General Relativity

Y. Ding
October 1, 2025

Contents

1 Introduction 7

1.1 Gravity vs. Spacetime Curvature . . . 7

1.1.1 Gravity . . . 7

1.1.2 Geometry and Curvature . . . 7

1.2 Newtonian Mechanics and Gravity . . . 9

1.2.1 Inertial Frames and Galilean Transformation . . . 9

1.2.2 Newtonian Gravity . . . 10

1.3 Relativistic Gravity . . . 11

2 Special Relativity 12

2.1 Spacetime Geometry . . . 12

2.1.1 Minkowski Spacetime . . . 12

2.1.2 Time Dilation . . . 13

2.2 Relativistic Kinematics . . . 13

2.2.1 Lorentz Transformation . . . 13

2.3 Relativistic Dynamics . . . 14

2.3.1 Four-Vectors . . . 14

2.3.2 4-Velocity and 4-Momentum . . . 15

2.3.3 Action and Lagrangian for a Free Paticle . . . 16

2.3.4 Observers and Observations . . . 18

2.4 Vectors and Tensors . . . 19

2.4.1 Vectors . . . 20

2.4.2 Dual Vectors . . . 20

2.4.3 Tensors . . . 21

3 Perfect Fluids in Special Relativity 23

3.1 Dust . . . 23

3.1.1 Number Density and Flux . . . 23

3.1.2 The Number-Flux Four-Vector $\vec{N}$ . . . 23

3.1.3 A One-Form Defines a Surface . . . 24

3.2 The Stress-energy tensor . . . 24

3.3 General Fluids . . . 25

3.3.1 First Law of Thermodynamics . . . 25

3.4 Perfect Fluids . . . 26

3.4.1 The Stress-Energy Tensor . . . 26
3.4.2 The Conservation Laws . . . 26
3.5 Further Concepts . . . 27
3.5.1 Importance for GR . . . 27
3.5.2 The Gauss' Law . . . 28

4 Curved Manifolds **29**
4.1 On the Relation of Gravitation to Curvature . . . 29
4.1.1 The Principle of Equivalence . . . 29
4.2 The Metric Tensor . . . 29
4.2.1 Properties of the Metric Tensor . . . 30
4.2.2 Local Inertial Frames . . . 30
4.3 Differentiable Manifolds and Riemannian Manifolds . . . 31
4.3.1 Differentiable Manifolds . . . 32
4.3.2 Riemannian Manifolds . . . 32
4.3.3 Length, Areas, and Volumes . . . 32
4.3.4 Geodesic Equation . . . 34
4.4 Covariant Derivatives . . . 35
4.4.1 Derivatives of Vectors . . . 36
4.4.2 Mathematical Origin of the Covariant Derivative . . . 37
4.5 Christoffel Symbols Revisited . . . 38
4.5.1 Covariant Derivative of a Metric . . . 38
4.5.2 Calculating Christoffel Symbols from the Metric . . . 39
4.6 Parallel-Transport, Geodesics, and Curvature . . . 39
4.6.1 Parallel Transport . . . 39
4.6.2 Geodesics . . . 40
4.7 The Curvature Tensor . . . 40
4.7.1 Riemann Curvature Tensor . . . 40
4.7.2 Geodesic Deviation . . . 42
4.8 Bianchi Identities, Ricci and Einstein Tensors . . . 42

5 A Curved Spacetime **44**
5.1 The Transition from Differential Geometry to Gravity . . . 44
5.2 Physics in Slightly Curved Spacetimes . . . 44
5.2.1 Newton's Equations of Motion . . . 45
5.2.2 Gravitational Time Dilation . . . 46
5.3 Conserved Quantities . . . 47

5.3.1 Killing Vector Fields . . . 48

6 The Field Equations **50**

6.1 The Purpose and Justification of the Field Equations . . . 50

6.2 Einstein's Equations . . . 50

7 Weak Fields and Gravitational Radiation **51**

7.1 Einstein's Equations for Weak Gravitational Fields . . . 51

7.1.1 Background Lorentz Transformations and Gauge Transformations . . . 51

7.1.2 Weak-Field Einstein Equations . . . 52

7.2 Newtonian Gravitational Fields . . . 53

7.2.1 Newtonian Limit . . . 53

7.2.2 The Far Field of Stationary Relativistic Sources and its Mass . . . 53

7.3 Gravitational Waves . . . 54

7.3.1 The Transverse-Traceless Gauge . . . 55

7.3.2 The Effects of Waves on Free Particles . . . 55

7.4 The Generation of Gravitational Waves . . . 56

7.4.1 Slow Motion Wave Generation . . . 56

8 Schwarzschild Solutions **57**

8.1 Coordinates for Spherically Symmetric Spacetime . . . 57

8.1.1 Flat Space in Spherical Coordinates and Two-Spheres . . . 57

8.1.2 Spherically Symmetric Spacetime . . . 57

8.2 Static Spherically Symmetric Spacetime . . . 57

8.2.1 The Metric . . . 57

8.2.2 The Einstein Tensor . . . 58

8.3 Static Perfect Fluid Einstein Equations . . . 58

8.3.1 The Stress-Energy Tensor . . . 58

8.3.2 Einstein Equations . . . 59

8.4 The Exterior Geometry . . . 59

8.4.1 Schwarzschild Metric . . . 59

8.4.2 Gravitational Redshift . . . 60

8.5 Orbits in Schwarzschild Spacetime . . . 61

8.5.1 The Effective Potential . . . 61

8.5.2 Radial Plunge Orbits . . . 63

8.5.3 Circular Orbits . . . 64

8.5.4 Elliptical Orbits and Precession . . . 66

8.5.5 Photon Orbits . . . 66

8.5.6 Gravitational Lensing . . . 68
8.5.7 The Shapiro Delay . . . 69
8.6 The Interior Structure of the Star . . . 70
8.6.1 General Rules for Integrating the Equations . . . 70
8.6.2 Newtonian Stars . . . 71
8.7 Exact Interior Solutions . . . 71
8.7.1 The Schwarzschild Constant-Density Interior Solution . . . 71
8.7.2 Buchdahl's Interior Solution . . . 72
8.8 Realistic Stars and Gravitational Collapse . . . 72
8.8.1 Buchdahl's Theorem . . . 72
8.8.2 Formation of Stellar-Mass Black Holes . . . 72

9 Black Holes **74**
9.1 The Surface $r = 2M$ . . . 74
9.1.1 Eddington-Finkelstein Coordinates . . . 74
9.1.2 Kruskal-Szekeres Coordinates . . . 75
9.2 General Black Holes . . . 76
9.2.1 General Properties of Black Holes . . . 77
9.2.2 Kerr Black Hole . . . 77
9.2.3 Ergoregion . . . 78
9.2.4 The Kerr Horizon . . . 78
9.3 Real Black Holes in Astronomy . . . 78
9.4 Quantum Mechanical Emission of Radiation by Black Holes . . . 79

10 Cosmology **81**
10.1 Cosmological Kinematics: Observing the Expanding Universe . . . 81
10.1.1 Homogenety and Isotropy of the Universe . . . 81
10.1.2 Models of the Universe: the Cosmological Principle . . . 81
10.1.3 Spacetime and Relativity . . . 81
10.1.4 Symmetric Three-Spaces . . . 82
10.1.5 Robertson-Walker Metric . . . 83
10.2 Kinematics . . . 83
10.2.1 Geodesics . . . 83
10.2.2 Redshift . . . 86
10.3 Dynamics . . . 86
10.3.1 Perfect Fluids . . . 86
10.3.2 Matter and Radiation . . . 88

10.3.3 Dark Energy . . . 89
10.3.4 Spacetime Curvature . . . 89
10.3.5 Friedmann Equations . . . 90
10.3.6 Exact Solutions . . . 91

A Proofs and Derivations **95**

A.1 Lorentz Transformation . . . 95
A.2 Geodesic Equation . . . 96
A.3 The Periapsis Precession . . . 97
A.3.1 Direct Integration . . . 97
A.3.2 Perturbation in Circular Orbit . . . 98
A.4 The Shapiro Delay Integral . . . 98
A.5 Black Holes . . . 100
A.5.1 Kruskal-Szekeres Metric . . . 100
A.5.2 Unphysical Regions . . . 101

To use this notes well, one should be familiar with various math tools and areas of physics:

1. Vector Calculus: partial derivatives, multiple integrals, the divergence theorem.
2. Analytical mechanics: calculus of variations, Lagrangian, principle of extremal action.
3. Other physics: Newtonian gravity, acquaintance with special relativity, electromagnetism.

Tensor analysis and some differential geometry will be introduced in this notes.

1 Introduction

1.1 Gravity vs. Spacetime Curvature

1.1.1 Gravity

There are four fundamental forces or interactions in nature: gravity, electromagnetism (EM), strong, and weak. In terms of interaction range, gravity is a long-range force like EM. In terms of strength, gravity is the weakest among the four: about $\sim 10^{30}$ weaker than the other three. For example, the ratio of the gravitational force to electric repulsion between two protons is

$$\frac{F_g}{F_e} = \frac{Gm_p^2/r^2}{e^2/4\pi\epsilon_0 r^2} = \frac{4\pi\epsilon_0 G m_p^2}{e^2} \sim 10^{-36}.$$

according to Newtonian gravity. Thus, gravity does not play a role in quantum physics because the masses of fundamental particles are very small. However, unlike charges of EM, which can be positive or negative, the charges of gravity, namely mass and energy, are always positive. On large length scales, objects are neutral, so the effect of EM weakens. Strong and weak interactions are short-range forces, so they don't play a role on these scales at all. But masses build up on these scales—as gravity is always attractive, it dominates on these scales and is important in astrophysics and cosmology.

Unlike the other three interactions, gravity does not have a well-posed, self-consistent, experimentally testable quantum theory. The understanding of gravity is a classical, but a large subject, known as Einstein's general theory of relativity, or **general relativity** (GR). In general relativity, gravity is no longer a force, but represents a curvature of spacetime. The mathematical tools necessary to study general relativity are tensor calculus and differential geometry. They are to general relativity as vector calculus is to Maxwell's theory of electromagnetism.

1.1.2 Geometry and Curvature

A curve is one-dimensional, meaning that it can be parametrized by only one parameter or variable. For example, in a three-dimensional Euclidean space with coordinates (x, y, z), a parametric curve can be expressed as

$$\mathbf{r}(t) = x(t)\hat{\mathbf{x}} + y(t)\hat{\mathbf{y}} + z(t)\hat{\mathbf{z}}.$$

The **curvature** κ at any point with $\dot{\mathbf{r}} \neq 0$ is defined as

$$\kappa = \frac{|\ddot{\mathbf{r}} \times \dot{\mathbf{r}}|}{|\dot{\mathbf{r}}|^3}.$$

There is not much geometry going on in one dimension, so let's turn to two-dimensional spaces: a surface. A two-dimensional Euclidean surface is a flat space with zero curvature. The Euclidean axioms and geometry works on this surface. For example, the sum of the interior angles of a triangle is π; the ratio of the circumference of a circle to its radius is 2π.

Then what are some surface with non-Euclidean geometry, or nonzero curvature?

Example 1.1. A Two-Sphere

A **two-sphere** is just a fancy name of the surface of a sphere embedded in 3D Euclidean space. The "two" stands for two-dimensional—the surface of a sphere is parametrized by two variables. When we are working with curved space, we should be careful about what is meant by some geometric object. Common questions can be: what defines a straight line, or what defines a triangle?

A **straight line** is defined as the path of shortest distance between two points. A straight line on a two-sphere is a segment of a **great circle**. A great circle is the largest circle you can draw on a sphere. It is the intersection of the sphere with a plane containing the center of the sphere. Longitude lines are great circles; latitude lines except the equator are not great circles. A **triangle** is defined as a three-sided figure with three non-collinear points connected by straight lines. Thus, a triangle on a two-sphere is called a spherical triangle. Its sides are paths along great circles.

All triangles on a two-sphere has sum of the interior angles greater than π. For instance, you start from the north pole and walk south along the $0°$ longitude line. Once you reach the equator (call it point A), turn west (this is a $90°$ or $\pi/2$ turn) and walk towards the $90°$W longitude along the equator. Once you reach the $90°$W longitude (call it point B), you turn north (this is another $90°$ or $\pi/2$ turn). Eventually, you will return to the north pole along the $90°$W longitude line, but you start with walking along the $0°$ longitude line. This means these two lines are also at angles $90°$ or $\pi/2$ (call this point C). The sum of the interior angles at A, B, and C is $3\pi/2 > \pi$. Note that this is a triangle because all its sides are segments of great circles (two longitudes and one equator).

In fact, the sum of the interior angles of a spherical triangle is related to its area S by

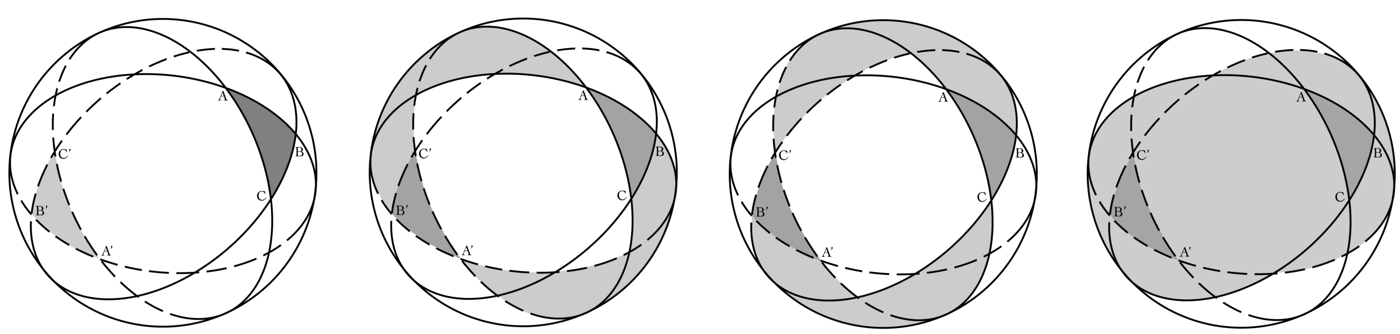
$$\sum_{\text{vertices}} \begin{pmatrix} \text{interior} \\ \text{angles} \end{pmatrix} = \pi + \frac{S}{R^2}, \tag{1.1}$$

where R is the radius of the 2-sphere, the distance from every point on the sphere to its center. All cicles that are not great circles have the ratio of its circumference to radius

$$\frac{C}{r} = 2\pi \frac{\sin(r/R)}{r/R}.$$

Note that the radius r is defined as the distance of a straight line (defined on the two-sphere) connecting one of the point on the circle to its center. The C/r ratio is easy to prove by making use of arc length formula $R\theta = r$ in spherical coordinates (θ, ϕ) and some trigonometry. Here we will prove (1.1).

Proof. First, we know that a spherical triangle is made from three great circles. The three great circles will intesect at six points, A, B, C, A', B', and C'. From the figure below, we can see that ABC forms a triangle, and $A'B'C'$ forms another congruent triangle opposite to $\triangle ABC$ by reflection symmetry around the center of the sphere.



Next, choose one vertex of $\triangle ABC$ and its corresponding reflection. For example, the second figure above chooses A and A'. The two great circles intersecting at A and A' encloses two congruent shaded area, each containing one of $\triangle ABC$ or $\triangle A'B'C'$. These shaded area are proportional to $\angle A$: if we take A to be the north pole ($\theta = 0$) and A' to be the south pole ($\theta = \pi$), then the shaded area is

$$S_A = S_{\text{sphere}} \cdot \frac{2\angle A}{2\pi} = 4\pi R^2 \frac{\angle A}{\pi} = 4R^2 \angle A.$$

We can do the same for B and B' (the fourth figure), and C and C' (the third figure),

$$S_B = 4R^2 \angle B, \quad S_C = 4R^2 \angle C.$$

If we sum up these shaded area, they will cover the entire sphere. But we are overcounting $\triangle ABC$ and $\triangle A'B'C'$ for additional two times, resulting in an additional area of $4S_{ABC}$. This means

$$S_A + S_B + S_C - 4S_{ABC} = S_{\text{sphere}} \quad \Longrightarrow \quad 4R^2(\angle A + \angle B + \angle C) = 4\pi R^2 + 4S_{ABC}.$$

Divide both sides by $4R^2$, we have

$$\angle A + \angle B + \angle C = \pi + \frac{S_{ABC}}{R^2}.$$

□

Curvature and geometry in higher-dimensional spaces are hard to visualize. To study general relativity, one needs to get used to four-dimensional spacetime. Unfortunately, we are unable to visualize a four-dimensional object embedded in three-dimensional spaces, which is the space we live in. To work with four-dimensional spacetime, we need some intrinsic properties of curvature that does not require or depend on the existence of embeddings. We do this by studying the **Riemann curvature tensor**, $R^{\mu}{}_{\nu\alpha\beta}$, a rank-4 tensor whose indices μ, ν, α, and β can take on 1 to N for an N-dimensional space. This means there are N^4 components in this tensor, hence 256 components for a 4D spacetime. Luckily, most components are dependent on each other and are related by symmetry. In general, the number of independent components is $N^2(N^2-1)/12$. There is only 1 in 2D (which is called the *Gaussian curvature*), 6 in 3D, and 20 in 4D. We will see the Riemann curvature tensor in Chapter 4.

Mathematically, the Riemann curvature tensor is related to the second derivatives of the **metric tensor**, $g_{\mu\nu}$. The key thing about the metric tensor provides a way to define a line element. A **line element** ds^2 is a measure of the infinitesimal distance in terms of coordinates. In 3D Euclidean flat space with Cartesian coordinates (x, y, z), the line element is given by

$$ds^2 = dx^2 + dy^2 + dz^2,$$

which is the famous Pythagorean theorem. Integrating $ds = \sqrt{dx^2 + dy^2 + dz^2}$ along a curve gives the length of the curve. In principle, with the line element, we are able to know everything about the geometry of a space. Of course e are not restricted to this coordinate system. Another useful coordinate to describe 3D Euclidean space is the spherical coordinates (r, θ, ϕ). They are related to Cartesian coordinates by

$$x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\theta.$$

Working out the differentials, one can find the line element to be

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\, d\phi^2.$$

One thing to remember here is that even though ds^2 *looks* different in different coordinates, it is an invariant, scalar quantity that does not depend the coordinate you choose. One can derive the circumference of a circle of radius r to be $2\pi r$, no matter which coordinates they choose. Coordinates are *representations* of quantities, and we prefer one coordinate over the other for a certain problem because good coordinates make a problem simpler.

1.2 Newtonian Mechanics and Gravity

1.2.1 Inertial Frames and Galilean Transformation

Newtonian mechanics assumes that the geometry is flat and independent of time. Time is a universal quantity—all clocks tick at the same rate. In Cartesian coordinates (x, y, z), the flat geometry is described by

$$ds^2 = dx^2 + dy^2 + dz^2.$$

The laws of Newtonian mechanics work the simplest in **inertial reference frames**. In an inertial frame, all free particles move in a straight line with constant velocity.

Suppose we find one inertial frame S and define an origin $\mathcal{O}$ in that frame. To measure time, we will also put a clock at the origin. We can find other inertial frames based on S. One (different) inertial frame S' can be a frame in which all axes (x', y', z') are parallel to to axes (x, y, z) respectively in S. Moreover, the x'-axis is aligned with the x-axis. The clock in both frames are synchronized when their origins are at the same point. The frame S' is also moving with a velocity v with respect to S. All the above descriptions can be summarized to a coordinate transformation:

$$t' = t, \quad x' = x - vt, \quad y' = y, \quad z' = z.$$

This is a **Galilean transformation**.

In an inertial frame, if you do not have any communication with or observation from other frames, you cannot tell whether your frame is moving or not. Moreover, all laws of physics are the same in all inertial frames. This is known as the **principle of relativity**. By "same laws of physics" we mean the Lagrangian/Action is invariant, or the equations of motion are *covariant*. (Basically, covariant means that certain quantities change in the equations of motion, but the form of equations stays the same. For example, the electric or magnetic fields may look different in two inertial frames, but Maxwell's equations are still true in both frames.)

1.2.2 Newtonian Gravity

In Newtonian gravity, we work with the law of universal gravitation:

$$\mathbf{F} = -\frac{GMm}{r^2}\hat{\mathbf{r}}. \tag{1.2}$$

It describes the gravitational force $\mathbf{F}$ between two masses M and m separated by $\mathbf{r}$, and $G \approx 6.67 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}$ is the gravitational constant. The negative sign in $-\hat{\mathbf{r}} = -\mathbf{r}/r$ indicates that gravity is always attractive.

Newtonian gravity works nearly identical with electrostatics govern by Coulomb's law, basically because they are both inverse-square laws.

1. Field: In electrostatics, we define the electric field $\mathbf{E}$ to be the force per unit charge on a particle. In Newtonian gravity, the gravitational field (or acceleration) $\mathbf{g}$ is defined as the force per unit mass:

$$\mathbf{g} = -\frac{GM}{r^2}\hat{\mathbf{r}}. \tag{1.3}$$

2. Potential: Both gravity and electrostatic force are conservative forces. In electrostatics, the electric potential ϕ is defined as the potential energy per unit charge, related to $\mathbf{E}$ by $-\nabla\phi = \mathbf{E}$. In Newtonian gravity, the gravitational poential Φ is defined as

$$\Phi = -\frac{GM}{r}, \tag{1.4}$$

with a similar relation to gravitational field:

$$\mathbf{g} = -\nabla\Phi \quad \Longleftrightarrow \quad \Phi(\mathbf{r}_b) - \Phi(\mathbf{r}_a) = -\int_{\mathbf{r}_a}^{\mathbf{r}_b} \mathbf{g}\cdot d\boldsymbol{\ell}.$$

Since gravity is conservative, the path integral of Φ (which is the work done on the particle) from $\mathbf{r}_a$ to $\mathbf{r}_b$ is path independent.

3. Linearity: Newtonian gravity is linear, so the gravitational field obeys the principle of superposition. The total potential due to a mass distribution is equal to the sum of the potentials from individual small masses:

$$\Phi(\mathbf{r}) = -\sum_i \frac{GM}{|\mathbf{r}-\mathbf{r}_i|} \qquad \text{and} \qquad \Phi(\mathbf{r}) = -\int \frac{G\rho(\mathbf{r}')d^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|},$$

where ρ stands for the mass density of the distribution.

4. Gauss's law and Poisson's equation: these are the gravitational counterpart of Gauss's law and Poisson's equation for electrostatics. Gauss's law:

$$\nabla\cdot\mathbf{g} = -4\pi G\rho \quad \Longleftrightarrow \quad \oint_S \mathbf{g}\cdot d\mathbf{S} = -4\pi GM. \tag{1.5}$$

Poisson's eqution:

$$\nabla^2\Phi = 4\pi G\rho. \tag{1.6}$$

Later, we will see that principle of superposition does not hold for general relativity. Nonetheless, to be a good theory, general relativity agrees with Newtonian gravity for weak gravitational field limit.

1.3 Relativistic Gravity

General relativity incorporates special relativity and gravity. There are basically two fundamental constants involved in general relativity: the speed of light c (which characterizes relativity) and the gravitational constant G (which characterizes gravity),

$$c \approx 2.998 \times 10^8\,\mathrm{m/s}, \quad G \approx 6.67 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}.$$

By dimensional analysis, we can guess the relativistic length scale of a mass M to be

$$R_g = \frac{GM}{c^2}, \tag{1.7}$$

which is known as the **gravitational radius** of M. When the typical size R of mass M is comparable to R_g (which is equivalent to the gravitational potential energy of some mass m being comparable to its rest mass energy), general relativity comes into play,

$$\frac{R_g}{R} = \frac{GM}{Rc^2} = \frac{GMm/R}{mc^2} \sim 1 \quad \Longrightarrow \quad \text{general relativity matters.}$$

For example,

1. The sun: $M_\odot = 1.989 \times 10^{30}\,\mathrm{m}$, $R_\odot = 6.96 \times 10^8\,\mathrm{m}$,

$$\frac{GM_\odot}{R_\odot c^2} \simeq 2 \times 10^{-6}.$$

2. The Earth: $M_\oplus = 5.974 \times 10^{24}\,\mathrm{kg}$, $R_\oplus = 6.38 \times 10^6\,\mathrm{m}$,

$$\frac{GM_\oplus}{R_\oplus c^2} \simeq 7 \times 10^{-10}.$$

   This means Newtonian gravity is a good approximation of the Sun and the solar system.

3. Neutron stars: $M \simeq 1.4\,M_\odot$, $R \simeq 10\,\mathrm{km}$,

$$\frac{GM}{Rc^2} \sim 0.2.$$

   General relativity is very important to the structure of neutron stars. In fact, the precise equation of state of neutron stars is not yet determined.

4. Black holes: a non-spinning, chargeless black hole has the size of the **Schwarzschild radius**, $R = 2GM/c^2$, so

$$\frac{GM}{Rc^2} = 0.5.$$

   General relativity is very important.

5. The universe: the mass-energy density of the universe is about $\rho \simeq 9 \times 10^{-27}\,\mathrm{kg/m^3}$, while the observable universe has a size of $R \sim 10^{26}\,\mathrm{m}$. This gives

$$\frac{GM}{Rc^2} \sim \frac{4\pi G R^3 \rho}{3Rc^2} \sim 0.7.$$

   The understanding of overall dynamics of the universe requires general relativity.

2 Special Relativity

In 1887, the Michelsom and Morley experiment shows the constancy of the speed of light. There are two postulates in special relativity:

1. The laws of physics take the same form in all inertial reference frames. All inertial frames are fundamentally indistinguishable based on experiments in that frame only.
2. The speed of light (and other massless particles) is the maximum speed in the universe and has the same value in all inertial reference frames,
$$c = 2.99792458 \times 10^8 \text{ m/s}.$$

Note that this number is exact because modern physics define the "meter" using the speed of light in vacuum. With these assumptions, many odd things would occur: Galilean transformation is no longer true, clocks do not run universally, etc. In this chapter, we will review special relativity using two approaches: one is using an introductory spacetime geometry, and the other is more mathematical, using Lorentz transformation formulas.

2.1 Spacetime Geometry

2.1.1 Minkowski Spacetime

In special relativity, time and space are somewhat mixed up in coordinate transformations. Thus, they are unified into a four-dimensional entity called **spacetime**. To be precise, let's list properties of an inertial frame in special relativity:

1. Clocks that have been synchronized at some time will remain synchronized.
2. The properties of points in *space* is described by Euclidean geometry.
3. The spatial separation between points in space with specified coordinates does not change with time.

When one tranform one inertial frame to another, simultaneous events in one inertial reference frame need not be simultaneous in other frames. Events that occur at the same place need not be at the same place in other frames.

An event in an inertial frame can be labeled with four coordinates known as **Minkowski coordinates**: (ct, x, y, z). (In spacetime coordinates, time is conventionally the zeroth coordinate and is multiplied by a c to ensure the dimension is right. Later, we will set $c = 1$ for convenience so that the time coordinate is just t, but for now let's stick with ct.) The geometry of spacetime in an inertial frame is given by the line element
$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2. \tag{2.1}$$
(We will prove it later using Lorentz transformations.) It is the square of the **spacetime interval** ds between two infinitesimally separated events. This spacetime is called **Minkowski spacetime**. The line element is a scalar quantity, invariant under any coordinate transformations. In another frame with Minkowski coordinates (ct', x', y', z'), the line element has exactly the form
$$ds^2 = -c^2dt'^2 + dx'^2 + dy'^2 + dz'^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2.$$

Note the minus sign in front of c^2dt^2. This indicates that the spacetime is flat, but non-Euclidean. The line element ds^2 can be positive, negative, or zero. We can put any two infinitesimally separated events into these three categories:

1. $ds^2 > 0$: the two events are **spacelike** separated. The distance separation dominates over the time separation. For example, we can have $dt = 0$ in some frame. Then the two events occur at the same time, or simultaneously.

2. $ds^2 < 0$: the two events are **timelike** separated. The time separation dominates over the distnace separation. For example, setting $dx = dy = dz = 0$ makes two events occur at the same point in space.

3. $ds^2 = 0$: in this case, $c^2dt^2 = dx^2 + dy^2 + dz^2$, or $d\mathbf{x}^2/dt^2 = c^2$. This describes a photon, so such events are called **lightlike** separated or **null** separated.

Since ds^2 is invariant, these qualitative result (spacelike, null, or timelike) are also independent of reference frame or coodinate system. In otherwords, if two events are spacelike separated (not necessarily occur at the same time), one can always find a reference frame such that these two events occur simultaneously. Similarly, if two events are timelike separated, one can always find a reference frame such that they occur at the same place.

A particle's trajectory through spacetime is called its **world line**. Any material particle ($m \neq 0$) follows a timelike world line—$ds^2 < 0$ along the world line. This is equivalent to say that their speed cannot exceed the speed of light. Since $ds^2 < 0$, we will define the quantity $d\tau$, called the **proper time** of that particle, to be

$$d\tau^2 = -\frac{ds^2}{c^2}. \tag{2.2}$$

Note that in the particle's rest frame, $dx = dy = dz = 0$, so the proper time is $d\tau^2 = -(c^2dt^2)/c^2$, or $d\tau = dt$. Thus, the proper time of a particle measures the infinitesimal time interval of a clock that is instantaneously comoving with the particle.

2.1.2 Time Dilation

With the proper time and the spacetime interval, we are able to produce a famous result of special relativity: time dilation. Suppose two events A and B occurs along the world line of a particle as seen from our rest frame with coordinates (ct, x, y, z). We would like to know in the particle's frame, what is the time separation between A and B. The time separation observed in the particle's frame is just its proper time separation integrated along the world line

$$\begin{aligned}
\tau_{AB} &= \int_A^B d\tau = \int_A^B \frac{1}{c}\sqrt{-ds^2} = \frac{1}{c}\int_A^B \sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2} \\
&= \frac{1}{c}\int_{t_A}^{t_B} \sqrt{c^2 - \left(\frac{dx}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2 - \left(\frac{dz}{dt}\right)^2}\, dt \\
&= \int_{t_A}^{t_B} \left[1 - \frac{\mathbf{v}(t)^2}{c^2}\right]^{1/2} dt = \int_{t_A}^{t_B} \frac{dt}{\gamma(t)},
\end{aligned}$$

where

$$\gamma = \left[1 - \frac{\mathbf{v}(t)^2}{c^2}\right]^{-1/2}$$

is the instantaneous **Lorentz factor** of the particle measured in our rest frame. Because $|\mathbf{v}| < c$, the Lorentz factor is always greater than 1. Therefore, $\tau_{AB} = \tau_B - \tau_A \leq t_B - t_A$: moving clocks run slow. We derive this entirely based on spacetime geometry.

2.2 Relativistic Kinematics

2.2.1 Lorentz Transformation

In Newtonian mechanics, suppose an observer in frame S assigns an event A with coordinates (t, x, y, z) and another observer in a frame S' that has a relative velocity v in x-direction to S. Assume both observers are in inertial frames, and their x-axes are aligned. Also, assume that $t = 0$ represents the point when the origins of S and S' (denoted by $\mathcal{O}$ and $\mathcal{O}'$ respectively) overlap. Then according to Galilean transformation, the coordinates (t', x', y', z') of the event A in S' frame will be

$$\begin{aligned} t' &= t, \\ x' &= x - vt, \\ y' &= y, \\ z' &= z. \end{aligned}$$

This is the usual coordinate transformation between inertial frames in classical mechanics. In special relativity, S' will assign the coordinates of A according to the **Lorentz transformation** from S to S',

$$\boxed{\begin{aligned} ct' &= \gamma(ct - \beta x), \\ x' &= \gamma(x - \beta ct), \\ y' &= y, \\ z' &= z, \end{aligned}} \tag{2.3}$$

where c is the speed of light, $\beta = v/c$, and $\gamma = 1/\sqrt{1-\beta^2}$ is the Lorentz factor. A detailed derivation of the Lorentz transformation is given in Appendix A.1. Taking the non-relativistic limit ($v/c \to 0$) reduces the Lorentz transformation to the Galilean transformation. One can also show that the line element, $ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2$, is invariant under Lorentz transformation. The inverse Lorentz transformation (from S' to S) can evidently obtained by replacing β with $-\beta$ (reversing the velocity).

Example 2.1. Time Dilation

The proper time τ of a clock fixed at $\mathcal{O}'$ is just the time t' measured by that clock in S' frame,

$$t' = \gamma\left(t - \frac{vx}{c^2}\right) \equiv \tau.$$

In S frame, the clock is moving at a velocity of v, so its position is $x = vt$. Hence the proper time of the clock is

$$\tau = \gamma\left(t - \frac{v^2t}{c^2}\right) = \gamma\left(1 - \frac{v^2}{c^2}\right)t = \frac{t}{\gamma}.$$

This agrees with the integral $\int dt/\gamma(t)$ when γ is constant.

2.3 Relativistic Dynamics

From now on, we will use the natural units $c = 1$. The Lorentz transformation is given by

$$t' = \gamma(t - vx), \quad x' = \gamma(x - vt), \quad y' = y, \quad z' = z.$$

It looks more symmetric for t' and x'. Also, the line element and proper time are essentially the same,

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 = -d\tau^2.$$

To return to SI units, we shall use dimensional analysis to find out where should the c go. For now, let's develop some math tools that are useful for relativity.

2.3.1 Four-Vectors

In special relativity, we often use Minkowski coordinates with basis vectors $\{\boldsymbol{e}_\mu\} = \{\boldsymbol{e}_t, \boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$. The basis vectors always have *subscript* indices. Conventionally, Greek indices represent numbers $(0, 1, 2, 3) \to (t, x, y, z)$, while Latin indices represent $(1, 2, 3) \to$

(x, y, z). The vector pointing from the origin to an event in S frame can be written as:

$$\boldsymbol{x} = x^0\boldsymbol{e}_0 + x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2 + x^3\boldsymbol{e}_3 = \sum_{\mu=0}^{3} x^\mu \boldsymbol{e}_\mu = (t, x, y, z).$$

Note that all components $\{x^\mu\}$ are labeled with *superscript* indices. To avoid writing the summation symbol everywhere, we will adopt the **Einstein summation convention**: repeated subscript and superscript indices are implicitly summed over. They are called dummy indices. Indices that does not repeat are known as free indices. Also in this notes, I will write four-vectors as bolded-italic (e.g. $\boldsymbol{x}, p, u$) or simply italic (x, p, u) to distinguish with just bolded three-vectors (e.g. $\mathbf{x}, \mathbf{p}, \mathbf{u}$). Therefore, a four-vector can be written compactly as

$$\boldsymbol{x} = x^\mu \boldsymbol{e}_\mu.$$

By definition, the component of any four-vector $\boldsymbol{a}$ (not just position) transform from one inertial frame to another according to the Lorentz transformation:

$$\begin{aligned} a'^0 &= \gamma(a^0 - va^1), \\ a'^1 &= \gamma(a^1 - va^0), \\ a'^2 &= a^2, \\ a'^3 &= a^3. \end{aligned}$$

A four-vector have a "length", or **magnitude**, defined by

$$\boldsymbol{a} \cdot \boldsymbol{a} = -a^0a^0 + a^1a^1 + a^2a^2 + a^3a^3 \begin{cases} < 0, & \implies \text{timelike vector}, \\ = 0, & \implies \text{null/lightlike vector}, \\ > 0, & \implies \text{spacelike vector}. \end{cases}$$

The magnitude is an invariant quantity, just like the line element. (However, the position or displacement four-vectors can only be defined in flat spacetime. In curved spacetime, four-vectors need to be defined in a local, flat, tangent space to the curved spacetime.) The definition of a magnitude can be generalized to a **scalar product** of four-vectors:

$$\boldsymbol{a} \cdot \boldsymbol{b} = -a^0b^0 + a^1b^1 + a^2b^2 + a^3b^3 = \boldsymbol{b} \cdot \boldsymbol{a} \quad \text{or} \quad \boldsymbol{a} \cdot \boldsymbol{b} = \eta_{\mu\nu} a^\mu b^\nu.$$

Here $\eta_{\mu\nu}$ is called the **Minkowski metric**. The $\{\eta_{\mu\nu}\}$ are components of the **metric tensor** in the (t, x, y, z) Minkowski coordinate basis of an inertial frame in special relativity. It can be represented by a diagonal matrix:

$$[\eta_{\mu\nu}] = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.4}$$

How is the metric$\eta_{\alpha\beta}$ related to basis vectors? We can write $\boldsymbol{a} \cdot \boldsymbol{b} = (a^\mu \boldsymbol{e}_\mu) \cdot (b^\nu \boldsymbol{e}_\nu) = a^\mu b^\mu (\boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu)$. Hence

$$\eta_{\mu\nu} = \boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu.$$

This implies that the Minkowski coordinate basis vectors satisfy $\boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu = 0$ if $\mu \neq \nu$—it is an orthogonal basis. Furthermore, they are also unit vectors because all nonzero components of $\eta_{\mu\nu}$ are only -1 or 1. In conclusion, the Minkowski coordinate basis vectors form an *orthonormal* set of basis vectors. Now with the metric, we can write down the most usual notation of the line element:

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu.$$

This is why $\eta_{\mu\nu}$ is called the matric tensor. It enables one to compute the infinitesimal spacetime interval ds between events with infinitesimal differences in coordinates. As we will see later, all geometric properties are stored in the metric tensor.

