. These observers are in free-fall, so gravity is 'transformed away' (locally at least). So the change in the frequency of the photon is just the 1st order Doppler shift. It follows that the fractional change in the wavelength is equal to the fractional change in their separation and, by extension, the fractional change in the scale factor.

- in reality there are no reflecting walls
- and surely the gravitational redshift has to be involved at some level
- and has been challenged (see papers on 'redshift-remapping')

- a rigorous way to prove this is illustrated in the centre panel of figure 15. One imagines a finite wavelength change $\lambda_{\rm obs}/\lambda_{\rm em}$ as being the product of a set of infinitesimal shifts.
  - by virtue of the fact that these observers are in free fall
  - and that in a locally freely falling frame the effect of gravity is 'transformed away'
  - one can be confident that the only effect is the 1st order Doppler effect $1 + \delta\lambda/\lambda = 1 + \delta v/c = 1 + H\delta x/c = 1 + H\delta\tau$ where $H = \dot{a}/a$ is the expansion rate and δt is the time elapsed as the photon makes its way between the two neighbouring observers
  - but, since $H\delta\tau = (\dot{a}/a)\delta\tau = \delta a/a$, it follows that
  - $\boxed{\delta\lambda/\lambda = \delta a/a}$
  - which we can integrate up to get $\lambda_{\rm obs}/\lambda_{\rm em} = a_{\rm obs}/a_{\rm em}$
  - this would appear to be a serious challenge to proponents of 'redshift-remapping'
- with a high resolution spectrograph, redshifts of galaxies can be measured with great precision
- and even with 'broad-band colours' one can obtain quite good accuracy (though one needs to beware of 'outliers')
- but these require that there be features in the spectrum whose 'rest-frame' wavelength is known

- a counter-example is the cosmic background radiation (CMB)
  - this has a thermal, or 'black-body', spectrum
  - which has the property that in an expanding universe it remains thermal even in the absence of interaction of matter
  - so there is nothing about the CMB that tells us at what redshift it was last interacting with matter
  - but we know that this was at $z \simeq 1100$ from Saha's equation

4.3.2 Conformal distance-redshift relation

Having established how the redshift can be measured, we now want to relate this to conformal distance. Not that this is our ultimate goal – which is to relate redshift to apparent (luminosity and angular diameter) distances.

From the metric, for light, which follows null trajectories: $ds = 0$, propagating radially from the origin (i.e. with θ, ϕ fixed, so $d\theta = d\phi = 0$) we have

$$\cancel{ds}^2 = \underbrace{-c^2 d\tau^2 + a(\tau)^2 (d\chi^2}_{d\chi = -c d\tau / a} + S_k(\chi)(\cancel{d\theta}^2 + \sin^2\theta \cancel{d\phi}^2) \tag{89}$$

and a useful chain of relations follows from the differential redshift:

$$dz = \underbrace{d(1+z) = a_0 da^{-1}}_{1+z \equiv a_0/a} = -\frac{a_0}{a^2} da = -\frac{a_0}{a^2}\dot{a} d\tau = \underbrace{-\frac{a_0}{a} H d\tau = \frac{a_0}{c} H d\chi}_{c d\tau = -a d\chi} \tag{90}$$

from which we can extract

$$d\tau = \frac{dz}{(1+z)H(z)} \tag{91}$$

which can be integrated to get the *'lookback time'*, or the age of the universe when the photons we see were emitted, and

$$a_0 d\chi = \frac{c dz}{H(z)}. \tag{92}$$

Integrating the latter we obtain the present epoch proper distance to redshift z

$$\boxed{a_0 \chi(z) = c \int_0^z \frac{dz}{H(z)}} \tag{93}$$

or, using the expression obtained above for $H(z)$

$$\boxed{a_0 \chi(z) = a_0 \int_0^{\chi(z)} d\chi = \frac{c}{H_0} \int_0^z \frac{dz}{[\Omega_{\mathrm{m}0}(1+z)^3 + \Omega_{\mathrm{r}0}(1+z)^4 + \Omega_\Lambda + \Omega_{k0}(1+z)^2]^{1/2}}} \tag{94}$$

which is a nasty integral, but something one can readily evaluate numerically given as input the values of the density parameters.

It should be noted that the integral here converges if there is any matter or radiation in the universe. In a completely empty universe (so $\Omega_{k0} = 1$) the conformal distance diverges logarithmically and in a universe with only dark energy and $\Omega_{k0} = 0$ the divergence is stronger. But these are not realistic options.

– The Einstein-de Sitter model

As an illustrative example – albeit a not an entirely realistic one – for the so-called *'Einstein-de Sitter'* model: $\Omega_m = 1$, all others zero, we obtain

$$a_0 \chi(z) = (2c/H_0)[1 - (1+z)^{-1/2}] \tag{95}$$

regarding which we note the following:

- in this spatially flat model ($\Omega_k = 0$), as in any spatially flat model, the present day scale factor a_0 (being the radius of curvature) is formally infinite, while $\chi(z)$ is formally zero, but the product $a_0\chi$ is finite
- the proper distance to infinite redshift – the *horizon* – is finite
- the distance grows linearly at low-z: $a_0\chi(z \ll 1) \simeq cz/H_0$
- but this growth tapers off:
  - by $z = 3$ we are half-way to the horizon
  - by $z \simeq 1000$ (the redshift at which the photons of the CMB were last scattered) we are about 97% of the way to the horizon (see figure 17)

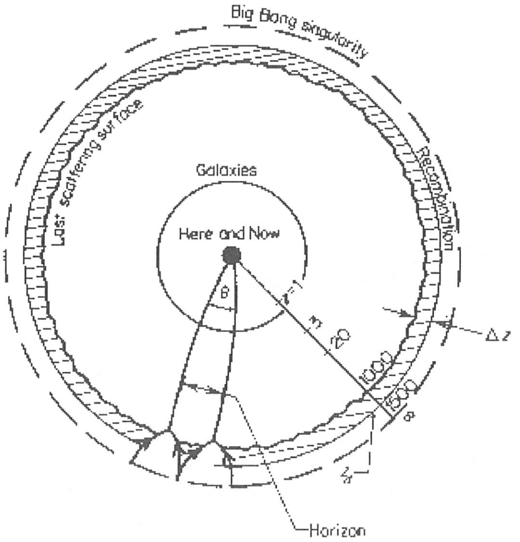

Figure 17: An equatorial slice through our universe showing the surfaces of constant redshift in a plot where radius is conformal distance χ. Galaxies can be seen out to $z \sim 10$, which is a good fraction of the volume within the entire region we can observer. Note that there is no 'edge' to the universe in this model. The density of matter, galaxies etc. is assumed to extend without limit – though we will only ever see galaxies below the redshift at which they formed (thought to be around $z \sim 20$). The horizon – the dashed circle – is simply the limit imposed by the fact that there is a maximum conformal distance that any information can have propagated in the age of the universe. The arrows labelled 'Horizon' in this plot indicate the horizon size back at $z \simeq 1000$. Parts of the sky with separation bigger than this were not, in the big-bang model, in causal contact with one-another when the photons were released, yet they have almost identical temperatures.

We can use (
refeq:ConformalDistanceFromRedshift) to compute the distance $a_0\chi(z)$ – at the present epoch – to a source with some measured redshift (assuming we know H_0 and the density parameters). This is often quoted in news articles (usually in light years), but it is not very useful. It is the integral of the distance on the hypersurface $\tau = \tau_0$, but the photons came to us on our *past light cone*. More useful are the *apparent distances* that depend also on the cosmological parameters and which tell us, assuming we know the intrinsic size or luminosity of an object, how far away would it have to be – in a fictitious empty non-expanding universe – to have the observed properties.

4.3.3 The angular diameter distance

Let's suppose we are observing a galaxy or some other extended object at redshift $z_{\rm em}$ and it subtends an angle on the sky $d\theta$

- how can we infer from this the physical linear size of the galaxy dl?

to do this, we consider two emission events that happened at the same cosmic time, so $d\tau = 0$, and at the same distance from us, so $d\chi = 0$ an let them lie in the equatorial plane, so $d\phi = 0$.

The metric then tells us that the proper size of the galaxy is $dl = ds$, where

$$ds^2 = -c^2 \cancel{d\tau^2} + a(\tau)^2(\cancel{d\chi^2} + S_k(\chi)(d\theta^2 + \sin^2\theta \cancel{d\phi^2}) \tag{96}$$

so $dl = a(z_{\rm em})S_k(\chi(z_{\rm em}))d\theta$.