2.3.2 4-Velocity and 4-Momentum

An important four-vector is the **4-velocity** of a particle along its world line. The conventional spatial three-velocity is tangent to the particle's path that is parametrized by time t. Similarly, we define the 4-velocity u to be a vector tangent to the world line parametrized by its proper time τ. Its component is

$$u^\mu \equiv \frac{dx^\mu}{d\tau}. \tag{2.5}$$

Writing them out explicitly:

$$u^0 = \frac{dt}{d\tau} = \gamma, \quad u^1 = \frac{dx}{d\tau} = \frac{dx}{dt}\frac{dt}{d\tau} = \gamma v^x, \quad u^2 = \gamma v^y, \quad u^3 = \gamma v^z.$$

Theorem 2.1.

The 4-velocity u is a unit timelike vector:

$$u \cdot u = \eta_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = \frac{\eta_{\mu\nu}dx^\mu dx^\nu}{d\tau^2} = \frac{ds^2}{d\tau^2} = -1.$$

As the Minkowski coordinate basis vectors for time is also a timelike unit vector, $\boldsymbol{e}_0 \cdot \boldsymbol{e}_0 = -1$, it suggests that the 4-velocity might be useful as a basis vector in some reference frame.

For a uniformly moving particle, we can define an inertial frame S' in which it is at rest. Then the 4-velocity is in the direction of $\boldsymbol{e}'_0$, the basis vector of time. For an accelerated particle, there is no inertial frame in which it is always at rest, but there is a frame called **momentarily comoving reference frame** (MCRF) which is a momentary inertial frame for the particle. The world line of the particle is curved, but the 4-velocity is still tangent to the world line, or tangent to $\boldsymbol{e}'_0$.

The three momentum in classical mechanics is defined as $m\mathbf{v}$. In special relativity, the **4-momentum** p is defined as the rest mass m times the 4-velocity,

$$p \equiv mu \quad \Longleftrightarrow \quad p^\mu = mu^\mu. \tag{2.6}$$

Since $u = (\gamma, \gamma\mathbf{v})$, the components of the 4-momentum are $(\gamma m, \gamma m\mathbf{v})$. The time component is the relativistic energy of an object $E = \gamma mc^2$ (with $c = 1$), while the spatial components are the relativistic three-momentum $\mathbf{p} = \gamma m\mathbf{v}$. The 4-momentum is a four-vector, so its magnitude should be invariant. In fact, its magnitude is just the rest mass,

$$p \cdot p = -E^2 + \mathbf{p}^2 = \eta_{\mu\nu}p^\mu p^\nu = m^2\eta_{\mu\nu}u^\mu u^\nu = -m^2.$$

Putting the c's back in, we recover the equation of relativistic energy:

$$E^2 = m^2c^4 + \mathbf{p}^2c^2. \tag{2.7}$$

2.3.3 Action and Lagrangian for a Free Paticle

Some review: in classical mechanics, we define the Lagrangian to be $\mathcal{L} = T - V$, where T and P are the kinetic and potential energy, respectively, of the system. The action along a trajectory from point a to point b is defined as

$$S[q_i(t)] = \int_{t_a}^{t_b} dt\, \mathcal{L}(q_i, \dot{q}_i),$$

where q_i is some generalized coordinates that specify every point on the trajectory. By Hamilton's principle (or principle of stationary action), the action of a actual physical trajectory is extremized among all possible trajectories going from a to b. We then extremize the action using calculus of variations. It helps obtain the equation of motion for each $q_i(t)$, which is the Euler-Lagrange equation

$$\frac{\partial\mathcal{L}}{\partial q} - \frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{q}_i} = 0.$$

For a non-relativistic free particle of mass m, if q_i represents (x, y, z), then

$$T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2), \quad V = 0.$$

The Euler-Lagrange equation says that

$$0 = \frac{d}{dt}\frac{\partial\mathcal{L}}{\partial q_i} = \frac{d}{dt}(m\dot{q}_i) = m\ddot{q}_i.$$

This is just Newton's second law for a free particle $m\mathbf{a} = 0$. It says that a free particle should travel at constant velocity in space.

The Lagrangian/action is the most fundamental one among all physical quantities. It tells which physical path a particle will take. Even though observers in each frame can use different coordinates to represent the trajectory, every one should agree with the actual physical trajectory it takes. Hence the Lagrangian/action should be the "most" invariant quantity, so it needs to be Lorentz invariant in special relativity.

The goal now is to find $\mathcal{L}$ for a relativistic particle. The classical Lagrangian cannot do the job because: 1. it has no c and hence no speed limit; 2. it is not Lorentz invariant. Note that the action has dimensions of angular momentum, $[S] = [E \cdot T]$. We want to have some Lorentz-invariant quantities that make the units right. There are three fundamental constants we can use: the speed of light c, Planck constant $\hbar$, and gravitational constant G. (In this section we will write c's out explicitly.) Planck constant $\hbar$ has the unit of angular momentum, but we cannot use it because there is no quantum mechanics going on yet. Of course G is not useful since there is no gravity. The only available constants now are c and m. We have no choice but to set $\mathcal{L} \propto mc^2$. To get dimensions of T, the only Lorentz-invariant candidate is the proper time τ. In conclusion, the action for a relativistic free particle must be of the form

$$S \propto mc^2 \int_{\mathrm{WL}} d\tau,$$

where WL stands for world line. For future convenience, we will choose the constant of proportionality to be -1. To make it an integral over dt in general reference frame, use the relation $d\tau = dt/\gamma$. Then the action along the world line is

$$S = -mc^2 \int_{t_a}^{t_b} \frac{dt}{\gamma} = -mc^2 \sqrt{1-\beta^2} \int_{t_a}^{t_b} dt,$$

where $\beta = v/c$ indicates the speed of this free particle in some inertial frame. We can identify the Lagrangian to be

$$\mathcal{L} = -mc^2 \sqrt{1-\beta^2}. \tag{2.8}$$

Checking the non-relativistic limit:

$$\mathcal{L} = -mc^2 \left(1 - \frac{1}{2}\beta^2 + \cdots\right) = -mc^2 + \frac{1}{2}mv^2 + \mathcal{O}(v^4).$$

Comparing to $\mathcal{L} = T - V$ in classical mechanics, the rest mass energy acts like a potential energy of the particle.

In the last section we stated the definition of 4-momentum without a reason, and we did not prove its conservation. Now with the Lagrangian, it is time to compute the conserved quantities. We know that if a coordinate q_i is cyclic (which means it does not appear in the Lagrangian), then its conjugate generalized momentum $\partial\mathcal{L}/\partial\dot{q}_i$ is conserved. In the free-particle case, there is no explicit x, y, z. The conserved momentum in x is

$$p_x = \frac{\partial\mathcal{L}}{\partial\dot{x}} = -mc^2 \frac{\partial}{\partial\dot{x}} \left(1 - \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{c^2}\right)^{-1/2} = -mc^2 \left(\frac{\dot{x}/c^2}{\sqrt{1-(v/c)^2}}\right) = mc\frac{\dot{x}/c}{\sqrt{1-(v/c)^2}} = \gamma mc\beta_x.$$

The same holds for p_y and p_z. Note that the Lagrangian has no explicit time-dependence, which means the Hamiltonian/total energy is also conserved. The Hamiltonian is defined as

$$\begin{aligned}
H &= \sum_i p_i \dot{q}_i - \mathcal{L} \\
&= (mc\boldsymbol{\beta}\gamma)(c\boldsymbol{\beta}) - \left(-mc^2\sqrt{1-\beta^2}\right) \\
&= mc^2 \left(\gamma\beta^2 + \sqrt{1-\beta^2}\right) \\
&= mc^2 \left[\frac{\beta^2}{\sqrt{1-\beta^2}} + \frac{1-\beta^2}{\sqrt{1-\beta^2}}\right] \\
&= \gamma mc^2.
\end{aligned}$$

This is the conserved relativistic energy. We have proved almost everything. The only problem now is the massless particle, for which $\mathcal{L} = 0$. By the invariant mass

$$p_\mu p^\mu = -E^2 + p^2 = -m^2 = 0.$$

This forces $E = p$ for a massless particle. Then how do we know E, for example, for a photon? It is given by Planck's formula

$E = \hbar\omega = h\nu$.

2.3.4 Observers and Observations

Consider an arbitrary observer moving on a generic timelike world line through spacetime. A clock and coordinates are well defined in their local frame S' Assume the basis vectors in this frame are orthonormal $\boldsymbol{e}'_\mu \cdot \boldsymbol{e}'_\nu = \eta_{\mu\nu}$. Note that the observer's frame S' need not be an inertial frame. There will always be an inertial frame at any moment which is instantaneously at rest (the MCRF) with S'.

The observer is always at rest in S', so their 4-velocity is $u = \boldsymbol{e}'_0$. Suppose the observer measures a particle with 4-momentum p, we want to know the energy and momentum of this particle in their frame. Recall that the energy is the zeroth *component* of the momentum. To obtain components, we just project a vector onto the basis vector, just like we do for ordinary 3-vectors. Thus, the energy of that particle in frame S' is

$$E' = -p \cdot \boldsymbol{e}'_0 = -p \cdot u.$$

Note that there is a minuc sign before the dot product. It comes from the metric tensor $\eta_{\mu\nu}$ (see the following theorem).

Theorem 2.2. Projection of 4-momentum

An observer with an instantaneous orthonormal basis with $\boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu = \eta_{\mu\nu}$ measures the components p^ν of a particle's 4-momentum. This component p^ν satisfy

$$\eta_{\mu\nu} p^\nu = p \cdot \boldsymbol{e}_\mu \tag{2.9}$$

Explicitly,

$$-p^0 = p \cdot \boldsymbol{e}_0, \quad p^1 = p \cdot \boldsymbol{e}_1, \quad p^2 = p \cdot \boldsymbol{e}_2, \quad p^3 = p \cdot \boldsymbol{e}_3.$$

Proof. Let the 4-momentum be expressed as components and basis vectors:

$$p = p^0\boldsymbol{e}_0 + p^1\boldsymbol{e}_1 + p^2\boldsymbol{e}_2 + p^3\boldsymbol{e}_3 = p^\nu \boldsymbol{e}_\nu.$$

Take the inner product with $\boldsymbol{e}_\mu$,

$$p \cdot \boldsymbol{e}_\mu = p^\nu \boldsymbol{e}_\nu \cdot \boldsymbol{e}_\mu = p^\mu \eta_{\nu\mu}.$$

Since the metric tensor is symmetric, $\eta_{\mu\nu} = \eta_{\nu\mu}$, this proves the theorem,

$$\eta_{\mu\nu} p^\nu = p \cdot \boldsymbol{e}_\mu.$$

In fact, the projection method can be generalized to any four-vector. It is not limited to 4-momentum. □

Example 2.2. Projection vs. Lorentz transformation

Consider an inertial frame S with orthonormal basis $\boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu = \eta_{\mu\nu}$. The observer is traveling at an instantaneous 4-velocity

$$u = \gamma_{\text{ob}}\boldsymbol{e}_0 + \gamma_{\text{ob}} v_{\text{ob}} \boldsymbol{e}_1,$$

where γ_{ob} is the (instantaneous) Lorentz factor of the observer, and v_{ob} is their spatial velocity. (We also assume that at the instant the observer is traveling in the x-direction.) The particle's 4-momentum as seen in frame S is given by

$$p = m\gamma_p \boldsymbol{e}_0 + m\gamma_p v_p \boldsymbol{e}_1,$$

where γ_p and v_p are the particle's Lorentz factor and spatial velocity, respectively. Then according to Theorem 2.2, the observer will measure the energy of the particle to be

$$E' = -p \cdot u = -(-m\gamma_p\gamma_{\text{ob}} + m\gamma_p\gamma_{\text{ob}} v_p v_{\text{ob}}) = m\gamma_p\gamma_{\text{ob}}(1 - v_p v_{\text{ob}}) = \frac{m(1 - v_p v_{\text{ob}})}{\sqrt{1 - v_p^2}\sqrt{1 - v_{\text{ob}}^2}}.$$

As a check, we will use the Lorentz transformation to derive E':

$$E' = \gamma_{\text{ob}}(p^0 - v_{\text{ob}} p^1) = \gamma_{\text{ob}}(m\gamma_p - v_{\text{ob}} m\gamma_p v_p) = m\gamma_p\gamma_{\text{ob}}(1 - v_p v_{\text{ob}}) = \frac{m(1 - v_p v_{\text{ob}})}{\sqrt{1 - v_p^2}\sqrt{1 - v_{\text{ob}}^2}}.$$

It seems like the Lorentz transformation is more direct and efficient. It turns out that the projection method works better in curved spacetime in general relativity.

Example 2.3. Photons

A photon with energy E has a 4-momentum $p = (E, \mathbf{p})$ and $|\mathbf{p}| = E$. It is possible to construct another 4-vector by dividing p by $\hbar$. According to quantum mechanics, the energy of a photon is related to its (angular) frequency ω by $E = \hbar\omega = h\nu$. The spatial momentum is related to its wave vector $\mathbf{k}$ by $\mathbf{p} = \hbar\mathbf{k}$. Hence

$$k = \frac{p}{\hbar} = (\omega, \mathbf{k})$$

is also a 4-vector, called the **wave 4-vector**. Since $k^2 = p^2/\hbar^2 = 0$, we must have $|\mathbf{k}| = \omega$. Suppose in frame S, an observer is moving at constant speed v along the x-axis. There is a source shooting photons in a straight line with an angle α with the x-axis and it meets the observer. We want to find the relativistic Doppler shift measured by the observer. Since the Doppler shift is directly related to ω, so the problem turns to finding the timelike component of the wave 4-vector.

We know the timelike basis vector of the observer is $u = (\gamma, \gamma v, 0, 0)$, and the wave 4-vector of this photon is

$$k = (\omega, \omega\cos\alpha, \omega\sin\alpha, 0),$$

both in frame S. Let the observer's basis vector be orthonormal, $\boldsymbol{e}'_\mu \cdot \boldsymbol{e}'_\nu = \eta_{\mu\nu}$. To find the timelike component of k in the observer's frame S', just take the inner product with the timelike basis vector u of the observer and multiply by a minus sign:

$$\omega' = -k \cdot u = +\gamma\omega - \gamma v\omega\cos\alpha = \gamma\omega(1 - v\cos\alpha).$$

The minus sign comes from

$$k \cdot u = (k'^\mu \boldsymbol{e}'_\mu) \cdot \boldsymbol{e}'_0 = k'^\mu \eta_{\mu 0} = k'^0 \eta_{00} = -\omega'.$$

The same as in Theorem 2.2. The ratio of ω' and ω defines the Doppler shift:

$$\frac{\omega'}{\omega} = \gamma(1 - v\cos\alpha) = \frac{1 - v\cos\alpha}{\sqrt{1 - v^2}}.$$

To check whether this makes sense, let's consider some special cases. First, let $\alpha = 0$ and $v > 0$, i.e. the photon and the observer are moving in the same $+x$-direction. The Doppler shift is $\sqrt{(1 - v)/(1 + v)} < 1$; the photon appears redshifted. If $\alpha = 0$ and $v < 0$, then $\sqrt{(1 - v)/(1 + v)} > 1$ and the photon is blueshifted. Next, let $\alpha = \pi/2$, which means the photon is in the transverse direction ($+y$-direction). The Doppler shift is $1/\sqrt{1 - v^2} = \gamma$. This is because the observer is time-dilated relative to the photon source, $dt' = dt/\gamma$. The photon then appears blueshifted according to the observer as it receives more wave crests per unit time.

Example 2.4. Relativistic Beaming

To be written.

2.4 Vectors and Tensors

This section deals with more formal operations on vectors and tensors.

2.4.1 Vectors

A **vector space** is a set of objects called vectors that are linear under addtion and scalar multiplication. Consider two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in a vector space V and two scalars a and b. Then

$$(a+b)(\boldsymbol{u}+\boldsymbol{v}) = a\boldsymbol{u} + b\boldsymbol{u} + a\boldsymbol{v} + b\boldsymbol{v}.$$

Vectors can be decomposed into components with respect to some set of basis vectors. A basis $\{\boldsymbol{e}_\mu\}$ is any set of linearly independent vectors that span the vector space. This means any vector can be written as a linear combination of basis vectors, and no vector in the basis is a linear combination of other basis vectors. We write the decomposition of a vector into basis vectors as

$$\boldsymbol{u} = u^\mu \boldsymbol{e}_\mu.$$

For a loose notation, we often (and will) call u^μ as the vector (but we should always remember that they are components). The number of vectors in a basis is the **dimension** of the vector space. We shall use Minkowski spacetime as an example.

Example 2.5. Minkowski Spacetime

Consider a vector $\boldsymbol{v}$ at a particular point p in spacetime. We write its coordinates as (v^0, v^1, v^2, v^3). Here v^μ are the components. The abstract vector $\boldsymbol{v}$ can be decomposed into

$$\boldsymbol{v} = v^0\boldsymbol{e}_0 + v^1\boldsymbol{e}_1 + v^2\boldsymbol{e}_2 + v^3\boldsymbol{e}_3.$$

Hence the vector space (called the **tangent space** T_p) around a point in Minkowski spacetime is 4-dimensional. If we write $\boldsymbol{v}$ as column vectors, then the Lorentz transformation would be

$$v^{\mu'} = \Lambda^{\mu'}{}_\nu v^\nu \quad \Longleftrightarrow \quad \begin{bmatrix} v^{0'} \\ v^{1'} \\ v^{2'} \\ v^{3'} \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v^0 \\ v^1 \\ v^2 \\ v^3 \end{bmatrix}.$$

Note that the abstract vector itself does not transform. It is the components v^μ that transform. Thus, we can work out the transformation rule of basis vectors,

$$\boldsymbol{v} = v^\mu \boldsymbol{e}_\mu = v^{\nu'}\boldsymbol{e}_{\nu'} = \Lambda^{\nu'}{}_\mu v^\mu \boldsymbol{e}_{\nu'}.$$

The basis vectors transform under the inverse of the Lorentz transformation:

$$\boldsymbol{e}_\mu = \Lambda^{\nu'}{}_\mu \boldsymbol{e}_{\nu'} \quad \Longrightarrow \quad \boldsymbol{e}_{\nu'} = \Lambda^\mu{}_{\nu'}\boldsymbol{e}_\mu.$$

2.4.2 Dual Vectors

For a vector space V, there is another associated vector space called the **dual vector space** V^*. It is the space of all linear maps from the original vector space to real numbers. Let $\omega \in V^*$, $\boldsymbol{u}, \boldsymbol{v} \in V$, and $a, b \in \mathbb{R}$, then

$$\omega(a\boldsymbol{u} + b\boldsymbol{v}) = a\omega(\boldsymbol{u}) + b\omega(\boldsymbol{v}) \in \mathbb{R}.$$

The dual vector space is also a vector space, so

$$(a\omega + b\eta)(\boldsymbol{u}) = a\omega(\boldsymbol{u}) + b\eta(\boldsymbol{u}).$$

We can always find a set of basis dual vectors $\{\boldsymbol{e}^{*\mu}\}$ such that

$$\boldsymbol{e}^{*\nu}(\boldsymbol{e}_\mu) = \delta^\nu{}_\mu.$$

Each dual vector in V^* can be written as a linear combination of basis vectors:

$$\omega = \omega_\mu \boldsymbol{e}^{*\mu}.$$

Again, for a loose notation, we often call ω_μ as the dual vector (but they are in fact components). The action of a dual vector on a vector, in component notation, is

$$\omega(\boldsymbol{v}) = \omega_\mu \boldsymbol{e}^{*\mu}(v^\nu \boldsymbol{e}_\nu) = \omega_\mu v^\nu \boldsymbol{e}^{*\mu}(\boldsymbol{e}_\nu) = \omega_\mu v^\nu \delta^\mu{}_\nu = \omega_\mu v^\mu \in \mathbb{R}.$$

Then the components can be equivalently written as $\omega_\mu = \omega(\boldsymbol{e}_\mu)$ because

$$\omega(\boldsymbol{e}_\mu) = \omega_\nu \boldsymbol{e}^{*\nu}(\boldsymbol{e}_\mu) = \omega_\nu \delta^\nu{}_\mu = \omega_\mu.$$

Because ω_μ and v^μ are just numbers, we can equally write $\omega(\boldsymbol{v}) = v^\mu \omega_\mu$. This suggests that we can think of vectors as linear maps from the dual vector space to real numbers:

$$\boldsymbol{v}(\omega) = v^\mu \omega_\mu = \omega_\mu v^\mu = \omega(\boldsymbol{v}).$$

In other words, the dual space $(V^*)^*$ to the dual vector space V^* is just V. The transformation of dual vectors is the inverse of the transformation of original vectors. This is required such that the scalar $\omega_\mu v^\mu$ to remain unchanged in transformation. The basis dual vectors, transform the same way as the transformation of original vectors. We will conclude them here:

$$v^{\mu'} = \Lambda^{\mu'}{}_\nu v^\nu, \quad \boldsymbol{e}_{\mu'} = \Lambda^\nu{}_{\mu'} \boldsymbol{e}_\nu, \quad \omega_{\mu'} = \Lambda^\nu{}_{\mu'} \omega_\nu, \quad \boldsymbol{e}^{*\mu'} = \Lambda^{\mu'}{}_\nu \boldsymbol{e}^{*\nu}.$$

These relations are completely consistent with index notation.

2.4.3 Tensors

Tensors are generalization of vectors and dual vectors. A **tensor** T of rank (k, l) is a multilinear map from a collection of k dual vectors and l vectors to $\mathbb{R}$:

$$T(\omega_1, \omega_2, \dots, \omega_k, \boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_l) \in \mathbb{R}.$$

Note that $\omega_1, \dots, \omega_k$ and $\boldsymbol{v}_1, \dots, \boldsymbol{v}_l$ are distinct dual vectors and vectors, not components. Multilinearity means that a tensor acts linearly in each of its arguments. For example, a rank $(1, 1)$ tensor has multilinearity

$$T(a\omega + b\eta, c\boldsymbol{u} + d\boldsymbol{v}) = acT(\omega, \boldsymbol{u}) + adT(\omega, \boldsymbol{v}) + bcT(\eta, \boldsymbol{u}) + bdT(\eta, \boldsymbol{v}).$$

To be less abstract, let's view some examples:

- A $(0, 0)$ tensor is a scalar in $\mathbb{R}$.
- A $(0, 1)$ tensor is a dual vector: it maps 1 vector to real numbers.
- A $(1, 0)$ tensor is a vector: it maps 1 dual vector to real numbers.
- A $(0, 2)$ tensor can be the metric tensor, which takes 2 vectors and map them to real numbers by via inner product $\boldsymbol{u} \cdot \boldsymbol{v}$. The inner product, by definition, is linear in both its argument.

We can manipulate tensors by usual addition and scalar multiplication. Note that only tensors of the same rank can be added together, e.g. you cannot add a scalar to a vector. If T is a (k, l) tensor and S is an (m, n) tensor, then their **tensor product** is a $(k + m, l + n)$ tensor $T \otimes S$, defined by

$$T \otimes S(\omega_1, \dots, \omega_k, \dots, \omega_{k+m}, \boldsymbol{v}_1, \dots, \boldsymbol{v}_l, \dots, \boldsymbol{v}_{l+n}) = T(\omega_1, \dots, \omega_k, \boldsymbol{v}_1, \dots, \boldsymbol{v}_l) \times S(\omega_{k+1}, \dots, \omega_{k+m}, \boldsymbol{v}_{l+1}, \dots, \boldsymbol{v}_{l+n}),$$

where $\times$ means ordinary multiplication of scalars. The tensor product helps us construct basis from vectors and dual vectors. In particular, the basis for a (k, l) tensor can be constructed by taking the tensor product of the basis vectors of k dual vectors and l vectors:

$$\boldsymbol{e}_{\mu_1} \otimes \cdots \otimes \boldsymbol{e}_{\mu_k} \otimes \boldsymbol{e}^{*\nu_1} \otimes \cdots \otimes \boldsymbol{e}^{*\nu_l}.$$

The components of tensors then follows

$$T = T^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l}\boldsymbol{e}_{\mu_1}\otimes\cdots\otimes\boldsymbol{e}_{\mu_k}\otimes\boldsymbol{e}^{*\nu_1}\otimes\cdots\otimes\boldsymbol{e}^{*\nu_l}.$$

The components can be equivalently written as

$$T^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l} = T(\boldsymbol{e}^{*\mu_1},\ldots,\boldsymbol{e}^{*\mu_k},\boldsymbol{e}_{\nu_1},\ldots,\boldsymbol{e}_{\nu_l}).$$

The transformation rule of tensors follows

$$T^{\mu_1'\cdots\mu_k'}{}_{\nu_1'\cdots\nu_l'} = \Lambda^{\mu_1'}{}_{\mu_1}\cdots\Lambda^{\mu_k'}{}_{\mu_k}\Lambda^{\nu_1}{}_{\nu_1'}\cdots\Lambda^{\nu_l}{}_{\nu_l'}T^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l}.$$

That is, each upper index transform like a vector, and each lower index transform like a dual vector. If we define tensors via the transformation rule, a $(1,1)$ tensor can also be viewed as a map from vectors to vectors:

$$v^{\nu} = T^{\mu}{}_{\nu}v^{\nu},$$

and so on for other tensors because they all satisfy the transformation rule. A special operation of tensors is **contraction**. We can sum over one upper index and one lower index by

$$S^{\mu\alpha}{}_{\nu\alpha} = T^{\mu}{}_{\nu}.$$

3 Perfect Fluids in Special Relativity

A **continuum** is a collection of particles so numerous that the dynamics of individual particles cannot be followed. A **fluid** is a continuum that flows. A fluid is characterized by the weakness of antislipping forces compared to the direct push-pull force, the pressure. A **perfect fluid** is defined as one in which all antislipping forces are zero, and the only force between enighboring fluid element is pressure.

3.1 Dust

A **dust** is a simplest fluid, defined to be a collection of particles, all of which are at rest in one Lorentz frame.

3.1.1 Number Density and Flux

First, we are interested in number density of these particles, the number of particles per unit volume. Define the **number density** to be n,

$$n = \text{number density in the MCRF of the element.}$$

In a moving frame $\bar{\mathcal{O}}$ with velocity v relative to MCRF frame. The number of particles in MCRF should be the same, but they seem not to occupy the same volume. Consider a rectangular solid of dimension $\Delta x \Delta y \Delta z$. It will be Lorentz contracted to $\Delta x \Delta y \Delta z \sqrt{1-v^2}$. Therefore, the number density in $\bar{\mathcal{O}}$ should be

$$\bar{n} = \frac{n}{\sqrt{1-v^2}}.$$

Next, we want to know how many particles are moving in a certain direction. Define the **flux** of particles across a surface to be the number crossing a unit area of that surface in a unit time. In the rest frame, the flux of dust is zero because all particles are at rest. In frame $\bar{\mathcal{O}}$, suppose the particles all move with velocity v in $\bar{x}$-direction through a surface $\mathcal{S}$ perpendicular to $\bar{x}$. In time interval $\Delta \bar{t}$, the particles travelling through ΔA of $\mathcal{S}$ is in volume $v \Delta \bar{t} \Delta A$. Thus, the flux of the particles is

$$\Phi^{\bar{x}} = \frac{nv}{\sqrt{1-v^2}}.$$

3.1.2 The Number-Flux Four-Vector $\vec{N}$

Consider the vector $\vec{N} = N^\mu$ defined by

$$N^\mu = n u^\mu$$

where u^μ is the 4-velocity of all particles and n is the number-density.

In frame $\bar{\mathcal{O}}$ in which the particles have a velocity (v^x, v^y, v^z), the four velocity have components

$$[u^\mu]_{\bar{\mathcal{O}}} = \left(\frac{1}{\sqrt{1-v^2}}, \frac{v^x}{\sqrt{1-v^2}}, \frac{v^y}{\sqrt{1-v^2}}, \frac{v^z}{\sqrt{1-v^2}} \right),$$

so

$$[N^\mu]_{\bar{\mathcal{O}}} = \left(\frac{n}{\sqrt{1-v^2}}, \frac{nv^x}{\sqrt{1-v^2}}, \frac{nv^y}{\sqrt{1-v^2}}, \frac{nv^z}{\sqrt{1-v^2}} \right),$$

Same as the 4-velocity, the four vector N^μ is frame-independent. The four vector N^μ also satisfies

$$\begin{aligned}\eta_{\mu\nu}N^\mu N^\nu &= -\frac{n^2}{1-v^2} + \frac{n^2(v^x)^2}{1-v^2} + \frac{n^2(v^y)^2}{1-v^2} + \frac{n^2(v^z)^2}{1-v^2} \\ &= -\frac{n^2}{1-v^2} + \frac{v^2}{1-v^2} \\ &= -\frac{n^2(1-v^2)}{1-v^2} \\ &= -n^2.\end{aligned}$$

3.1.3 A One-Form Defines a Surface

A surface can be defined as the solution to the equation

$$\phi(t, x, y, z) = \text{const.}.$$

The gradient of ϕ, $\tilde{d}\phi$, has direction normal to the surface (so as any multiple of $\tilde{d}\phi$). When the surface is not null, we can define the unit-normal one-form as

$$\tilde{n} = \frac{\tilde{d}\phi}{\left|\tilde{d}\phi\right|},$$

where $\left|\tilde{d}\phi\right|$ is the magnitude of $\tilde{d}\phi$,

$$\left|\tilde{d}\phi\right| = \sqrt{\eta^{\mu\nu}(\partial_\mu\phi)(\partial_\nu\phi)}.$$

As in three-dimensional vector calculus, the surface element is defined as the unit normal times the area element in the surface,

$$d\mathbf{S} = \hat{\mathbf{n}}\, dA = \hat{\mathbf{n}}\, dxdy.$$

Same idea in SR, we can have a volume element represented by

$$\tilde{n}\, dx^\alpha dx^\beta dx^\gamma$$

with α, β, γ all distinct and dx's infinitesimals.

i

3.2 The Stress-energy tensor

GR is not about how many dust particles. It is about energy and momentum which can generate gravitational field.

Assume all the dust particles have the same rest mass m. In the MCRF, the energy of each particle is m, so the energy per unit volume is

$$\rho = mn.$$

In general fluids, there will be kinetic energy so this definition is not precise.

In frame $\bar{\mathcal{O}}$, the number density is $n/\sqrt{1-v^2}$ and the energy of each particle is $m/\sqrt{1-v^2}$. Thus, the energy density is

$$\bar{\rho} = \frac{\rho}{1-v^2}.$$

Since this transformation is of two factors of $(1-v^2)^{-1/2}$, it is a component of a $\binom{2}{0}$ tensor instead of a vector. This tensor is called the **stress energy tensor**,

$$T^{\mu\nu} = \{\text{flux of } \mu \text{ component of momentum across a surface of constant } x^\nu\}.$$

T^{00} is the flux of energy across a surface with t constant—this is the energy density.
T^{0i} is the flux of energy across a surface with x^i constant.

T^{i0} is the flux of ith momentum across the surface $t = \text{const}$.
T^{ij} is the flux of ith momentum across j surface.

For dust, there is no motion of particles. Therefore, in the MCRF we have

$$T^{00} = \rho = mn, \quad T^{0i} = T^{i0} = T^{ij} = 0.$$

It is obvious that in the MCRF, the tensor $T^{\mu\nu}$ satisfies

$$T^{\mu\nu} = p^\mu N^\nu$$

where $p^\mu = mu^\mu$ and $N^\nu = nu^\nu$. Hence, the stress energy tensor can also be expressed as

$$T^{\mu\nu} = \rho u^\mu u^\nu.$$

In frame $\bar{\mathcal{O}}$, we have

$$\begin{aligned}
T^{\bar{0}\bar{0}} &= \rho U^{\bar{0}} U^{\bar{0}} = \rho/(1-v^2), \\
T^{\bar{0}\bar{i}} &= \rho U^{\bar{0}} U^{\bar{i}} = \rho v^i/(1-v^2) \\
T^{\bar{i}\bar{0}} &= \rho U^{\bar{i}} U^{\bar{0}} = \rho v^i/(1-v^2) \\
T^{\bar{i}\bar{j}} &= \rho U^{\bar{i}} U^{\bar{j}} = \rho v^i v^j/(1-v^2).
\end{aligned}$$

Notice that $T^{\mu\nu}$ is symmetric. This property is not just for dust; it is true in general.

3.3 General Fluids

General fluids contains particles that move with random velocities and there may be forces that raise potential energy between particles.

All scalar quantities associated with a fluid element in relativity (such as number density, energy density, and temperature) are defined to be their values in the MCRF. To avoid interpenetrating flows within the fluid, we consider the one that contain only one kind of particle.

3.3.1 First Law of Thermodynamics

The fluid element can exchange energy with its surroundings in two ways: by heat conduction (δQ) and by work ($P\delta V$). Let E be the total energy of the element, then δQ is energy gained and $P\delta V$ is energy lost, they are related by

$$\delta E = \delta Q - P\delta V.$$

If there are N particles in the fluid, we can write

$$V = \frac{N}{n}, \quad \delta V = -\frac{N}{n^2}\delta n.$$

From the definition of energy density ρ,

$$E = \rho V = \rho\frac{N}{n} \quad \Longrightarrow \quad \delta E = \rho\delta V + V\delta\rho.$$

These two results imply

$$\delta Q = \frac{N}{n}\delta\rho - N(\rho + P)\frac{\delta n}{n^2}.$$

Defining $q = Q/N$ to be heat absorbed per particle and taking the infinitesimal change,

$$n\, dq = d\rho - \frac{\rho + P}{n}dn.$$

According to the general theory of first-order differential equation, there exist two functions A and B of only ρ and n, such that

$$d\rho - \frac{\rho + P}{n}dn \equiv A\, dB.$$

In thermodynamics, we define temperature to be A/n and specific entropy B:

$$d\rho - \frac{\rho + P}{n} dn = nT\, dS, \tag{3.1}$$

or, more familiarly

$$\delta Q = T\delta S.$$

3.4 Perfect Fluids

A **perfect fluid** in relativity is defined as a fluid that has no viscosity and no heat conduction in the MCRF. These two restrictions simplify the stress-energy tensor.

No heat conduction means that $T^{0i} = T^{i0} = 0$ because energy can flow only if particles flow. Viscosity is a force parallel to the interface between particles, so no viscosity means all forces are perpendicular to the interface. This means $T^{ij} = 0$ unless $i = j$, or T^{ij} should be a diagonal matrix. Moreover, "no viscosity" is a statement independent of the spatial axes, so T^{ij} should be diagonal in all MCRF frames. Hence, T^{ij} is a multiple of the identity matrix. The quantity on the diagonal is the momentum flux—the pressure P.

3.4.1 The Stress-Energy Tensor

In the MCRF, $T^{\mu\nu}$ has the components

$$[T^{\mu\nu}] = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix}.$$

We can also verify that $T^{\mu\nu} = (\rho + P)u^\mu u^\nu + P\eta^{\mu\nu}$. This is the stress-energy tensor of a perfect fluid. The dust is a special case of a pressure-free perfect fluid ($P = 0$). This just simply means a perfect fluid can be pressure free only if its particles have no random motion.

3.4.2 The Conservation Laws

In relativity with stress-energy tensor, the conservation of energy and momentum is

$$\partial_\nu T^{\mu\nu} = 0, \tag{3.2}$$

with $\mu = 0$ as the conservation of energy and $\mu = i$ as the one of momentum.

For a perfect fluid, we have

$$\partial_\nu T^{\mu\nu} = \partial_\nu[(\rho + P)u^\mu u^\nu + P\eta^{\mu\nu}] = 0.$$

Assume the particles are conserved,

$$\partial_\nu N^\nu = \partial_\nu(nu^\nu) = 0.$$

Write the first term of 3.2 as

$$\partial_\nu[(\rho + P)u^\mu u^\nu] = \partial_\nu\left[\frac{\rho + P}{n}u^\mu nu^\nu\right] = nu^\nu \partial_\nu\left(\frac{\rho + P}{n}u^\mu\right).$$

Moreover, $\eta^{\mu\nu}$ is a constant matrix, so $\partial_\gamma \eta^{\mu\nu} = 0$, which follows that

$$\partial_\nu u^\mu U_\mu = 0.$$

Proof. Recall that

$$u^\mu U_\mu = -1 \quad \Longrightarrow \quad \partial_\nu(u^\mu U_\mu) = 0.$$

Also,

$$0 = \partial_\nu(u^\mu U_\mu) = \partial_\nu(u^\mu U^\gamma \eta_{\mu\gamma}) = \partial_\nu(u^\mu U^\gamma)\eta_{\mu\gamma} = 2\partial_\nu u^\mu U^\gamma \eta_{\alpha\gamma}.$$

The last step follows from the symmetry of $\eta_{\mu\gamma}$ that $\partial_\nu u^\mu U^\gamma \eta_{\mu\gamma} = u^\mu \partial_\mu U^\gamma \eta_{\mu\gamma}$. The last expression converts to $2\partial_\mu u^\mu U_\mu$ and it is zero. □

After the above analysis, the original equation reads

$$\partial_\nu T^{\mu\nu} = nu^\nu \partial_\nu \left(\frac{\rho + P}{n} u^\mu\right) + \partial_\nu P \eta^{\mu\nu} = 0.$$

Multiply by U_μ and sum on μ gives the time component in MCRF,

$$nu^\nu U_\mu \partial_\nu \left(\frac{\rho + P}{n} u^\mu\right) + \partial_\nu \eta^{\mu\nu} U_\mu = 0.$$

Using $\partial_\nu u^\mu U_\mu = 0$ and $u^\mu U_\mu = -1$, we get

$$u^\nu \left[-n\partial_\nu \left(\frac{\rho + P}{n}\right) + \partial_\nu P\right] = 0.$$

A little algebra gives

$$-u^\nu \left[\partial_\nu \rho - \left(\frac{\rho + P}{n}\right)\partial_\nu n\right] = 0.$$

Using the properties of derivatives along the world line ($u^\nu \partial_\nu \rho = d\rho/d\tau$, etc.), we can write this equation in another way

$$\frac{d\rho}{d\tau} - \frac{\rho + P}{n}\frac{dn}{d\tau} = 0.$$

Comparing it with 3.1, this means

$$u^\mu \partial_\mu S = \frac{dS}{d\tau} = 0 \tag{3.3}$$

In conclusion, the flow of a particle-conserving perfect fluid conserves specific entropy and this is called **adiabatic**. Now we write the equation before the consideration of time components to obtain spatial components. Again,

$$nu^\nu \partial_\nu \left(\frac{\rho + P}{n} u^\mu\right) + \partial_\nu P \eta^{\mu\nu} = 0.$$

In the MCRF, $U^i = 0$ but $\partial_\nu U^i \neq 0$. If we only need ith components, we have

$$nu^\nu \partial_\nu \left(\frac{\rho + P}{n} U^i\right) + \partial_\nu P \eta^{i\nu} = 0.$$

Since $U^i = 0$, the ν derivative of $(\rho + P)/n$ vanishes, then

$$(\rho + P)\partial_\nu U^i u^\nu + \partial_\nu P \eta^{i\nu} = 0.$$

Lowering the index changes nothing and since $\eta_i^\nu = \delta_i^\nu$,

$$(\rho + P)\partial_\nu U_i u^\nu + \partial_i P = 0.$$

Recall that $\partial_\nu U_i u^\nu$ is the four-acceleration a_i, so

$$(\rho + P)a_i + \partial_i P = 0 \tag{3.4}$$

3.5 Further Concepts

3.5.1 Importance for GR

The stress-energy tensor plays an important role in GR. Instead of using mass-density ρ_0 as a source of gravitational field, we should use T^{00}. However, T^{00} is only one component of the stress-energy tensor and only using this will end up in a noninvariant theory of gravity. Therefore, Einstein guessed that all the components (energy, stresses, pressures and momenta) are sources of gravity. Combining these with curvature gives the GR.

Pressure is more significant in GR than in Newtonian theory because first, it is a source of field. Second, the gravitational field of Newtonian stars are nearly the same as in GR when $P \ll \rho$. Things change whn $P \approx \rho$—this increase in pressure will end up in a gravitational collapse as we will see later. This will occur for very dense material like neutron stars or a “relativistic gas” with particles moving in the speed of light.

3.5.2 The Gauss' Law

Similar to three-dimensional vector calculus, we have the relation between divergence and the “surface integral” in spacetime:

$$\int_V \partial_\mu F^\mu \, d^4x = \oint_{\partial V} F^\mu \tilde{n}_\mu d^3x,$$

where $\tilde{n}$ is the unit-normal one-form and ∂V is the three-volume of the three-dimensional hypersurface bounding the four-dimensional volume of integration.

4 Curved Manifolds

4.1 On the Relation of Gravitation to Curvature

The difference between special relaticity and general relativity is that the latter takes gravitational fields into account. We already know that in special relativity, things happens differently in frames with relative velocity such as time dilation. This is the case where we neglect gravitational fields. With a non-uniform gravitational field in one frame, clocks do not run at the same rate either.

4.1.1 The Principle of Equivalence

Gravity is the special one among all forces. Objects with the same velocity follow the same trajectory in a gravitational field, regardless of their internal composition. Other forces does not have this property: electromagnetism affects only charged particles and their trajectory depends on the charge-mass ratio. Strong and weak interactions are similar—different particles act differently. Only gravity acts on all particles. Alternatively, we can think of *mass* as the "charge" of gravity so the trajectory depends on the "mass-mass" ratio, which obviously does not affect the trajectory of an object.

A frame that falls freely in a gravitational field is a frame in which particles keep a uniform velocity. This works locally when the gravitational field is not uniform. Also, there are infinitely many freely falling frames at any point because they can differ in velocities and in orientation. A free fall frame is actually the "true" inertial frame and Einstein came up with general relativity using this hypothesis. Consider two frames, each of which has an identical particle in it:

1. A frame with a uniform downward gravitational field g.
2. A frame with a uniform upward acceleration with magnitude $a = g$, e.g. a rocket in free space accelerating.

The second frame is a "non-inertial" frame in the sense of special relativity. If we stick to this frame, we will see the particle accelerating downward, as if it is in a gravitational field. One in the frame, with measurements all solely based on experiments in this frame, cannot tell whether the frame is accelerated or is in a gravitational field. This leads to the **Principle of Equivalence**:

The Principle of Equivalence

There is no experiment that locally can distinguish a uniform gravitational field from a laboratory which undergoes uniform acceleration.

Now consider the same two frame, but now there is a laser beam on the left wall that emits light horizontally. If you are in the second frame, light bends downward because the frame is accelerating upward. By the equivalence principle, the light beam in the frame with gravity should also bend downward. Since the two frames are equivalent, we arrive at a very astonishing conclusion: light in a gravitational field follows a curved path. Because light travels along null geodesics, this is saying that gravity curves space or spacetime.

Minkowski space is a flat space with Euclid's parallelism axiom but it is not a Euclidean space. For example, photons with straight world lines have zero proper length. Thus, SR has a flat, non-Euclidean geometry.

In a nonuniform gravitational field, parallel world lines of two particles do not generally remain parallel, so gravitational spacetime is not flat. Einstein proposed that objects follows the geodesics that appears to be locally straight parallel lines but not globally. Hence the theory of gravity uses a curved spacetime to represent the effects of gravity on particles' trajectory. We need to first study the mathematics of curvature—differential geometry.

4.2 The Metric Tensor

4.2.1 Properties of the Metric Tensor

The line element in curved spacetime is defined by the metric tensor $g_{\mu\nu}$, as in Minkowski spacetime,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu.$$

What becomes complicated in general relativity is that now elements of $g_{\mu\nu}$ are in general functions of spacetime coordinates. Even in Minkowski spacetime, $g_{\mu\nu}$ can be functions of coordinates. For example, if we express the 3-space using spherical coordinates, then the line element is

$$ds^2 = -dt^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2 \quad \Longleftrightarrow \quad g_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{bmatrix}.$$

Here $g_{\theta\theta} = r^2$ and $g_{\phi\phi} = r^2 \sin^2\theta$ are functions of coordinates. Because ds^2 depends only on the geometry of spacetime, the coordinates we choose can have no physical meaning. To transform coordinates, we use the change rule,

$$dx'^\mu = \frac{\partial x'^\mu}{\partial x_\alpha} dx^\alpha = \frac{\partial x'^\mu}{\partial x^0} dx^0 + \frac{\partial x'^\mu}{\partial x^1} dx^1 + \frac{\partial x'^\mu}{\partial x^2} dx^2 + \frac{\partial x'^\mu}{\partial x^3} dx^3.$$

To transform the components of the metric tensor from one set of coordinates to the other, we use the line element

$$ds^2 = g_{\mu\nu} dx'^\mu dx'^\nu = g_{\mu\nu} \frac{\partial x'^\mu}{\partial x^\alpha} \frac{\partial x'^\nu}{\partial x^\beta} dx^\alpha dx^\beta \quad \Longrightarrow \quad \boxed{g_{\mu\nu} = g_{\alpha\beta} \frac{\partial x'^\alpha}{\partial x^\mu} \frac{\partial x'^\beta}{\partial x^\nu}.} \tag{4.1}$$

Also note that $dx^\mu dx^\nu = dx^\nu dx^\mu$. This is simply saying that matrix multiplication is commutative. But it also indicates that

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = g_{\mu\nu} dx^\nu dx^\mu \quad \Longrightarrow \quad g_{\mu\nu} = g_{\nu\mu}.$$

In other words, the metric tensor is a symmetric tensor. Its matrix representation can always be seen as a symmetric matrix. Hence this 4×4 matrix will have ten independent entries (4 diagonal and 6 off-diagonal). Given that we are allowed to transform coordinates using four functions $x'^\mu(x^0, x^1, x^2, x^3)$, there will be only 6 independent degrees of freedom left without changing any physics.

It is useful to define the **inverse metric** $g^{\mu\nu}$ such that

$$g^{\mu\alpha} g_{\alpha\nu} = \delta^\mu{}_\nu \quad \text{where} \quad \delta^\mu{}_\nu = \begin{cases} 1, & \mu = \nu, \\ 0, & \mu \neq \nu. \end{cases}.$$

In matrices, $g^{\mu\nu}$ and $g_{\mu\nu}$ multiplying together will give the identity matrix. For example, the inverse metric of Minkowski spacetime in spherical coordinates is

$$g^{\mu\nu} = \text{diag}\left(-1, 1, \frac{1}{r^2}, \frac{1}{r^2 \sin^2\theta}\right).$$

The elements of metric tensor is directly related to basis vectors $\{\boldsymbol{e}_\mu\}$ by

$$ds^2 = dx \cdot dx = (dx^\mu \boldsymbol{e}_\mu) \cdot (dx^\nu \boldsymbol{e}_\nu) = dx^\mu dx^\nu (\boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu) = g_{\mu\nu} dx^\mu dx^\nu,$$

or

$$g_{\mu\nu} = \boldsymbol{e}_\mu \cdot \boldsymbol{e}_\nu. \tag{4.2}$$

4.2.2 Local Inertial Frames

By the principle of equivalence, any local effects of gravity can be transformed to a free-falling frame by some coordinate transformations. Mathematically, we can always find a set of coordinates $\{x'^\mu\}$ in which the metric is Minkowski at a specific event in space. We say that a curved spacetime is *locally* flat. In matrix language, $g_{\mu\nu}$ is a symmetric, real matrix. It can always be diagonalized at a point by an orthogonal matrix with columns being eigenvectors, and the transformed matrix is diagonal with entries being its eigenvalues. Both the eigenvalues and eigenvectors are real. Then rescaling the coordinate can make the diagonal entries $(-1, 1, 1, 1)$.

Proof. The transformation of the metric tensor (4.1) can be written as

$$g'_{\mu\nu} = g_{\alpha\beta}\frac{\partial x'^\alpha}{\partial x^\mu}\frac{\partial x'^\beta}{\partial x^\nu} = g_{\alpha\beta}\Lambda^\alpha{}_\mu\Lambda^\beta{}_\nu = \Lambda^\alpha{}_\mu g_{\alpha\beta}\Lambda^\beta{}_\nu.$$

In matrix language,

$$g' = \Lambda^T g\Lambda.$$

Since g can be diagonalized by an orthogonal matrix O, we have

$$g' = O^{-1}gO = O^T gO,$$

where $O^T = O^{-1}$ or $O^T O = I$ for an orthogonal matrix. We see that $\Lambda = O$. Now suppose the diagonal metric tensor g' has diagonal entries $\lambda_0, \lambda_1, \lambda_2, \lambda_3$, which means

$$ds^2 = \lambda_0(dx'^0)^2 + \lambda_1(dx'^1)^2 + \lambda_2(dx'^2)^2 + \lambda_3(dx'^3)^2$$

for some coordinates $\{dx'^\mu\}$. Then we can always rescale dx'^μ to make $g'_{\mu\nu} = \eta_{\mu\nu}$:

$$ds^2 = -(\sqrt{-\lambda_0}\,dx'^0)^2 + (\sqrt{\lambda_1}\,dx'^1)^2 + (\sqrt{\lambda_2}\,dx'^2)^2 + (\sqrt{\lambda_3}\,dx'^3)^2.$$

For physical coordinates, we must require one of the λ's to be negative (in this case λ_0). Otherwise, we cannot rescale a positive eigenvalue to a negative one with dx'^μ squared. □

The **local inertial reference frame** has an even stronger statement.

Theorem 4.1. Local inertial reference frame

It is always possible to find a coordinate system $\{x'^\mu\}$ in which

$$g'_{\mu\nu}\Big|_P = \eta_{\mu\nu} \qquad \text{and} \qquad \frac{\partial g'_{\mu\nu}}{\partial x'^\alpha}\Big|_P = 0 \quad \forall\ \alpha, \mu, \nu \tag{4.3}$$

at any particular point P in spacetime.

Though this is true for any point in spacetime, different points will in general require different new coordinate transformations. This local inertial reference frame is the frame of a free-falling observer at that point.

Now because there always a local inertial frame, the causal structure at a particular point is the same as it is in special relativity,

$$ds^2 \begin{cases} < 0, & \text{timelike,} \\ = 0, & \text{null/lightlike,} \\ > 0, & \text{spacelike.} \end{cases}$$

This has significant consequence. Since ds^2 is frame invariant, the type of separation is preserved under all coordinate transformations. Then we can integrate along world line of particles to find quantities we want. First, light rays always travel between events that are null-separated: $ds^2 = 0$ along its world lines no matter how it is curved in spacetime. Next, for material particles that travel below the speed of light, its proper time is the integral

$$\tau = \int_A^B \sqrt{-g_{\mu\nu}dx^\mu dx^\nu} = \int_A^B d\lambda\,\sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}$$

for any **affine parameter** λ. The choice of affine parameter will not affect any of the physic. We will see that they always cancel in physical equations such as the geodesic equation.

4.3 Differentiable Manifolds and Riemannian Manifolds

4.3.1 Differentiable Manifolds

A **Manifold** is any set that can be continuously parametrized. The number of independent parameters is the **dimension** of the manifold, while the parameters are the **coordinates**. The manifold looks locally like Euclidean space, but globally it may not. For example, the surface of a torus is not Euclidean, but locally it can be mapped one-to-one into the plane tangent to it.

Differentiable manifolds are spaces that are continuous and differentiable. Nearly all manifolds useful in physics are differentiable everywhere. The surface of a sphere is differentiable; a cone except its apex is differentiable; the curved spacetimes of GR are differentiable.

4.3.2 Riemannian Manifolds

A differentiable manifold on which a symmetric $\binom{0}{2}$ tensor field $\mathbf{g}$ acts as the metric at each point is called a **Riemannian manifold**. There is a more strict definition: only if the metric is positive-definite, $\mathbf{g}(\mathbf{v}, \mathbf{v}) > 0$ for all $\mathbf{v} \neq 0$, is called Riemannian. Indefinite metrics like SR and GR are called pseudo-Riemannian. We will see that the metric $\mathbf{g}$ defines the curvature of the manifold.

Now, recall that the metric $\mathbf{g}$ has components $g_{\mu\nu}$ and an inverse of $g^{\mu\nu}$. We can use it to raise of lower index between vector components and corresponding one-form components,

$$V_\mu = g_{\mu\nu} V^\nu.$$

In SR we have studied Lorentz frames, which are inertial. However, the GR does not allow global Lorentz frames because of gravity so we need transformations of $g_{\mu\nu}$. The matrix $[g_{\mu\nu}]$ is a symmetric matrix. According to matrix algebra, there always exist a transformation matrix that transform a symmetric matrix into a diagonal matrix with main diagonal entry $+1$, -1 or 0. The number of $+1$ means the number of positive eigenvalues of $[g_{\mu\nu}]$; the number of -1 means the number of negative eigenvalues. Therefore, if we choose $\mathbf{g}$ to have three positive eigenvalues and one negative, we can find a matrix $\Lambda^{\mu'}{}_{\nu}$ such that

$$[g_{\mu'\nu'}] = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = [\eta_{\mu\nu}].$$

4.3.3 Length, Areas, and Volumes

We already know how to calculate lengths,

$$s_{AB} = \int_A^B ds = \int_A^B d\lambda \sqrt{g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}},$$

assuming that the length is spacelike separated, $ds^2 = g_{\mu\nu} dx^\mu dx^\nu > 0$. Areas and volumes are more complicated.

Example 4.1. Diagonal metric

For simplicity, consider a diagonal metric, ($\{\boldsymbol{e}_\mu\}$ are orthogonal basis vectors),

$$ds^2 = g_{00}(dx^0)^2 + g_{11}(dx^1)^2 + g_{22}(dx^2)^2 + g_{33}(dx^3)^2,$$

with $g_{00} < 0$. A two-dimensional surface of spacetime can be described by the set of events on two constant coordinates. For example, let $dx^0 = dx^1 = 0$. Because we are using a orthogonal coordinate system and spacetime is locally flat, the infinitesimal area dA can be seen as a small rectangle of side lengths $d\ell^2$ and $d\ell^3$,

$$dA = d\ell^2 d\ell^3 \quad \text{where} \quad d\ell^2 = dx|_{\text{fixed } x^0, x^1, x^3} = \sqrt{g_{22}}\, dx^2, \quad d\ell^3 = ds|_{\text{fixed } x^0, x^1, x^2} = \sqrt{g_{33}}\, dx^3.$$

Thus, $dA = \sqrt{g_{22}g_{33}}\, dx^2 dx^3$. 3-volumes is similar. For example, events with the same x^0 constitutes a volume of

$$dV = \sqrt{g_{11}g_{22}g_{33}}\, dx^1 dx^2 dx^3,$$

and finally, the 4-volume

$$dV = \sqrt{-g_{00}g_{11}g_{22}g_{33}}\, dx^0 dx^1 dx^2 dx^3.$$

Example 4.2. Spatial geometry in Schwarzschild spacetime

The Schwarzschild metric describes all spacetime *outside* a spherically symmetric object of mass M (like a star or a black hole). The line element is given by

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta\, d\phi^2.$$

We will derive this metric from Einstein field equation in Chapter 8. For now we will require $r > 2GM$. Consider events that occur at the same coordinate time t. A sphere of fixed r-coordinate has surface area

$$A(r) = \int \sqrt{g_{\theta\theta}}\sqrt{g_{\phi\phi}}\, d\theta\, d\phi = \int \sqrt{r^2}\sqrt{r^2 \sin^2\theta}\, d\theta\, d\phi = r^2 \int_0^{\pi} d\theta\, \sin\theta \int_0^{2\pi} d\phi = 4\pi r^2.$$

This is consistent with Euclidean geometry, but we will see that the spatial geometry is not Euclidean. The problem is with the radial distance separating two spheres at $r = r_1$ and $r = r_2 > r_1$,

$$\Delta R = \int_{r_1}^{r_2} \sqrt{g_{rr}}\, dr = \int_{r_1}^{r_2} \left(1 - \frac{2GM}{r}\right)^{-1/2} dr > \int_{r_1}^{r_2} dr = r_2 - r_1$$

for $r_1, r_2 > 2GM$ (outside the Schwarzschild radius). We see that even though the surface areas of the two spheres are $4\pi r_1^2$ and $4\pi r_2^2$, the radial distance between them is greater than $r_2 - r_1$.

In general, $g_{\mu\nu}$ is not diagonal. We can still get the four-volume. In Minkowski spacetime, the four-volume element is $dx^0 dx^1 dx^2 dx^3$. In any coordinate system $\{x'^\mu\}$, we use the Jacobian to calculate the transformed volume element,

$$dx^0 dx^1 dx^2 dx^3 = \frac{\partial(x^0, x^1, x^2, x^3)}{\partial(x'^0, x'^1, x'^2, x'^3)} dx'^0 dx'^1 dx'^2 dx'^3,$$

where $\partial(\ \ )/\partial(\ \ )$ is the Jacobian of the transformation:

$$\frac{\partial(x^0, x^1, x^2, x^3)}{\partial(x^{0'}, x^{1'}, x^{2'}, x^{3'})} = \det(\Lambda^\mu{}_\nu).$$

Calculating this determinant of this 4×4 matrix is tedious. There is an easier way using the metric. We know that $g = [g_{\mu\nu}]$ can always be diagonalized as

$$\eta = \Lambda^T g \Lambda,$$

where η is the Minkowski metric. Taking the determinant on both sides,

$$\det \eta = (\det \Lambda^T)(\det g)(\det \Lambda) = (\det \Lambda)^2 \det g.$$

Since $\det \eta = -1$, we have

$$\det(\Lambda^\mu{}_\nu) = \sqrt{-\det g}.$$

4.3.4 Geodesic Equation

The world lines of free particles in a gravitational field are **geodesics** of the spacetime. To find the trajectory of a free particle from events A and B, we extremize its proper time:

$$\tau = \int_A^B d\lambda \sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}.$$

The Euler-Lagrange equation reads,

$$\frac{\partial \mathcal{L}}{\partial x^\mu} - \frac{d}{d\lambda}\left[\frac{\partial \mathcal{L}}{\partial(dx^\mu/d\lambda)}\right] = 0, \quad \text{where} \quad \mathcal{L} = \sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}} = \frac{d\tau}{d\lambda}$$

can be seen as the Lagrangian. Before we go on, we should look at a simple example.

Example 4.3. Geodesic equation in cylindrical polar coordinates

Consider a flat spacetime in cylindrical polar coordinates:

$$d\tau^2 = dt^2 - dr^2 - r^2 d\phi^2 - dz^2.$$

The Lagrangian is

$$\mathcal{L} = \frac{d\tau}{d\lambda} = \left[\left(\frac{dt}{d\lambda}\right)^2 - \left(\frac{dr}{d\lambda}\right)^2 - r^2\left(\frac{d\phi}{d\lambda}\right)^2 - \left(\frac{dz}{d\lambda}\right)^2\right]^{1/2}.$$

Then calculate the derivatives needed in the Euler-Lagrange equation. For example,

$$\frac{\partial \mathcal{L}}{\partial(dt/d\lambda)} = \frac{1}{2\mathcal{L}}\frac{2\,dt}{d\lambda} = \frac{d\lambda}{d\tau}\frac{dt}{d\lambda} = \frac{dt}{d\tau}.$$

Other derivatives are similar,

$$\frac{\partial \mathcal{L}}{\partial(dr/d\lambda)} = -\frac{dr}{d\tau}, \quad \frac{\partial \mathcal{L}}{\partial(d\phi/d\lambda)} = -r^2\frac{d\phi}{d\tau}, \quad \frac{\partial \mathcal{L}}{\partial(dz/d\lambda)} = -\frac{dz}{d\tau}.$$

$$\frac{\partial \mathcal{L}}{\partial t} = \frac{\partial \mathcal{L}}{\partial \phi} = \frac{\partial \mathcal{L}}{\partial z} = 0, \quad \frac{\partial \mathcal{L}}{\partial r} = -\frac{r}{\mathcal{L}}\left(\frac{d\phi}{d\lambda}\right)^2 = -r\frac{d\lambda}{d\tau}\left(\frac{d\phi}{d\lambda}\right)^2.$$

We can immediately see three almost trivial equations. For example, the Lagrangian does not depend on t, so

$$0 = \frac{d}{d\lambda}\left(\frac{dt}{d\tau}\right) \implies \frac{d\lambda}{d\tau}\frac{d}{d\lambda}\left(\frac{dt}{d\tau}\right) = 0 \implies \boxed{\frac{d^2 t}{d\tau^2} = 0.}$$

The ϕ- and z- equations are similar,

$$0 = \frac{d}{d\lambda}\left(-r^2\frac{d\phi}{d\tau}\right) \implies \boxed{r\frac{d^2\phi}{d\tau^2} + 2\frac{dr}{d\tau}\frac{d\phi}{d\tau} = 0,} \quad \text{and} \quad \frac{d}{d\lambda}\left(-\frac{dz}{d\tau}\right) = 0 \implies \boxed{\frac{d^2 z}{d\tau^2} = 0,}$$

The only nontrivial equation is the r-equation,

$$-r\frac{d\lambda}{d\tau}\left(\frac{d\phi}{d\lambda}\right)^2 - \frac{d}{d\lambda}\left(-\frac{dr}{d\tau}\right) = 0 \implies \boxed{\frac{d^2 r}{d\tau^2} - r\left(\frac{d\phi}{d\tau}\right)^2 = 0.}$$

The r-, ϕ-, and z-equations are the usual acceleration components in cylindrical polar coordinates in Newtonian mechanics, with dt replaced by $d\tau$,

$$\mathbf{a} = (\ddot{r} - r\dot{\phi}^2)\hat{\mathbf{r}} + (r\ddot{\phi} + 2\dot{r}\dot{\phi})\hat{\boldsymbol{\phi}} + \ddot{z}\hat{\mathbf{z}} = 0.$$

The Lagrangian helps find geodesic equations quickly even in curved spacetime. The general **geodesic equation** is written as

$$\boxed{\frac{d^2 x^\alpha}{d\tau^2} + \Gamma^\alpha{}_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0 \iff \frac{du^\alpha}{d\tau} + \Gamma^\alpha{}_{\mu\nu}u^\mu u^\nu.} \tag{4.4}$$

where $\Gamma^\mu{}_{\nu\alpha}$ are known as **Christoffel symbols**. They are related to the metric tensor $g_{\mu\nu}$ by

$$\boxed{\Gamma^\alpha{}_{\mu\nu} = \frac{1}{2} g^{\alpha\beta} \left(\frac{\partial g_{\beta\nu}}{\partial x^\mu} + \frac{\partial g_{\beta\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\beta} \right)} \tag{4.5}$$

See Appendix (A.2) for proof. From (4.5), we can see that $\Gamma^\alpha{}_{\mu\nu} = \Gamma^\alpha{}_{\nu\mu}$: exchanging μ and ν does nothing to the sum of first two terms, and $g_{\mu\nu} = g_{\nu\mu}$ for the third term. Hence instead of $4 \times 4 \times 4 = 64$ components in $\Gamma^\alpha{}_{\mu\nu}$, there are only 4×10 components (4 from α, and 10 from symmetric μ, ν). The Christoffel symbols play important roles in general relativity, not just in geodesics. Their mathematical meaning in geometry will appear in the next section. We will also derive the geodesic equation using *parallel transport* in Section 4.6 after we have enough mathematical tools.

Example 4.4. Christoffel symbols in cylindrical polar coordinates

Consider the metric

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + dr^2 + r^2 d\phi^2 + dz^2, \quad g_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad g^{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/r^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We won't calculate all 40 of the Christoffel symbols because most of them are zero, as you can observe from the geodesic equations in cylindrical polar coordinates. There are only three nonzero Christoffel symbols:

$$\Gamma^r{}_{\phi\phi} = \frac{1}{2} g^{r\beta} \left(\cancel{\frac{\partial \gamma_{\beta\phi}}{\partial \phi}} + \cancel{\frac{\partial g_{\beta\phi}}{\partial d\phi}} - \frac{\partial g_{\phi\phi}}{\partial x^\beta} \right) = \frac{1}{2} g^{rr} \left(-\frac{\partial g_{\phi\phi}}{\partial r} \right) = \frac{1}{2}(1)(-2r) = -r.$$

$$\Gamma^\phi{}_{\phi r} = \Gamma^\phi{}_{r\phi} = \frac{1}{2} g^{\phi\beta} \left(\frac{\partial g_{\beta\phi}}{\partial r} + \cancel{\frac{\partial g_{\beta r}}{\partial \phi}} - \cancel{\frac{\partial g_{\phi r}}{\partial x^\beta}} \right) = \frac{1}{2} g^{\phi\phi} \frac{\partial g_{\phi\phi}}{\partial r} = \frac{1}{2} \frac{1}{r^2}(2r) = \frac{1}{r}.$$

The geodesic equation is

$$\frac{d^2 x^\alpha}{d\tau^2} + \Gamma^\alpha{}_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0.$$

Check that the Christoffel symbols produce the geodesic equations for cylindrical polar coordinates:

$$\begin{aligned} \frac{d^2 t}{d\tau^2} &= 0 & &\iff & &\Gamma^t{}_{\mu\nu} = 0, \\ \frac{d^2 r}{d\tau^2} - r\left(\frac{d\phi}{d\tau}\right)^2 &= 0 & &\iff & &\Gamma^r{}_{\phi\phi} = -r, \text{ others vanish} \\ \frac{d^2 \phi}{d\tau^2} + \frac{2}{r}\frac{dr}{d\tau}\frac{d\phi}{d\tau} &= 0 & &\iff & &\Gamma^\phi{}_{\phi r} = \Gamma^\phi{}_{r\phi} = \frac{1}{r}, \text{ others vanish} \\ \frac{d^2 z}{d\tau^2} &= 0 & &\iff & &\Gamma^z{}_{\mu\nu} = 0. \end{aligned}$$

(Note the factor of 2 in the third equation because $\Gamma^\phi{}_{\phi r} = \Gamma^\phi{}_{r\phi}$.)

If we are interested in geodesics of massless particles such as a photon, the proper time τ in (4.4) should be replaced with an affine parameter λ because τ is undefined,

$$\frac{d^2 x^\alpha}{d\lambda^2} + \Gamma^\alpha{}_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = 0.$$

This also works for material particles, but the equation is more physically meaningful with derivatives with respect to τ.

4.4 Covariant Derivatives

In curved coordinates like polar coordinates, the basis vectors are not constant everywhere. Differentiating the components of a vector will not give the derivative of the vector. We need to differentiate the basis vectors as well.

4.4.1 Derivatives of Vectors

First, we will find the derivatives of the basis vector $\boldsymbol{e}_r$ in polar coordinates.

$$\begin{aligned}\frac{\partial}{\partial r}\boldsymbol{e}_r &= \frac{\partial}{\partial r}\left(\boldsymbol{e}_x\cos\theta + \boldsymbol{e}_y\sin\theta\right) = 0,\\ \frac{\partial}{\partial \theta}\boldsymbol{e}_r &= \frac{\partial}{\partial \theta}\left(\boldsymbol{e}_x\cos\theta + \boldsymbol{e}_y\sin\theta\right) = -\boldsymbol{e}_x\sin\theta + \boldsymbol{e}_y\cos\theta = \frac{1}{r}\boldsymbol{e}_\theta.\end{aligned}$$

Similarly, for $\boldsymbol{e}_\theta$,

$$\begin{aligned}\frac{\partial}{\partial r}\boldsymbol{e}_\theta &= -\boldsymbol{e}_x\sin\theta + \boldsymbol{e}_y\cos\theta = \frac{1}{r}\boldsymbol{e}_\theta\\ \frac{\partial}{\partial \theta}\boldsymbol{e}_\theta &= -r\boldsymbol{e}_x\cos\theta - r\boldsymbol{e}_y\sin\theta = -r\boldsymbol{e}_r.\end{aligned}$$

A general vector $\vec{V}$ has components (V^r, V^θ) with polar basis. To differentiate this vector, we need to obtain the derivatives of both its components and respective bases. The derivative of V^μ with respect to r is

$$\frac{\partial\vec{V}}{\partial r} = \frac{\partial}{\partial r}\left(V^r\boldsymbol{e}_r + V^\theta\boldsymbol{e}_\theta\right) = \frac{\partial V^r}{\partial r}\boldsymbol{e}_r + V^r\frac{\partial\boldsymbol{e}_r}{\partial r} + \frac{\partial V^\theta}{\partial r}\boldsymbol{e}_\theta + V^\theta\frac{\partial\boldsymbol{e}_\theta}{\partial r}.$$

A similar method goes for $\partial_\theta V^\mu$. In index notation for general result, we write

$$\frac{\partial\vec{V}}{\partial x^\nu} = \frac{\partial V^\mu}{\partial x^\nu}\boldsymbol{e}_\mu + V^\mu\frac{\partial\boldsymbol{e}_\mu}{\partial x^\nu}.$$

Example 4.5. The derivative of $\boldsymbol{e}_x$ with respect to θ
The basis $\boldsymbol{e}_x$ can be expressed as

$$\boldsymbol{e}_x = \boldsymbol{e}_r\cos\theta - \frac{1}{r}\boldsymbol{e}_\theta\sin\theta.$$

Thus,

$$\begin{aligned}\frac{\partial}{\partial\theta}\boldsymbol{e}_x &= \frac{\partial}{\partial\theta}\boldsymbol{e}_r + \cos\theta\frac{\partial}{\partial\theta} - \frac{\partial}{\partial\theta}\left(\frac{1}{r}\sin\theta\right)\boldsymbol{e}_\theta - \frac{1}{r}\sin\theta\frac{\partial}{\partial\theta}\boldsymbol{e}_\theta\\ &= -\boldsymbol{e}_r\sin\theta + \cos\theta\left(\frac{1}{r}\boldsymbol{e}_\theta\right) - \frac{1}{r}\boldsymbol{e}_\theta\cos\theta - \frac{1}{r}\sin\theta(-r\boldsymbol{e}_r)\\ &= 0.\end{aligned}$$

This is intuitively true since $\boldsymbol{e}_x$ is a constant basis independent of any parameters.

The final term of the equation is significant. Notice that $\partial\boldsymbol{e}_\nu/\partial x^\nu$ itself is a vector, so it can be written as a linear combination of basis vectors. We will use the **Christoffel symbols** ${\Gamma^\alpha}_{\mu\nu}$ to denote the coefficients of this combination:

$$\frac{\partial\boldsymbol{e}_\mu}{\partial x^\nu} = {\Gamma^\alpha}_{\mu\nu}\,\boldsymbol{e}_\alpha.$$

The indices in this definition has the following meaning: μ is the basis vector being differentiated; ν is the coordinate with respect to which it is being differentiated; α is the component of the resulting derivative vector. We now have the derivative of the general vector in the form with Christoffel symbols:

$$\frac{\partial\vec{V}}{\partial x^\nu} = \frac{\partial V^\mu}{\partial x^\nu}\boldsymbol{e}_\mu + V^\alpha{\Gamma^\mu}_{\alpha\nu}\,\boldsymbol{e}_\mu = \left(\partial_\nu V^\mu + V^\alpha{\Gamma^\mu}_{\alpha\nu}\right)\boldsymbol{e}_\mu.$$

We can conclude that the vector field $\partial_\nu\vec{V}$ has components

$$\partial_\nu V^\mu + V^\alpha{\Gamma^\mu}_{\alpha\nu}.$$

We call this component the **covariant derivative**, $\nabla\vec{V}$,

$$\nabla_\nu V^\mu = \partial_\nu V^\mu + V^\alpha{\Gamma^\mu}_{\alpha\nu}.$$

The covariant derivative is different from the partial derivative because the basis vectors are changing.

4.4.2 Mathematical Origin of the Covariant Derivative

The covariant derivative satisfies five conditions:

1. Linearity: for any tensors $A, B \in \mathcal{T}(m,n)$ and scalars $a, b \in \mathbb{R}$,
$$\nabla_\mu (aA^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n} + bB^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n}) = a\nabla_\mu A^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n} + b\nabla_\mu B^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n}.$$
2. Leibnitz rule: for all $A \in \mathcal{T}(m,n)$, $B \in \mathcal{T}(m',n')$,
$$\nabla_\mu (A^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n} B^{\gamma_1\cdots\gamma_{m'}}{}_{\delta_1\cdots\delta_{n'}}) = (\nabla_\mu A^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n}) B^{\gamma_1\cdots\gamma_{m'}}{}_{\delta_1\cdots\delta_{n'}} + A^{\alpha_1\cdots\alpha_m}{}_{\beta_1\cdots\beta_n} (\nabla_\mu B^{\gamma_1\cdots\gamma_{m'}}{}_{\delta_1\cdots\delta_{n'}}).$$
3. Commutativity with contraction: for all $A \in \mathcal{T}(m,n)$,
$$\nabla_\mu (A^{\alpha_1\cdots\nu\cdots\alpha_m}{}_{\beta_1\cdots\nu\cdots\beta_n}) = \nabla_\mu A^{\alpha_1\cdots\nu\cdots\alpha_m}{}_{\beta_1\cdots\nu\cdots\beta_n}.$$
4. For all scalar fields $f \in \mathcal{F}$ and all tangent vectors $t^\mu \in V_p$,
$$t(f) = t^\mu \nabla_\mu f.$$
5. Torsion Free (in most cases): for all $f \in \mathcal{F}$,
$$\nabla_\mu \nabla_\nu f = \nabla_\nu \nabla_\mu f.$$

To be written.