The factor multiplying $d\theta$ is, by definition, the angular diameter distance, so

$$\boxed{D_{\rm a}(z) = a(\tau(z))S_k(\chi(z))} \tag{97}$$

since, in empty space, an object of proper size dl subtends an angle $d\theta = dl/D$.

If we assume a spatially flat universe ($k = 0$), which is believed to be a good approximation in reality, we have $S_k(\chi) = \chi$ so the angular diameter distance is $D_{\rm a}(z) = a_{\rm em}\chi(z_{\rm em}) = (a_0/(1+z))\chi(z_{\rm em})$ or

$$\boxed{D_{\rm a}(z) = \frac{c}{1+z}\int_0^z \frac{dz}{H(z)}.} \tag{98}$$

This is a very powerful result. It turns out there is a rather accurate 'standard ruler' known as the *'baryon acoustic oscillation'* (or BAO) scale, which is a feature imprinted in the spatial distribution of galaxies and which enables a measure of $D_a(z)$ and hence a powerful test of cosmology (see figure 19). This reinforces the evidence for dark energy and helps determine the cosmological parameters.

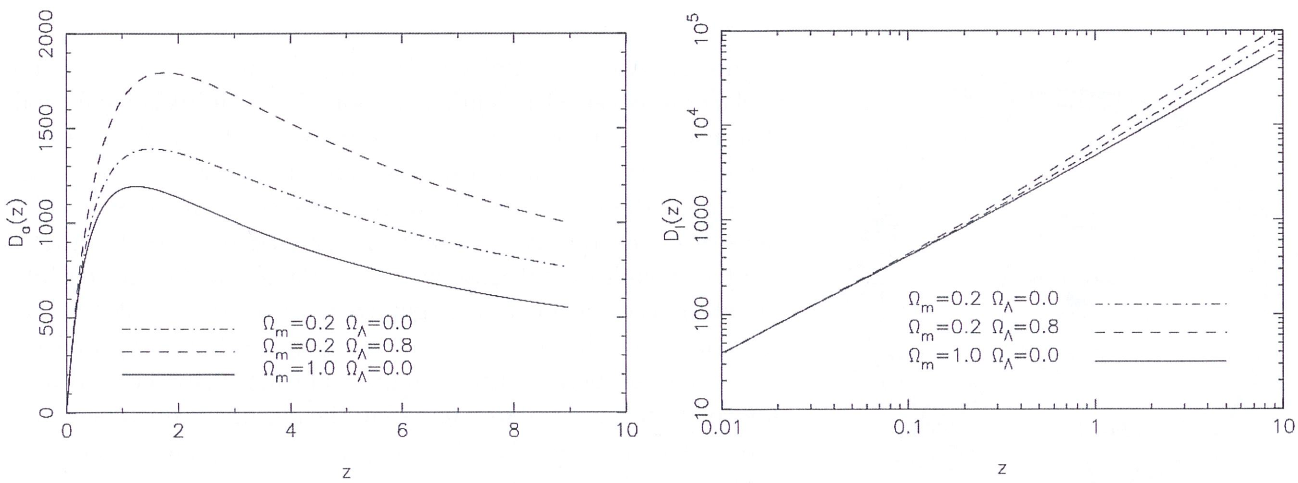

Figure 18: Angular diameter distance (left) and luminosity distance (right) as a function of redshift for various cosmological models.

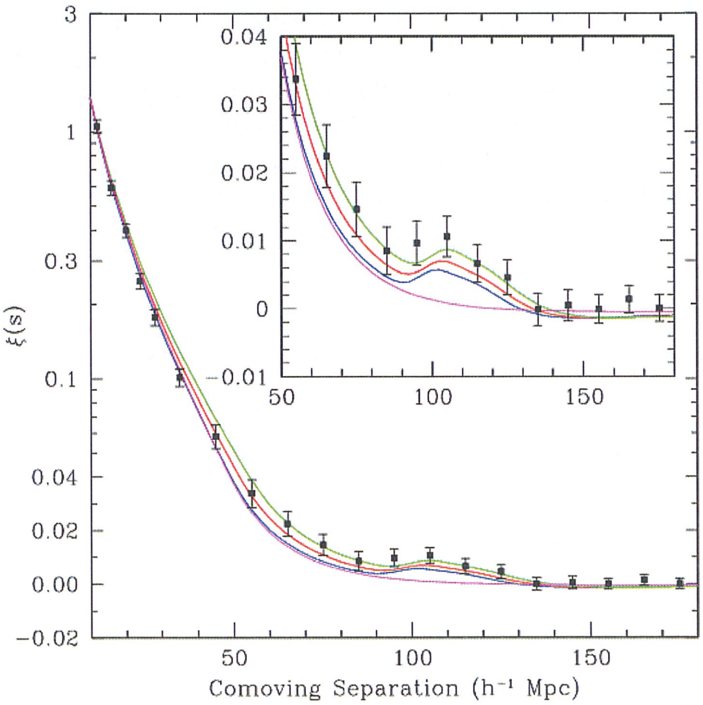

Figure 19: Baryon acoustic oscillations. In the conventional model for the origin of cosmological structure, the 'seeds' of structure were laid down during inflation in the very early universe, with a nearly scale-invariant spectrum. During the period immediately before the plasma reionized these triggered sound waves in the plasma-radiation fluid. This imprinted a feature in the spectrum of density fluctuations that emerged and later developed into fluctuations in the observed large-scale structure traced by galaxies. The feature is rather weak, but it is highly valuable as its scale is predicted from CMB observations (it is essentially the 'sound-horizon'; the product of the sound speed and the age of the universe at the decoupling epoch). Massive redshift surveys were able to measure this feature by means of the 'two-point' function characterising the galaxy distribution at left. This provides a powerful constraint on the cosmological model.

4.3.4 The luminosity distance

- An analogous apparent distance D_L can be defined that uses flux density F of 'standard candles' of luminosity L
  - flat space : $F = L/4\pi D_L^2 \quad \Rightarrow \quad \boxed{D_L(z) \equiv \sqrt{L/4\pi F}}$
- to calculate $D_L(z)$ we consider a source located at $\chi = 0$
- and co-moving observers on a spherical shell at distance χ (see figure 20)

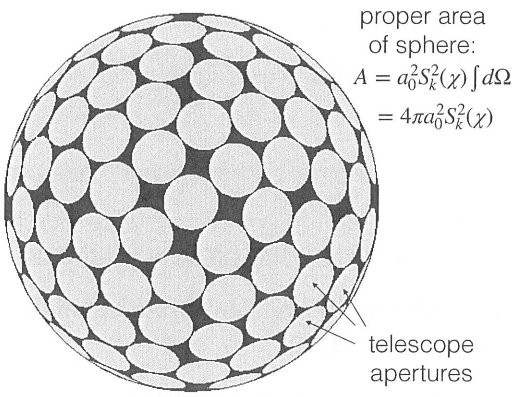

Figure 20: To calculate the luminosity distance we consider a source at the origin of our coordinate system $\chi = 0$ and we consider a sphere at distance χ covered with observers, each of whom would observe the source to have the same flux density and redshift. If the source emits N photons of frequency ν per period $\tau = \nu^{-1}$ (so the proper luminosity is $L = Nh\nu^2$) then conservation of photons implies that there must be N photons of energy $h\nu/(1+z)$ crossing the sphere per red-shifted period $\tau = (1+z)\nu^{-1}$. The energy flux density is therefore $F = (L/(1+z)^2)/A = L/4\pi a_0^2 S_k(\chi)^2(1+z)^2$ and so the luminosity distance is $D_L(z) \equiv \sqrt{L/4\pi F} = a_0 S_k(\chi(z))(1+z)$.

- let the source emit N photons of frequency $\nu_{\rm em}$ per (rest-frame) period $\tau_{\rm em} = 1/\nu_{\rm em}$
- that means $L_{\rm em} = Nh\nu_{\rm em}/\tau_{\rm em} = Nh\nu_{\rm em}^2$
- conservation of photons implies that N photons cross the shell per red-shifted period $\tau_{\rm obs} = (1+z)\tau_{\rm em}$
- and these photons have energy: $h\nu_{\rm obs} = h\nu_{em}/(1+z)$
- so the *energy flux* (energy per unit time) across the surface is $L_{\rm obs} = Nh\nu_{\rm obs}/\tau_{\rm obs} = Nh\nu_{\rm obs}^2$ or
  - $\boxed{L_{\rm obs} = L_{\rm em}/(1+z)^2}$
- but area is $A = 4\pi a_0^2 S_k^2(\chi) = 4\pi((1+z)a_{\rm em})^2 S_k^2(\chi)$
- so the *energy flux density* F (energy per unit time per unit area) is $F = L_{\rm obs}/A = (1+z)^{-4}/4\pi S_k^2(\chi)a(\tau_{\rm em})^2$ and hence
  - $\boxed{D_L = a_{\rm em} S_k(\chi_{\rm em})(1+z)^2}$
- or, comparing with the angular diameter distance, $D_{\rm a} = a_{\rm em} S_k(\chi_{\rm em})$ we have
  - $\boxed{D_L = D_{\rm a}(1+z)^2}$
  - which is called *Etherington's reciprocity relation*
- aside:
  - if the source has size dl then it subtends a solid angle $d\Omega \sim dl^2/D_{\rm a}^2$
  - while the intensity is $I \sim F/d\Omega \sim (L/D_L^2)/d\Omega$
  - this, for a given object (fixed L, dl) the surface brightness must vary with redshift as $I \propto D_{\rm a}^2/D_L^2 \propto (1+z)^{-4}$
  - so the bolometric intensity (or surface brightness) suffers a $(1+z)^{-4}$ *'surface-brightness dimming'*
  - this is consistent with Liouville's theorem, which says $I_\nu/\nu^3 =$ constant along any ray since if we integrate this we get a bolometric surface brightness $I = \int d\nu I_\nu \propto \nu^4 \propto (1+z)^{-4}$
- taking the logarithm of $D_L/10{\rm pc}$ (and multiplying by 5) we get the predicted *distance modulus* $m - M$ which can be compared with data for 1a SN (see figure 21)

4.3.5 The deceleration parameter

- FRW models predict linear relation $D \propto z$– i.e. Hubble's law – for $z \ll 1$
  - all observers perceive themselves to be at the 'centre of the universe'
  - analogous to ants on an expanding balloon
  - key observable is 'Hubble parameter' H_0 – essentially the inverse of the age of the universe
- going to small but finite z we start to probe departures from linearity
  - at lowest order this is parameterised by the *deceleration parameter*
    - $\boxed{q_0 \equiv (-a\ddot{a}/\dot{a}^2)_0}$
  - This led Alan Sandage, who studied with Hubble, to famously state that *'cosmology is the search for two numbers'*, these being
    1. how fast the universe is expanding (H_0)
    2. and how fast that expansion is slowing down (q_0)
- in Sandage's time, even the first was quite uncertain and the second largely a matter of speculation
- this changed at the end of the '90s when two groups (led by Saul Perlmutter and by Brian Schmidt) obtained the famous *'type 1a supernova Hubble diagram'* shown in figure 21
  - this leap forward was the result of careful 'standardisation' of the 'candles' in question
  - such supernovae taken as whole actually having a range of intrinsic luminosities
  - but that variation, it turns out, is a function of colour and of the duration of the supernova and so can be corrected for

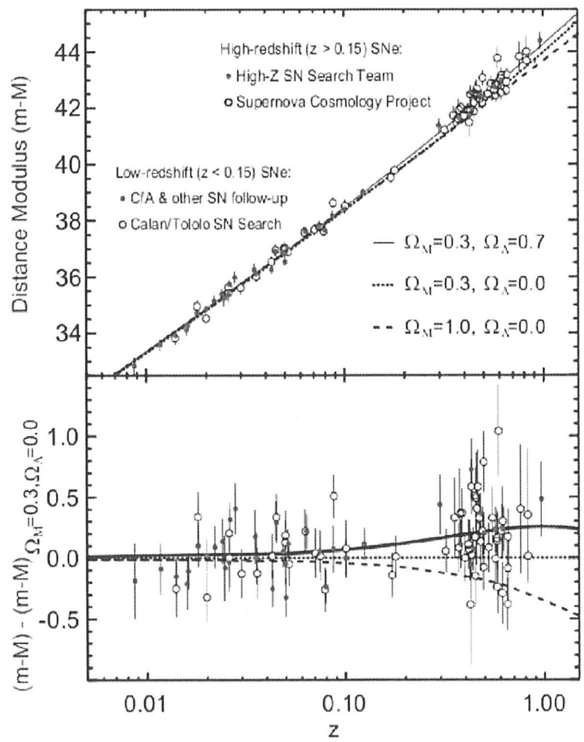

Figure 21: Hubble diagram for type 1a supernovae (contains data shown earlier, but extended here to higher redshift sources). The background to these Nobel prize winning observations is the following: By the '80s there was strong evidence that there was significant non-baryonic dark matter: much more than the roughly 5% of critical density for normal matter allowed by big-bang nucleosynthesis. At the same time, the idea that inflation predicted that the universe should have closure density and be spatially flat had firmly taken hold. For a while the CDM model with $\Omega_{\rm CDM} \simeq 1$ seemed the natural solution. But there were various problems: the predicted age of the universe was uncomfortably short; the dynamical evidence was for $\Omega_{\rm m} \simeq 0.3$, not 1; and evolution of galaxy clusters seemed slower than predicted. At the same time observations of the CMB were very hard to reconcile with an open universe with $\Omega_{\rm m} \simeq 0.3$ as the negative curvature would make the predicted scale of the ripples too small. Adding Λ to the cosmic mix, while repugnant, was becoming widely promoted. The clincher came with the high-z SN data that indicated that the universe was accelerating.

- Figure 21 shows that the supernovae at high-z have a greater apparent distance than predicted in cosmological models containing only normal matter and the observed distance is better fit by models with a cosmological constant Λ (or with 'dark energy') with $\Omega_\Lambda \simeq 0.7$.
  - the present-epoch deceleration parameter in these models is *negative*:

  - the dark energy is causing the expansion rate to *increase* with time
  - i.e. we inhabit an *accelerating universe*

- why acceleration causes an increase in apparent distance can be understood as follows
  - all distances (comoving/conformal, luminosity, angular diameter) involve $\chi(z) = (c/a_0)\int dz/H(z)$
  - and in flat models (a class to which we believe our universe closely approximates) they are simply proportional to this times factors of $1+z$
  - imagine we compare two models
    1. a 'fiducial' model with some expansion history $H_1(z)$
    2. a 'relatively accelerating' model with idential expansion history for z below some redshift $z_\star$ but $H_2(z) < H_1(z)$ for $z > z_\star$
  - evidently, the relatively accelerating model will have greater apparent distances for sources at $z > z_\star$

4.4 The closed FLRW models

The closed model is finite, yet has no boundary. However, at least if we restrict attention to zero-pressure equation of state, we are free to take only a finite part of the total solution $\chi < \chi_{\rm max}$. This is a spherically symmetric mass configuration, and so should match onto the Schwarzschild solution for a point mass m, for which the space-time metric is (in units such that $c = G = 1$)

$$-d\tau^2 = -\left(1 - \frac{2m}{r}\right)dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \qquad (99)$$

Comparing the angular part of the metric it is apparent that the Schwarzschild radial coordinate r and the FLRW 'development angle' χ are related by $r = a\sin\chi$. Now a particle at the edge of the FLRW model can equally be considered to be a radially moving test particle in the Schwarzschild geometry. We found that in Schwarzschild geometry, the normalisation of the 4-velocity for a radially moving particle implies

$$\left(\frac{dr}{d\tau}\right)^2 = \frac{2Gm}{r} + \text{constant}. \qquad (100)$$

Compare this with the energy equation

$$\left(\frac{da}{d\tau}\right)^2 = \frac{2GM}{a} + \text{constant} \qquad (101)$$

where we have defined the mass parameter $M = 4\pi\rho a^3/3$. With $r = a\sin\chi$ this implies that the Schwarzschild mass parameter is

$$m = M\sin^3\chi. \qquad (102)$$

This is interesting. For $\chi \ll 1$, the mass increases as χ^3 as expected. However, the mass is maximized for a model with a development angle of $\pi/2$, or half of the complete closed model. If we take a larger development angle, and therefore include more proper-mass, the Schwarzschild mass parameter decreases. To the outside world, this positive addition of proper mass has negative total energy. This means that the negative gravitational potential energy outweighs the rest-mass energy. The gravitating mass shrinks to zero as $\chi \to \pi$. Evidently a nearly complete closed model with $\chi = \pi - \epsilon$ looks, to the outside world, like a very low mass, that of a much smaller closed model section with $\chi = \epsilon$.