The covariant derivative of a scalar is the same as its partial derivative since the scalar is independent of coordinates. In this case, the covariant derivative represent the gradient of the scalar field,

$$\nabla_\mu \phi = \partial_\mu \phi.$$

With this idea, we now come to the divergence. The divergence is a scalar obtained by contracting $\partial_\nu V^\mu$ on its two indices, which then will give $\partial_\mu V^\mu$. The covariant derivative and the partial derivative is the same in Cartesian coordinates, so does their contraction process. The contraction is also frame independent, so when we are contracting the covariant derivative, we get the divergence in other coordinates. The divergence is independent of coordinates (since it is a scalar). Hence we have

$$\partial_\mu V^\mu = \nabla_\nu V^\nu,$$

where μ is for cartesian coordinates and ν is for arbitrary curvilinear coordinates.

Example 4.6. The Divergence in Polar Coordinates:
We first write down the divergence contraction form:

$$\nabla_\mu V^\mu = \partial_\mu V^\mu + \Gamma^\mu{}_{\alpha\mu} V^\mu.$$

It is easy to calculate that

$$\Gamma^\mu{}_{r\mu} = 1/r, \quad \Gamma^\mu{}_{\theta\mu} = 0.$$

Therefore, we have

$$\begin{aligned} \nabla_\mu V^\mu &= \frac{\partial V^r}{\partial r} + \frac{\partial V^\theta}{\partial \theta} + \frac{1}{r} V^r \\ &= \frac{1}{r}\frac{\partial}{\partial r}(rV^r) + \frac{\partial}{\partial \theta} V^\theta. \end{aligned}$$

The covariant derivative of the general tensor is

$$\nabla_\beta T_{\mu\nu} = \partial_\beta T_{\mu\nu} - T_{\alpha\nu}\Gamma^\alpha{}_{\mu\beta} - T_{\mu\alpha}\Gamma^\alpha{}_{\nu\beta}. \tag{4.6}$$

$$\nabla_\beta A^{\mu\nu} = \partial_\beta A^{\mu\nu} + A^{\alpha\nu}\Gamma^{\mu}{}_{\alpha\beta} + A^{\mu\alpha}\Gamma^{\nu}{}_{\alpha\beta}. \tag{4.7}$$

$$\nabla_\beta B^{\mu}{}_{\nu} = \partial_\beta B^{\mu}{}_{\nu} + B^{\alpha}{}_{\nu}\Gamma^{\mu}{}_{\alpha\beta} - B^{\mu}{}_{\alpha}\Gamma^{\alpha}{}_{\nu\beta}. \tag{4.8}$$

4.5 Christoffel Symbols Revisited

4.5.1 Covariant Derivative of a Metric

In Cartesian coordinates, the components of the one-form and its corresponding vector are equal. Since ∇ is the differentiation of components, the components of the covariant derivatives of the one-form and vector must be equal. This means that for a vector $\vec{V}$ and its related one-form $\tilde{V} = \mathbf{g}(\vec{V}, \quad)$,

$$\nabla_\mu \tilde{V} = \mathbf{g}(\nabla_\mu \vec{V}, \quad).$$

This equation is a tensor equation so it works in all coordinates. Therefore,

$$\nabla_\nu V_\mu = g_{\mu\alpha}\nabla_\nu V^\alpha.$$

The following is the formal proof for this statement.

Proof. Let unprimed indices denote Cartesian coordinates and primed indices arbitrary coordinates. The index lowering equation is valid in all coordinates:

$$V_{\mu'} = g_{\mu'\nu'}V^{\nu'},$$

and in Cartesian coordinates,

$$g_{\mu\nu} = \delta_{\mu\nu}, \quad V_\mu = V^\mu.$$

Also in Cartesian coordinates, the Christoffel symbol vanishes so that,

$$\nabla_\alpha V_\mu = \partial_\alpha V_\mu \qquad \text{and} \qquad \nabla_\alpha V^\mu = \partial_\alpha V^\mu.$$

Thus, only in Cartesian coordinates we have

$$\nabla_\alpha V_\mu = \nabla_\alpha V^\mu.$$

Now is to convert the equation into one that satisfy all coordinates. Note that in Cartesian coordinates

$$\nabla_\alpha V^\mu = g_{\mu\nu}\nabla_\alpha V^\nu \quad \Longrightarrow \quad \nabla_\alpha V_\mu = g_{\mu\nu}\nabla_\alpha V^\nu.$$

This second equation is a tensor equation (so it is valid in all coordinates). If we take the α' covariant derivative in arbitrary coordinates for the index lowering equation, then we have

$$\nabla_{\alpha'}V_{\mu'} = \nabla_{\alpha'}(g_{\mu'\nu'}V^{\nu'}) = (\nabla_{\alpha'}g_{\mu'\nu'})V^{\nu'} + g_{\mu'\nu'}\nabla_{\alpha'}V^{\nu'}.$$

□

Furthermore, this means that in all coordinate system (we use normal index to represent all coordinates),

$$\nabla_\alpha g_{\mu\nu} = 0.$$

Let us see polar coordinates as an example:

Example 4.7. We use the equation 4.6,

$$\nabla_\alpha g_{\mu\nu} = \partial_\alpha g_{\mu\nu} - \Gamma^{\beta}{}_{\mu\alpha}g_{\beta\nu} - \Gamma^{\beta}{}_{\mu\alpha}g_{\mu\beta}. \tag{4.9}$$

Let $\mu = \nu = \alpha = r$, we have $\partial_r g_{rr} = 0$ and $\Gamma^{\beta}{}_{rr} = 0$. Hence, trivially,

$$\nabla_r g_{rr} = \partial_r g_{rr} - \Gamma^{\beta}{}_{rr}g_{\beta r} - \Gamma^{\beta}{}_{rr}g_{r\beta} = 0.$$

Let $\mu = \nu = \theta, \alpha = r$, then with $g_{\theta\theta} = r^2$, $\Gamma^\theta{}_{\theta r}$ and $\Gamma^r{}_{\theta r} = 0$,

$$\begin{aligned}\nabla_r g_{\theta\theta} &= \partial_r g_{\theta\theta} - \Gamma^\beta{}_{\theta r} g_{\beta\theta} - \Gamma^\beta{}_{\theta r} g_{\theta\beta} \\ &= \partial_r (r^2) - \frac{1}{r}(r^2) - \frac{1}{r}(r^2) \\ &= 0.\end{aligned}$$

4.5.2 Calculating Christoffel Symbols from the Metric

First, Christoffel symbols are "commutive" in any coordinates:

$$\Gamma^\alpha{}_{\mu\nu} = \Gamma^\alpha{}_{\nu\mu}. \tag{4.10}$$

Then write three permutations of indices for equation 4.9 and it should be equal to zero:

$$\begin{aligned}\partial_\alpha g_{\mu\nu} &= \Gamma^\beta{}_{\mu\alpha} g_{\beta\nu} + \Gamma^\beta{}_{\nu\alpha} g_{\mu\beta}, \\ \partial_\nu g_{\mu\alpha} &= \Gamma^\beta{}_{\mu\nu} g_{\beta\alpha} + \Gamma^\beta{}_{\alpha\nu} g_{\mu\beta}, \\ -\partial_\mu g_{\nu\alpha} &= -\Gamma^\beta{}_{\nu\mu} g_{\beta\alpha} - \Gamma^\beta{}_{\alpha\mu} g_{\nu\alpha}.\end{aligned}$$

Adding them up and using the symmetry of Γ, we simplify the equation to

$$\partial_\alpha g_{\mu\nu} + \partial_\nu g_{\mu\alpha} - \partial_\mu g_{\nu\alpha} = 2 g_{\mu\beta} \Gamma^\beta{}_{\nu\alpha}.$$

Dividing by 2 and multiplying $g^{\mu\gamma}$, we have

$$\frac{1}{2} g^{\mu\gamma} (\partial_\alpha g_{\mu\nu} + \partial_\nu g_{\mu\alpha} - \partial_\mu g_{\nu\alpha}) = \Gamma^\gamma{}_{\nu\alpha} \tag{4.11}$$

4.6 Parallel-Transport, Geodesics, and Curvature

There are two kinds of curvature, the extrinsic curvature and intrinsic curvature. For example, a cylinder might be thought of as curved in three-dimensional flat space. This is its **extrinsic curvature**. On the other hand, you can roll a flat piece of paper to make a cylinder without tearing it or crumpling it. The **intrinsic curvature** is the curvature of that original paper, which is flat. This means that all Euclid's axioms hold for the surface of the cylinder, such as the parallel lines on the cylinder will continue to be parallel.

However, a sphere has an intrinsically curved surface. It cannot be made by flat surface. Parallel lines may intersect on the surface of a sphere. For example, two lines perpendicular to the equator will intersect at the poles.

4.6.1 Parallel Transport

Define a vector field $\vec{V}$ on the sphere. If the vectors $\vec{V}$ at infinitesimally closed points of the curve are parallel and of equal length, then $\vec{V}$ are said to be parallel-transported along the curve. Consider a curve with parameter λ, and $u = d\vec{x}/d\lambda$ tangent to the curve. In a locally inertial coordinate system at point P, the components of $\vec{V}$ must be constant along the curve at P,

$$\frac{dV^\mu}{d\lambda} = 0 \quad \text{at } P,$$

or

$$\frac{dV^\mu}{d\lambda} = u^\nu \partial_\nu V^\mu = u^\nu \nabla_\nu V^\mu = 0 \quad \text{at } P.$$

The second equality means $\Gamma^\mu{}_{\alpha\beta} = 0$ at P in locally inertial frame. The third inequality is frame-invariant so it holds in any basis. Thus, the parallel-transport of $\vec{V}$ along u has a frame invariant definition

$$u^\nu \nabla_\nu V^\mu = 0 \quad \iff \quad \frac{d}{d\lambda}\vec{V} = \nabla_u \vec{V} = 0, \tag{4.12}$$

where $\nabla_u \vec{V} = u^\nu \nabla_\nu V^\mu$.

4.6.2 Geodesics

The lines drawn by parallel-transport of the tangent vector are called **geodesics**:

$$\nabla_{\vec{U}} u = 0.$$

In terms of components,

$$u^\nu \nabla_\nu u^\mu = u^\nu \partial_\nu u^\mu + \Gamma^\mu{}_{\alpha\nu} u^\alpha u^\nu = 0.$$

If again λ is the parameter of the curve, then $u^\mu = dx^\mu/d\lambda$ and $u^\nu \partial_\nu = d/d\lambda$:

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^\mu{}_{\alpha\nu} \frac{dx^\alpha}{d\lambda} \frac{dx^\nu}{d\lambda} = 0. \tag{4.13}$$

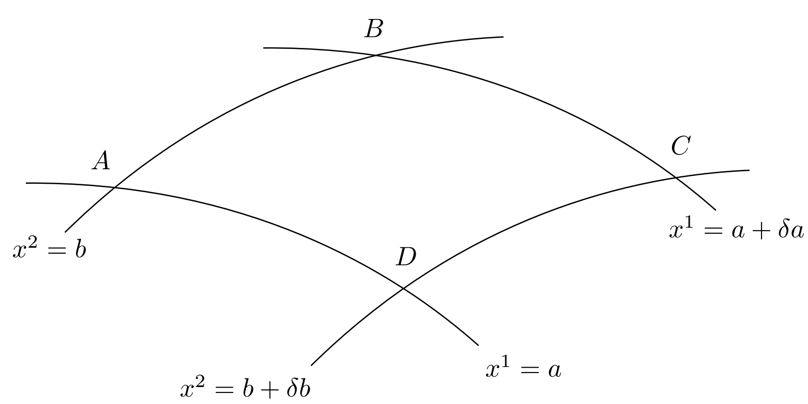
This is a nonlinear, second-order differential equation because $\Gamma^\mu{}_{\alpha\nu}$ depends on the coordinates $\{x^\mu\}$ itself. When initial conditions $x_0^\mu = x^\mu(\lambda_0)$ and $U_0^\mu = (dx^\mu/d\lambda)_{\lambda_0}$ are given, this equation has unique solution. Also, this equation is valids for a new parameter $\phi = a\lambda + b$ (the linear transformations of λ). The parameters ϕ and λ are called **affine parameters**.

4.7 The Curvature Tensor

4.7.1 Riemann Curvature Tensor

In a manifold, define a small closed loop whose four sides are coordinate lines,

$$x^1 = a, \quad x^1 = a + \delta a, \quad x^2 = b, \quad x^2 = b + \delta b.$$



Suppose a vector $\vec{V}$ at A is parallel transported to B with $\nabla_{e_1} \vec{V} = 0$, so the component form is

$$\frac{\partial V^\mu}{\partial x^1} = -\Gamma^\mu{}_{\alpha 1} V^\alpha. \tag{4.14}$$

Integrating from A to B gives

$$V^\mu(B) = V^\mu(A) + \int_A^B \frac{\partial V^\mu}{\partial x^1} dx^1 = V^\mu(A) - \int_{x^2=b} \Gamma^\mu{}_{\alpha 1} V^\alpha dx^1.$$

Similarly, the parallel transport along the loop satisfies

$$V^\mu(C) = V^\mu(B) - \int_{x^1=a+\delta a} \Gamma^\mu{}_{\alpha 2} V^\alpha dx^2,$$
$$V^\mu(D) = V^\mu(C) + \int_{x^2=b+\delta b} \Gamma^\mu{}_{\alpha 1} V^\alpha dx^1,$$
$$V^\mu(A_{\text{final}}) = V^\mu(D) + \int_{x^1=a} \Gamma^\mu{}_{\alpha 2} V^\alpha dx^2.$$

The change in sign before integral is because of the integration along negative x^1 or x^2 direction. We then get the net change of $V^\mu(A)$,

$$\begin{aligned}\delta V^\mu &= V^\mu(A_{\text{final}}) - V^\mu(A_{\text{initial}}) \\ &= \int_{x^1=a} \Gamma^\mu{}_{\alpha 2} V^\alpha dx^2 - \int_{x^1=a+\delta a} \Gamma^\mu{}_{\alpha 2} V^\alpha dx^2 \\ &+ \int_{x^2=b+\delta b} \Gamma^\mu{}_{\alpha 1} V^\alpha dx^1 - \int_{x^2=b} \Gamma^\mu{}_{\alpha 1} V^\alpha dx^1.\end{aligned}$$

Combine these integrals in pairs,

$$\begin{aligned}\delta V^\mu &\simeq -\int_b^{b+\delta b} \delta a\, \partial_1(\Gamma^\mu{}_{\alpha 2} V^\alpha) dx^2 + \int_a^{a+\delta a} \delta b\, \partial_2(\Gamma^\mu{}_{\alpha 1} V^\alpha) dx^1 \\ &\approx \delta a\, \delta b\, [-\partial_1(\Gamma^\mu{}_{\alpha 2} V^\alpha) + \partial_2(\Gamma^\mu{}_{\alpha 1} V^\alpha)]\,.\end{aligned}$$

The derivatives of V^μ can be turn into Christoffel symbols by Eq. 4.14,

$$\delta V^\mu = \delta a\, \delta b\, [\partial_2 \Gamma^\mu{}_{\alpha 1} - \partial_1 \Gamma^\mu{}_{\alpha 2} + \Gamma^\mu{}_{\nu 2}\Gamma^\nu{}_{\alpha 1} - \Gamma^\mu{}_{\nu 1}\Gamma^\nu{}_{\alpha 2}] V^\alpha.$$

Notice that 1 and 2 are antisymmetric because interchanging them means integrate in the opposite direction. We can use general coordinates x^σ and x^λ for δV^μ,

$$\delta V^\mu = \delta a\, \delta b\, [\partial_\lambda \Gamma^\mu{}_{\alpha\sigma} - \partial_\sigma \Gamma^\mu{}_{\alpha\lambda} + \Gamma^\mu{}_{\nu\lambda}\Gamma^\nu{}_{\alpha\sigma} - \Gamma^\mu{}_{\nu\sigma}\Gamma^\nu{}_{\alpha\lambda}] V^\alpha.$$

The Γs in the square parenthesis is known as the **Riemann curvature tensor** and it is a $\binom{1}{3}$ tensor,

$$R^\mu{}_{\nu\alpha\beta} = \partial_\alpha \Gamma^\mu{}_{\nu\beta} - \partial_\beta \Gamma^\mu{}_{\nu\alpha} + \Gamma^\mu{}_{\gamma\alpha}\Gamma^\gamma{}_{\nu\beta} - \Gamma^\mu{}_{\gamma\beta}\Gamma^\gamma{}_{\nu\alpha}. \tag{4.15}$$

In a locally inertial frame, the christoffel symbol $\Gamma^\mu{}_{\alpha\beta} = 0$ at point P, but its derivative is not:

$$\partial_\sigma \Gamma^\mu{}_{\alpha\beta} = \frac{1}{2} g^{\mu\nu}(\partial_{\beta\sigma} g_{\nu\alpha} + \partial_{\alpha\sigma} g_{\nu\beta} - \partial_{\nu\sigma} g_{\alpha\beta}).$$

There is no product rule for $g^{\mu\nu}$ because the covariant derivative of $g^{\mu\nu}$ proved to be always zero. The Christoffel symbol vanishes at P, which means the partial derivative $\partial_\sigma g^{\mu\nu}$ must also be 0 by Eq. 4.6. The second derivatives of $g_{\mu\nu}$ do not vanish so at P,

$$R^\mu{}_{\nu\alpha\beta} = \frac{1}{2} g^{\mu\sigma}(\partial_{\beta\alpha} g_{\sigma\nu} + \partial_{\nu\alpha} g_{\sigma\beta} - \partial_{\sigma\alpha} g_{\nu\beta} - \partial_{\alpha\beta} g_{\sigma\nu} - \partial_{\nu\beta} g_{\sigma\alpha} + \partial_{\sigma\beta} g_{\nu\alpha}).$$

Using the commutive property of partial derivatives, we can cancel out two terms,

$$R^\mu{}_{\nu\alpha\beta} = \frac{1}{2} g^{\mu\sigma}(\partial_{\nu\alpha} g_{\sigma\beta} - \partial_{\nu\beta} g_{\sigma\alpha} + \partial_{\sigma\beta} g_{\nu\alpha} - \partial_{\sigma\alpha} g_{\nu\beta}).$$

If we lower the index μ,

$$R_{\mu\nu\alpha\beta} = g_{\mu\lambda} R^\lambda{}_{\nu\alpha\beta} = \frac{1}{2}(\partial_{\nu\alpha} g_{\mu\beta} - \partial_{\nu\beta} g_{\mu\alpha} + \partial_{\mu\beta} g_{\nu\alpha} - \partial_{\mu\alpha} g_{\nu\beta}). \tag{4.16}$$

Then we have the following identities

$$R_{\mu\nu\alpha\beta} = -R_{\nu\mu\alpha\beta} = -R_{\mu\nu\beta\alpha} = R_{\alpha\beta\mu\nu} \tag{4.17}$$

$$R_{\mu\nu\alpha\beta} + R_{\mu\beta\nu\alpha} + R_{\mu\alpha\beta\nu} = 0 \tag{4.18}$$

In conclusion, $R_{\mu\nu\alpha\beta}$ is antisymmetric on the first pair and on the second pair of indices, and symmetric on exchange of the two pairs. A flat manifold with global parallelism has Riemann curvature tensor 0.

$$R^\mu{}_{\nu\alpha\beta} = 0 \quad \Longleftrightarrow \quad \text{flat manifold}.$$

The curvature tensor is important when we take second covariant derivative of a vector field $\vec{V}$. The second derivative (by Eq. 4.8) is

$$\nabla_\alpha(\nabla_\beta V^\mu) = \partial_\alpha(\nabla_\beta V^\mu) + \Gamma^\mu{}_{\nu\alpha}\nabla_\beta V^\nu - \Gamma^\nu{}_{\beta\alpha}\nabla_\nu V^\mu.$$

In locally inertial coordinates, all the Γs vanish, but not their partial derivatives. At P,

$$\nabla_\alpha \nabla_\beta V^\mu = \partial_{\beta\alpha} V^\mu + \partial_\alpha \Gamma^\mu{}_{\nu\beta} V^\nu.$$

If α and β exchanges,

$$\nabla_\beta \nabla_\alpha V^\mu = \partial_{\alpha\beta} V^\mu + \partial_\beta \Gamma^\mu{}_{\nu\alpha} V^\nu.$$

Subtract these two gives the commutator of ∇_α and ∇_β,

$$[\nabla_\alpha, \nabla_\beta] V^\mu = \nabla_\alpha \nabla_\beta V^\mu - \nabla_\beta \nabla_\alpha V^\mu = (\partial_\alpha \Gamma^\mu{}_{\nu\beta} - \partial_\beta \Gamma^\mu{}_{\nu\alpha}) V^\nu.$$

With all Γs being 0, using Eq. 4.15, we obtain

$$[\nabla_\alpha, \nabla_\beta] V^\mu = R^\mu{}_{\nu\alpha\beta} V^\nu. \tag{4.19}$$

This is a tensor equation so it works in any coordinate system. The Riemann curvature tensor gives the commutator of covariant derivatives. The commutator is not zero means that we need to be careful about the order of covariant derivatives since they do not commute.

4.7.2 Geodesic Deviation

To be written.

4.8 Bianchi Identities, Ricci and Einstein Tensors

If we differentiate the curvature tensor with respect to x^λ,

$$\partial_\lambda R_{\mu\nu\alpha\beta} = \frac{1}{2}(\partial_{\nu\alpha\lambda} g_{\mu\beta} - \partial_{\nu\beta\lambda} g_{\mu\alpha} + \partial_{\mu\beta\lambda} g_{\nu\alpha} - \partial_{\mu\alpha\lambda} g_{\nu\beta}).$$

Using the symmetry $g_{\mu\nu} = g_{\nu\mu}$ and commutative property of partial derivatives,

$$\partial_\lambda R_{\mu\nu\alpha\beta} + \partial_\beta R_{\mu\nu\lambda\alpha} + \partial_\alpha R_{\mu\nu\beta\lambda} = 0,$$

and the local coordinates that have Γs vanishing gives

$$\nabla_\lambda R_{\mu\nu\alpha\beta} + \nabla_\beta R_{\mu\nu\lambda\alpha} + \nabla_\alpha R_{\mu\nu\beta\lambda} = 0.$$

This is a tensor equation valid in any system called **Bianchi identities**.

Before using Bianchi identities, define the **Ricci tensor** $R_{\mu\nu}$ as

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu} = R_{\nu\mu} \tag{4.20}$$

and the **Ricci scalar** as

$$R = g^{\mu\nu} R_{\mu\nu} = g^{\mu\nu} g^{\alpha\beta} R_{\alpha\mu\beta\nu}. \tag{4.21}$$

Apply the Ricci contraction to the Bianchi identities,

$$g^{\mu\alpha} \left[\nabla_\lambda R_{\mu\nu\alpha\beta} + \nabla_\beta R_{\mu\nu\lambda\alpha} + \nabla_\alpha R_{\mu\nu\beta\lambda}\right] = 0.$$

Using the fact that $\nabla_\alpha g^{\mu\nu} = 0$, we get

$$\nabla_\lambda R_{\nu\beta} + (-\nabla_\beta R_{\nu\lambda}) + \nabla_\alpha R^\alpha{}_{\nu\beta\lambda}. \tag{4.22}$$

The negative sign comes from the fact that

$$g^{\mu\alpha} \nabla_\beta R_{\mu\nu\lambda\alpha} = -g^{\mu\alpha} \nabla_\beta R_{\mu\nu\alpha\lambda} = -\nabla_\beta R_{\nu\lambda}.$$

Eq.4.22 is called the contracted Bianchi identites. If we contract this again,

$$g^{\nu\beta}\left[\nabla_\lambda R_{\nu\beta} - \nabla_\beta R_{\nu\lambda} + \nabla_\alpha R^\alpha{}_{\nu\beta\lambda}\right] = 0.$$

By similar method, we obtain

$$\nabla_\lambda R - \nabla_\alpha R^\alpha{}_\lambda + (-\nabla_\alpha R^\alpha_\lambda) = 0.$$

$$\nabla_\alpha(2R^\alpha{}_\lambda - \delta^\alpha{}_\lambda R) = 0.$$

Note that R is a scalar so that $\nabla_\lambda = \partial_\lambda R$ in all coordinates. This is the twice-contracted Bianchi identites. Now define a symmetric tensor

$$G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = G^{\nu\mu}. \tag{4.23}$$

and according to the Bianchi identities,

$$\nabla_\nu G^{\mu\nu} = 0.$$

The tensor $G^{\mu\nu}$ is known as the **Einstein tensor**.

5 A Curved Spacetime

5.1 The Transition from Differential Geometry to Gravity

Now we have developed mathematical representation of the curvature, which depends on the metric. We have also known that there is no global inertial frame in nonuniform gravitational fields. In conclusion, we obtain

- Spacetime is a four-dimensional manifold with a matric.
- The metric is measured by rods and clocks. The distance along a rod between two nearby points is $|d\mathbf{x} \cdot d\mathbf{x}|^{1/2}$. The time measured by a clock that experiences two events closely separated in time is $|-d\mathbf{x} \cdot d\mathbf{x}|^{1/2}$.
- The metric of spacetime can be put in Lorentz form $\eta_{\mu\nu}$ at any particular event by an appropriate choice of coordinate system.

What to do now is first to identify how physical objects behave in a curved spacetime. Second is to know how the curvature is generated by the objects in the spacetime.

The acceleration of a particle in a gravitational field is independent of the mass of it, so it is always possible to set a freely falling frame with no acceleration of nearby particles and thus a locally inertial frame. The particles follow straight lines locally. The straight lines in a locally inertial frame are the definition of geodesics in a full curved manifold. Then we arrive at the first postulate of how particles move:

weak equivalence principle (WEP):
Freely falling particles move on timelike geodesics of the spacetime.

The term "free falling" means the particles are unaffected by other forces and there are always particles unaffected by other forces (like neutron unaffected by EM). This is why gravity is special and WEP can be tested to high accuracy.

However, WEP only applies to particles. There are other kinds of matter like fluids, so we need a generalization:

Einstein Equivalence Principle:
Any local physical experiment not involving gravity will have the same result if performed in a freely falling inertial frame as if it were performed in the flat spacetime of special relativity.

From what we know now, gravity introduces nothing new locally, so we can generalize many laws of conservation as in SR by replacing partial derivatives to covariant derivatives. In local inertial frame:

$$\begin{cases} \nabla_\mu (n u^\mu) = 0 & \text{[conservation of particles]}, \\ \nabla_\nu (T^{\mu\nu}) = 0 & \text{[conservation of energy and momentum]}, \\ u^\mu \nabla_\mu S = 0 & \text{[conservation of entropy]}. \end{cases}$$

Note that entropy is a scalar, so he law is unchanged in a curved spacetime. Also, the stress-energy tensor is defined as before,

$$T^{\mu\nu} = (\rho + P) u^\mu u^\nu + P g^{\mu\nu}.$$

5.2 Physics in Slightly Curved Spacetimes

A **weak gravitational field** is characterized by particles with gravitational potential energy much less than its rest-mass energy, $|m\Phi| \ll m$, or $|\Phi| \ll 1$. Newtonian potential Φ determines the metric,

$$ds^2 = -(1 + 2\Phi) dt^2 + (1 - 2\Phi)(dx^2 + dy^2 + dz^2). \tag{5.1}$$

This will be proved later when we study how a metric is generated. In most cases, the weak field limit is useful for Newtonian gravitational potential $\Phi = -GM/r$. In ordinary units, this is

$$ds^2 = -\left(1 + \frac{2\Phi}{c^2}\right)c^2dt^2 + \left(1 - \frac{2\Phi}{c^2}\right)(dx^2 + dy^2 + dz^2). \tag{5.2}$$

The differential dt should be perceived as the infinitesimal increment of time where there is no gravitational field at all. In this case, it is measured by a clock at rest at infinity where $\Phi \to 0$.

5.2.1 Newton's Equations of Motion

By Proper Time. In the non-relativistic limit, this weak-field metric should produce Newton's equations of motion. The proper time of a particle is

$$d\tau = \sqrt{-ds^2} = \sqrt{(1+2\Phi)dt^2 - (1-2\Phi)(dx^2+dy^2+dz^2)} = dt\sqrt{(1+2\Phi) - (1-2\Phi)v^2},$$

where we let $(dx^2 + dy^2 + dz^2)/dt^2 = v^2$. In the weak-field limit $|\Phi| \ll 1$ and non-relativistic limit $v \ll 1$, we have $2\Phi v^2 \lll 1$. The last term in the square root is negligible. Now Taylor expand the proper time,

$$d\tau \simeq dt\sqrt{1 + 2\Phi - v^2} \simeq dt\left(1 + \Phi - \frac{1}{2}v^2\right).$$

For a free particle along its world line from A to B, the proper time elasped is

$$\tau_{AB} = \int_A^B dt\left(1 + \Phi - \frac{1}{2}v^2\right).$$

We know that a free particle's proper time is extremized along its geodesic. In other words, the Lagrangian of this particle is the integrand (multiplied by m),

$$\mathcal{L} = m\left[1 + \Phi(x,y,z) - \frac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)\right].$$

Now apply Lagrange's equation, we obtain Newton's equations of motion:

$$\frac{\partial \mathcal{L}}{\partial x^i} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}^i} = 0 \quad\Longrightarrow\quad m\frac{\partial \Phi}{\partial x^i} + m\frac{d}{dt}\dot{x}^i = 0 \quad\Longrightarrow\quad -\nabla\Phi = \frac{d^2\mathbf{r}}{dt^2}.$$

By Geodesic Equation. A particle's path follow the geodesic. The proper time is an affine parameter on geodesic so the four velocity u must satisfy

$$\nabla_{\bar{U}} u = 0.$$

Any constant times proper time is an affine parameter: τ/m is one. Thus we have

$$\mathbf{p} = \frac{d\mathbf{x}}{d(\tau/m)} \quad\Longrightarrow\quad \nabla_{\bar{p}}\mathbf{p} = 0.$$

This equation is also valid for photons since we can define $\mathbf{p}$ but not u. If the particle has a nonrelativistic velocity, we can find an approximation for equation of geodesics. Consider the zero component of the equation:

$$m\frac{dp^0}{d\tau} + \Gamma^0{}_{\mu\nu}p^\mu p^\nu = 0.$$

Since the particle is not relativistic, or $p^0 \gg p^i$, the equation can be approximated as

$$m\frac{dp^0}{d\tau} + \Gamma^0{}_{00}(p^0)^2 = 0.$$

Compute ${\Gamma^0}_{00}$,

$$\begin{aligned}{\Gamma^0}_{00} &= \frac{1}{2}g^{0\mu}(\partial_0 g_{\mu 0} + \partial_0 g_{\mu 0} - \partial_\mu g_{00}) \\ &= \frac{1}{2}g^{00}\partial_0 g_{00} = \frac{1}{2}\frac{1}{(-1-2\phi)}\partial_0(-1-2\phi) \\ &\approx \partial_0\phi + \mathcal{O}(\phi^2).\end{aligned}$$

The second equality is because $g^{0\alpha}$ is nonzero only when $\mu = 0$. Plug in ${\Gamma^0}_{00}$, we obtain

$$\frac{dp^0}{d\tau} = -m\frac{\partial\phi}{\partial\tau}.$$

Recall that p^0 is the energy of the particle. This equation means the energy is conserved unless the gravitational field depends on time.

The spatial version of the geodesic equation is

$$p^\mu\partial_\mu p^i + {\Gamma^i}_{\mu\nu}p^\mu p^\nu \approx m\frac{dp^i}{d\tau} + {\Gamma^i}_{00}(p^0)^2 = 0.$$

Put $(p^0)^2 = m^2$ gives the approximation

$$\frac{dp^i}{d\tau} = -m{\Gamma^i}_{00}.$$

Calculate the Christoffel symbol:

$$\begin{aligned}{\Gamma^i}_{00} &= \frac{1}{2}g^{i\mu}(\partial_0 g_{\mu 0} + \partial_0 g_{\mu 0} - \partial_\mu g_{00}) \\ &= \frac{1}{2(1-2\phi)}\delta^{ij}(2\partial_0 g_{j0} - \partial_j g_{00}).\end{aligned}$$

Notice that $g_{j0} = 0$ so

$${\Gamma^i}_{00} \approx -\frac{1}{2}\partial_j g_{00}\delta^{ij} + \mathcal{O}(\phi^2) = -\frac{1}{2}\partial_j(1-2\phi)\delta^{ij}.$$

Thus,

$$\frac{dp^i}{d\tau} = -m\partial_j\phi\delta^{ij}.$$

This is equivalent to Newton's second law and that the force of a gravitational field is $-m\nabla\phi$. This means that general relativity in weak gravitational field predicts the Kepler's motion of most planets. Note that the weak gravitational field has two characteristics: one is the metric was nearly the Minkowski metric ($|\phi| \ll 1$), and the particle's velocity is nonrelativistic ($p^0 \gg p^i$). It will turn out that not only for particles, but also for perfect fluids, the physics in weak gravitational field will work out.

5.2.2 Gravitational Time Dilation

Again, if we are in the weak field limit $\Phi \ll 1$ and non-relativistic limit $v \ll 1$, the proper time $d\tau$ is related to the time at infinity dt by

$$d\tau \simeq dt\left(1 + \Phi - \frac{1}{2}v^2\right). \tag{5.3}$$

For a clock placed in a gravitational field, its time is in general not the same as a clock outside the field. The infinitesimal proper time depends on the position of the clock in a gravitational field. We will discuss one thought experiment first, and then turn to a very important application.

Example 5.1. A parabolic trajectory

Consider a ball tossed with an initial velocity v_0 and angle α with respect to the horizontal. Assume that the downward gravitational acceleration is a constant and equal to g in magnitude. There is an internal clock within the ball that measures its proper time after it is launched and return to ground. Another clock is also measuring the process but it is placed on the ground. When the ball is above the ground, gravitational time dilation should be less, but it is also moving and moving clocks run slow. We want to find the angle α at which these two effects cancel out.

Let the motion be in the xy-plane. Suppose the initial height of the ball is $y = 0$. The proper time of the clock at rest on the ground is

$$\tau_{\text{rest}} = T[1 + \Phi(0)],$$

where T is the coordinate time measured by a clock at rest at infinity. For the traveling ball, its proper time is given by the integral

$$\tau_{\text{ball}} = \int_0^T dt \left(1 + \Phi - \frac{1}{2}v^2\right) = \int_0^T \left[1 + \Phi(0) + gy - \frac{1}{2}v^2\right] = \tau_{\text{rest}} + \int_0^T dt \left(gy - \frac{1}{2}v^2\right).$$

By basic Newtonian kinematics (ignoring the difference between t and τ), we have

$$v_x = v_0 \cos\alpha, \quad v_y = v_0 \sin\alpha - gt, \quad y = v_t \sin\alpha - \frac{1}{2}gt^2 \quad \Longrightarrow \quad T = \frac{2v_0 \sin\alpha}{g}.$$

Then compute the integral

$$\begin{aligned}
\tau_{\text{ball}} &= \tau_{\text{rest}} + \int_0^T dt \left[gv_0 t \sin\alpha - \frac{1}{2}g^2t^2 - \frac{1}{2}v_0^2 \cos^2\alpha - \frac{1}{2}(v_0 \sin\alpha - gt)^2\right] \\
&= \tau_{\text{rest}} + \int_0^T dt \left[gv_0 t \sin\alpha - \frac{1}{2}g^2t^2 - \frac{1}{2}v_0^2 + gtv_0 \sin\alpha - \frac{1}{2}g^2t^2\right] \\
&= \tau_{\text{rest}} + \int_0^T dt \left[2gv_0 t \sin\alpha - g^2t^2 - \frac{1}{2}v_0^2\right] \\
&= \tau_{\text{rest}} + \left[gv_0 T^2 \sin\alpha - \frac{1}{3}g^2T^3 - \frac{1}{2}v_0^2 T\right] \\
&= \tau_{\text{rest}} + \left[gv_0 \frac{4v_0^2 \sin^2\alpha}{g^2} \sin\alpha - \frac{1}{3}g^2 \frac{8v_0^3 \sin^3\alpha}{g^3} - \frac{1}{2}v_0^2 \frac{2v_0 \sin\alpha}{g}\right] \\
&= \tau_{\text{rest}} + \frac{v_0^3}{g}\left(\frac{4}{3}\sin^3\alpha - \sin\alpha\right).
\end{aligned}$$

Note that $\tau_{\text{ball}} = \tau_{\text{rest}}$ when $\alpha = \pi/3$. For $\alpha > \pi/3$, $\tau_{\text{ball}} > \tau_{\text{rest}}$: gravitational time dilation dominates over special relativistic time dilation. For $\alpha < \pi/3$, $\tau_{\text{ball}} < \tau_{\text{rest}}$: special relativistic time dilation wins.