The total energy of a complete closed universe is therefore zero. Zel'dovich, and many others subsequently, have argued that this is therefore a natural choice of world model if, for instance, one imagines that the Universe is created by some kind of quantum mechanical tunnelling event. To be consistent with the apparent flatness of the Universe today one would need to assume that the curvature scale has been inflated to be much larger than the present apparent horizon size.

It is interesting to compare the external gravitational mass with the total proper mass. The volume element of the parallelepiped with legs $d\chi$, $d\theta$, $d\phi$ is

$$d^3x = (ad\chi) \times (a\sin\chi d\theta) \times (a\sin\chi\sin\theta d\phi), \qquad (103)$$

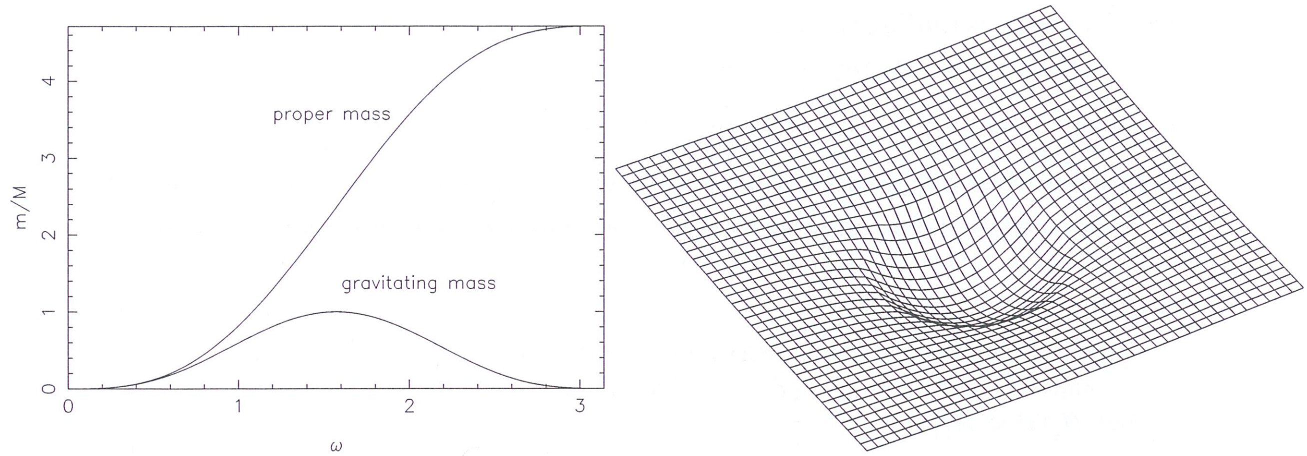

Figure 22: On the left, the proper-mass and gravitational mass for a partial closed FRW cosmology surrounded by Schwarzschild geometry are plotted against the development angle $\omega = \chi$. On the right is shown an embedding diagram. The interior is part of a sphere and the outside is like a trumpet horn. Here the 'development angle' χ is less than $\pi/2$. For $\chi > \pi/2$ we have more than half of the total FLRW model and the radius r is decreasing with increasing χ. It is still possible to match on to an exterior Schwarzschild geometry, but the embedding diagram then has a 'throat'.

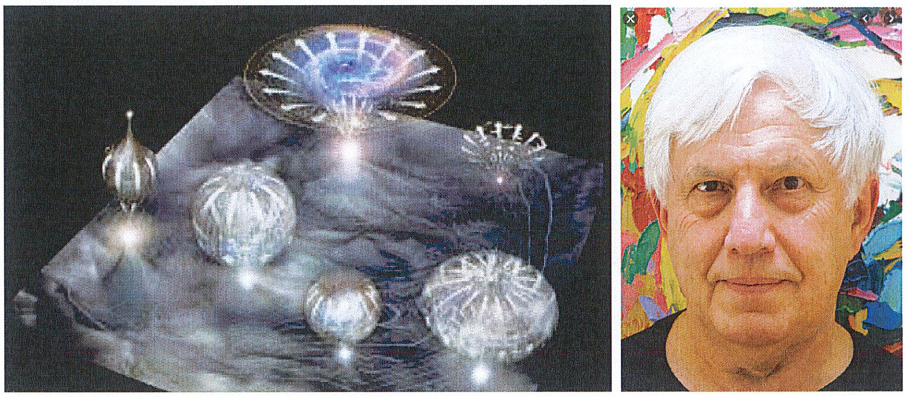

Figure 23: The fact that a – potentially large – closed universe can be matched onto an external universe is the physics behind the cartoon shown at the left which is meant to indicate how in the early universe a multitude of universes could 'herniate' from a parent universe – or from each other in Andrei-Linde's 'chaotic inflation' models.

so the total mass interior to χ is

$$M_{\rm proper} = \rho a^3 \int\limits_0^\chi d\chi \, \sin^2\chi \int d\theta \, \sin\theta \int d\phi = \frac{3}{2} M \left[\chi - \frac{\sin 2\chi}{2} \right] \tag{104}$$

The gravitational mass (102) and proper mass (104) are shown on the left in figure 22.

One can make an embedding diagram for this combined FLRW + Schwarzschild space-time. This is shown pm the right in figure 22.

These partial closed FRW models start from a singularity of infinite density and then expand, passing through the Schwarzschild radius $r = 2Gm/c^2$. With $r = a \sin\chi$, $m = M \sin^3\chi$, and $a = M(1 - \cos\eta)$, particles on the exterior cross the Schwarzschild radius at conformal time η when $1 - \cos\eta = 2\sin^2\chi$. For $\chi \ll 1$, this occurs when $\eta = 2\chi$. Such solutions spend the great majority of their time outside the Schwarzschild radius. For the case $\chi = \pi/2$ — i.e. half of the complete solution — the exterior particles just reach the Schwarzschild radius. It may seem strange that the matter in these models can expand from within the Schwarzschild radius, but this is indeed the case. If one considers only the collapsing phase of these models then one has the classic model for black-hole formation as developed by Oppenheimer and Snyder. The spherical mass collapses to a point, and photons leaving the surface can only escape to infinity if they embark on their journey while the radius exceeds the Schwarzschild radius. The expanding phase of these models is just the time reverse of such models; what we have is a 'white-hole' solution The initial singularity is visible to the outside world (eventually) just as photons from the outside can fall in to the final singularity.

A The Friedmann equation from the Einstein field equation

It is convenient to use the (τ, r, θ, ϕ) metric in the form

$$ds^2 = -d\tau^2 + a(\tau)^2\left(\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right) \tag{105}$$

from which we find that the non-vanishing Christoffel symbols are

$$\begin{array}{ccc} \Gamma^0{}_{jk} = \frac{\dot{a}}{a}g_{ij} & \Gamma^j{}_{0k} = \frac{\dot{a}}{a}\delta^i_k & \Gamma^r{}_{rr} = \frac{kr}{1-kr^2} \\ \Gamma^r{}_{\theta\theta} = -r(1-kr^2) & & \Gamma^r{}_{\phi\phi} = -r(1-kr^2)\sin^2\theta \\ \Gamma^\theta{}_{r\theta} = \frac{1}{r} & \Gamma^\theta{}_{\phi\phi} = \sin\theta\cos\theta & \Gamma^\phi{}_{\theta\phi} = \cot\theta \end{array} \tag{106}$$

The non-vanishing components of the Ricci tensor are then found from

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu} \tag{107}$$

and

$$R^\alpha{}_{\mu\beta\nu} = \Gamma^\alpha{}_{\mu\nu,\beta} - \Gamma^\alpha{}_{\mu\beta,\nu} + \Gamma^\alpha{}_{\gamma\beta}\Gamma^\gamma{}_{\mu\nu} - \Gamma^\alpha{}_{\gamma\nu}\Gamma^\gamma{}_{\mu\beta} \tag{108}$$

to be

$$\begin{aligned} R_{\tau\tau} &= -3\frac{\ddot{a}}{a} \\ R_{rr} &= (a\ddot{a} + 2\dot{a}^2 + 2k)/(1-kr^2) \\ R_{\theta\theta} &= (a\ddot{a} + 2\dot{a}^2 + 2k)r^2 \\ R_{\phi\phi} &= (a\ddot{a} + 2\dot{a}^2 + 2k)r^2\sin^2\theta \end{aligned} \tag{109}$$

from which the Ricci scalar is

$$R = g^{\mu\nu}R_{\mu\nu} = 6(a\ddot{a} + \dot{a}^2 + k)/a^2. \tag{110}$$

The stress energy tensor for a homogeneous universe containing dust or radiation can be taken to be that of a perfect fluid

$$T_{\mu\nu} = (\rho + P)U^\mu U^\nu + g^{\mu\nu}P \tag{111}$$

where the 4-velocity is that of a *'fundamental observer'* who is moving with the fluid and has 4-velocity $U^\mu = (1, 0, 0, 0)$ (this is properly normalised as $g_{\tau\tau} = -1$).

The Einstein field equations can be written as

$$R_{\mu\nu} = 8\pi G(T_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}T) \tag{112}$$

The $\tau\tau$ component of which is the *acceleration equation*

$$\boxed{\ddot{a} = -\frac{4\pi G}{3}(\rho + 3P)a} \tag{113}$$

which is the same as what we found for the geodesic deviation equation in the centre of a star.

Any one of the spatial equations gives $a\ddot{a} + 2\dot{a}^2 + 2k = 4\pi G(\rho - P)$ which, with the acceleration equation gives the *energy equation*

$$\boxed{\dot{a}^2 = \frac{8\pi G}{3}\rho a^2 - k.} \tag{114}$$

which are the Friedmann equations.

Differentiating the latter and combining with the former gives the *continuity equation*

$$\boxed{\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + P)} \tag{115}$$

which can also be obtained as the $\mu = \tau$ component of the energy momentum conservation law $T^{\nu\mu}{}_{;\nu} = 0$ and is, in essence, the 1st law of thermodynamics as it says that the rate at which a volume element is losing mass-energy is equal to the PdV work it is doing in the process of expansion.

One could work backwards from the latter, using the acceleration equation to obtain the energy equation. But while that would tell you that $\dot{a}^2 - (8\pi G/3)\rho a^2$ is constant it would not pin down the value of k. One can, however, appeal to the Milne model, which is the limiting case of the FRW models as $\rho \to 0$ and for which $k = -1$ and $a = \tau$ to make the connection between the curvature constant and the (minus) sign of the 'total energy' $\dot{a}^2 - (8\pi G/3)\rho a^2$.

B The local flatness theorem

If we make a 2nd order Taylor series expansion:

$$x^{\alpha'}(x^\alpha) = \Lambda^{\alpha'}{}_\alpha x^\alpha + \tfrac{1}{2}\Lambda^{\alpha'}{}_{\alpha\beta}x^\alpha x^\beta \tag{116}$$

where $\Lambda^{\alpha'}{}_{\alpha\beta} = x^{\alpha'}{}_{,\alpha\beta}(\vec{x}_0)$, then the infinitesimal primed-frame displacement

$$dx^{\alpha'} \equiv x^{\alpha'}(x^\alpha + dx^\alpha) - x^{\alpha'}(x^\alpha) \tag{117}$$

corresponding to an un-primed dx^α is

$$dx^{\alpha'} = (\Lambda^{\alpha'}{}_\alpha + \Lambda^{\alpha'}{}_{\alpha\beta}x^\beta)dx^\alpha \tag{118}$$

so we have a linear transformation, as before, but one where the matrix effecting the transformation depends on position.

In order to express the line element in terms of the primed differentials we would like to turn this around and write the dx^α (that appears in ds^2) as some matrix times $dx^{\alpha'}$. To do this, we first multiply by $\Lambda^\gamma{}_{\alpha'}$ to obtain

$$\Lambda^\gamma{}_{\alpha'}dx^{\alpha'} = (\delta^\gamma_\alpha + \Lambda^\gamma{}_{\alpha'}\Lambda^{\alpha'}{}_{\alpha\beta}x^\beta)dx^\alpha \tag{119}$$

and then observe that, at linear order in x^β, the inverse of the matrix $M^\gamma{}_\alpha = \delta^\gamma_\alpha + \Lambda^\gamma{}_{\alpha'}\Lambda^{\alpha'}{}_{\alpha\beta}x^\beta$ appearing on the right is $(\mathbf{M}^{-1})^\alpha{}_\gamma = \delta^\alpha_\gamma - \Lambda^\alpha{}_{\alpha'}\Lambda^{\alpha'}{}_{\gamma\beta}x^\beta$. So, multiplying the above by this gives the un-primed in terms of primed differential displacement as

$$dx^\alpha = (\delta^\alpha_\gamma - \Lambda^\alpha{}_{\alpha'}\Lambda^{\alpha'}{}_{\gamma\beta}x^\beta + \ldots)\Lambda^\gamma{}_{\mu'}dx^{\mu'}. \tag{120}$$

And in terms of this, with judicious choice of dummy variables, we can write the line element $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$ as

$$\begin{aligned} ds^2 &= \overbrace{\left(g_{\alpha\beta} + g_{\alpha\beta,\gamma}x^\gamma\right)}^{g_{\alpha\beta}(\vec{x})}\overbrace{\left(\delta^\alpha_\mu - \Lambda^\alpha{}_{\alpha'}\Lambda^{\alpha'}{}_{\mu\sigma}x^\sigma\right)\Lambda^\mu{}_{\mu'}dx^{\mu'}}^{dx^\alpha}\overbrace{\left(\delta^\beta_\nu - \Lambda^\beta{}_{\beta'}\Lambda^{\beta'}{}_{\nu\tau}x^\tau\right)\Lambda^\nu{}_{\nu'}dx^{\nu'}}^{dx^\beta} \\ &= \Lambda^\mu{}_{\mu'}\Lambda^\nu{}_{\nu'}dx^{\mu'}dx^{\nu'}\left[g_{\mu\nu} + x^\gamma\left\{g_{\mu\nu,\gamma} - g_{\mu\beta}\Lambda^\beta{}_{\alpha'}\Lambda^{\alpha'}{}_{\nu\gamma} - g_{\nu\beta}\Lambda^\beta{}_{\alpha'}\Lambda^{\alpha'}{}_{\mu\gamma}\right\} + \ldots\right]. \end{aligned} \tag{121}$$

Thus if, given the curvilinear frame metric components $g_{\mu\nu}$ and their derivatives $g_{\mu\nu,\delta}$ at $\vec{x}_0$, we can find a set of transformation coefficients $\Lambda^{\alpha'}{}_{\mu\nu}$ that make the quantity in parentheses $\{\ldots\}$ vanish then we will not only have $g_{\mu'\nu'} = \delta_{\mu'\nu'}$ at $\vec{x}_0$ but its derivatives $g_{\mu'\nu',\gamma'}$ will vanish there also.

But we know how to solve $\{\ldots\} = 0$. If we define

$$\Gamma^\beta{}_{\nu\gamma} \equiv \Lambda^\beta{}_{\alpha'}\Lambda^{\alpha'}{}_{\nu\gamma} = \frac{\partial x^\beta}{\partial x^{\alpha'}}\frac{\partial^2 x^{\alpha'}}{\partial x^\nu \partial x^\gamma} \tag{122}$$

in terms of which $\Lambda^{\alpha'}{}_{\nu\gamma} = \Lambda^{\alpha'}{}_\beta\Gamma^\beta{}_{\nu\gamma}$, then the equation we need to solve is

$$g_{\mu\nu,\delta} - \Gamma^\beta{}_{\nu\gamma}g_{\mu\beta} - \Gamma^\beta{}_{\mu\gamma}g_{\nu\beta} = 0. \tag{123}$$

But we see that (122) is none other than the formula for the Christoffel symbols, and (123) is the statement that the covariant derivative of $\mathbf{g}$ vanishes: $g_{\mu\nu;\delta} = 0$. And the explicit solution for the Christoffels in terms of the metric (rather than in terms of the transformation matrices) is, as usual,

$$\Gamma^\beta{}_{\mu\nu} = \tfrac{1}{2}g^{\beta\gamma}(g_{\gamma\mu,\nu} + g_{\gamma\mu,\nu} - g_{\mu\nu,\gamma}). \tag{124}$$

This is what is meant by *local flatness* and what we have shown above is often called the *local flatness theorem*. Close to some point $\vec{x}_0$ – which was arbitrary, we could have chosen any point as the origin – we can always find primed frame coordinates – in fact a family of such frames, as they can be rotated with respect to each other – such that the metric is exactly Euclidean at $\vec{x}_0$ and where any corrections are at most of second order in distance from $\vec{x}_0$. This has the implication that, if we find the equations of motion of a curve of extremal length – i.e. a curve for which $\delta\int d\lambda L(x^{\alpha'}, \dot{x}^{\alpha'}) = 0$ with $L(x^{\alpha'}, \dot{x}^{\alpha'}) = \sqrt{-g_{\alpha'\beta'}(x^{\alpha'})\dot{x}^{\alpha'}\dot{x}^{\beta'}}$ and where $\dot{x}^{\alpha'} = dx^{\alpha'}/d\lambda$ and the parameter λ is chosen to be 'affine', so it measures distance along the path – then close to $\vec{x}_0$ the 'generalised force' $\partial L/\partial x^{\gamma'} = -\frac{1}{2}g_{\alpha'\beta',\gamma'}(x^{\alpha'})\dot{x}^{\alpha'}\dot{x}^{\beta'}/L$ vanishes, and the Euler-Lagrange equations are $\ddot{x}^{\gamma'} = 0$; the equation of a straight path in Euclidean space.