Example 5.2. The Global Positioning System (GPS)

The GPS is based on a network of 24 satellites in 12-hour orbits around the Earth. By receiving signals from multiple satellites, one's location in space can be determined within an error of 10 meters, or a timing accuracy of $10\,\text{m}/c \simeq 30\,\text{ns}$. By the centripetal force formula

$$\frac{GM_\oplus}{r_s^2} = \frac{v_s^2}{r_s} \qquad \text{and} \qquad v_s = \frac{2\pi r_s}{T} \quad \Longrightarrow \quad v_s = \left(\frac{2\pi GM_\oplus}{T}\right)^{1/3}.$$

Plugging in the period $T = 12\,\text{hr}$, mass of the Earth $M_\oplus = 5.972 \times 10^{24}\,\text{kg}$, we can get the speed of the satellite $v \approx 1.3 \times 10^{-5}c$ and $r_s \approx 2.7 \times 10^7\,\text{m}$. (We will neglect the speed of objects on Earth because it is much smaller.) Now use (5.3),

$$d\tau \simeq dt\left(1 + \frac{\Phi}{c^2} - \frac{v^2}{2c^2}\right) = dt\left(1 - \frac{GM_\oplus}{rc^2} - \frac{v^2}{2c^2}\right).$$

For objects on the ground, they have $r = R_\oplus = 6.378 \times 10^6\,\text{m}$ and $v \approx 0$. The proper times of objects on the ground and the satellites are

$$d\tau_{\text{ground}} = dt\left(1 - 6.937 \times 10^{-10}\right), \quad d\tau_s = dt\left(1 - 2.484 \times 10^{-10}\right).$$

The clock on satellites run faster by $4.453 \times 10^{-10}\,dt$. If the GPS does not account for this effect, then after $\Delta t = 1\,\text{min}$, the discrepancy between the ground and the satellites is $\Delta\tau = 2.672 \times 10^{-8}\,\text{s}$. The timing accuracy changes from 30 ns to about 57 ns. This corresponds to an additional distance error of about 10 meters in one minute. We see that accounting for relativistic effects (mostly general relativistic) are crucial in the GPS.

5.3 Conserved Quantities

Usually, the components of momentum is not constant along the trajectory of a particle, but there is an important exception. Write the geodesic equation in lowered components of p^μ,

$$p^\mu \partial_\mu p_\nu - \Gamma^\alpha{}_{\nu\mu} p^\mu p_\alpha = 0 \quad \text{or} \quad m\frac{dp_\nu}{d\tau} = \Gamma^\alpha{}_{\nu\mu} p^\mu p_\alpha.$$

Calculate the Christoffel symbol,

$$\begin{aligned}
\Gamma^\alpha{}_{\mu\nu} p^\mu p_\alpha &= \frac{1}{2} g^{\alpha\beta}(\partial_\mu g_{\beta\nu} + \partial_\nu g_{\beta\mu} - \partial_\beta g_{\mu\nu}) p^\mu p_\alpha \\
&= \frac{1}{2}(\partial_\mu g_{\beta\nu} + \partial_\nu g_{\beta\mu} - \partial_\beta g_{\mu\nu}) g^{\alpha\beta} p_\alpha p^\mu \\
&= \frac{1}{2}(\partial_\mu g_{\beta\nu} + \partial_\nu g_{\beta\mu} - \partial_\beta g_{\mu\nu}) p^\beta p^\mu.
\end{aligned}$$

Notice that $p^\beta p^\mu$ is symmetric on β and μ, while the first and third term in the parenthesis are antisymmetric on β and μ, so they cancel. Now only the second term remains,

$$\Gamma^\alpha{}_{\nu\mu} p^\mu p_\alpha = \frac{1}{2}\partial_\nu g_{\beta\mu} p^\beta p^\mu.$$

The geodesic equation is then

$$m\frac{dp_\nu}{d\tau} = \frac{1}{2}\partial_\nu g_{\beta\mu} p^\beta p^\mu.$$

This equation means that if all the components $g_{\mu\beta}$ are independent of x^ν for some fixed ν, then p_ν is constant along any particle's trajectory. For example, if a gravitational field is independent of time, then p_0 is conserved. This is the energy conservation. This is reflected in usual "laboratory frame" on Earth, we will now show that the energy $(m + T + V)$ of a particle is conserved in weak gravitatioal field. Consider the equation

$$\mathbf{p} \cdot \mathbf{p} = -m^2 = g_{\mu\nu} p^\mu p^\nu = -(1+2\phi)(p^0)^2 + (1-2\phi)(p^2)$$

$$(p^0)^2 = [m^2 + (1-2\phi)p^2](1+2\phi)^{-1},$$

where p^2 is for magnitude of spatial momentum squared. With the approximation $|\phi| \ll 1$ and $\left|p^i\right| \ll m$, we can simplify this to

$$(p^0)^2 \approx m^2(1 - 2\phi + p^2/m^2) \quad \Longrightarrow \quad p^0 = m(1 - \phi + p^2/2m^2).$$

Lower the index:

$$p_0 = g_{0\mu} p^\mu = g_{00} p^0 = -(1+2\phi)p^0.$$

$$-p_0 \approx m(1 + \phi + p^2/2m^2) = m + m\phi + p^2/2m.$$

It shows that if p_0 is constant, then the energy (rest mass, potential energy and kinetic energy) is conserved.

Similarly, if a metric is axially symmetric ($g_{\mu\nu}$ independent of angle ψ around the axis) then the angular momentum p_ψ will be conserved.

$$p_\psi = g_{\psi\psi} p^\psi \approx g_{\psi\psi} m \frac{d\partial}{dt} = m g_{\psi\psi}\omega$$

where ω is the angular velocity. For a nearly flat metric,

$$g_{\psi\psi} = \boldsymbol{e}_\psi \cdot \boldsymbol{e}_\psi \approx r^2.$$

Thus, the conserved quantity is

$$p_\psi = mr^2\omega,$$

the angular momentum.

5.3.1 Killing Vector Fields

For a particular coordinate system, if the components of the metric tenor are independent of a particular coordinate x^A, then the corresponding coordinate basis vector $\boldsymbol{e}_A$ is a Killing vector (named after mathematician Wilhelm Killing). Killing vectors are useful to find conserved quantities. Recall that the Euler-Lagrange equation says

$$\frac{d}{d\lambda}\frac{\partial \mathcal{L}}{\partial(dx^A/d\lambda)} - \frac{\partial \mathcal{L}}{\partial x^A} = 0 \qquad \text{where} \qquad \mathcal{L} = \sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}} = \frac{d\tau}{d\lambda}.$$

If $g_{\mu\nu}$ is independent of x^A, then $\partial\mathcal{L}/\partial x^A = 0$, which means

$$\frac{\partial\mathcal{L}}{\partial(dx^A/d\lambda)} \text{ is conserved along the geodesic.}$$

This statement depends on the choice of coordinates. A coordinate-independent way to find a conserved quantity is to use the Killing vector. Let $\xi = \boldsymbol{e}_A = \delta^\mu{}_A \boldsymbol{e}_\mu$ be a Killing vector, so the components of it are $\xi^\mu = \delta^\mu{}_A$. Then we can say

$$\frac{\partial\mathcal{L}}{\partial(dx^A/d\lambda)} = \delta^\mu_A \frac{\partial\mathcal{L}}{\partial(dx^\mu/d\lambda)} = \xi^\mu \frac{\partial\mathcal{L}}{\partial(dx^\mu/d\lambda)} \text{ is conserved along the geodesic.}$$

Writing out this derivative explicitly,

$$\xi^\mu \left(\frac{2g_{\mu\nu} dx^\nu/d\lambda}{2\mathcal{L}} \right) = -\xi^\mu g_{\mu\nu} \frac{dx^\nu}{d\tau} = -g_{\mu\nu}\xi^\mu u^\nu.$$

In other words,

$$\boxed{\xi \cdot u \text{ is conserved along the geodesic.}} \tag{5.4}$$

It is more useful to write it in terms of 4-momentum $p = mu$ when constructing conserved quantities,

$$\boxed{\xi \cdot p \text{ is conserved along the geodesic.}}$$

Note that $\xi \cdot p$ is a scalar, which depends only on the geometry of spacetime.

6 The Field Equations

6.1 The Purpose and Justification of the Field Equations

In Newtonian mechanics, the equation of gravity is

$$\nabla^2 \phi = 4\pi G \rho,$$

where ρ is the density and ϕ is the potential. The solution is

$$\phi = -\frac{Gm}{r}.$$

The source of the gravitational field in Newtonian mechanics is the mass density. In GR, the source is the whole of the stress-energy tensor $\mathbf{T}$. The equation will then have the form

$$\mathbf{O}(\mathbf{g}) = k\mathbf{T},$$

where k is a constant and $\mathbf{O}$ is a differential operator on the metric tensor $\mathbf{g}$. There will be 10 differential equations because $\mathbf{T}$ is symmetric instead of 16. The second-order differential operator $\mathbf{O}$ must produce a tensor of rank $\binom{2}{0}$. Its components $\{O^{\mu\nu}\}$ must be combinations of $\partial_{\lambda\sigma} g_{\mu\nu}$, $\partial_\lambda g_{\mu\nu}$ and $g_{\mu\nu}$. It turns out that the tensor of the form

$$O^{\mu\nu} = R^{\mu\nu} + \beta g^{\mu\nu} R + \Lambda g^{\mu\nu}$$

will satisfy the condition. The conservation of energy and momentum have

$$\nabla_\nu T^{\mu\nu} = 0 \quad \Longrightarrow \quad \nabla_\nu O^{\mu\nu} = 0.$$

Since $\nabla_\alpha g^{\mu\nu} = 0$, we find that

$$\partial_\nu (R^{\mu\nu} + \beta g^{\mu\nu} R) = 0.$$

From the twice-contracted Bianchi identity, we must have $\beta = -1/2$. Thus, by this chain of argument,

$$G^{\mu\nu} + \Lambda g^{\mu\nu} = kT^{\mu\nu} \tag{6.1}$$

In following sections, we use the geometrized units where $c = G = 1$ (G is the gravitational constant).

6.2 Einstein's Equations

Take $\Lambda = 0$ and $k = 8\pi$,

$$G^{\mu\nu} = 8\pi T^{\mu\nu} \tag{6.2}$$

The constant Λ is called the **cosmological constant**, which is small but not zero. The fact that $l = 8\pi$ is obtained by requiring that Einstein's equations must be consistent with previous theories like Newton's of Kepler's. The Einstein's equations have ten component equations, but Bianchi identities say that

$$\nabla_\nu G^{\mu\nu} = 0.$$

This means there are four identities, one for each value of μ, among the ten equations. Hence there are just six independent differential equations.

7 Weak Fields and Gravitational Radiation

7.1 Einstein's Equations for Weak Gravitational Fields

A weak gravitational field has spacetime "nearly" flat. The metric is defined as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{7.1}$$

where $|h_{\mu\nu}| \ll 1$ everywhere in spacetime. Basically, we consider $h_{\mu\nu}$ as a metric perturbation on Minkowski metric.

7.1.1 Background Lorentz Transformations and Gauge Transformations

The matrix of a Lorentz transformation in SR is

$$[\Lambda^{\bar{\mu}}{}_{\nu}] = \begin{pmatrix} \gamma & -v\gamma & 0 & 0 \\ -v\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \gamma = (1 - v^2)^{-1/2}.$$

For weak gravitational fields, define a "background Lorentz transformation" with the form

$$x^{\bar{\mu}} = \Lambda^{\bar{\mu}}{}_{\nu} x^{\nu}.$$

Then the transformation of the metric tensor will be

$$g_{\bar{\mu}\bar{\nu}} = \Lambda^{\alpha}{}_{\bar{\mu}}\Lambda^{\beta}{}_{\bar{\nu}} g_{\alpha\beta} = \Lambda^{\alpha}{}_{\bar{\mu}}\Lambda^{\beta}{}_{\bar{\nu}}\eta_{\alpha\beta} + \Lambda^{\alpha}{}_{\bar{\mu}}\Lambda^{\beta}{}_{\bar{\nu}} h_{\alpha\beta}.$$

The Lorentz transformation satisfy

$$\Lambda^{\alpha}{}_{\bar{\mu}}\Lambda^{\beta}{}_{\bar{\nu}}\eta_{\alpha\beta} = \eta_{\bar{\mu}\bar{\nu}},$$

so we get

$$g_{\bar{\mu}\bar{\nu}} = \eta_{\bar{\mu}\bar{\nu}} + h_{\bar{\mu}\bar{\nu}}, \quad \text{where} \quad h_{\bar{\mu}\bar{\nu}} = \Lambda^{\alpha}{}_{\bar{\mu}}\Lambda^{\beta}{}_{\bar{\nu}} h_{\alpha\beta}.$$

There is an important coordinate change that leaves 7.1 unchanged: a very small change in coordinates called the **gauge transformation** of the form

$$x^{\mu'} = x^{\mu} + \xi^{\mu}(x^{\nu}),$$

where vector ξ^{μ} is a function of position. If ξ^{μ} is small and $|\partial_{\nu}\xi^{\mu}| \ll 1$, then

$$\Lambda^{\mu'}{}_{\nu} = \frac{\partial x^{\mu'}}{\partial x^{\nu}} = \delta^{\mu}{}_{\nu} + \partial_{\nu}\xi^{\mu}.$$
$$\Lambda^{\mu}{}_{\nu'} = \frac{\partial x^{\mu}}{\partial x^{\nu'}} = \frac{\partial x^{\mu}}{\partial(x^{\nu} + \xi^{\mu})} = \frac{1}{(\delta^{\nu}{}_{\mu} + \partial_{\mu}\xi^{\nu})} = \delta^{\mu}{}_{\nu} - \partial_{\nu}\xi^{\mu} + \mathcal{O}(|\partial_{\nu}\xi^{\mu}|^2).$$

We can varify that to the first order in small quantities ($\partial_\nu \xi_\mu$ and $h_{\mu\nu}$),

$$\begin{aligned}
g_{\mu'\nu'} = \Lambda^\alpha{}_{\mu'}\Lambda^\beta{}_{\nu'}g_{\alpha\beta} &= (\delta^\alpha{}_\mu - \partial_\mu\xi^\alpha)(\delta^\beta{}_\nu - \partial_\nu\xi^\beta)g_{\alpha\beta} \\
&= g_{\mu\nu} - \delta^\alpha{}_\mu\partial_\nu\xi^\beta g_{\alpha\beta} - \delta^\beta{}_\nu\partial_\mu\xi^\alpha g_{\alpha\beta} + \mathcal{O}(\partial_\mu\xi^\alpha\partial_\nu\xi^\beta)g_{\alpha\beta} \\
&= g_{\mu\nu} - \partial_\nu\xi^\beta g_{\mu\beta} - \partial_\mu\xi^\alpha g_{\alpha\nu} \\
&= g_{\mu\nu} - \partial_\nu\xi^\beta(\eta_{\mu\beta} + h_{\mu\beta}) - \partial_\mu\xi^\alpha(\eta_{\alpha\nu} + h_{\alpha\nu}) \\
&= g_{\mu\nu} - \partial_\nu\xi^\mu - \partial_\mu\xi_\nu - \partial_\nu\xi^\beta h_{\mu\beta} - \partial_\mu\xi^\alpha h_{\alpha\nu} \\
&= \eta_{\mu\nu} + h_{\mu\nu} - \partial_\nu\xi_\mu - \xi_\mu\xi_\nu.
\end{aligned}$$

This means the coordinate change is to change

$$h_{\mu\nu} \to h_{\mu\nu} - \partial_\nu\xi_\mu - \partial_\mu\xi_\nu.$$

In nearly flat space defined by 7.1, some useful quantites are

- The Christoffel symbols:

$$\Gamma^\mu{}_{\nu\alpha} = \frac{1}{2}\eta^{\mu\beta}(\partial_\alpha h_{\beta\nu} + \partial_\nu h_{\beta\alpha} - \partial_\beta h_{\nu\alpha}) = \frac{1}{2}(\partial_\alpha h^\mu{}_\nu + \partial_\nu h^\mu{}_\alpha - \partial^\mu h_{\nu\alpha}).$$

- The Riemann curvature tensor (to the first order):

$$\begin{aligned}
R^\mu{}_{\nu\alpha\beta} &= \partial_\alpha\Gamma^\mu{}_{\nu\beta} - \partial_\beta\Gamma^\mu{}_{\nu\alpha} \\
&= \frac{1}{2}(\partial_\alpha\partial_\nu h^\mu{}_\beta + \partial_\beta\partial^\mu h_{\nu\alpha} - \partial_\alpha\partial^\mu h_{\nu\beta} - \partial_\beta\partial_\nu h^\mu{}_\alpha).
\end{aligned}$$

- The Ricci tensor:

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu} = \frac{1}{2}(\partial_\alpha\partial_\nu h^\alpha{}_\mu + \partial^\alpha\partial_\mu h_{\nu\alpha} - \Box h_{\mu\nu} - \partial_\mu\partial_\nu h),$$

  where $h = h^\mu{}_\mu$ (the trace) and $\Box = \partial_\mu\partial^\mu = \nabla^2 - \partial_t^2$.

- The Ricci scalar:

$$R = R^\mu{}_\mu = \partial_\nu\partial^\mu h^\nu{}_\mu - \Box h.$$

7.1.2 Weak-Field Einstein Equations

Define index-raised quantities

$$h^\mu{}_\beta = \eta^{\mu\alpha}h_{\alpha\beta}, \qquad \text{and} \qquad h^{\mu\nu} = \eta^{\nu\beta}h^\mu{}_\beta,$$

the trace

$$h = h^\mu{}_\mu,$$

and the "trace reverse" tensor of $h_{\mu\nu}$

$$\bar{h}^{\mu\nu} = h^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}h \quad \text{with property} \quad -\bar{h} = -\bar{h}^\mu{}_\mu = h.$$

The inverse of trace reverse tensor is the same,

$$h^{\mu\nu} = \bar{h}^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}\bar{h}.$$

With these definitions together with $\partial^\mu f = \eta^{\mu\nu}\partial_\nu f$, the Einstein tensor is

$$G_{\mu\nu} = -\frac{1}{2}[\partial^\alpha\partial_\alpha\bar{h}_{\mu\nu} + \eta_{\mu\nu}\partial^{\alpha\beta}\bar{h}_{\alpha\beta} - \partial^\alpha\partial_\nu\bar{h}_{\mu\alpha} - \partial^\alpha\partial_\mu\bar{h}_{\nu\alpha} + \mathcal{O}(h_{\mu\nu}^2)].$$

This will get simplified if require

$$\partial_\nu\bar{h}^{\mu\nu} = 0.$$

This is possible because there is always a gauge such that this requirement is true. This gauge condition is called **Lorentz gauge** condition.

Proof. Consider an arbitrary $\bar{h}_{\mu\nu}$ for which $\partial_\nu \bar{h}^{\mu\nu} \neq 0$. Under gauge transformation,

$$\bar{h}^{(\text{new})}_{\mu\nu} = \bar{h}^{(\text{old})}_{\mu\nu} - \partial_\nu \xi_\mu - \partial_\mu \xi_\nu + \eta_{\mu\nu}\partial_\alpha \xi^\alpha.$$

The divergence is then

$$\partial_\nu \bar{h}^{(\text{new})\mu\nu} = \partial_\nu \bar{h}^{(\text{old})\mu\nu} - \partial_\nu \partial^\nu \xi^\mu.$$

If the gauge $\partial_\nu \bar{h}^{(\text{new})\mu\nu} = 0$, then

$$\Box \xi^\mu = \partial_\nu \partial^\nu \xi^\mu = \partial_\nu \bar{h}^{(\text{old})\mu\nu}.$$

This is a three-dimensional inhomogeneous wave equation and it always has a solution ξ^μ. In fact, the solution is not unique. It can take the form$(\xi^\mu + \eta^\mu)$. □

In this gauge, the Einstein tensor becomes

$$G^{\mu\nu} = -\frac{1}{2}\Box \bar{h}^{\mu\nu}$$

and the weak-field Einstein equations are

$$\Box \bar{h}^{\mu\nu} = -16\pi T^{\mu\nu} \tag{7.2}$$

7.2 Newtonian Gravitational Fields

7.2.1 Newtonian Limit

Newtonian gravity satisfies $|\phi| \ll 1$ and $|v| \ll 1$. GR must predict the same result as Newtonian gravity. The velocities are small means the stress-energy tensor typically obeys $\left|T^{00}\right| \gg \left|T^{0i}\right| \gg \left|T^{ij}\right|$ and similar property for $h^{\mu\nu}$. We should expect the dominant Newtonian gravitational field comes from the dominant field equation

$$\Box \bar{h}^{00} = -16\pi\rho,$$

where $T^{00} = \rho + \mathcal{O}(\rho v^2)$. The time component of $\Box$ vanishes because $\partial_t = v\partial_x \approx 0$, so

$$\Box = \nabla^2 + \mathcal{O}(v^2\nabla^2).$$

The equation is then

$$\nabla^2 \bar{h}^{00} = -16\pi\rho.$$

Comparing this with the Newtonian equation:

$$\nabla^2\phi = 4\pi\rho \quad \Longrightarrow \quad \bar{h}^{00} = -4\phi.$$

With other components of $\bar{h}^{\mu\nu}$ negligible,

$$h = h^\mu{}_\mu = -\bar{h}^\mu{}_\mu = \bar{h}^{00}.$$

Then we can calculate $h^{\mu\nu}$:

$$\begin{aligned} h^{00} &= \bar{h}^{00} - \frac{1}{2}\eta^{00}\bar{h} = -4\phi - \frac{1}{2}(-1)(4\phi) = -2\phi, \\ h^{xx} &= h^{yy} = h^{zz} = 0 - \frac{1}{2}(4\phi) = -2\phi. \end{aligned}$$

or

$$ds^2 = -(1+2\phi)dt^2 + (1-2\phi)(dx^2 + dy^2 + dz^2).$$

7.2.2 The Far Field of Stationary Relativistic Sources and its Mass

For any source, when we go far away enough, the spacetime becomes asymptotically flat. We assume that $T^{\mu\nu}$ is stationary so that $h^{\mu\nu}$ is independent of time. Far away from the source, with no energy density presented, the linearized field equations can be written directly as

$$\nabla^2 \bar{h}^{\mu\nu} = 0,$$

with solution

$$\bar{h}^{\mu\nu} = \frac{A^{\mu\nu}}{r} + \mathcal{O}(r^{-2}).$$

where $A^{\mu\nu}$ is a constant. Also, the gauge condition must be satisfied:

$$0 = \partial_\nu \bar{h}^{\mu\nu} = \partial_j \bar{h}^{\mu j} = -\frac{A^{\mu j} n_j}{r^2} + \mathcal{O}(r^{-3}),$$

where n_j is the unit radial normal. This equation holds for all x^i, which means that all $A^{\mu j}$ vanishes and only $\bar{h}^{00}$ left, or

$$\left|\bar{h}^{00}\right| \gg \left|\bar{h}^{ij}\right|, \quad \left|\bar{h}^{00}\right| \gg \bar{h}^{0j}.$$

This shows that the far field behaves just like Newtonian field with potential

$$\phi = -\frac{M}{r} \quad \text{if} \quad A^{00} = 4M.$$

Any small body falls freely in the relativistic source's gravitational field but stay far way from it follow the geodesics of the metric. The geodesics obey Kepler's laws for the gravitational field of a body of **total mass** M. We can write for any stationary source (from far away):

$$ds^2 = -[1 - \frac{2M}{r} + \mathcal{O}(r^{-2})]\, dt^2 + [1 + \frac{2M}{r} + \mathcal{O}(r^{-2})](dx^2 + dy^2 + dz^2).$$

This mass of the relativistic source is not the integral of the source by adding up the particles. The mass is measured by observing the orbits of small bodies like planets. The behavior of a relativistic source, such as the one of black holes, may be different from Newtonian masses.

7.3 Gravitational Waves

Sometimes the gravitational field may be weak but not stationary. Consider the weak-field Einstein equations 7.2 in vacuum ($T^{\mu\nu} = 0$) far outside the source of the field:

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2\right) \bar{h}^{\mu\nu} = 0. \tag{7.3}$$

This the three-dimensional wave equation and it has a solution of the form

$$\bar{h}^{\mu\nu} = A^{\mu\nu} e^{ik_\alpha x^\alpha},$$

where k_μ are real constant components of some one-form and $A^{\mu\nu}$ the complex constant components of some tensor. Eq. 7.3 can be written as

$$\eta^{\alpha\beta} \partial_\alpha \partial_\beta \bar{h}^{\mu\nu} = 0$$

and from the solution, we have

$$\partial_\alpha \bar{h}^{\mu\nu} = ik_\alpha \bar{h}^{\mu\nu}.$$

Thus, in conclusion,

$$\eta^{\alpha\beta} \partial_\alpha \partial_\beta \bar{h}^{\mu\nu} = -\eta^{\alpha\beta} k_\alpha k_\beta \bar{h}^{\mu\nu} = 0 \quad \Longrightarrow \quad \eta^{\alpha\beta} k_\alpha k_\beta = k^\beta k_\beta = 0.$$

This means k^μ is a null vector, or tangent to the world line of a photon. The value of $\bar{h}^{\mu\nu}$ is constant on a hypersurface on which $k_\mu x^\mu$ is constant. The vector $k^\mu = (\omega, \mathbf{k})$ has k^0 as the frequency of the wave and $\mathbf{k}$ as the usual wavenumber. Consider a photon moving in the direction of $\vec{k}$ which travels on a curve

$$x^\mu(\lambda) = k^\mu \lambda + l^\mu,$$

where λ is the parameter and l^μ is the photon's position at $\lambda = 0$. This means

$$k_\mu x^\mu(\lambda) = k_\mu l^\mu = \text{const}.$$

The photon travels with the gravitational wave at the same phase. The gravitational wave travels at the speed of light. The gauge condition $\partial_\nu \bar{h}^{\mu\nu} = 0$ and $\partial_\alpha \bar{h}^{\mu\nu} = ik_\alpha \bar{h}^{\mu\nu}$ implies that

$$k_\nu A^{\mu\nu} = 0. \tag{7.4}$$

This is a restriction on $A^{\mu\nu}$: it is orthogonal to k^ν. The solution $A^{\mu\nu} e^{ik_\alpha x^\alpha}$ is a plane wave.

7.3.1 The Transverse-Traceless Gauge

We can impose more gauge condition by solving

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2\right) \xi_\mu = 0.$$

Similarly, the solution is in the form

$$\xi_\mu = B_\mu e^{ik_\nu x^\nu}.$$

This produces a change in $h^{\mu\nu}$ by

$$h_{\mu\nu}^{(\text{new})} = h_{\mu\nu}^{(\text{old})} - \partial_\nu \xi_\mu - \partial_\mu \xi_\nu \quad \text{or} \quad \bar{h}_{\mu\nu}^{(\text{new})} = \bar{h}_{\mu\nu}^{(\text{old})} - \partial_\nu \xi_\mu - \partial_\mu \xi_\nu + \eta_{\mu\nu} \partial_\alpha \xi^\alpha.$$

Factoring out all exponential term for $A_{\mu\nu}$ gives

$$A_{\mu\nu}^{(\text{new})} = A_{\mu\nu}^{(\text{old})} - iB_\mu k_\nu - iB_\nu k_\mu + i\eta_{\mu\nu} B^\alpha k_\alpha.$$

Carefully selecting B_μ will impose two restrictions on $A_{\mu\nu}^{(\text{new})}$:

$$A^\mu{}_\mu = 0 \qquad \text{and} \qquad A_{\mu\nu} u^\nu = 0,$$

where u is some fixed 4-velocity (any constant timelike unit vector). These two conditions together with (7.4) are called **transverse-traceless** (TT) gauge conditions. These conditions imply

$$\bar{h}_{\mu\nu}^{\text{TT}} = h_{\mu\nu}^{\text{TT}}.$$

Choose a background Lorentz frame and set $u^\mu = \delta^\mu{}_0$. The gauge conditions will give $A_{\mu 0} = 0$. Set the wave to travel in the z-direction, $k^\mu = (\omega, 0, 0, \omega)$. Then (7.4) gives $A_{\mu z} = 0$. In this way, only A_{xx}, A_{yy}, and $A_{xy} = A_{yx}$ are nonzero. Finally, the trace condition $A^\mu{}_\mu$ implies $A_{xx} = -A_{yy}$. In conclusion,

$$h_{\mu\nu}^{\text{TT}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & A_+ & A_\times & 0 \\ 0 & A_\times & -A_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} e^{i\omega(t-z)} \equiv \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where we replaced $A_{xx} = A_+$ and $A_{xy} = A_\times$, for reasons to be clear in a moment. Once we have the plane wave solutions, it is possible to construct arbitrary function of the form $h_+(t-z)$ and $h_\times(t-z)$ by Fourier analysis. The line element using the metric $h_{\mu\nu}^{\text{TT}}$ can be written as

$$ds^2 = -dt^2 + [1 + h_+]dx^2 + 2h_\times dx\, dy + [1 - h_+]dy^2 + dz^2. \tag{7.5}$$

7.3.2 The Effects of Waves on Free Particles

A free particle obeys the geodesic equation,

$$\frac{d}{d\tau} u^\mu + \Gamma^\mu{}_{\alpha\beta} u^\alpha u^\beta = 0.$$

Since the particle is initially at rest, the only non-trivial part of acceleration is

$$\frac{d}{d\tau} u^\mu = -\Gamma^\mu{}_{tt} = -\frac{1}{2} \eta^{\mu\nu} (\partial_t h_{\nu t} + \partial_t h_{t\nu} - \partial_\nu h_{tt}).$$

But because all $h_{\nu t}^{\text{TT}} = 0$, $\Gamma^\mu{}_{tt}$ also vanishes. Thus, the component of particle's acceleration is 0 forever, but this just means the particle stays "at rest", in *coordinate position*. To measure the real distance, we need to calculate the proper distance between particles. Consider one particle at the origin with coordinates $(t, 0, 0, 0)$ and another at $(t, x_b, y_b, 0)$.

7.4 The Generation of Gravitational Waves

7.4.1 Slow Motion Wave Generation

The object is to solve

$$\left(-\frac{\partial^2}{\partial t^2} + \nabla^2\right) \bar{h}_{\mu\nu} = -16\pi T_{\mu\nu}.$$

Assume that $T_{\mu\nu}$ is in sinusoical oscillation with frequency Ω,

$$T_{\mu\nu} = S_{\mu\nu}(x^i) e^{-i\Omega t}.$$

The region of space in which $S_{\mu\nu} \neq 0$ is small compared with the wavelength of the gravitational wave $2\pi/\Omega$. Suppose we are looking for the solution of the form

$$\bar{h}_{\mu\nu} = B_{\mu\nu}(x^i) e^{-\Omega t}.$$

Putting this and the assumption into the equation gives

$$(\nabla^2 + \Omega^2) B_{\mu\nu} = -16\pi S_{\mu\nu}. \tag{7.6}$$

Outside the source where $S_{\mu\nu} = 0$, the solution for $B_{\mu\nu}$ in spherical polar coordinate is

$$B_{\mu\nu} = \frac{A_{\mu\nu}}{r} e^{i\Omega r} + \frac{Z_{\mu\nu}}{r} e^{-i\Omega r}.$$

The term $e^{-i\Omega r}$ is a wave toward the origin and the other term is outword. If we only want wave emitted by the source, $Z_{\mu\nu} = 0$. Now is to determine $A_{\mu\nu}$. Consider a sphere of radius $\epsilon \ll 2\pi/\Omega$ only in which the source is nonzero. Integrate Eq. 7.6 gives

$$\int \Omega^2 B_{\mu\nu}\, dx^3 \leq \Omega^2 |B_{\mu\nu}|_{\max} \frac{4\pi\epsilon^3}{3}$$

and

$$\int \nabla^2 B_{\mu\nu}\, dx^3 = \oint \mathbf{n} \cdot \nabla B_{\mu\nu}\, dS = 4\pi\epsilon^2 \left[\frac{d}{dr} B_{\mu\nu}\right]_{r=\epsilon} \approx -4\pi A_{\mu\nu}.$$

The RHS is

$$J_{\mu\nu} = \int S_{\mu\nu}\, d^3x.$$

Taking the limit $\epsilon \to 0$ gives

$$A_{\mu\nu} = 4J_{\mu\nu} \quad \Longrightarrow \quad \bar{h}_{\mu\nu} = 4J_{\mu\nu} \frac{e^{i\Omega(r-t)}}{r}.$$

8 Schwarzschild Solutions

8.1 Coordinates for Spherically Symmetric Spacetime

Now we come to the study of strong gravitational fields in GR, in which we will use spherically symmetric systems. This is convenient, and also many celestial objects appear to be nearly spherical.

8.1.1 Flat Space in Spherical Coordinates and Two-Spheres

The line element of Minkowski space can be written as

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2).$$

Each surface of constant r and t is a two-sphere. Distance along curves with $dt = dr = 0$ is

$$d\ell^2 = r^2(d\theta^2 + \sin^2\theta\, d\phi^2) = r^2 d\Omega^2.$$

Any two-surface with line-element independent of θ and ϕ has the intrinsic geometry of a two-sphere. Every point of spacetime is on a two-surface which is a two-sphere with line element

$$d\ell^2 = f(r,t)(d\theta^2 + \sin^2\theta\, d\phi^2),$$

where $f(r,t)$ is an unknown function. The area of each sphere is $4\pi f(r,t)$.

8.1.2 Spherically Symmetric Spacetime

When t is constant, consider spheres at r and $r + dr$, each with a coordinate system (θ, ϕ). To define the orientation of spheres, we need $\theta = \text{const.}$ and $\phi = \text{const.}$ orthogonal to the two-spheres. This means

$$\boldsymbol{e}_r \cdot \boldsymbol{e}_\theta = \boldsymbol{e}_r \cdot \boldsymbol{e}_\phi = 0$$

or

$$g_{r\theta} = g_{r\phi} = 0.$$

When t is not constant, we must require the whole spacetime is spherically symmetric, so $r = \text{const.}$ is also orthogonal to the two-spheres. Now we have

$$ds^2 = g_{00}t^2 + 2g_{0r}dr\,dt + g_{rr}dr^2 + r^2 d\Omega^2,$$

where g_{00}, g_{0r} and g_{rr} are functions of r and t.

8.2 Static Spherically Symmetric Spacetime

8.2.1 The Metric

A **static spacetime** has two properties:

- All metric components are independent of t, i.e. stationary.
- The geometry is unchanged by time reversal, $t \to -t$.

A spacetime with the first property but not the second is said to be **stationary**. Consider a coordinate transformation $(t, r, \theta, \phi) \to (-t, r, \theta, \phi)$ that has $\Lambda^{\bar{0}}{}_0 = -1$ and $\Lambda^{\bar{i}}{}_j$,

$$\begin{cases} g_{\bar{0}\bar{0}} = (\Lambda^0{}_{\bar{0}})^2 g_{00} = g_{00} \\ g_{\bar{0}\bar{r}} = \Lambda^0{}_{\bar{0}}\Lambda^r{}_{\bar{r}} g_{0r} = -g_{0r} \\ g_{\bar{r}\bar{r}} = (\Lambda^r{}_{\bar{r}})^2 g_{rr} = g_{rr}. \end{cases}$$

To have the geometry unchanged we must have $g_{0r} = 0$. Thus, the metric of a static, spherically symmetric spacetime is

$$ds^2 = -e^{2\Phi(r)}dt^2 + e^{2\Lambda(r)}dr^2 + r^2 d\Omega^2.$$

Later we will se that this metric will hold inside stars but not black holes. Also we need the boundary condition that the spacetime is asymptotically flat far from the star, so

$$\lim_{r\to\infty} \Phi(r) = \lim_{r\to\infty} \Lambda(r) = 0.$$

There are some physical interpretation of the metric terms. First, the proper radial distance from radius r_1 to r_2 is

$$\ell_{12} = \int_{r_1}^{r_2} e^{\Lambda} dr.$$

Second, since the metric is independent of t, any particle following a geodesic will have a constant momentum component $p_0 = -E$. However, a local inertial observer momentarily *at rest* measures a different energy. The 4-velocity must have $U^i = 0$, so $u \cdot u = -1$ implies $U^0 = e^{-\Phi}$. The energy observed is

$$E_\infty = -u \cdot p = g_{\mu\nu} u^\mu p^\nu = g_{00} u^0 p^0 = (e^{2\Phi}) e^{-\Phi} E_{\text{em}} = e^{\Phi} E_{\text{em}},$$

where E_{em} is the emitted energy at coordinate radius r. This result is significant for photons. If a photon is emitted at r, its energy is $E_{\text{em}} = h\nu_r$. The received energy is then $E_\infty = h\nu_r e^{\Phi(r)}$ or frequency $\nu_\infty = \nu_r e^{\Phi(r)}$. The redshift of the photon is then

$$z = \frac{\lambda_\infty}{\lambda_r} - 1 = \frac{\nu_r}{\nu_\infty} - 1 = \frac{\nu_r}{\nu_r e^{\Phi(r)}} - 1 = e^{-\Phi(r)} - 1. \tag{8.1}$$

8.2.2 The Einstein Tensor

We can calculate from the metric, the nonzero components of the Einstein tensor is

$$\begin{aligned} G_{00} &= \frac{1}{r^2} e^{2\Phi} \frac{d}{dr}[r(1 - e^{-2\Lambda})], \\ G_{rr} &= -\frac{1}{r^2} e^{2\Lambda}(1 - e^{-2\Lambda}) + \frac{2}{r}\Phi', \\ G_{\theta\theta} &= r^2 e^{-2\Lambda}\left[\Phi'' + (\Phi')^2 + \frac{\Phi'}{r} - \Phi'\Lambda' - \frac{\Lambda'}{r}\right], \\ G_{\phi\phi} &= G_{\theta\theta}\sin^2\theta. \end{aligned}$$

8.3 Static Perfect Fluid Einstein Equations

8.3.1 The Stress-Energy Tensor

Static stars are fluids with no motion, so the only nonzero component of u is U^0. The normalization condition $u \cdot u = -1$ implies

$$U^0 = e^{-\Phi}, \quad U_0 = -e^{\Phi}.$$

The stress-energy tensor is then

$$T_{\mu\nu} = \begin{bmatrix} \rho e^{2\Phi} & 0 & 0 & 0 \\ 0 & Pe^{2\Lambda} & 0 & 0 \\ 0 & 0 & Pr^2 & 0 \\ 0 & 0 & 0 & Pr^2\sin^2\theta \end{bmatrix}.$$

The pressure P, energy density ρ and specific entropy S can be related by an equation of state,

$$P = P(\rho, S). \tag{8.2}$$

The entropy cna be considered negligibly small for many situations, so we have $P = P(\rho)$. The conservation laws by stress-energy tensor are

$$\nabla_\nu T^{\mu\nu} = 0.$$

There are four equations, but only when $\mu = r$ the equation is not trivial:

$$(\rho + P)\frac{d\Phi}{dr} = -\frac{dP}{dr}. \tag{8.3}$$

This equation means pressure gradient is needed to keep the fluid static in the gravitational field.

8.3.2 Einstein Equations

Replace $\Lambda(r)$ with function

$$m(r) = \frac{1}{2}r(1 - e^{-2\Lambda}),$$

so that

$$g_{rr} = e^{2\Lambda} = \left(1 - \frac{2m(r)}{r}\right)^{-1}.$$

The $(0,0)$ Einstein equation implies

$$\frac{dm(r)}{dr} = 4\pi r^2 \rho. \tag{8.4}$$

This is same as the Newtonian equation, where $m(r)$ is the mass function. With this definition, the (r, r) equation is

$$\frac{d\Phi}{dr} = \frac{m(r) + 4\pi r^3 P}{r[r - 2m(r)]}. \tag{8.5}$$

Here, Eq. 8.2, 8.3, 8.4, 8.5 are four equations for ρ, P, m, Φ. The (θ, θ) and (ϕ, ϕ) Einstein equations do not provide additional information.

8.4 The Exterior Geometry

8.4.1 Schwarzschild Metric

Outside the star, $\rho = P = 0$, so we get two equations

$$\frac{dm}{dr} = 0 \qquad \text{and} \qquad \frac{d\Phi}{dr} = \frac{m}{r(r - 2m)}.$$

They have solutions

$$m(r) = M = \text{const.} \qquad \text{and} \qquad e^{2\Phi} = 1 - \frac{2M}{r}.$$

This satisfies the requirement that $\Phi \to 0$ as $r \to \infty$. This exterior metric is called the **Schwarzschild metric**,

$$\boxed{ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2.} \tag{8.6}$$

The coordinates (t, r, θ, ϕ) are **Schwarzschild coordinates**. They *do not* have the same physical meaning as the Minkowski, spherical coordinates because spacetime is curved now. There are two interesting radius r that blows up the metric: $r = 0$ and $r = 2M$. We will analyze them in Chapter 9. For now, let's focus on the exterior geometry of the Schwarzschild metric.

Theorem 8.1. Birkhoff's Theorem

The spacetime outside any spherically symmetric, electrically neutral body is given by Schwarzschild geometry,

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2.$$

There is one parameter M that we identify with the mass of the central object. To justify this physical meaning, we shall show that at large r, we recover the Newtonian weak-field spacetime metric,

$$ds^2 = -(1 - 2\Phi)dt^2 + (1 + 2\Phi)(dr^2 + r^2 d\Omega^2).$$

Note that the Schwarzschild metric already resembles the weak-field metric if we take $r \to \infty$, but the angular part is not matching. Consider a coordinate transformation from Schwarzschild coordinates (t, r, θ, ϕ) to **isotropic coordinates** $(t, \bar{r}, \theta, \phi)$, with $\bar{r}$ defined by

$$r = \bar{r}\left(1 + \frac{M}{2\bar{r}}\right)^2 = \bar{r} + M + \frac{M^2}{4\bar{r}}, \qquad dr = \left(1 - \frac{M^2}{4\bar{r}^2}\right) d\bar{r} = \left(1 + \frac{M}{2\bar{r}}\right)\left(1 - \frac{M}{2\bar{r}}\right) d\bar{r}.$$

Then

$$1 - \frac{2M}{r} = 1 - \frac{2M}{\bar{r}(1 + M/2\bar{r})^2} = \frac{(1 + M/2\bar{r})^2 - 2M/\bar{r}}{(1 + M/2\bar{r})^2} = \frac{(1 - M/2\bar{r})^2}{(1 + M/2\bar{r})^2}.$$

The Schwarzschild metric in isotropic coordinates is

$$ds^2 = -\left(\frac{1 - M/2\bar{r}}{1 + M/2\bar{r}}\right)^2 dt^2 + \left(1 + \frac{M}{2\bar{r}}\right)^4 (d\bar{r}^2 + \bar{r}^2 d\Omega^2).$$

Now the spatial part looks like Euclidean space in spherical coordinates. So far the metric is still *exact.* Now as $r \to \infty$, or $M/\bar{r} \to 0$, we can Taylor expand the metric using $(1 + \epsilon)^n \simeq 1 + n\epsilon$:

$$\left(\frac{1 - M/2\bar{r}}{1 + M/2\bar{r}}\right)^2 \simeq \left(1 - \frac{M}{2\bar{r}}\right)^4 \simeq 1 - \frac{2M}{r}, \qquad \left(1 + \frac{M}{2\bar{r}}\right)^4 \simeq 1 + \frac{2M}{\bar{r}}.$$

This gives the Newtonian weak-field metric,

$$ds^2 \simeq (1 + 2\Phi)dt^2 + (1 - 2\Phi)(d\bar{r}^2 + \bar{r}^2 d\Omega^2),$$

where $\Phi = -GM/\bar{r}$ (expressing in ordinary units for physical meaning).

8.4.2 Gravitational Redshift

There are many ways to obtain the gravitational redshift in Schwarzschild spacetime. One method is to use (8.1), $z = e^{-\Phi(r)} - 1$, where we now know $e^{\Phi} = (1 - 2M/r)^{1/2}$. Hence a photon emitted at coordinate r is observed to have a redshift

$$z = \left(1 - \frac{2M}{r}\right)^{-1/2} - 1.$$

for a static observer at infinity.

By Time Dilation. We may also check this using gravitational time dilation directly. Since a photon's wave period is itself a clock, the gravitational time dilation therefore affects the frequency. We know that a static observer measures their proper time by

$$d\tau = \left(1 - \frac{2M}{r}\right)^{1/2} dt$$

in Schwarzschild spacetime. Suppose there is a static light source emitting two wave crests of a photon within a proper time period $\delta\tau = (1 - 2M/r)\,\delta t$, so the emitted frequency is $\nu_r = 1/\delta\tau$. A static observer at $r \to \infty$ measures the proper time interval between two wave crests to be δt, or a frequency $\nu_\infty = 1/\delta t$. Compare these frequencies,

$$\frac{\nu_\infty}{\nu_r} = \frac{\delta\tau}{\delta t} = \left(1 - \frac{2M}{r}\right)^{1/2} \quad \Longrightarrow \quad z = \frac{\lambda_\infty}{\lambda_r} - 1 = \frac{\nu_r}{\nu_\infty} - 1 = \left(1 - \frac{2M}{r}\right)^{-1/2} - 1.$$

By Killing Vectors. We can check gravitational redshift using Killing vectors for conserved quantities. For a static observer at radius r, its 4-velocity satisfy $u \cdot u = -1$, and its spatial component is zero. In Schwarzschild metric, it means

$$g_{\mu\nu} u^\mu u^\nu = g_{tt}(u^t)^2 = -1 \quad \Longrightarrow \quad u^t = \left(1 - \frac{2M}{r}\right)^{-1/2}.$$

Suppose there is an orthonormal basis in the observer's frame an u is its timelike basis vector. Then the energy measured by the observer will satisfy $\eta_{\mu\nu} p^t = p \cdot u$, or

$$E = -p \cdot u = -g_{\mu\nu} p^\mu u^\nu = -g_{tt} p^t u^t = \left(1 - \frac{2M}{r}\right) p^t \left(1 - \frac{2M}{r}\right)^{-1/2} = \left(1 - \frac{2M}{r}\right)^{1/2} p_t.$$

In other words,

$$h\nu_r = \left(1 - \frac{2M}{r}\right)^{1/2} p^t \bigg|_r, \quad h\nu_\infty = p^t \bigg|_\infty.$$

The goal now is to relate $p^t|_r$ to $p^t|_\infty$. It is achieved by realizing that there exists a timelike Killing vector ξ since the metric tensor does not depend on time. Then $\xi \cdot p = g_{\mu\nu} \xi^\mu p^\nu = g_{tt}(1) p^t$ is conserved. Hence

$$-\left(1 - \frac{2M}{r}\right) p^t \bigg|_r = -p^t \bigg|_\infty \quad \Longrightarrow \quad \left(1 - \frac{2M}{r}\right)^{1/2} \nu_r = \nu_\infty,$$

which is the desired gravitational redshift.

8.5 Orbits in Schwarzschild Spacetime

In Newtonian mechanics, we know that a spherically symmetric object gives rise to a $-1/r$ potential, and the orbits are conic sections (circles, ellipses, parabolas, and hyperbolas). In general relativity, the orbits of free particles are geodesics in spacetime. We will see that many quantities in Schwarzschild spacetime resemble those in Newtonian gravity.

8.5.1 The Effective Potential

We don't need to solve the geodesic equations. Instead, we can invoke conserved quantities using Killing vectors. Since the metric tensor is independent of t and ϕ, there are two conserved quantities,

$$e \equiv -\xi \cdot u \qquad \text{and} \qquad L \equiv \eta \cdot u,$$

where $\xi^\mu = \delta^\mu{}_t$ and $\eta^\mu = \delta^\mu{}_\phi$. These inner products (in Schwarzschild geometry) are

$$e = \left(1 - \frac{2M}{r}\right) \frac{dt}{d\tau} \qquad \text{and} \qquad L = r^2 \sin^2\theta \frac{d\phi}{d\tau}.$$

They are like conserved energy and conserved angular momentum per unit mass, and are integrals of motion for us. We argue that the particle should move in a plane, just like in Newtonian mechanics. It is always possible to orient the coordinates such that $\phi = 0$ and $u^\phi = d\phi/d\tau = 0$ at some instant. Then because $L = r^2 \sin^2\theta u^\phi = 0$ is conserved, the particle will continue to move along $\phi = 0$ with $d\phi/d\tau = 0$. Thus, without loss of generality, we can choose a convenient coordinates such that the particle always move in the $\theta = \pi/2$ plane with $u^\theta = 0$. Now every $\sin\theta$ becomes a 1. In particular, $L = r^2 d\phi/d\tau$. The last integral of motion is $u \cdot u = -1$, which is true in all geometries,

$$-\left(1 - \frac{2M}{r}\right)\left(\frac{dt}{d\tau}\right)^2 + \left(1 - \frac{2M}{r}\right)^{-1}\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\phi}{d\tau}\right)^2 = -1.$$

Multiplying both sides by $(1-2M/r)$ and plugging in the conserved quantities, we get an equation for $r(\tau)$,

$$-e^2 + \left(\frac{dr}{d\tau}\right)^2 + \left(1+\frac{L^2}{r^2}\right)\left(1-\frac{2M}{r}\right) = 0.$$

Multiplying both sides by 1/2 and then minus 1/2,

$$\frac{e^2-1}{2} = \frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 + \frac{1}{2}\left(1+\frac{L^2}{r^2}\right)\left(1-\frac{2M}{r}\right) - \frac{1}{2} = \frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 + \frac{L^2}{2r^2} - \frac{M}{r} - \frac{L^2 M}{r^3}.$$

Now the equation looks more familiar if we write

$$\boxed{\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 + V_{\text{eff}}(r) = \mathcal{E} \qquad \text{where} \qquad V_{\text{eff}}(r) = -\frac{M}{r} + \frac{L^2}{2r^2} - \frac{L^2M}{r^3}, \quad \mathcal{E} \equiv \frac{e^2-1}{2}.} \tag{8.7}$$

The quantity $\mathcal{E}$ is not the total energy. If we write $e = 1+E$, then E is the energy per unit mass without the rest mass energy, and

$$\mathcal{E} = \frac{e^2-1}{2} = \frac{E^2-2E}{2} = E + \frac{1}{2}E^2 \simeq E$$

for $E \ll 1$. This means if the total energy of the particle is much less than its rest mass energy, the quantity $\mathcal{E}$ is like the Newtonian-limit energy E_{Newt}. Observing the effective potential V_{eff}: the first term is the gravitational potential, while the second is the centrifugal barrier. If we were in Newtonian gravity, these two terms will produce orbits in conic sections. If the particle is too close to $r=0$, the centrifugal barrier will prevent it from falling into the central object. One can find that there exists a stable equilibrium of $V_{\text{eff,Newton}}$ at $r = L^2/M$. However, in GR, there is a third, *negative* term that dominates at small r. It enhance the gravitational attraction of the central mass M. This means that if the particle is too close to the central object, it will definitely falls into it. Another difference of this equation of motion in GR from Newtonian mechanics is that the time-derivative is with respect to τ, but not t. Moreover, the coordinate radius r should not be taken as the real radial distance from the center to a sphere of radius r. Figure 8.1 contains V_{eff} in Newtonian and GR pictures for different parameters M and L that we shall analyze below.

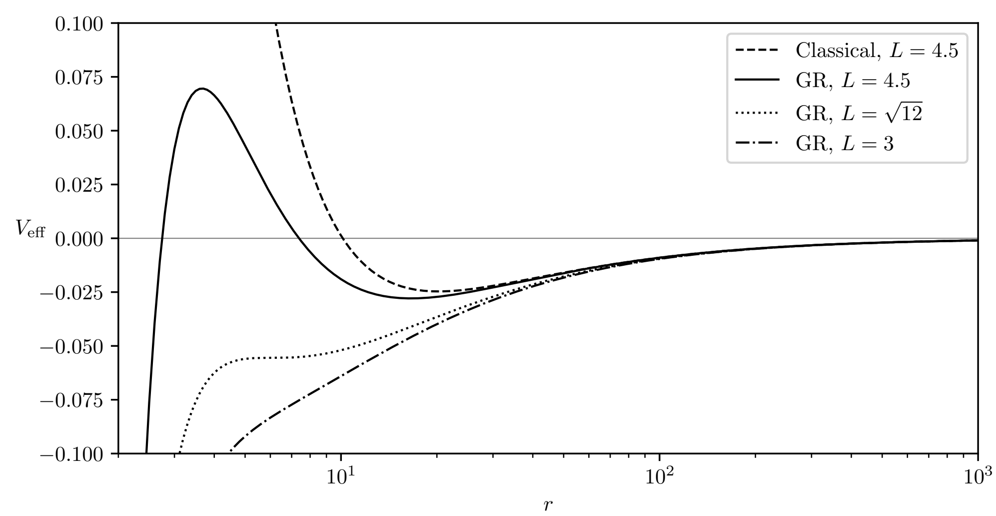


Figure 8.1: Newtonian and GR effective potential with $M = 1$. In Newtonian gravity, there always exist a stable equilibrium (minimum in V_{eff}) at which the orbit is circular. There is a centrifugal barrier. In GR, the shape of V_{eff} depends on the relative size of M and L. For $L^2 > 12M^2$, there exists both a minimum and a maximum. For $L^2 = 12M^2$, there is an innermost stable circular orbit at $r = 6M$. For $L^2 < 12M^2$, there are no stable circular orbits.

The Schwarzschild radius $r = 2M$ is not surprisingly also a special point for V_{eff}. At $r = 2M$,

$$V_{\text{eff}} = -\frac{M}{2M} + \frac{L^2}{8M^2} - \frac{ML^2}{8M^3} = -\frac{1}{2},$$

independent of L. Next, the maximum and minimum of V_{eff} are at

$$r_{\min} = \frac{L^2}{2M}\left[1 + \sqrt{1 - 12\left(\frac{M}{L}\right)^2}\right] \qquad \text{and} \qquad r_{\max} = \frac{L^2}{2M}\left[1 - \sqrt{1 - 12\left(\frac{M}{L}\right)^2}\right].$$

Note that $r_{\max}$ occurs closer to the central object. Because of the square root in r_1 and $r_{\max}$, there are three cases of V_{eff}.

- $L^2 < 12M^2$: V_{eff} has no extrema. There is no circular orbits.
- $L^2 = 12M^2$: for this particular angular momentum, $r_1 = r_{\max} = 6M$. At this point, V_{eff} has a point of inflection, so $r = 6M$ is usually referred to as the **ISCO**, or **innermost stable circular orbit**.
- $L^2 > 12M^2$: $V_{\text{eff}}(r)$ has both a minimum and a maximum, with the maximum closer to the origin. If $L^2 = 16M^2$, then $r_{\max} = 4M$ and $V_{\text{eff}}(4M) = 0$. It turns out that for $L^2 < 16M^2$, the effective potential is negative everywhere. Even if a particle starts from rest at infinity ($e = 1$ or $\mathcal{E} = 0$), it will reach the central object. Since $L^2 = 12M^2$ sets the boundary of $r_{\min} = r_{\max} = 6M$, any $L^2 \geq 12M^2$ will have $r_{\min} \geq 6M$ and $r_{\max} \leq 6M$. As $L \to \infty$,

  $$r_{\max} = \frac{L^2}{2M}\left[1 - \sqrt{1 - 12\left(\frac{M}{L}\right)^2}\right] = \frac{L^2}{2M}\left[1 - 1 + 6\left(\frac{M}{L}\right)^2\right] = 3M.$$

  Hence there is a lower bound for $r_{\max}$ at $3M$. The range of $r_{\max}$ is $3M < r_{\max} \leq 6M$.

8.5.2 Radial Plunge Orbits

The simplest "orbits" are orbits with zero angular momentum and the particle starts from rest at infinity. This implies $L = 0$, $e = 1$, and $\mathcal{E} = 0$. We would like to know how long does it take the particle to reach $r = 2M$. The energy equation (8.7) reads

$$\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 - \frac{M}{r} = 0 \quad \Longrightarrow \quad \frac{dr}{d\tau} = \pm\sqrt{\frac{2M}{r}}.$$

This looks like a particle in a Newtonian gravitational field, but note that t is replaced by τ. Since we are interested in inward-falling particles, we shall choose $u^r = dr/d\tau = -\sqrt{2M/r}$. The solution is

$$\tau(r) = \tau_\star - \frac{2}{3}\sqrt{\frac{r^3}{2M}} \quad \Longleftrightarrow \quad r(\tau) = \left(\frac{9M}{2}\right)^{1/3}(\tau_\star - \tau)^{2/3}, \tag{8.8}$$

where $\tau_\star$ is some integration constant. It takes infinite proper time for the particle to reach coordinate radius r if it starts at infinity because $\tau \to -\infty$ as $r \to \infty$. If the particle moves from a finite r_i to another finite r_f, the proper time elapsed is finite. This is important because it holds true even when $r_f = 2M$.

For coordinate time t, things get more complicated. From the definition of the conserved quantity e (which we define to be 1 in Plunge orbits),

$$e = 1 = \left(1 - \frac{2M}{r}\right)\frac{dt}{d\tau} = \left(1 - \frac{2M}{r}\right)\frac{dt}{dr}\frac{dr}{d\tau} = -\sqrt{\frac{2M}{r}}\left(1 - \frac{2M}{r}\right)\frac{dt}{dr}.$$

The resulting differential equation is

$$\frac{dt}{dr} = -\left(\frac{2M}{r}\right)^{-1/2}\left(1 - \frac{2M}{r}\right)^{-1}.$$

The solution is

$$t = t_\star + 2M\left[-\frac{2}{3}\left(\frac{r}{2M}\right)^{3/2} - 2\left(\frac{r}{2M}\right)^{1/2} + \ln\left|\frac{(r/2M)^{1/2} + 1}{(r/2M)^{1/2} - 1}\right|\right], \tag{8.9}$$

where $t_\star$ is another integration constant. Again, it takes infinite coordinate time for the particle to reach r if it starts at infinity, $t \to -\infty$ as $r \to \infty$. However, it also takes infinite coordinate time for the particle to reach $r = 2M$ from any finite radius because the logarithm blows up at $r \to 2M$. This is another special property of the Schwarzschild radius.

Example 8.1. Speed of a particle measured by a static observer

We know the particle's 4-velocity in a radial plunge orbit,

$$u = \left(\frac{dt}{d\tau}, \frac{dr}{d\tau}, 0, 0\right) = \left(\left[1 - \frac{2M}{r}\right]^{-1}, -\sqrt{\frac{2M}{r}}, 0, 0\right), \quad u \cdot u = -1.$$

For the observer, construct their orthonormal basis $\{\boldsymbol{e}'_\mu\}$. The timelike basis vector is the 4-velocity of the observer, u_{ob}. The only nonzero component is u^t_{ob},

$$g_{tt}(u^t_{\text{ob}})^2 = -1 \quad \Longrightarrow \quad u^t_{\text{ob}} = \left(1 - \frac{2M}{r}\right)^{-1/2} = (\boldsymbol{e}'_t)^t.$$

(The notation $(\boldsymbol{e}'_t)^t$ means the time component of the observer's timelike basis vector $\boldsymbol{e}'_t$ in the coordinate basis.) We can construct other orthonormal basis. The four orthonormal basis vectors are

$$\boldsymbol{e}'_t = \left(1 - \frac{2M}{r}\right)^{-1/2} \boldsymbol{e}_t \quad \boldsymbol{e}'_r = \left(1 - \frac{2M}{r}\right)^{1/2} \boldsymbol{e}_r, \quad \boldsymbol{e}'_\theta = \frac{1}{r}\boldsymbol{e}_\theta, \quad \boldsymbol{e}'_\phi = \frac{1}{r\sin\theta}\boldsymbol{e}_\phi. \tag{8.10}$$

It is straightforward to check that $\boldsymbol{e}'_\mu \cdot \boldsymbol{e}'_\nu = \eta_{\mu\nu}$. The velocity of the particle measured by the observer is then

$$v = \frac{dr'}{dt'} = \frac{dr'/d\tau'}{dt'/d\tau'} = \frac{u \cdot \boldsymbol{e}'_r}{-u \cdot \boldsymbol{e}'_t} = \frac{g_{rr}u^r(\boldsymbol{e}'_r)^r}{-g_{tt}u^t(\boldsymbol{e}'_t)^t} = \frac{(1-2M/r)^{-1}(-\sqrt{2M/r})\sqrt{1-2M/r}}{-(2M/r-1)(1-2M/r)^{-1}(1-2M/r)^{-1/2}} = -\sqrt{\frac{2M}{r}}.$$

Here the primed quantities $dr'/d\tau'$ belong to the particle, but measured in the observer's frame. The radial velocity is negative because it is falling inward. In fact, recall that the 4-velocity of a particle in an orthonormal basis is just $u = (\gamma, \gamma v, 0, 0)$. Thus, the Lorentz factor is $\gamma = -u \cdot \boldsymbol{e}'_t = (1-2M/r)^{-1/2} = 1/\sqrt{1-v^2}$. These are all consistent in a locally Minkowski spacetime (the observer's frame). To perceive this velocity, we turn the particle to an outgoing particle. Because the particle reaches infinity with zero velocity, $v = \sqrt{2M/r}$ is the *escape velocity*. Note that this is identical to the Newtonian limit, but with real radius replaced by the Schwarzschild coordinate radius r.

For a particle at infinity starting with a generic e, its radial velocity measured by a stationary observer at coordinate radius r is

$$v = -\frac{1}{e}\sqrt{e^2 - 1 + \frac{2M}{r}}.$$

This reduces to the result above if we plug in $e = 1$.

8.5.3 Circular Orbits

Circular orbits occur at $dV_{\text{eff}}/dr = 0$ only when $L^2 \geq 12M^2$, which gives an equation and the solution for the angular momentum L:

$$Mr^2 - L^2 r + 3ML^2 = 0 \quad \Longrightarrow \quad L = \pm r\sqrt{\frac{M}{r-3M}}.$$

We will take the positive solution; the particle is moving in counterclockwise direction.

By the energy equation (8.7), we can calculate the energy e for a circular orbit ($dr/d\tau = 0$):

$$\frac{e^2-1}{2} = \cancel{\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2} + \left(-\frac{M}{r} + \frac{L^2}{2r^2} - \frac{ML^2}{r^3}\right) \quad \Longrightarrow \quad e = \frac{r-2M}{\sqrt{r(r-3M)}}.$$

From Figure 8.2, there are two possible circular orbits for every $L > 2\sqrt{3}M$. The unstable orbit is inside $r = 6M$, while the stable one is outside $r = 6M$. There is no circular orbits inside $r = 3M$, no matter how large the angular momentum is. We will see why in later sections. From the graph of $e(r)$, we see that the ISCO at $r = 6M$ is the most tightly bound orbit because e is at minimum. The energy (per unit mass) is $e = 4/3\sqrt{2} \approx 0.9428$.

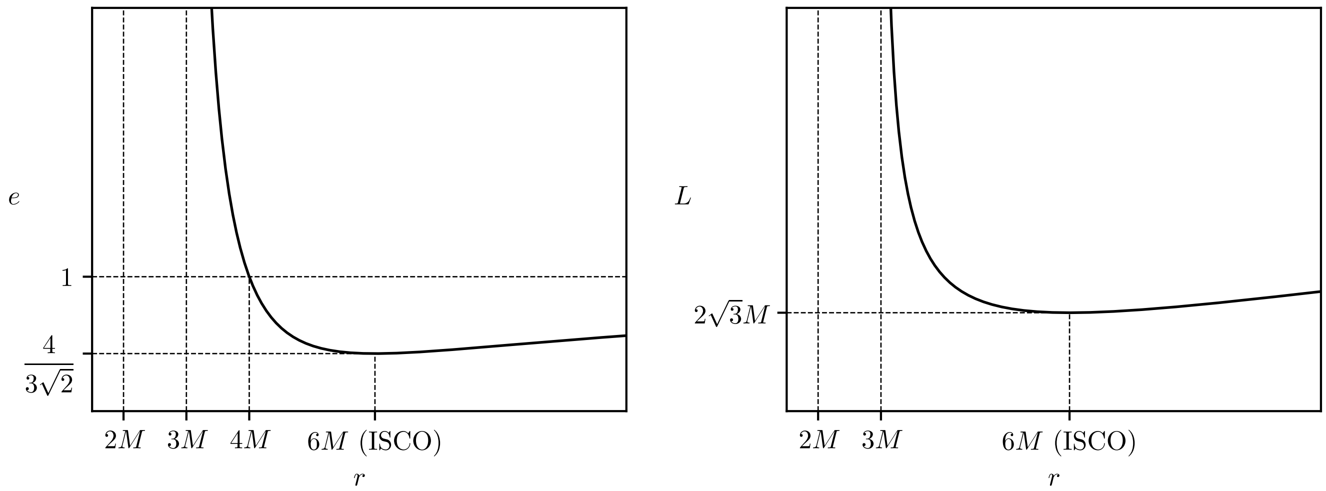


Figure 8.2: A plot of $e(r)$ and $L(r)$. There are several important radii on each graph.

Example 8.2. Brightest source in the universe

In black hole physics, the ISCO is an important radius. It is around where the accretion disk is formed. The energy per unit mass $e \approx 0.9428$ means that if an object starts at rest and come to the ISCO, it will release an per unit mass of $1 - e \approx 0.057$. Here 1 (or mc^2/m) can be seen as the rest mass energy per unit mass. Thus, an obsect reaching the ISCO releases the equivalent of 5.7% of its rest mass energy before falling into the black hole since there is no stable circular orbits inward. 5.7% does not sound much, but it is an enormous amount of energy. The nuclear fusion rate of hydrogen into helium (the current power source of the Sun) only releases 0.7% of the rest mass energy of the reactants. Moreover, if the black hole is a Kerr black hole (a spinning one), the object will release as high as 42% of its rest mass energy. Therefore, the accretion disk around a black hole is one of the most luminous source in the universe.

Example 8.3. Period of a circular orbit

For a circular orbit,

$$e = \left(1 - \frac{2M}{r}\right)\frac{dt}{d\tau} = \frac{(r-2M)}{\sqrt{r(r-3M)}}, \quad L = r^2\frac{d\phi}{d\tau} = r\sqrt{\frac{M}{r-3M}}. \tag{8.11}$$

We would like to know the period of a particle revolving around the central object at radius r. There are at least three interpretation of this period: the period in coordinate time measured by a static observer at infinity, the period in the proper time of the particle, and the period measured by a static observer at radius r in the gravitational field. We first consider the coordinate time period T_t. The coordinate angular velocity (the angular velocity measured by a static observer at infinity) and the coordinate time period are

$$\omega_t = \frac{d\phi}{dt} = \frac{d\phi/d\tau}{dt/d\tau} = \frac{L/r^2}{e/(1-2M/r)} = \sqrt{\frac{M}{r^3}} \quad \Longrightarrow \quad T_t = 2\pi r\sqrt{\frac{r}{M}}.$$

The period measured by the particle gives the proper time period,

$$T_\tau = \frac{2\pi}{\omega_\tau} = \frac{2\pi}{d\phi/d\tau} = \frac{2\pi}{L/r^2} = 2\pi r\sqrt{\frac{r-3M}{M}}.$$

The period measured by a static observer at r can be calculated from the coordinate time period by gravitational time dilation because a circular orbit with fixed speed is a perfect clock:

$$d\tau_{\rm ob} = \sqrt{g_{tt}}\,dt \quad \Longrightarrow \quad T_{\rm ob} = \sqrt{g_{tt}}\,T_t = 2\pi r\sqrt{\frac{r-2M}{M}}.$$

Alternatively, we can compute the speed of the particle in the observer's frame. The 4-velocity of the particle in its circular orbit is

$$u = \left(\frac{dt}{d\tau}, 0, 0, \frac{d\phi}{d\tau}\right) = \left(\frac{e}{1-2M/r}, 0, 0, \frac{L}{r^2}\right).$$

Using the same approach as in Example 8.1, the period of the particle measured by the observer is

$$v = \left|\frac{u \cdot \boldsymbol{e}'_\phi}{-u \cdot \boldsymbol{e}'_t}\right| = \left|\frac{g_{\phi\phi}(d\phi/d\tau)(g_{\phi\phi})^{-1/2}}{-g_{tt}(dt/d\tau)(-g_{tt})^{-1/2}}\right| = \sqrt{\frac{M}{r-2M}} \quad \Longrightarrow \quad T_{\text{ob}} = \frac{2\pi r}{v} = 2\pi r\sqrt{\frac{r-2M}{M}}.$$

In fact, we can compute the Lorentz factor and find the desired relation between T_{ob} and T_τ:

$$\gamma = \frac{1}{\sqrt{1-v^2}} = \sqrt{\frac{r-2M}{r-3M}}, \quad T_{\text{ob}} = \gamma T_\tau.$$

The hierarchy of the three periods is $T_t > T_{\text{ob}} > T_\tau$. The first inequality comes from gravitational time dilation, while the second comes from special relativistic time dilation. The particle's proper time for one revolution is the extremal (in this case minimum) period because it is a free particle in Schwarzschild spacetime.

8.5.4 Elliptical Orbits and Precession

There are only two potentials for which all bound orbits are closed: Newtonian $-1/r$ potential and the harmonic oscillator potential r^2. These bound orbits are ellipses. In relativity, the $1/r$ potential no longer give rise to a closed orbit in general. In Schwarzschild spacetime, it is mainly because of the $-ML^2/r^3$ term in the effective potential. Hence a particle following an elliptical orbit does not have a closed track, but instead precesses in the same direction as the orbital direction. This is known as **apsidal precession**, or **precession of the perihelion** for planets. The apsidal angle $\delta\phi$ is defined as $\Delta\phi - 2\pi$, where $\Delta\phi$ is the angular increment of the particle over one revolution. (If there is no precession, then $\Delta\phi = 2\pi$, the definition of one revolution.) To find $\delta\phi$, we compute

$$\frac{d\phi}{dr} = \frac{d\phi/d\tau}{dr/d\tau} = \frac{L/r^2}{\pm\sqrt{e^2-1-2V_{\text{eff}}}} = \pm\frac{L}{r^2}\left(e^2 - 1 + \frac{2M}{r} - \frac{L^2}{r^2} + \frac{2ML^2}{r^3}\right)^{-1/2},$$

where we replace $dr/d\tau$ by (8.7). Then over one revolution,

$$\Delta\phi = 2\int_{r_1}^{r_2}\frac{d\phi}{dr}\,dr = 2L\int_{r_1}^{r_2}\frac{dr}{r^2}\left(e^2 - 1 + \frac{2M}{r} - \frac{L^2}{r^2} + \frac{2ML^2}{r^3}\right)^{-1/2}.$$

Here r_1 and r_2 are periapsis and apoapsis where $dr/d\tau = 0$. To lowest order of relativistic effects, we will find (see derivation in Appendix A.3)

$$\Delta\phi = 2\pi + \delta\phi, \quad \delta\phi \simeq \frac{6\pi M^2}{L^2} = \frac{6\pi M}{a(1-\epsilon^2)}, \tag{8.12}$$

where ϵ is the eccentricity of the ellipse, defined as $\epsilon = \sqrt{1-b^2/a^2}$ where b is the semiminor axis and a is the semimajor axis. The apsidal angle is large for small semimajor axis because the particle experience stronger gravity. It is also large for large eccentricity because the periapsis is in region with stronger gravity.