At no point in the above did we invoke any fictitious, perfectly flat tangent space. The primed coordinate system is only locally flat; there are non-vanishing second and higher order derivatives of the metric. We have shown how, given the metric measured in some arbitrary coordinate system, we can find the coefficients ${\Lambda^{\alpha'}}_{\nu\gamma}$ of a quadratic transformation (116) that makes the 1st derivatives of the metric in the new frame vanish. As we will shortly show, this cannot, in general, be extended to the 2nd derivatives of the primed frame metric. The reason for this is very simple; there are not enough parameters in a cubic transformation to fix all of the 2nd derivates of the primed frame metric.

C Milne coordinates

Rindler space-time is Minkowski space as viewed from a coordinate system with spatial coordinates tied to accelerated observers. The lines of constant X-coordinate were lines of fixed proper distance from the origin of Minkowski space and lines of constant T coordinate were orthogonal to these and were actually geodesics (straight lines) in Minkowski space. This provided a foliation of space-time, but only contained those regions with space-like separation from the origin.

An alternative, and mathematically very similar, coordinate system is that proposed by Milne in 1933 [?] as way to help visualise and understand the geometry of Friedmann-Lemaitre-Robertson-Walker (FLRW) cosmological models. This is again Minkowski space-time, but viewed from the perspective of a set of expanding observers; specifically a family of inertial observers who emerge from an explosion at $\vec{x}_0 \to (ct_0, \mathbf{x}_0)$ – which is arbitrary, and which we take, for simplicity, to be the origin – with all velocities $\mathbf{v}$ with $|\mathbf{v}| < c$. This is illustrated in figure 24. In the *Milne model* events on a surface of constant time coordinate are at constant proper time from the origin. And lines orthogonal to these – the particle trajectories – are geodesics. This coordinate system foliates Minkowski space, but only contains the region in which events have a time-like separation from the origin. It is therefore complementary to Rindler space-time as it covers the region of space-time that was not covered by the *Rindler wedge*.

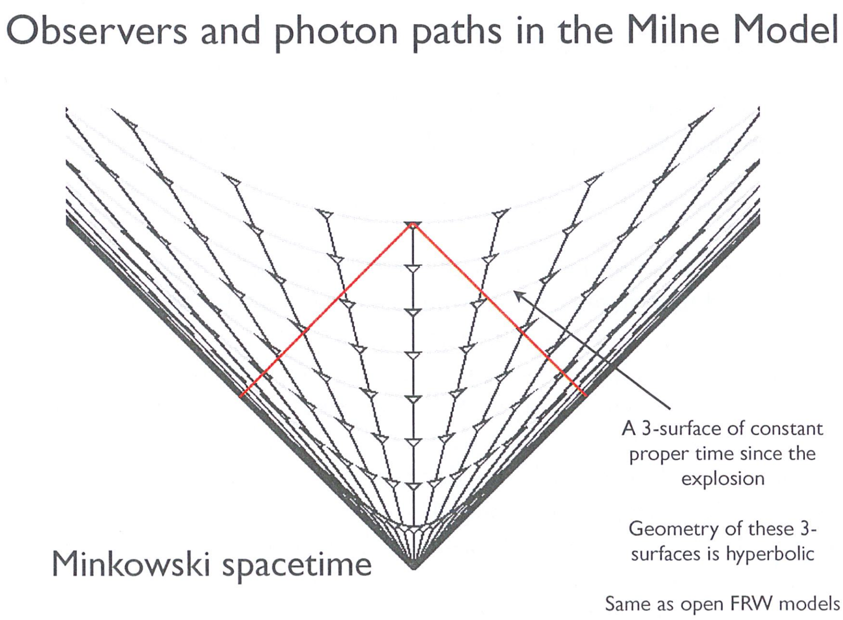

Figure 24: Milne's cosmological model is constructed by considering an explosion at some point in Minkowski space-time from which massless particles emerge with all velocities less than the speed of light. The metric is then written in coordinates where time is proper time since the big bang – which, because of time dilation is a hyperbola – and the spatial radial coordinate χ is a function of velocity (so particles maintain constant χ, θ, & ϕ). The result is formally identical to the FLRW metric (??).

C.1 The Milne metric

The time coordinate is taken to be the *proper time from the big-bang* and is denoted by τ. The hypersurfaces of constant proper time τ are, like the paths of Rindler observers, hyperbolae, but now

$$c^2t^2 - |\mathbf{x}|^2 = c^2\tau^2 \tag{125}$$

and the trajectories of the particles can be written parametrically as

$$\begin{aligned} ct &= c\tau \cosh \chi \\ \mathbf{x} &= \hat{\boldsymbol{\chi}} c\tau \sinh \chi \end{aligned} \tag{126}$$

where $\boldsymbol{\chi}$, in terms of which $\chi \equiv |\boldsymbol{\chi}|$ and $\hat{\boldsymbol{\chi}} \equiv \boldsymbol{\chi}/\chi$, labels the particle. Since $\cosh^2 \chi - \sinh^2 \chi = 1$, this clearly satisfies (125) and, along a world-line (constant $\boldsymbol{\chi}$), $c^2dt^2 - |d\mathbf{x}|^2 = c^2d\tau^2$ also. The velocity of a particle (also constant since they are inertial) is

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} = c\hat{\boldsymbol{\chi}} \tanh \chi \tag{127}$$

so

$$\chi = \tanh^{-1} |\mathbf{v}|/c. \tag{128}$$

Writing $C = \cosh \chi$ and $S = \sinh \chi$, the differentials cdt and $d\mathbf{x}$ are, from (125),

$$\begin{aligned} dt &= Cd\tau + S\tau d\chi \\ d\mathbf{x} &= c\hat{\boldsymbol{\chi}}(Sd\tau + C\tau d\chi) + c\tau S d\hat{\boldsymbol{\chi}} \end{aligned} \tag{129}$$

squaring these, and noting that $\hat{\boldsymbol{\chi}}$, being a vector of fixed length, has $\hat{\boldsymbol{\chi}} \cdot d\hat{\boldsymbol{\chi}} = 0$, gives the line element (i.e. the proper separation of two infinitesimally separated events)

$$ds^2 = -c^2dt^2 + |d\mathbf{x}|^2 = -c^2d\tau^2 + (c\tau)^2(d\chi^2 + \sinh^2 \chi |d\hat{\boldsymbol{\chi}}|^2) \tag{130}$$

or, using polar coordinates, so $\hat{\boldsymbol{\chi}} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, and defining $a(\tau) = c\tau$

$$\boxed{ds^2 = -c^2d\tau^2 + a(\tau)^2(d\chi^2 + \sinh^2 \chi(d\theta^2 + \sin^2\theta d\phi^2)).} \tag{131}$$

The is the *line element in Milne coordinates* from which we can read off the components of the *metric in Milne coordinates*, defined, as usual, such that $ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta$. With $x^\alpha = (\tau, \chi, \theta, \phi)$ it is

$$g_{\alpha\beta} = \text{diag}(-c^2, a^2, a^2\sinh^2\chi, a^2\sinh^2\chi\sin^2\theta). \tag{132}$$

It is complicated, and the hypersurfaces of constant τ actually have negative curvature[4]. But one must keep in mind that it is just Minkowski space-time, but written in a coordinate system with spatial coordinates tied to a family of expanding observers . It is identical in form to the *open FLRW metric*, where the scale factor is $a(\tau) = c\tau$ in the limit that the density of matter ρ is such that $G\rho \ll (\dot{a}/a)^2$. In such models the observers are called *comoving observers* as they are expanding with the matter; the flux density of matter they would measure vanishes. But here there is no matter, so they aren't expanding with anything.