8.5.5 Photon Orbits

For photon orbits, the conserved quantities are similar:

$$e = -\xi \cdot u = \left(1 - \frac{2M}{r}\right)\frac{dt}{d\lambda}, \quad L = \eta \cdot u = r^2\frac{d\phi}{d\lambda}, \quad u \cdot u = 0,$$

where again we are considering photon orbit in a plane, so $\sin\theta = 1$. Note that all τ is replaced by the affine parameter λ, and $u \cdot u$ is null instead of -1. The equation $u \cdot u = 0$ gives

$$0 = -\left(1-\frac{2M}{r}\right)\left(\frac{dt}{d\lambda}\right)^2 + \left(1-\frac{2M}{r}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 + r^2\left(\frac{d\phi}{d\lambda}\right)^2 = -\left(1-\frac{2M}{r}\right)^{-1}e^2 + \left(1-\frac{2M}{r}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 + \frac{L^2}{r^2},$$

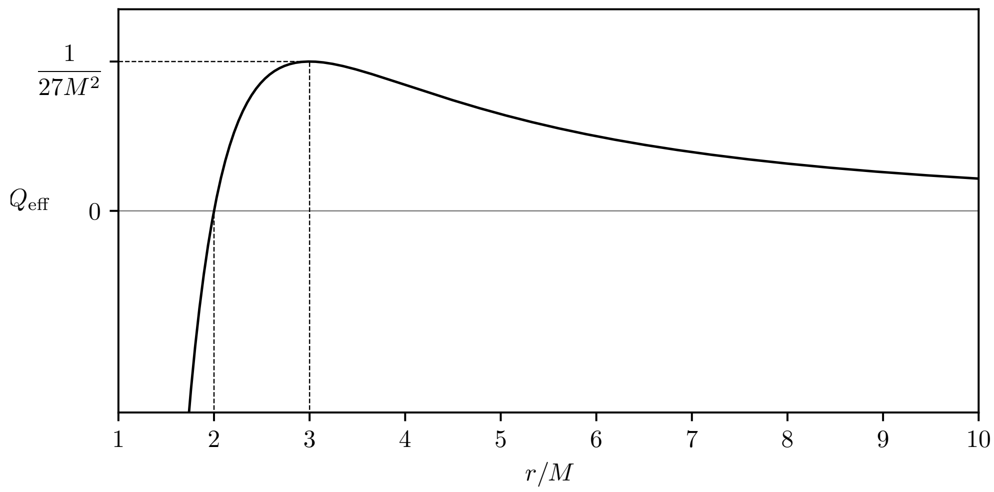


Figure 8.3: Q_{eff} as a function of r. There is only one unstable circular orbit at $3M$.

or

$$\boxed{\frac{1}{L^2}\left(\frac{dr}{d\lambda}\right)^2 + Q_{\text{eff}}(r) = \frac{1}{b^2} \qquad \text{where} \qquad b = \frac{L}{e}, \quad Q_{\text{eff}}(r) = \frac{1}{r^2}\left(1 - \frac{2M}{r}\right).} \tag{8.13}$$

Here Q_{eff} is like an effective potential. The shape of Q_{eff} is shown in Figure 8.3. Note that Q_{eff} approaches 0 from above as $r \to \infty$. It is zero at $r = 2M$, and approaches $-\infty$ as $r \to 0$. Recall that for material particles, we arrived at the conclusion that there is no circular orbit at or below $r = 3M$ no matter how large their angular momentum is. This is because the circular orbit at $r = 3M$ is for photons. A material particle must travel faster than the speed of light to have circular orbit at $r < 3M$, since,

$$v = \sqrt{\frac{M}{r - 2M}} \to 1 \quad \text{as} \quad r \to 3M.$$

Now consider a photon approaching the central object from infinity. If $1/b^2 < Q_{\text{eff}}(3M) = 1/27M^2$, then it will reach a minimum radius and travel to infinity. If $1/b^2 > Q_{\text{eff}}(3M)$, then it will overcome the centrifugal barrier and fall into the object. The quantity b has a more physical meaning. At infinity, the spacetime is flat, so

$$b = \frac{L}{e} = r^2\left(1 - \frac{2M}{r}\right)^{-1}\frac{d\phi}{dt} \to r^2\frac{d\phi}{dt}.$$

In the xy-plane ($\theta = \pi/2$ plane), $y = r\sin\phi \simeq r\phi$ as $r \to \infty$ or $\phi \to 0$, where y is some small distance compared to r.

$$\phi = \frac{y}{r}, \quad \frac{d\phi}{dr} = -\frac{y}{r^2} \quad \implies \quad b = r^2\left(-\frac{y}{r^2}\right)\frac{dr}{dt} = -y\frac{dr}{dt}.$$

But again, at infinity, $dr/dt = -1$, which means $b = y$ is the **impact parameter** of the photon (see Figure 8.4).

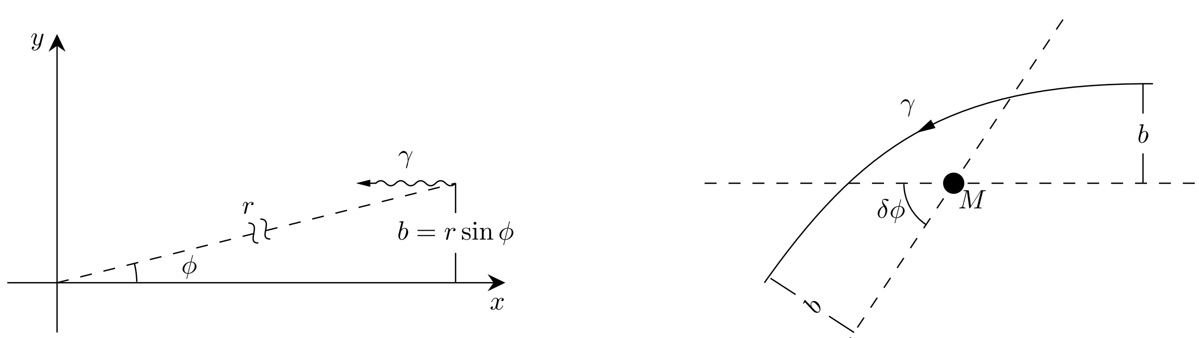


Figure 8.4: Interpretation of b as the impact parameter for a photon from infinity (left) and gravitational light bending (right).

Because any photon with the impact parameter $b_{\text{capture}} < 3\sqrt{3}M$ will be captured by the central mass, there is a capture cross-section:

$$\sigma = \pi b_{\text{capture}}^2 = 27\pi M^2. \tag{8.14}$$

8.5.6 Gravitational Lensing

Consider a photon coming from infinity with an impact parameter $b \gg 3\sqrt{3}M$ so that its trajectory is weakly bent. We want to calculate the deflection angle $\delta\phi$ (see Figure 8.4, right). From (8.13),

$$\frac{1}{b^2} = \frac{1}{L^2}\left(\frac{dr}{d\lambda}\right)^2 + \frac{1}{r^2}\left(1 - \frac{2M}{r}\right) = \frac{1}{r^4}\left(\frac{dr}{d\phi}\right) + \frac{1}{r^2}\left(1 - \frac{2M}{r}\right).$$

Let $u \equiv M/r$, $dr = -M\,du/u^2$,

$$\frac{1}{b^2} = \frac{u^4}{M^4}\frac{M^2}{u^4}\left(\frac{du}{d\phi}\right)^2 + \frac{u^2}{M^2}(1-2u) \quad \Longrightarrow \quad \left(\frac{du}{d\phi}\right)^2 + u^2 - \frac{M^2}{b^2} = 2u^3.$$

Note that if $b \gg 3\sqrt{3}M$, the closest approach of the photon to the central mass is $r_1 \sim b$, or $u_{\max} \sim M/b \ll 1$ at the point of closest approach. The term $2u^3$ on the RHS is third order in u, which is much smaller than terms on the LHS. Neglecting $2u^3$, we have a harmonic oscillator equation,

$$\left(\frac{du}{d\phi}\right)^2 + u^2 \simeq \frac{M^2}{b^2}.$$

The general solution is $u = A\cos\phi + B\sin\phi$. For a finite impact parameter b, we have $u = M/r \to 0$ as $r \to \infty$ and $\phi \to 0$. This forces $A = 0$. Substituting $u = B\sin\phi$ back into the equation gives $B = M/b$. It turns out that we will have the trivial (or classical) result:

$$u = \frac{M}{r} = \frac{M}{b}\sin\phi \quad \Longrightarrow \quad b = r\sin\phi = \text{const}.$$

This makes sense because the term we neglected $2u^3 \propto 1/r^3$ is the one arising from general relativity. Neglecting it means returning back to classical orbital mechanics. Now we should include $2u^3$, but keeping it small as a perturbation:

$$\left(\frac{du}{d\phi}\right)^2 + u^2 - \frac{M^2}{b^2} = 2u^3. \tag{8.15}$$

We expect the solution be perturbed by a small amount $u_1 \ll M/b$ so that $u = (M/b)\sin\phi + u_1$. Plugging this into (8.15), and keeping terms up to linear order in u_1 on the LHS but zeroth order on the RHS because $2u^3$ is already small,

$$\frac{M^2}{b^2}\cos^2\phi + \frac{2M}{b}\cos\phi\frac{du_1}{d\phi} + \frac{M^2}{b^2}\sin^2\phi + u_1\frac{2M}{b}\sin\phi - \frac{M^2}{b^2} = 2\frac{M^3}{b^3}\sin^3\phi \quad \Longrightarrow \quad \cos\phi\frac{du_1}{d\phi} + u_1\sin\phi = \frac{M^2}{b^2}\sin^3\phi.$$

To solve this equation, we write the LHS as a total derivative,

$$\cos^2\phi\frac{d}{d\phi}\left(\frac{u_1}{\cos\phi}\right) = \frac{M^2}{b^2}\sin^3\phi \quad \Longrightarrow \quad \frac{u_1}{\cos\phi} = \frac{M^2}{b^2}\int\tan^2\phi\sin\phi\,d\phi = \frac{M^2}{b^2}\left(\frac{1}{\cos\phi} + \cos\phi\right) + C.$$

The initial condition $u_1 = 0$ as $r \to \infty$ and $\phi \to 0$ requires $C = -2M^2/b^2$. The final solution of u is

$$u = \frac{M}{b}\sin\phi + \frac{M^2}{b^2}(1-\cos\phi)^2 + \mathcal{O}\left(\frac{M^3}{b^3}\right). \tag{8.16}$$

Now, $u \to 0$ ($r \to \infty$) should also be satisfied as $\phi \to \pi + \delta\phi$ (see Figure 8.4, right) after it bypasses the mass. In the weak bending limit $\delta\phi \ll 1$,

$$0 = \frac{M}{b}\sin(\pi + \delta\phi) + \frac{M^2}{b^2}[1 - \cos(\pi+\delta\phi)]^2 \simeq -\frac{M}{b}\delta\phi + \frac{4M^2}{b^2},$$

or

$$\boxed{\delta\phi = \frac{4M}{b}.} \tag{8.17}$$

The classical (exact) deflection angle of a *material* particle passing through a $1/r$ potential such as in Rutherford scattering or in a gravitational potential $\Phi = -GM/r$,

$$\delta\phi = 2\tan^{-1}\left(\frac{GM}{v_0^2 b}\right) \to \frac{2GM}{v_0^2 b}$$

for large b, where v_0 is the speed at infinity. It seems reasonable to put $v_0 = c$ for photon, but it is only half the value obtained in general relativity. Thus, gravitational light bending is a excellent test of GR. In fact, it was a test of gravitational lensing in 1919 that made Einstein and his theory of general relativity famous.

Example 8.4. Gravitational lensing: a Schwarzschild lens

Consider the **Schwarzschild lens**, where a background point source at large impact parameter b is lensed by an isolated mass M. The lensing event always has the source and lens close together on the sky. From Figure (8.5), there are many angles related:

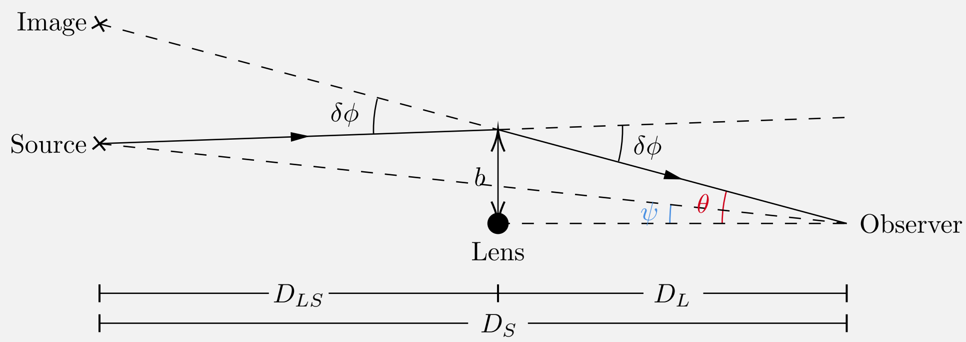


Figure 8.5: An illustration of gravitational lensing. The trajectory of photon can be approximated as straight lines because the size of the lens is small compared to D_S, just like in geometric optics. Note that θ is not necessarily equal to $\delta\phi$.

The angle ψ is the true angular separation, while θ is the observed angular separation between the source and the lens. The bending angle $\delta\phi$ are related to ψ and θ. In small angle approximation, the source-image distance in the source plane is $\delta\phi D_{LS} = (\theta - \psi)D_S$. This provides a relation between ψ, θ, and $\delta\phi$,

$$\psi = \theta - \frac{D_{LS}}{D_S}\delta\phi = \theta - \frac{D_{LS}}{D_S D_L}\frac{4M}{\theta},$$

where we used the fact that $b = D_L\theta$ and (8.17). We can rewrite this equation as

$$\theta^2 - \psi\theta - \theta_{\mathrm{E}}^2 = 0, \quad \text{where} \quad \theta_{\mathrm{E}} \equiv \sqrt{\frac{4MD_{LS}}{D_S D_L}}.$$

This θ_{E} is known as the **Einstein radius**. Its physical meaning is the following: if a source is *directly* behind the lens so that $\psi = 0$, the source will be lensed into a ring around the lens with angular radius $\theta = \theta_{\mathrm{E}}$. This is known as an **Einstein ring**, one of the most spectacular observation and test of GR. If $\psi \neq 0$, we need to solve the quadratic equation above,

$$\theta_\pm = \frac{1}{2}\left(\psi \pm \sqrt{\psi^2 + 4\theta_{\mathrm{E}}^2}\right).$$

Both θ and θ_{E} are observable quantities, so the true angular separation can be found from the solution $\theta_\pm$.

8.5.7 The Shapiro Delay

Returning to (8.13),

$$\frac{1}{b^2} = \frac{1}{L^2}\left(\frac{dr}{d\lambda}\right)^2 + Q_{\mathrm{eff}}(r), \quad \text{where} \quad Q_{\mathrm{eff}}(r) = \frac{1}{r^2}\left(1 - \frac{2M}{r}\right).$$

This time we change $dr/d\lambda$ to dr/dt using the definition of e for photon,

$$\frac{1}{b^2} = \frac{e^2}{L^2}\left(1 - \frac{2M}{r}\right)^{-2}\left(\frac{dr}{dt}\right)^2 + Q_{\mathrm{eff}}(r) = \frac{1}{b^2}\left(1 - \frac{2M}{r}\right)^{-2}\left(\frac{dr}{dt}\right)^2 + Q_{\mathrm{eff}}(r).$$

Solving for dt/dr,

$$\frac{dt}{dr} = \pm\frac{1}{b}\left(1-\frac{2M}{r}\right)^{-1}\left(\frac{1}{b^2}-Q_{\text{eff}}\right)^{-1/2}\begin{cases}+, & \text{for outward motion,}\\ -, & \text{for inward motion.}\end{cases}$$

Let r_1 be the radius of closest approach, so

$$\frac{1}{b^2} = Q_{\text{eff}}(r_1) = \frac{1}{r^2}\left(1-\frac{2M}{r_1}\right).$$

When the photon travels from r_1 to $r > r_1$, the Schwarzschild coordinate time elapsed is

$$t(r, r_1) = \int_{r_1}^{r} dr'\frac{1}{b}\left(1-\frac{2M}{r'}\right)^{-1}\left[\frac{1}{b^2}-Q_{\text{eff}}(r')\right]^{-1/2}.$$

To first order in M/r, this integral (see derivation in Appendix A.4) becomes

$$t(r, r_1) = \sqrt{r^2-r_1^2} + 2M\ln\left(\frac{r+\sqrt{r^2-r_1^2}}{r_1}\right) + M\left(\frac{r-r_1}{r+r_1}\right)^{1/2}. \tag{8.18}$$

The first term is the classical time elapsed (with $c = 1$) by the Pythagorean theorem. The other two terms are all positive, meaning that the actual travel time of the photon is longer than classical prediction. This effect is called the **Shapiro Delay**.

8.6 The Interior Structure of the Star

8.6.1 General Rules for Integrating the Equations

Combine Eq. 8.3 and Eq. 8.5 gives the **Tolman-Oppenheimer-Volkov** (T-O-V) equation:

$$\frac{dP}{dr} = -\frac{(\rho+P)(m+4\pi r^3 P)}{r(r-2m)}. \tag{8.19}$$

This equation, Eq. 8.4 and Eq. 8.2 gives three equations for m, ρ and P. There are two first-order differential equations, so they require two constants of integration: $m(r=0)$ and $P(r=0)$. First, $m(r=0)=0$ and more importantly, $\lim_{r\to 0}(m/r) = 0$. Second, set $P(r=0) = P_c$. Once $m(r)$, $P(r)$ and $\rho(r)$ are known, the surface of the star is defined as the place where $P = 0$. The reason is P must be continuous everywhere, or else the pressure gradient or the forces on the fluid would be infinite. Since $P = 0$ outside of the star, the boundary or the surface must have $P = 0$. The geometry must also be smooth: the interior geometry at the surface should be the same as Schwarzschild metric.

Inside the star,

$$g_{rr} = \left(1-\frac{2m(r)}{r}\right)^{-1}.$$

Outside the star,

$$g_{rr} = \left(1-\frac{2M}{r}\right)^{-1}.$$

The continuity means that at $r = R$,

$$M = m(R) \qquad \text{and} \qquad M = \int_0^R 4\pi r^2 \rho\, dr,$$

which is the same as Newtonian theory. We have now obtained M, so g_{00} outside the star and at the surface should be

$$g_{00}(r=R) = -\left(1-\frac{2M}{R}\right).$$

This is useful when we want to solve for Φ in Eq. 8.3.

8.6.2 Newtonian Stars

For Newtonian situations, $P \ll \rho$, or $4\pi r^3 P \ll m$. The metric should also be nearly flat, so according to g_{rr}, we require $m \ll r$. This simplify the T-O-V equation to

$$\frac{dP}{dr} = -\frac{\rho m}{r^2}.$$

This is exactly the sae as the equation of hydrostatic equilibrium for Newtonian stars. Comparing this to the original T-O-V equation, we see that the pressure gradient is smaller in Newtonian stars. This means in GR, stronger internal forces are required to hold the star's shape.

8.7 Exact Interior Solutions

The differential equations are hard to solve for stars in Newtonian theory, and even more in GR. For exact solutions, we need to make some assumptions.

8.7.1 The Schwarzschild Constant-Density Interior Solution

The assumption is $\rho = \text{const}$. This does not have any physial justification, the speed of sound is even infinite inside the star. Nonetheless, neutron stars have nearly uniform density so the solution is of our interest.

Integrate Eq. 8.4, we get

$$m(r) = \frac{4\pi\rho r^3}{3}, \quad r \le R,$$

where R is the star's radius. For continuity of g_{rr}, or continuity of $m(r)$ at R, this implies

$$m(r) = \frac{4\pi\rho R^3}{3} = M, \quad r \ge R.$$

We can also solve the T-O-V equation:

$$\frac{dP}{dr} = -\frac{4}{3}\pi r \frac{(\rho + P)(\rho + 3P)}{1 - 8\pi r^2 \rho/3}.$$

With arbitrary central pressure P_c, integrating the equation gives

$$\frac{\rho + 3P}{\rho + P} = \frac{\rho + 3P_c}{\rho + P_c}\left(1 - \frac{2m}{r}\right)^{1/2}.$$

This follows that

$$R^2 = \frac{3}{8\pi\rho}\left[1 - \frac{(\rho + P_c)^2}{\rho + 3P_c}^2\right] \quad \text{or} \quad P_c = \rho\frac{1 - (1 - 2M/R)^{1/2}}{3(1 - 2M/R)^{1/2} - 1}.$$

Notice that $P_c \to \infty$ as $M/R \to 4/9$. Replacing P_c in left equation gives

$$P = \rho\frac{(1 - 2Mr^2/R^3)^{1/2} - (1 - 2M/R)^{1/2}}{3(1 - 2M/R)^{1/2} - (1 - 2Mr^2/R^3)^{1/2}}.$$

Now we can solve for Φ. The continuity of g_{00} at R implies

$$g_{00}(R) = -\left(1 - \frac{2M}{R}\right).$$

With this initial value, we find

$$e^{\Phi} = \frac{3}{2}\left(1 - \frac{2M}{R}\right)^{1/2} - \frac{1}{2}\left(1 - \frac{2Mr^2}{R^3}\right)^{1/2}, \quad r \le R.$$

8.7.2 Buchdahl's Interior Solution

Buchdahl found a solution for the equations of state

$$\rho = 12(P_\star P)^{1/2} - 5P,$$

where $P_\star$ is an arbitrary constant. For small P it reduced to

$$\rho = 12(P_\star P)^{1/2}.$$

It is required that

$$P < P_\star, \quad \rho < 7P_\star.$$

If we have a different radial coordinate r' implicitly defined as

$$r(r') = r'\frac{1-\beta+u(r')}{1-2\beta}$$

with

$$u(r') = \beta\frac{\sin Ar'}{Ar'}, \quad A^2 = \frac{288\pi P_\star}{1-2\beta},$$

where β is another arbitrary constant and u is a function of r'. We shall not show how to obtain the solution here, but just look at the metric functions as a result. For $Ar' \leq \pi$,

$$\begin{aligned} e^{2\Phi} &= \frac{(1-2\beta)(1-\beta-u)}{1-\beta+u}, \\ e^{2\Lambda} &= \frac{(1-2\beta)(1-\beta+u)}{(1-\beta-u)(1-\beta+\beta\cos Ar')^2}, \\ P(r) &= \frac{A^2(1-2\beta)u^2}{8\pi(1-\beta+u)^2}, \\ \rho(r) &= \frac{2A^2(1-2\beta)u(1-\beta-3u/2)}{8\pi(1-\beta+u)^2}. \end{aligned}$$

At the surface where $P=0$, or $u=0$,

$$e^{2\Phi} = e^{-2\Lambda} = 1-2\beta \qquad \text{and} \qquad R = r(R') = \frac{\pi(1-\beta)}{A(1-2\beta)}.$$

Hence β is the valueof M/R on the surface. The limit $\beta \to 0$ is a nonrelativistic limit, the mass of the star is

$$M = \frac{\pi\beta(1-\beta)}{(1-2\beta)A} = \left[\frac{\pi}{288P_\star(1-2\beta)}\right]^{1/2}\beta(1-\beta).$$

8.8 Realistic Stars and Gravitational Collapse

8.8.1 Buchdahl's Theorem

There are no uniform-density stars with radii smaller than $9M/4$ of any stellar model.

8.8.2 Formation of Stellar-Mass Black Holes

The luminosity and temperature of a star depends on its mass. An ordinary star like our Sun has nuclear reactions in its core, converting hydrogen to helium. It will burn hydrogen for the order of 10^{10} years. A more massive star can remain steady for only a million years. These steady stars are called **main sequence stars**.

When all hydrogen turns into helium, the core starts to shrink. This process will heat up the core. If the temperature is high enough, it will trigger further nuclear reaction, converting helium into carbon and oxygen and releasing more energy. As a result, the luminosity will increase dramatically as well. Moreover, the outer layers of the star have to expand and the surface temperature will drop. Such a star is called a **red giant**.

As the star uses up all its helium, its fate depends on what mass it has left. If the mass is small, then the star will cool off and turn into a **white dwarf**. If the mass is sufficient for the core to have higher temperature, nuclear reaction will happen again, turning carbon into silicon and finally to iron. Iron is the most stable nucleus, which means any reaction converting iron into something else absorbs energy rather than releasing it. The subsequent evolution of the star depends mainly on four things: the star's mass, rotation, magnetic field and chemical composition.

- The Mass: for slowly rotating stars, a star of the Sun's mass or smaller will become a white dwarf. If the star's mass is between $15 \sim 20M_\odot$, the strong nuclear repulsion forces may stop the collapse when the mean density reaches the density of an atomic nucleus. The infalling matter bounces mack and is expelled in Type II supernova. The compact core left behind is a neutron star. If the original star is more massive, the collapse will result in a black hole. A stellar-mass black holes may have masses from $5 \sim 60M_\odot$.
- Rotation and magnetic field: in the collapse phase, rotation becomes important if angular momentum is conserved by the collapsing core. Substantial magnetic field may allow transfer of angular momentum from the core to the rest of the star, permitting a more spherical collapse.
- Chemical composition: the first generation of stars were composed of pure hydrogen and helium. These stars could be more massive with $M > 100M_\odot$. They would evolve rapidly to the point of gravitational collapse and became intermidiate-mass black holes. We use the word "intermediate" because there are supermassive black holes in the centers of most galaxies.

9 Black Holes

In the late 1700's, John Michell and Pierre Laplace have already thought of a star with escape velocity larger than the speed of the light. According to conservation of energy, the escape velocity is related to the mass M and the radius R of the star by

$$\frac{1}{2}v^2 = \frac{GM}{R}.$$

Setting $v = c$ gives the radius of the "invisible star":

$$R = \frac{2GM}{c^2}.$$

It is surprising that this result is exactly the formula for the radius of a Schwarzschild black hole in GR. However, the interpretations of Newtonian black holes and GR black holes are different. Newtonian one says that the light will still leave the star but will finally pulled back by gravity. In GR, everything will never leave the surface of the black hole once it is inside $R = 2M$. The surface itself is also just empty space, but not the edge of a massive body.

9.1 The Surface $r = 2M$

Consider the Schwarzschild metric

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2.$$

At $r = 2M$ and $r = 0$, there must be something wrong with the line element. Especially at $r = 2M$, when we cross this surface, g_{rr} gets negative and g_{tt} gets positive. The coordinate radius becomes a timelike coordinate, while the coordinate time becomes a spacelike coordinate.

It turns out that $r = 2M$ is merely a coordinate singularity, which can be removed by choosing another appropriate coordinate system. When we analyze the radial plunge orbit of a particle trajectory in Section 8.5.2, we find that it requires an infinite coordinate time t for the particle to reach singularity. However, the particle itself, only experiences a finite proper time in this process. This indicates that there is something wrong with our coordinates. On the other hand, the $r = 0$ singularity is a true singularity where the spacetime curvature blows up. In this section, we will talk about three systems of coordinates that remove Schwarzschild coordinate singularity and study the interior of a black hole.

9.1.1 Eddington-Finkelstein Coordinates

Define a new radial coordinate (also known as the **tortoise coordinate**):

$$\frac{dr^*}{dr} = \left(1 - \frac{2M}{r}\right)^{-1} \qquad \Longrightarrow \qquad r^* = r + 2M \ln|r - 2M| + C.$$

We choose the integration constant to be $C = -2M \ln 2M$ so that

$$r^* = r + 2M \ln\left|\frac{r}{2M} - 1\right|.$$

Define another coordinate $v \equiv t + r^*$, $dv = dt + dr^*$, and $dt = dv - dr(1 - 2M/r)^{-1}$. The tortoise coordinate r^* acts like an intermediate coordinate. Computing the line element, we find that the coordinate singularity at $r = 2M$ goes away

$$\begin{aligned} ds^2 &= -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \\ &= -\left(1 - \frac{2M}{r}\right)\left[dv^2 - 2\left(1 - \frac{2M}{r}\right)^{-1} dv\,dr + \cancel{\left(1 - \frac{2M}{r}\right)^{-2} dr^2}\right] + \cancel{\left(1 - \frac{2M}{r}\right)^{-1} dr^2} + r^2 d\Omega^2 \end{aligned}$$

This set of coordinates is known as **Eddington-Finkelstein coordinates**, first proposed by Eddington in 1924. The line element in Eddington-Finkelstein coordinates (v, r, θ, ϕ) is

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dv^2 + 2\, dv\, dr + r^2 d\Omega^2. \tag{9.1}$$

Note that Eddington-Finkelstein metric is not diagonal. Because of this property, the inverse metric also does not have a singularity at $r = 2M$. We will find the shape of light cones in Eddington-Finkelstein coordinates to study the causal structure of spacetime around $r = 2M$. Consider radial light rays ($d\theta = d\phi = 0$). Photons travel along null geodesics,

$$0 = -\left(1 - \frac{2M}{r}\right) dv^2 + 2\, dv\, dr.$$

Solving this equation, either $dv = 0$ ($v =$ const.) or

$$\frac{dv}{dr} = \frac{2r}{r - 2M} = 2\frac{dr^*}{dr} \quad \Longrightarrow \quad v = 2r^* + \text{const.} = 2r + 4M \ln\left|\frac{r}{2M} - 1\right| + \text{const.}$$

These are the equations for light cones. Figure 9.1 shows photon trajectories on a spacetime diagram of v vs. r.

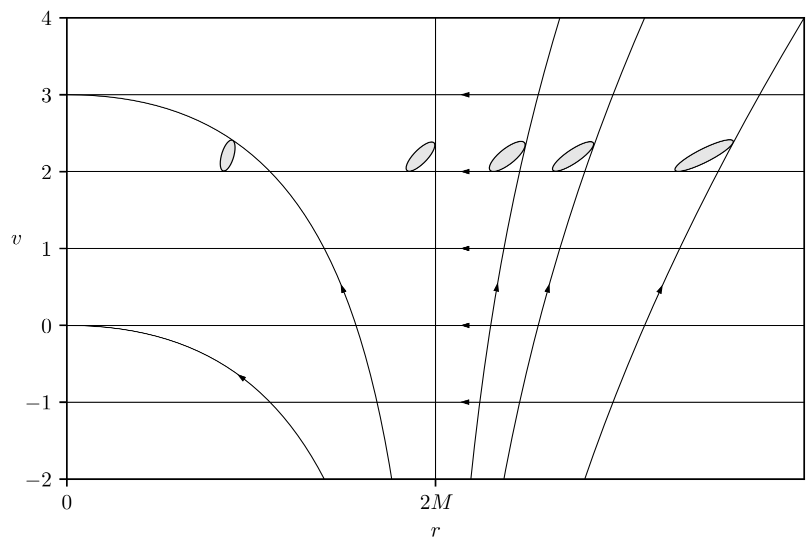


Figure 9.1: Photon world lines in (v, r) spacetime diagram. Several future light cones are drawn. The arrows indicates the direction of world lines in increasing t: the horizontal lines represent $v = t + r^* =$ const., where increasing t means decreasing r^* or r; the curved lines represent $v = t + r^* = 2r^* +$ const., or $t - r^* =$ const. Increasing t means increasing r^*, or increasing r for $r > 2M$, and decreasing r for $r < 2M$.

From Figure 9.1, we see that material particles following timelike world lines must following trajectories with decreasing r once they are inside $r = 2M$. Eventually, they will reach the $r = 0$ singularity. The $r = 2M$ surface is a null surface: an outward propagating radial light ray stays at this surface. Anything information inside $r = 2M$ thus cannot reach the outside world, which means it is impossible to probe the interior of a black hole from outside.

9.1.2 Kruskal-Szekeres Coordinates

Another set of good coordinates is known as **Kruskal-Szekeres coordinates**, discovered in 1960. The coordinates are defined as

$$u = \left(\frac{r}{2M} - 1\right)^{1/2} e^{r/4M} \cosh\left(\frac{t}{4M}\right), \quad v = \left(\frac{r}{2M} - 1\right)^{1/2} e^{r/4M} \sinh\left(\frac{t}{4M}\right),$$

for $r > 2M$ and

$$u = \left(1 - \frac{r}{2M}\right)^{1/2} e^{r/4M} \sinh\left(\frac{t}{4M}\right), \quad v = \left(1 - \frac{r}{2M}\right)^{1/2} e^{r/4M} \cosh\left(\frac{t}{4M}\right),$$

for $r > 2M$. The θ and ϕ coordinates are the same as in Schwarzschild coordinates. The metric in Kruskal-Szekeres coordinates is

$$ds^2 = -\frac{32M^3}{r}e^{-r/2M}(dv^2 - du^2) + r^2 d\Omega, \tag{9.2}$$

where r and t are now function of u and v, defined implicitly by

$$\left(\frac{r}{2M} - 1\right)e^{r/2M} = u^2 - v^2, \quad \tanh\left(\frac{t}{4M}\right) = \begin{cases} v/u, & r > 2M, \\ u/v, & r < 2M. \end{cases} \tag{9.3}$$

The derivation of (9.2) is available in Appendix A.5.1. There is no singularity at $r = 2M$ in this metric but there is one at $r = 0$ as it should be. The Kruskal-Szekeres metric is diagonal. From (9.3), constant r lines are hyperbolae $u^2 - v^2 = \text{const.}$, while constant t lines are straight lines passing through the origin $(u, v) = (0, 0)$. Radial null lines ($d\theta = d\phi = ds = 0$) satisfy $dv = \pm du$, or $v = \pm u + \text{const.}$, which looks like null lines in Minkowski spacetime. Material particles have $u \cdot u = -1$,

$$\frac{32M^3}{r}e^{-r/2M}\left[-\left(\frac{dv}{d\tau}\right)^2 + \left(\frac{du}{d\tau}\right)^2\right] = -1.$$

In order for the LHS to be negative, we must have $(dv/d\tau)^2 > (du/d\tau)^2$, or $|dv/du| > 1$, again, similar to material particle world lines in Minkowski spacetime. Figure 9.2 shows constant r, constant t lines, and $r = 0$ singularity.

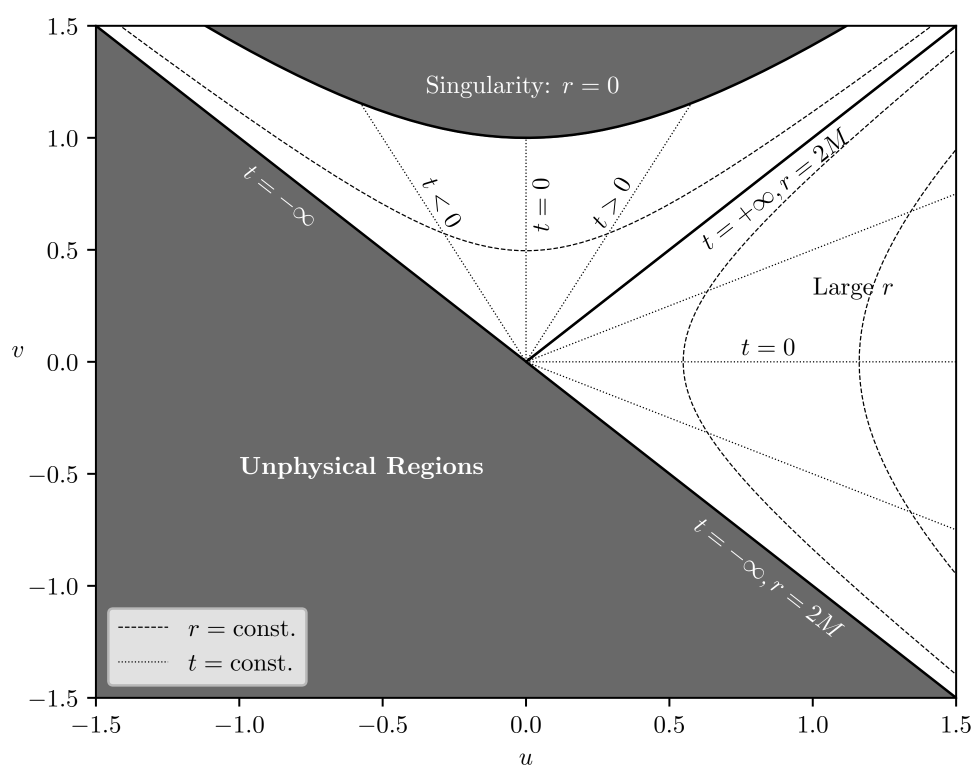


Figure 9.2: Constant r, constant t lines, and $r = 0$ singularity in (u, v) coordinates.

From Figure 9.2, clearly $r = 2M$ is a null surface. It is also clear that any photon emitted inside $r = 2M$ always stays inside $r = 2M$, and eventually runs into the singularity, because $r = 0$ hyperbola is asymptotically the same as $u = \pm v$. Interestingly, the $r = 0$ singularity is a spacelike surface. It is to be interpreted as a moment in time in language of relativity. There are more interesting things happening in the unphysical regions, see Appendix A.5.2.

9.2 General Black Holes

The horizon itself is composed of null world lines, the ones that neither move away to infinity or fall inwards. This definition fits the

Schwarzschild horizon.

There must be a process for the horizon to grow from zero to its full size. Consider a collapsing star. Photon (a) gets out with little trouble, photon (b) with some delay, photon (c) as the marginal one, which just trapped and remains on the Schwarzschild horizon. Anything later than (c) is permanently trapped. Therefore, photon (c) represent the horizon at all times.

9.2.1 General Properties of Black Holes

- Any horizon will eventually become stationary. A stationary vacuum black hole is characterized by its total mass M and total angular momentum L. This is a remarkable property. A massive black hole, whose history as a gas may have included complex gas motions, shock waves, magnetic fields, nucleosynthesis, and all kinds of other complications, is described fully and exactly by just two numbers: the mass and spin.
- If the black hole is not in vacuum, it may carry an electric charge Q and a magnetic monopole momentum F, even though magnetic monopoles are not found in nature. There can also be distortion by tidal effect of matter surrounding the black hole.
- If the gravitational collapse is nearly spherical, then all nonspherical parts of the mass distribution are radiated away in gravitational waves. A stationary Kerr black hole is left behind. If there is no angular momentum, a Schwarzshild hole is left behind.
- The area of the black hole cannot decrease with time. Two black holes can merge into one, but it is impossible for one black hole to separate into two smaller ones spontaneously. This holds if the local energy density of the black hole is positive. Entropy is associated with the area of the horizon so the entropy of the black hole always increases. This area theorem may be violated by quantum effects because in quantum mechanics, the energy is not necessarily positive. Hawking radiation may occur.
- There are curvature singlarities inside the horizon.
- Some also proposed that there are naked singularities, the ones outside horizons. If the universe's fate is a Big Crunch, then it is a nakes singularity.