The model as developed here is somewhat different from the Rindler model developed above in that there we considered observers accelerating only in the x-direction. We chose that option as we wanted to obtain the metric of space-time in a rocket. In Rindler's original formulation, the observers are accelerating radially from the origin. This is readily analysed by making the X-coordinate in (??) a 3-vector $\mathbf{X}$ in a manner analogous to the coordinate $\boldsymbol{\chi}$ in (126) as we will show later.

[4]This can be seen by looking at a circle in the equatorial plane $\sin^2\theta = 1$. This has radial length $r = \int ds_\parallel = a\int d\chi = a\chi$, but circumference $\int ds_\perp = a\sinh\chi \int d\phi = 2\pi a \sinh\chi \simeq 2\pi r(1 + \chi^2/6)$ which is the hallmark of negative curvature. This was for a circle drawn around the origin, but the same is true regardless of location of the centre.

C.2 Expanding radiation in a 1-dimensional Milne model

Maxwell's equations, as well as other field equations like the Klein-Grdon equation, when written in cosmological coordinates, contain a damping term, which seems to sap energy from fluctuations in the field, giving rise to redshifting of the energy density of e.g. the cosmic microwave background.

This term is often said to arise because of the *coupling of the EM – or other – field to the gravitational field of an expanding universe.* But this phenomenon appears also in the Milne model – which is just Minkowski space – in which there is no gravity and we know that energy is conserved. The resolution of this apparent paradox is that the solutions of these wave equations – wave-like solutions with comoving wavenumber k constant, and therefore with physical wavelength increasing with time and energy red-shifting – must also be solutions of the field equations in the original non-expanding coordinates.

To see this graphically, consider a coordinate system that is expanding in the manner of the Milne model, but only in one dimension (say the x-direction). If we make the transformation from (ct, x, y, z) coordinates to $(c\tau(x,t), \chi(x,t), y, z)$ where $c^2\tau^2 = c^2t^2 - x^2$ – so τ measures proper time since the explosion – and where $\chi = \tanh^{-1}(x/ct)$ then the metric is $ds^2 = -c^2d\tau^2 + c^2\tau^2d\chi^2 + dy^2 + dz^2$ and the field equation for a field that is only a function of τ and χ is simply

$$\ddot{\phi} + \dot{\phi}/\tau - \phi''/\tau^2 + \mu^2\phi = 0 \tag{133}$$

where now dot denotes $d/d\tau$ and prime denotes $d/d\chi$ and where we see that there is again a 'Hubble damping term' – now $H\dot{\tau}$ since $a = c\tau$, so $H = \dot{a}/a = 1/\tau$ – and which is the analogue of (??) to the case that the expansion is only in one direction.

Let's consider, for simplicity, waves of very high spatial frequency $\mu\lambda \ll 1$, the wave-equation becomes effectively massless ($\mu = 0$) and, in the original Minkowski coordinates, is

$$\frac{1}{c^2}\ddot{\phi} - \phi'^2 = 0. \tag{134}$$

This allows *d'Alembertian solutions*, $\phi = f_+(x - ct) + f_-(x + ct)$ where f_+ and f_- are arbitrary functions.

Consider the super-position of two oppositely propagating *logarithmic chirp waves*:

$$\phi(x,t) = \cos(\kappa \ln(ct + x)) + \cos(\kappa \ln(ct - x)) \tag{135}$$

which are illustrated in figure 25.

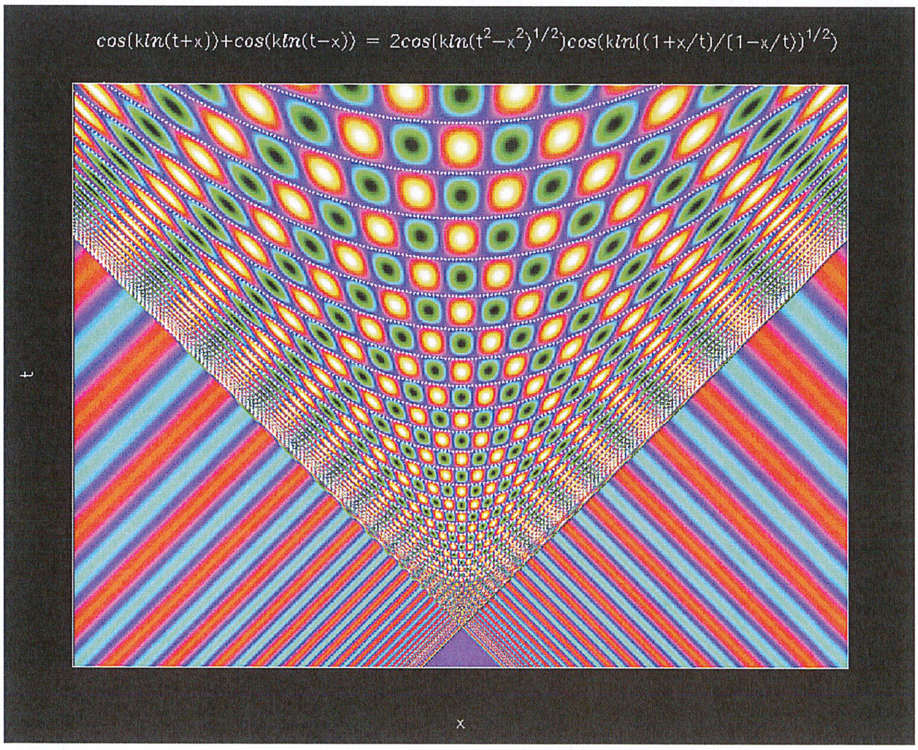

Figure 25: Expanding radiation in a 1-dimensional version of Milne's model. This shows a solution of Maxwell's equations in Minkowski coordinates. It is the sum of two oppositely propagating semi-infinite 'logarithmic-chirps'. It is at the same time a 'standing wave' solution of Maxwell's equations written in expanding coordinates. Milne observers would say that the energy flux density in their frame vanishes, and they would find the energy density of the radiation decreasing.

Using $\cos A + \cos B = 2\cos((A+B)/2)\cos((A-B)/2)$ this is equivalent to

$$\begin{aligned}\phi(x,t) &= 2\cos(\kappa[\ln(ct+x)+\ln(ct-x)]/2)\cos(\kappa[\ln(ct+x)-\ln(ct-x)]/2)\\ &= 2\cos\left(\kappa\ln\sqrt{(ct+x)(ct-x)}\right)\cos\left(\kappa\ln\sqrt{(ct+x)/(ct-x)}\right)\\ &= 2\cos\left(\kappa\ln\sqrt{c^2t^2-x^2}\right)\cos\left(\kappa\ln\sqrt{(\cosh\chi+\sin\chi)/(\cosh\chi-\sin\chi)}\right)\\ &= 2\cos(\kappa\ln(c\tau))\cos(\kappa\chi)\end{aligned} \tag{136}$$

and differentiating this shows that it is a solution of (133) (with $\mu = 0$).

If there were a family of *Milne observers* – expanding inertially from an initial explosion at the 'focal point' – where the chirps first overlap they would find that the energy density of the radiation is decreasing as $\mathcal{E} \propto 1/t^2$. They would measure zero energy flux density, but they would find that the spatial divergence of the flux density is positive (consistent, in the equation of continuity of energy, with the decreasing energy density). So this is an expanding '1-dimensional fire-ball' of radiation in which the radiation is everywhere redshifting and apparently losing energy. But the decrease of energy density is not being caused by any coupling to anything; and certainly not to the 'gravitational field of the expanding universe'. It is important to realise that the only thing that is expanding here – indeed the only thing that exists – is the radiation itself.

D Why pressure gravitates in GR – from weak-field theory

When developing the *Newtonian-limit* in weak-field gravity, it is usual to impose the condition that the trace-reversed metric perturbations be divergence free condition $\overline{h}_{\mu\nu}{}^{,\nu} = 0$. One then finds that the solution of the resulting field equations, for a static, pressure-free source $T_{\mu\nu} = \rho(\mathbf{x})c^2\delta^0_\mu\delta^0_\nu$ was $\overline{h}_{\mu\nu} = -4\Phi(\mathbf{x})\delta^0_\mu\delta^0_\nu$, which obeys the gauge condition, and, when trace unreversed, gives $h_{\mu\nu} = -2\Phi(\mathbf{x})\delta_{\mu\nu}$.

It is interesting to turn this around and ask, if we assume $h_{\mu\nu} = -2\Phi(\vec{x})\delta_{\mu\nu}$ – i.e. that it remain diagonal, but we drop the condition that the potential be static – what matter source this would correspond to. The result is given by $\mathbf{T} = (8\pi\kappa)^{-1}\mathbf{G}$ with the components of the Einstein tensor given, in a general weakly perturbed coordinate system, from (??) as

$$\begin{aligned}G_{\alpha\beta} &= 2[\delta^0_\alpha\delta^0_\beta\Phi_{,\mu}{}^{,\mu} + \eta_{\alpha\beta}\delta^0_\mu\delta^0_\nu\Phi^{,\mu\nu} - \delta^0_\alpha\delta^0_\mu\Phi_{,\beta}{}^{,\mu} - \delta^0_\beta\delta^0_\mu\Phi_{,\alpha}{}^{,\mu}]\\ &= 2[\delta^0_\alpha\delta^0_\beta\Phi_{,\mu}{}^{,\mu} + \eta_{\alpha\beta}\Phi^{,00} - \delta^0_\alpha\Phi_{,\beta}{}^{,0} - \delta^0_\beta\Phi_{,\alpha}{}^{,0}]\end{aligned} \tag{137}$$

from which we find

$$\begin{aligned}G_{00} &= 2\nabla^2\Phi\\ G_{0i} = G_{i0} &= 2c^{-1}\dot{\Phi}_{,i}\\ G_{ij} &= 2c^{-2}\delta_{ij}\ddot{\Phi}\end{aligned} \tag{138}$$

so, if we demand that $\dot{\Phi} = 0$ at some time, the metric we have postulated describes the gravitational field of matter with non-zero, but isotropic, pressure ($G_{ij} = 8\pi\kappa P\delta_{ij}$) and we would then have a metric looking just like the Newtonian limit metric, but with a non-zero time variation of the potential being driven by the pressure of the matter. This is known in cosmology as the longitudinal gauge. It isn't really so much a choice of gauge as a restriction on the metric. Nonetheless, it allows one to describe somewhat more general matter than just pressure free 'dust'.

To illustrate this, consider the case of spatially uniform density $T_{00} = \rho c^2$ and pressure $T_{ij} = P\delta_{ij}$, in which case we must have

$$\begin{aligned}\nabla^2\Phi &= 4\pi G\rho/c^2\\ \Phi_{,i} &= 0\\ \ddot{\Phi} &= 4\pi GP/c^2\end{aligned} \tag{139}$$

which admit the solution

$$\Phi = \frac{2\pi G}{3c^2}(\rho|\mathbf{x}|^2 + 3Pt^2) \tag{140}$$

giving the metric expressed through the line element

$$ds^2 = -(1 + 2\Phi(|\mathbf{x}|, t))c^2dt^2 + (1 - 2\Phi(|\mathbf{x}|, t))(dx^2 + dy^2 + dz^2) \tag{141}$$

or, if we preferred, in terms of polar spatial coordinates.

$$ds^2 = -(1 + 2\Phi(r, t))c^2dt^2 + (1 - 2\Phi(r, t))(dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)). \tag{142}$$

This metric could describe the space-time inside, for instance, a balloon containing pressurised gas. It would not apply throughout all of space, as in the walls of the balloon the stress tensor would be anisotropic. But it would apply within the body of the gas. If could also apply near the centre of a star. It may seem strange that the metric one finds for a static source here is actually time varying but as we are working in perturbation theory, and if we don't consider times that are too long this appears to be physically reasonable.

D.1 Geodesic equation – non-relativistic particles

It is interesting to look at the deviation of geodesics in this space time. If we consider two initially stationary particles with a separation $\vec{\xi} \longrightarrow (0, \xi, 0, 0)$, as illustrated on the left in figure 26, the geodesic deviation equation

$$\frac{d^2\vec{\xi}}{d\tau^2} = \mathbf{R}(\ ,\vec{U}, \vec{\xi}, \vec{U}). \tag{143}$$

gives, for the rate of change of the x-component of their separation

$$\left(\frac{d^2\vec{\xi}}{d\tau^2}\right)^x = R^x{}_{\mu\beta\nu}U^\mu\xi^\beta U^\nu = c^2 R^x{}_{0x0}\xi. \tag{144}$$

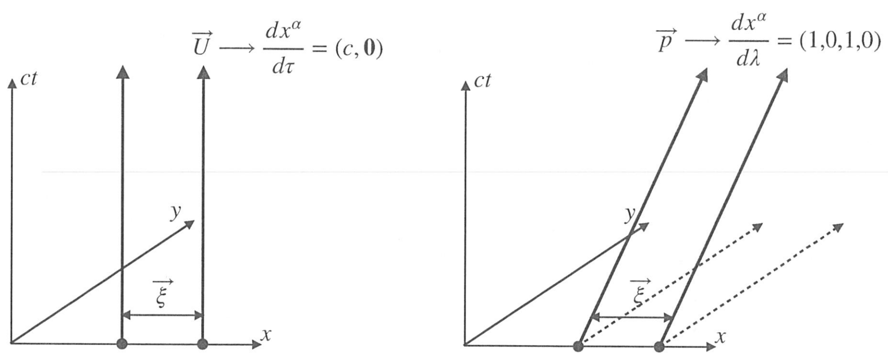

Figure 26: On the left are shown the world lines for two initially stationary particles with coordinate separation along the x-axis. Their zeroth order 4-velocity is as indicated. On the right are shown two photons propagating along the y-axis.

The relevant component of the curvature tensor is readily found to be

$$R^x{}_{0x0} = \Gamma^x{}_{00,x} - \Gamma^x{}_{0x,0} = \tfrac{1}{2}(h_{00,xx} + h_{xx,00}) = -\frac{4\pi G}{3c^2}(\rho + 3P/c^2) \tag{145}$$

so the 2nd rate of change of the physical separation is

$$\boxed{\ddot{\xi} = -\frac{4\pi G}{3}(\rho + 3P/c^2)\xi} \tag{146}$$

where the first term is the Newtonian acceleration two freely falling particles would have towards each other inside a uniform density sphere. We see here that the pressure enhances the gravitational tidal field; a result of considerable importance in cosmology and in the theory of stellar structure for relativistic stars such as neutron stars where the pressure may be comparable to ρc^2.

D.2 Geodesic equation – massless particles

We can also calculate the geodesic deviation for two light rays, again with separation $\vec{\xi}$ along the x-axis and propagating along the y-direction as illustrated on the right of figure 26. In this case we have

$$\left(\frac{d^2\vec{\xi}}{d\lambda^2}\right)^x = R^x{}_{\mu\beta\nu}p^\mu\xi^\beta p^\nu = (R^x{}_{0x0} + R^x{}_{0xy} + R^x{}_{yx0} + R^x{}_{yxy})\xi \tag{147}$$

where the path parameter λ is affine distance, which, for our choice of a unit-$|\mathbf{p}|$ photon, is the same as physical path distance. The two central terms above vanish and, with the addition of the last, we get

$$\boxed{\frac{d^2\xi}{d\lambda^2} = -\frac{4\pi G}{c^2}(\rho + P/c^2)\xi} \tag{148}$$

which shows that it is a different combination of energy density and pressure – actually the *enthalpy* – that causes focussing of light rays.

An interesting application of this is to the case $P = -\rho c^2$ which can arise, in late-time inflation, if the density of the universe becomes dominated by a scalar *quintessence field* or if there is a cosmological constant Λ (as this behaves like matter with $P = -\rho c^2$). In that case there is no focussing of light rays; if initially parallel they remain equidistant from one another.

Note however that, just as we found with Geraint's tunnel, if one were to measure light ray paths on a rigid but light photographic plate (so its gravity is negligible) then one would find that rays that do not pass through the centre of the plate would be found to be locally curved and bending outwards (as compared to straight lines). Yet their distance from a ray passing through the centre would be unchanging. This is because the rays, while being geodesics of the 4-geometry described by (141) are not geodesics of the spatial 2-geometry $dl^2 = (1 - 2\Phi)(dx^2 + dy^2)$. The latter is positively curved, so initially parallel spatial geodesics would be focussed towards one another. The question of whether Λ affects gravitational lensing has been debated; we see here that Λ certainly does cause light bending if you measure it locally, but it does not cause any global focussing of light rays.

$\nabla P = a(\rho + P/c^2).\ dt$