9.2.2 Kerr Black Hole

The Kerr black hole is axially symmetric but not spherically. It is characterized by two parameters, M and L. L has dimension m^2 ao we define

$$a = \frac{L}{M},$$

with same dimensions as M. The line element is

$$ds^2 = -\frac{\Delta - a^2\sin^2\theta}{\rho^2}dt^2 - 2a\frac{2Mr\sin^2\theta}{\rho^2}dt\,d\phi + \frac{(r^2+a^2)^2 - a^2\Delta\sin^2\theta}{\rho^2}\sin^2\theta\,d\phi^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2,$$

where

$$\Delta = r^2 - 2Mr + a^2$$
$$\rho^2 = r^2 + a^2\cos^2\theta.$$

The metric for $a = 0$ is just the Schwarzschild metric. There is an off-diagonal term in the matric:

$$g_{t\phi} = -a\frac{2Mr\sin^2\theta}{\rho^2}.$$

This has new effects on particle trajectories. Since $g_{\mu\nu}$ is independent of ϕ, p_ϕ is still conserved, but now we have

$$p^\phi = g^{\phi\mu}p_\mu = g^{\phi\phi}p_\phi + g^{\phi t}p_t \qquad \text{and} \qquad p^t = g^{t\mu}p_\mu = g^{tt}p_t + g^{t\phi}p_\phi.$$

Consider a particle with $p_\phi = 0$. With definitions $p^t = m\,dt/d\tau$ and $p^\phi = m\,d\phi/d\tau$, the particle's trajectory has

$$\frac{d\phi}{dt} = \frac{p^\phi}{p^t} = \frac{g^{\phi t}}{g^{tt}} = \omega(r,\theta).$$

This means the particle has an angular velocity with zero angular momentum. We see now if the black hole has an angular momentum L, the particle dropped straight in from infinity is dragged by the influence of gravity so that it acquires an angular velocity. This effect is called the **dragging of inertial frames** and it weakens with distance roughly as $1/r^3$.

9.2.3 Ergoregion

Consider photons in the equatorial plane ($\theta = \pi/2$) at some r initially going $\pm\phi$-direction. They have only dt and $d\phi$ nonzero and sine $ds^2 = 0$ for photons,

$$0 = g_{tt}dt^2 + 2g_{t\phi}dt\,d\phi + g_{\phi\phi}d\phi^2 \quad \Longrightarrow \quad \frac{d\phi}{dt} = -\frac{g_{t\phi}}{g_{\phi\phi}} \pm \left[\left(\frac{g_{t\phi}}{g_{\phi\phi}}\right)^2 - \frac{g_{tt}}{g_{\phi\phi}}\right]^{1/2}.$$

If $g_{tt} = 0$:

$$\frac{d\phi}{dt} = 0 \qquad \text{and} \qquad \frac{d\phi}{dt} = -\frac{2g_{t\phi}}{g_{\phi\phi}}.$$

The first solution suggest that photons sending opposite to rotation will end up not moving at all because of dragging. The second solution is the photon sending in the same direction as the rotation. In fact, the surface where $g_{tt} = 0$ lies outside the horizon and it is called the **ergosphere**. Setting $g_{tt} = 0$ it means

$$r_{\text{ergosphere}} = M + \sqrt{M^2 - a^2\cos^2\theta}.$$

Inside this radius with $g_{tt} > 0$, all particles and photons must rotate with the hole.

9.2.4 The Kerr Horizon

In the Schwarzschild solution, the horizon is at $g_{tt} = 0$ and $g_{rr} = \infty$. The serr solution indicates that ergosphere is at g_{tt} and the horizon is at $g_{rr} = \infty$ or $\Delta = 0$:

$$r_{\text{Kerr}} = M + \sqrt{M^2 - a^2}.$$

The Kerr ergosphere lies outside the horizon except at the poles.

Now we want to calculate the area of the horizon because it is important to Hawking's area theorem. The horizon has $dt = dr = 0$ so

$$d\ell^2 = \frac{(r^2+a^2)^2 - a^2\Delta}{\rho^2}\sin^2\theta\,d\phi^2 + \rho^2 d\theta^2.$$

The proper area of this surface is by integrating the square root of the determinant of this metric:

$$\begin{aligned} A(r) &= \int_0^{2\pi}\phi\int_0^{\pi}d\theta\sqrt{(r^2+a^2)^2 - a^2\Delta}\sin\theta \\ &= 4\pi\sqrt{(r^2+a^2)^2 - a^2\Delta}. \end{aligned}$$

Since the horizon is defined by $\Delta = 0$, we have

$$A(\text{horizon}) = 4\pi(r^2 + a^2).$$

9.3 Real Black Holes in Astronomy

Black holes of stellar mass are formed by star collapse. A black hole formed by the collapse of a massive star is hard to identify. All known stellar-mass black holes are in binary systems, where the companion star is dumping gas onto the hole. Because of friction and angular momentum of the gas near the black hole, the gas may be heated up to 10^6 K. The peak of its emission spectrum is at X-ray wavelengths.

Many X-ray binary systems are known. However, some neutron star binary system can also have X-ray emission. The difference is: if the X-ray is emitted at a steady rate, then it is a pulsar and not a black hole.

There are also supermassive black holes in the universe. The best evidence is the supermassive black holes (with mass $4.3 \times 10^6\,M_\odot$)

in the center of the Milky Way. It is discovered by observing the motion near the center of the galaxy. The extremely high speeds of stars revolving the center indicate that there must be a supermassive black hole.

There are some questions associated with supermassive black holes in the center of many galaxies. It seems like these black holes have some relation with galaxy formation, but did the holes come first or they form as part of the process of galaxy formation? It is not clear that whether these holes formed with large mass or they started small and grew larger. If they grew, it is by meging with other black holes or by accreting gas and stars?

The supermassive black holes mentioned above are relatively small comparing to quasars typically with mass $10^9 M_\odot$. There are more quasars in the early universe. This suggests that black holes like quasars are not formed by merging black holes like ours—another unanswered question.

Finally, there exists intermediate-mass black holes with masses around $100 \sim 10^4 M_\odot$. They are formed in the early universe when there was only pure hydrogen and helium

9.4 Quantum Mechanical Emission of Radiation by Black Holes

In 1974, Stephan Hawking proved that black holes radiate energy continuously. The process occurs near the horizon of the black hole because of the uncertainty principle. One form of the uncertainty principle is

$$\Delta E \Delta t \geq \frac{\hbar}{2}.$$

One interpretation is the following: two photons with energy total ΔE can be created out of nowhere and recombine within time interval Δt. The energy conservation is violated under this small scale but only in small amount of time. Consider a fluctuation which produces two photons, one with energy E and the other with $-E$. Negative energy means the photon propagates backward in time. In flat spacetime these photons will recombine within time Δt. However, if these two photons are just produced outside the horizon, there is a chance of photons crossing the horizon before the time $\hbar/2E$ elapses. If one photon is inside the horizon, the other one can propagate freely to infinity.

Consider a simple Schwarzschild metric. For convenience, choose an observer on a trajectory with $p_0 = U^0 = 0$ going toward decreasing r inside the horizon. U^r is then the only nonzero component of u. By normalization condition $u \cdot u = 1$:

$$U^r = -\left(\frac{2M}{r} - 1\right)^{1/2}, \quad r < 2M.$$

Any photon orbit is allowed for which $-p \cdot u > 0$. Consider a photon with $L = 0$. By photon orbit equation, it has energy $E = \pm p^r$. Its energy relative to the observer is

$$-p \cdot u = -p^r U^r g_{rr} = -\left(\frac{2M}{r} - 1\right)^{-1/2} p^r.$$

This is possible if and only if the photon is also ingoing: $p^r < 0$, but there is no restriction on E. Photons may travel on null geodesics inside the horizon with either sign of E. Recall that t is the spatial coordinate inside the horizon so E is actually the spatial momentum componnt, which of course can be negative.

Now, if a fluctuation outside the horizon can put the negative-energy photon into the correct trajectory (towards the horizon). The positive-energy photon can escape to infinity without recombinination.

A frame momentarily at rest at $2M + \epsilon$ will immediately fall inwards, following the trajectory of a particle with $\tilde{L} = 0$ and $\tilde{E} = [1 - 2M/(2M + \epsilon)]^{1/2} \approx (\epsilon/2M)^{1/2}$. It reaches the horizon after a proper-time Δt by

$$\Delta\tau = -\int_{2M+\epsilon}^{2M} \left(\frac{2M}{r} - \frac{2M}{2M+\epsilon}\right)^{-1/2} dr.$$

To the first order in ϵ this is

$$\Delta\tau = 2(2M\epsilon)^{1/2}.$$

The energy $\mathcal{E}$ of the photon in this frame in time $\hbar/2\mathcal{E} >$ is

$$\mathcal{E} = \frac{1}{4}\hbar(2M\epsilon)^{-1/2}.$$

This is the energy of the photon in local inertial frame. The energy of it at infinity can be calculated by

$$\mathcal{E} = -p \cdot u \quad \text{with} \quad -U_0 = \tilde{E} \approx (\epsilon/2M)^{1/2}.$$

Hence

$$\mathcal{E} = -g^{00} p_0 U_0 = U_0 g^{00} E.$$

Using g^{00} at $2M + \epsilon$, we have

$$E = \mathcal{E} \left(\frac{\epsilon}{2M} \right)^{1/2} = \frac{\hbar}{8M}.$$

According to this energy, the blackbody radiation temperature from a black hole should be

$$kT_{\rm H} = \frac{\hbar}{8\pi M} \quad \text{or} \quad T_{\rm H} = \frac{\hbar c^3}{8\pi GkM} \approx 6.15 \times 10^{-8} \left(\frac{M_\odot}{M} \right) \text{K},$$

where $k \approx 1.38 \times 10^{-23}$ J/K is the Boltzmann constant. The energy (at the peak of the black-body spectrum) is

$$E = 4.965kT = 1.580 \frac{\hbar}{8M} \sim \frac{\hbar}{8M}.$$

Note that the Hawking temperature of the black hole is proportional to M^{-1}. The rate of radiation from a blackbody is proportional to AT^4 and the area A of the horizon is proportional to M^2. Combining these information, the luminosity of the hole is proportional to M^{-2}. This energy comes from the mass of the hole as negative-energy photons fall into it. Thus, we have

$$\frac{dM}{dt} \propto M^{-2} \quad \text{or} \quad \tau \propto M^3.$$

This means bigger black hole lives longer and has cooler temperature. If the black hole is small (like $\sim 10^{12}$ kg), the temperature will be 10^{11} K and such a black hole emits γ-rays. This type of black holes, called **primordial black holes**, may occur in the very early universe. We can estimate the evaporation time of a black hole by assuming the black hole radiates like a blackbody, with luminosity

$$L = 4\pi R_{\rm S}^2 \sigma T^4,$$

where $\sigma = \pi^2 k^4 / 60 c^2 \hbar^3 \approx 5.67 \times 10^{-8}\, \text{W m}^{-2}\, \text{K}^{-4}$ is the Stefan-Boltzmann constant. The evaporation time will be

$$t_{\rm evap} \sim \frac{Mc^2}{L} = \frac{15360\pi G^2 M^3}{\hbar c^4} = 2 \times 10^{76}\,\text{s} \left(\frac{M}{M_\odot} \right)^3,$$

much longer than the age of the universe for even a solar mass black hole.

Consider Hawking's area theorem,

$$\frac{dA}{dt} \geq 0.$$

For a Schwarzschild black hole,

$$A = 16\pi M^2 \quad \text{or} \quad dM = \frac{1}{32\pi M}\, dA = \frac{\hbar}{8\pi kM} d \left(\frac{kA}{4\hbar} \right).$$

Sine dM is the change in the black hole's total energy and $\hbar/8\pi kM$ is its Hawking temperature T_H, this may be written as

$$dE = T_{\rm H}\, dS.$$

By Hawking's area theorem, the quantity S (entropy) can never decrease. These are the first and second laws of thermodynamics and they can apply to black holes.

10 Cosmology

Cosmology is the study of the universe as a whole: its history, evolution, composition and dynamics. The reason why we need GR for cosmology is that a system shows relativistic effects when its mass M and its radius R satisfy the relation

$$M/R \sim 1.$$

Now, the observable universe by our telescopes is in the order of 10 Gpc. On this large scale, the universe appears to be **homogeneous** with density of mass-energy $\rho = 10^{-26}\,\mathrm{kg\,m^{-3}}$. We can calculate that when $R \sim 6$ Gpc, the mass $M = 4\pi\rho R^3/3$ is equal to R, which is within the observable universe. Therefore, to understand the universe as a whole, the key topic of cosmology, we need GR.

10.1 Cosmological Kinematics: Observing the Expanding Universe

10.1.1 Homogenety and Isotropy of the Universe

The universe is not only homogeneous, but also **isotropic** about *every* point: there is no consistently defined special direction. The observable universe is uniformly expanding. For example, the galaxies are receding from us at a speed proportional to their distance from us, which is called the **Hubble flow**. This recessional velocity and distance satisfy the relation

$$v = Hd \tag{10.1}$$

where H is the **Hubble's parameter**. Hubble's parameter changes with time and the present value is called **Hubble's constant**. It is measured to be $H_0 = (71\pm4)\,\mathrm{km\,s^{-1}\,Mpc^{-1}} = (2.3\pm0.1)\times10^{-18}\,\mathrm{s^{-1}}$. The associated **Hubble time** is $t_H = H_0^{-1} = (4.3\pm0.2)\times10^{17}\,\mathrm{s}$, about 14 billion years, same order of magnitude of the age of the universe.

10.1.2 Models of the Universe: the Cosmological Principle

Not all parts of the universe is observable or else we do not have the term *observable universe*. Thus, to make a model of the whole universe, we need to make some assumptions about the inaccessible regions. There are two types of inaccessible regions.

Some region is so distant that no information could reach us from it. Such region is everything outside our past light-cone called the **particle horizon**. As time passes, more and more of the previously unknown regions enters the particle horizon and becomes observable. This means the unknown regions have influence on our future. Another type of inaccessible regions are regions with lights that are so dim that we cannot observe them. We assume that these regions are very like what we have observed and they are homogeneous and isotropic.

10.1.3 Spacetime and Relativity

We have spacetime metric in special relativity as $\eta_{\mu\nu}$ which creates the invariant line element

$$ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu.$$

In general relativity, the spacetime metric $g_{\mu\nu} = g_{\mu\nu}(t, \mathbf{x})$ depends on the position in spacetime. This spacetime dependence incorporates the effects of gravity, the distribution of matter and energy in the universe.

The spatial homogeneity and isotropy of the universe results in the expression of four-dimentional line element as

$$ds^2 = -c^2dt^2 + a^2(t)d\ell^2, \tag{10.2}$$

where $d\ell^2 = \gamma_{ij}(x^k)dx^i dx^j$ is the spatial line element and $a(t)$ is the scale factor which describes the expansion of the universe. The spatial metric γ_{ij} depends on the allowed form of three-spaces.

10.1.4 Symmetric Three-Spaces

Homogeneous and isotropic three-spaces must have constant curvature with three options: 0, positive or negative.

- Flat space (0 curvature): the simplest Euclidean space E^3 with non-intersecting parallel lines.

$$d\ell^2 = d\mathbf{x}^2 = \delta_{ij}dx^i dx^j,$$

  which is invariant under spatial translations $x^i \mapsto x^i + a^i$ and rotations $x^i \mapsto R^i_k x^k$, with $\delta_{ij}R^i_k R^j_l = \delta_{kl}$.

- Spherical Space (constant positive curvature): represented by a three-sphere S^3 in four-dimensional Euclideans space E^4:

$$d\ell^2 = d\mathbf{x}^2 + du^2, \quad \mathbf{x}^2 + u^2 = R_0^2,$$

  where R_0 is the radius of the sphere. Parallel lines will eventually meet in spherical space.

- Hyperbolic space (constant negative curvature): a hyperboloid H^3 in four-dimensional Lorentzian space $\mathbb{R}^{1,3}$:

$$d\ell^2 = d\mathbf{x}^2 - du^2, \quad \mathbf{x}^2 - u^2 = -R_0^2,$$

  where $R_0^2 > 0$ is a constant determining the curvature of the hyperboloid.

Combine the spherical and hyperbolic line elements,

$$d\ell^2 = d\mathbf{x}^2 \pm du^2, \quad \mathbf{x}^2 \pm u^2 = \pm R_0^2.$$

Differentiating the condition $\mathbf{x}^2 \pm u^2 = \pm R_0^2$ gives

$$u\,du = \mp \mathbf{x} \cdot d\mathbf{x}.$$

In this way we can eliminate the dipendence on u,

$$d\ell^2 = d\mathbf{x}^2 \pm \frac{(\mathbf{x} \cdot d\mathbf{x})^2}{R_0^2 \mp \mathbf{x}^2}.$$

The final unified line element can be written as

$$d\ell^2 = d\mathbf{x}^2 + k\frac{(\mathbf{x} \cdot d\mathbf{x})^2}{R_0^2 - k\mathbf{x}^2}, \quad \text{for} \quad k = \begin{cases} 0 & \mathrm{E}^3 \\ +1 & \mathrm{S}^3 \\ -1 & \mathrm{H}^3 \end{cases}. \tag{10.3}$$

It is convenient to write the metric in spherical polar coordinates using conversions

$$\begin{aligned} d\mathbf{x}^2 &= dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2), \\ \mathbf{x} \cdot d\mathbf{x} &= r\,dr \\ \mathbf{x}^2 &= r^2 \end{aligned}.$$

Then the metric in 10.3 becomes

$$\begin{aligned} d\ell^2 &= dr^2 + \frac{kr^2\,dr^2}{R_0^2 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \\ &= \left(1 + \frac{kr^2}{R_0^2 - kr^2}\right) dr^2 + r^2 d\Omega^2 \\ &= \left(\frac{R_0^2}{R_0^2 - kr^2}\right) dr^2 + r^2 d\Omega^2 \\ &= \frac{dr^2}{1 - kr^2/R_0^2} + r^2 d\Omega^2 \end{aligned}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$.

10.1.5 Robertson-Walker Metric

Substitute the line elements into 10.2, it is the **Robertson-Walker metric**, or Friedmann-Robertson-Walker (FRW) metric, in spherical polar coordinates

$$ds^2 = -c^2 dt^2 + a^2(t)\left[\frac{dr^2}{1 - kr^2/R_0^2} + r^2 d\Omega^2\right]. \tag{10.4}$$

This result has some important features:

- The line element has a rescaling symmetry

$$a \to \lambda a, \quad r \to r/\lambda, \quad R_0 \to R_0/\lambda.$$

  This means we can set the scale factor today, at $t = t_0$, to be $a(t_0) = 1$. The scale R_0 is the physical curvature scale today.

- The coordinate r is called the **comoving coordinate** which can be changed using the rescaling factor λ so it is not a physical observable. The physical results only depend on the **physical coordinate**, $r_{\text{phys}} = a(t)r$.

  A galaxy with comoving coordinates $\mathbf{r}(t)$ and physical coordinates $\mathbf{r}_{\text{phys}} = a(t)\mathbf{r}$ has a physical velocity of

$$\mathbf{v}_{\text{phys}} = \frac{d\mathbf{r}_{\text{phys}}}{dt} = \frac{da}{dt}\mathbf{r} + a(t)\frac{d\mathbf{r}}{dt} = H\mathbf{r}_{\text{phys}} + \mathbf{v}_{\text{pec}},$$

  where H is the Hubble parameter

$$H = \frac{\dot{a}}{a}. \tag{10.5}$$

  The first term $H\mathbf{r}_{\text{phys}}$ is the Hubble flow, the velocity of the galaxy resulting from the expansion of the space between the origin and $\mathbf{r}_{\text{phys}}(t)$. The second term $\mathbf{v}_{\text{pec}} = a(t)\dot{\mathbf{r}}$ is the **peculiar velocity**, the velocity measured by a "comoving observer"(an observer who follows the Hubble flow).

- We may redefine the radial coordinate, $d\chi = dr/\sqrt{1 - kr^2/R_0^2}$, so that

$$ds^2 = -c^2 dt^2 + a^2(t)[d\chi^2 + S_k^2(\chi)d\Omega^2],$$

  where

$$S_k(\chi) = R_0 \begin{cases} \sinh(\chi/R_0) & k = -1 \\ \chi/R_0 & k = 0 \\ \sin(\chi/R_0) & k = +1 \end{cases}.$$

- Define the **conformal time** as

$$d\eta = \frac{dt}{a(t)}.$$

  Then the metric becomes

$$ds^2 = a^2(\eta)\left[-c^2 d\eta^2 + \left(d\chi^2 + S_k^2(\chi)d\Omega^2\right)\right].$$

  This form is useful when studying the propagation of light.

10.2 Kinematics

One key feature of GR is that freely-falling particles moves along **geodesics**.

10.2.1 Geodesics

A free particle means the particle experiences no forces, $d^2\mathbf{x}/dt^2 = 0$. In two-dimensional Cartesian coordinates it is

$$\frac{d^2 x^i}{dt^2} = 0.$$

In other coordinates, $\ddot{\mathbf{x}} = 0$ does not necessarily imply $\ddot{x}^i = 0$. For example, in polar coordinates we have

$$\ddot{r} = r\dot{\phi}^2 \quad \text{and} \quad \ddot{\phi} = -\frac{2}{r}\dot{r}\dot{\phi}.$$

The equation of motion of a free particle in an arbitrary coordinate system generally have metric $g_{ij} \neq \delta_{ij}$ with Lagrangian

$$\mathcal{L} = \frac{m}{2} g_{ij}(x^k)\dot{x}^i\dot{x}^j.$$

The Euler-Lagrange equation gives us

$$\frac{d^2x^i}{dt^2} = -\Gamma^i_{ab}\frac{dx^a}{dt}\frac{dx^b}{dt}, \tag{10.6}$$

where Γ^i_{ab} is the **Christoffel symbol** and

$$\Gamma^i_{ab} = \frac{1}{2}g^{ij}(\partial_a g_{jb} + \partial_b g_{ja} - \partial_j g_{ab}). \tag{10.7}$$

Example 10.1. The Lagrangian of a free particle in polar coordinates is

$$\mathcal{L} = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\phi}^2\right) \quad \text{with} \quad [g_{ij}] = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}.$$

We have $g^{ij} = \operatorname{diag}(1, r^{-2})$. Then compute the Christoffel symbol:

$$\begin{aligned}
\Gamma^r_{\phi\phi} &= \frac{1}{2}g^{rr}(\partial_\phi g_{r\phi} + \partial_\phi g_{r\phi} + \partial_r g_{\phi\phi}) = \frac{1}{2}(-2r) = -r \\
\Gamma^\phi_{r\phi} &= \frac{1}{2}g^{\phi\phi}(\partial_r g_{\phi\phi} + \partial_\phi g_{\phi r} - \partial_\phi g_{r\phi}) = \frac{1}{2r^2}(2r) = \frac{1}{r} \\
\Gamma^\phi_{\phi r} &= \frac{1}{2}g^{\phi\phi}(\partial_\phi g_{\phi r} + \partial_r g_{\phi\phi} - \partial_\phi g_{\phi r}) = \frac{1}{2r^2}(2r) = \frac{1}{r}
\end{aligned}$$

with other components zero. Therefore,

$$\frac{d^2r}{dt^2} = -\Gamma^r_{\phi\phi}\frac{d\phi}{dt}\frac{d\phi}{dt} = -(-r)\left(\frac{d\phi}{dt}\right)^2 = r\dot{\phi}^2$$

$$\frac{d^2\phi}{dt^2} = -\Gamma^\phi_{r\phi}\frac{dr}{dt}\frac{d\phi}{dt} - \Gamma^\phi_{\phi r}\frac{d\phi}{dt}\frac{dr}{dt} = -2\left(\frac{1}{r}\right)\left(\frac{dr}{dt}\frac{d\phi}{dt}\right) = -\frac{2}{r}\dot{r}\dot{\phi}.$$

Derivation 10.1. Christoffel Symbol

The Euler-Lagrange equation is

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}^k}\right) = \frac{\partial\mathcal{L}}{\partial x^k}.$$

We ignore the mass m because it will cancel on both sides. The LHS is

$$\begin{aligned}
\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}^k}\right) = \frac{d}{dt}(g_{ik}\dot{x}^i) &= g_{ik}\ddot{x}^i + \frac{\partial g_{ik}}{\partial x^j}\frac{dx^j}{dt}\dot{x}^i \\
&= g_{ik}\ddot{x}^i + \partial_j g_{ik}\dot{x}^i\dot{x}^j \\
&= g_{ik}\ddot{x}^j + \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik})\dot{x}^i\dot{x}^j.
\end{aligned}$$

The RHS is

$$\frac{\partial\mathcal{L}}{\partial x^k} = \frac{1}{2}\partial_k g_{ij}\dot{x}^i\dot{x}^j.$$

Combining the LHS and the RHS,

$$g_{ki}\ddot{x}^i = -\frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij})\dot{x}^i\dot{x}^j.$$

Multiplying both sides by $g^{lk} = (g_{ki})^{-1}$,

$$\ddot{x}^l = -\frac{1}{2}g^{lk}(\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij})\dot{x}^i\dot{x}^j \equiv -\Gamma^l_{ij}\dot{x}^i\dot{x}^j.$$

In curved spacetime, a geodesic is a timelike curve $x^\mu(\tau)$ which extremises the proper time $\Delta\tau$ between two points in the spacetime. The **geodesic equation** is

$$\frac{dx^\mu}{d\tau^2} = -\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}, \tag{10.8}$$

where the Crhistoffel symbol is

$$\Gamma^\mu_{\alpha\beta} \equiv \frac{1}{2}g^{\mu\lambda}(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta}). \tag{10.9}$$

The geodesic equation can also be written in terms of the four momentum

$$p^\mu = m\frac{dx^\mu}{d\tau}.$$

We have

$$\frac{d}{dt}p^\mu(x^\alpha(\tau)) = \frac{\partial p^\mu}{\partial x^\alpha}\frac{dx^\alpha}{d\tau} = \frac{p^\alpha}{m}\frac{\partial p^\mu}{\partial x^\alpha}.$$

Thus, the geodesic equation becomes

$$p^\alpha\frac{\partial p^\mu}{\partial x^\alpha} = -\Gamma^\mu_{\alpha\beta}p^\alpha p^\beta \quad \text{or} \quad p^\alpha(\partial_\alpha p^\mu + \Gamma^\mu_{\alpha\beta}p^\beta) = 0. \tag{10.10}$$

The term in brackets is the **covariant derivative** of p^μ. It can be denoted by $\nabla_\alpha p^\mu \equiv \partial_\alpha p^\mu + \Gamma^\mu_{\alpha\beta}p^\beta$, so we have

$$p^\alpha\nabla_\alpha p^\mu = 0.$$

This form of geodesic equation is convenient because it can be applied to massless particles which moves with the speed of light.

For free particles in FRW, we need Christoffel symbols for the FRW metric

$$ds^2 = -c^2dt^2 + a^2(t)\gamma_{ij}dx^i dx^j,$$

where γ_{ij} depends on the choice of spatial coordinates (such as Cartesian or polar) and on the curvature of the spatial slices. Let $g_{\mu\nu} = \mathrm{diag}\,(-1, a^2\gamma_{ij})$. All Chrisoffel symbols with two time indices are 0. The only nonzero components are

$$\begin{aligned}
\Gamma^0_{ij} &= c^{-1}a\dot{a}\gamma_{ij} \\
\Gamma^i_{0j} &= c^{-1}\frac{\dot{a}}{a}\delta^i_j \\
\Gamma^i_{jk} &= \frac{1}{2}\gamma^{il}(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk}),
\end{aligned}$$

or they can be related by symmetry ($\Gamma^\mu_{\alpha\beta} = \Gamma^\mu_{\beta\alpha}$).

The homogeneity of FRW background means $\partial_i p^\mu = 0$, so

$$p^0\frac{dp^\mu}{dx^0} = -\Gamma^\mu_{\alpha\beta}p^\alpha p^\beta.$$

Consider the $\mu = 0$ component. Since all Christoffel symbol with two time indices vanish, the RHS is $-\Gamma^0_{ij}p^ip^j$. Since $p^0 = E/c$ and $\Gamma^0_{ij} = c^{-1}a\dot{a}\gamma_{ij}$, we have

$$\frac{E}{c^3}\frac{dE}{dt} = -\frac{1}{c}a\dot{a}\gamma_{ij}p^ip^j. \tag{10.11}$$

For massless particles, the four momentum is $p^\mu = (E/c, p^i)$ and it must obey the constraint

$$g_{\mu\nu}p^\mu p^\nu = -\frac{E^2}{c^2} + a^2\gamma_{ij}p^ip^j = 0.$$

Combining with 10.11, we get

$$\frac{1}{E}\frac{dE}{dt} = -\frac{\dot{a}}{a}.$$

This means that the energy of a massless particle decays with the expansion of the universe:

$$E \propto a^{-1}.$$

For massive particles, they have the constraint

$$g_{\mu\nu}p^{\mu}p^{\nu} = -m^2c^2.$$

It can be derived that the physical three-momentum obeys $p \propto a^{-1}$

10.2.2 Redshift

The light waves we observe now is not the same as the original one emitted by the galaxy or other distant objects. This is because the universe is expanding, the wavelength of these lights will be stretched based on the scaling factor $a(t)$.

Quantum mechanically, recall that the wavelength of light is inversely proportional to its energy,

$$\lambda = \frac{h}{E}.$$

Since the energy satisfies $E \propto a^{-1}$, the wavelength is proportional to the scaling factor, $\lambda \propto a(t)$, or

$$\lambda_0 = \frac{a(t_0)}{a(t_1)}\lambda_1 \tag{10.12}$$

This change in wavelength is called the **redshift**.

In classical electromagnetism, since the spacetime is isotropic, we can set light to travel in the radial direction with $\theta = \phi = \text{const}$. Then the line element is

$$ds^2 = a^2(\eta)[-c^2d\eta^2 + d\chi^2] = 0$$

because light travels along the null geodesics. Therefore,

$$\Delta\chi(\eta) = \pm c\Delta\eta.$$

At time η_1, suppose a galaxy emits a signal of short conformal duration $\delta\eta$. The light arrives at time $\eta_0 = \eta_1 + d/c$ where d is the galaxy's distance from us. At the points of emission and detection, the physical time intervals are

$$\delta t_1 = a(\eta_1)\delta\eta \qquad \text{and} \qquad \delta t_0 = a(\eta_0)\delta\eta.$$

If δt is the period of the light wave, the light has wavelength emitted $\lambda_1 = c\delta t_1$ and observed $\lambda_0 = c\delta t_0$. Thus, we arrive at the same result as 10.12

$$\frac{\lambda_0}{\lambda_1} = \frac{a(\eta_0)}{a(\eta_1)}.$$

10.3 Dynamics

The evolution of scale factor follows from the **Einstein equation**

$$G_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{10.13}$$

This equation relates the **Einstein tensor** $G_{\mu\nu}$ to the **energy-momentum tensor** $T_{\mu\nu}$. We will see some possible forms of cosmological energy-momentum tensors.

10.3.1 Perfect Fluids

The matter consistent with homogeneous and isotropic constraints is a perfect fluid with energy-momentum tensor

$$T_{\mu\nu} = \left(\rho + \frac{P}{c^2}\right)U_{\mu}U_{\nu} + Pg_{\mu\nu} \tag{10.14}$$

where ρc^2 and P are the energy density and pressure in the rest frame of the fluid and u^{ν} is the 4-velocity relative to a comoving observer.

Recall that the four-vector N^μ is the number currnt. The N^0 component is the number density and N^i is the flux of the particles in direction x^i. Isotropy requires the mean of any three-vector to be 0. Homogeneity requires the mean of any three-scalar (such as N^0) to be only a function of time,

$$N^0 = cn(t), \quad N^i = 0.$$

where $n(t)$ is the number of galaxies per proper volume (here a galaxy is counted as one "particle"). An observer will measure the number current to be

$$N^\mu = nu^\mu = n\frac{dx^\mu}{d\tau}.$$

If the number of particles is conserved, the rate of change of number density should be equal to the divergence of the flux of the particles, so we have the continuity equation

$$\partial_0 N^0 = -\partial_i N^i \quad \text{or} \quad \partial_\mu N^\mu = 0,$$

which can be generalized to curved spacetimes as covariant derivative

$$\nabla_\mu N^\mu = 0 \tag{10.15}$$

and this form is independent of coordinates. The covariant derivative acts on four-vectors with formula

$$\nabla_\mu A^\nu = \partial_\mu A^\nu + \Gamma^\nu_{\mu\lambda} A^\lambda \tag{10.16}$$

$$\nabla_\mu B_\nu = \partial_\mu B_\nu - \Gamma^\lambda_{\mu\nu} B_\lambda \tag{10.17}$$

Combine these two,

$$\nabla_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\mu\lambda} T^\lambda_\nu - \Gamma^\lambda_{\mu\nu} T^\mu_\lambda \tag{10.18}$$

Combine equation 10.15 and 10.16,

$$\partial_\mu N^\mu = -\Gamma^\mu_{\mu\lambda} N^\lambda.$$

In rest frame of the fluid, substitute $N^0 = cn(t)$ and $N^i = 0$, we have

$$\frac{1}{c}\frac{dn}{dt} = -\Gamma^\mu_{\mu 0} n = -\frac{3}{c}\frac{\dot{a}}{a} n$$

because $\Gamma^0_{00} = 0$ and $\Gamma^i_{i0} = 3c^{-1}\dot{a}/a$. Therefore

$$\frac{\dot{n}}{n} = -3\frac{\dot{a}}{a} \quad \Longrightarrow \quad n(t) \propto a^{-3}.$$

This makes phyical sense because the proper volume $V \propto a^3$ and the number density decreases with $n(t) \propto a^{-3}$.

The energy-momentum tensor $T_{\mu\nu}$ can be decomposed into

$$T_{\mu\nu} = \left(\begin{array}{c|c} T_{00} & T_{0j} \\ \hline T_{i0} & T_{ij} \end{array}\right) = \left(\begin{array}{c|c} \text{energy density} & \text{momentum density} \\ \hline \text{energy flux} & \text{stress tensor} \end{array}\right).$$

As we have said, in homogeneous universe, the energy density can only depend on time, $T_{00} = \rho(t)c^2$. Isotropy says the mean of three-vectors is zero, $T_{i0} = T_{0j} = 0$. Moreover, isotropy around a point $\mathbf{x} = 0$ constrains the mean of any three-tensor to be proportional to δ_{ij}. Since the metric $g_{ij} = a^2\delta_{ij}$, we have

$$T_{ij}(\mathbf{x} = 0) \propto \delta_{ij} \propto g_{ij}(\mathbf{x} = 0).$$

The proportionality constant is only a function of time and since it is between two tensors, it is unaffected by transformations of the spatial coordinates. Hence we conclude that

$$T_{00} = \rho(t)c^2, \quad T_{i0} = T_{0j} = 0, \quad T_{ij} = P(t)g_{ij}(t, \mathbf{x}).$$

Then we can write it in matrix

$$T^\mu_\nu = g^{\mu\lambda} T_{\lambda\nu} = \begin{pmatrix} -\rho c^2 & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix},$$

or explicitly as equation

$$T_{\mu\nu} = \left(\rho + \frac{P}{c^2}\right) U_\mu U_\nu + P g_{\mu\nu}$$

where, for a comoving observer, $u^\mu = (c, 0, 0, 0)$.

In Minkowski space, energy and momentum are conserved, so they satisfies the **continuity equations**,

$$\dot{\rho} = -\partial_i \pi^i, \quad \dot{\pi}_i = \partial_i P.$$

These laws can be combined into component conservation equation for the energy-momentum tensor

$$\partial_\mu T^\mu_\nu = 0,$$

or in general relativity according to 10.18

$$\nabla_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\mu\lambda} T^\lambda_\nu - \Gamma^\lambda_{\mu\nu} T^\mu_\lambda = 0.$$

There are four separate equations, one for each ν. When $\nu = 0$, it is the equation for energy density,

$$\partial_\mu T^\mu_0 + \Gamma^\mu_{\mu\lambda} T^\lambda_0 - \Gamma^\lambda_{\mu 0} T^\mu_\lambda = 0.$$

By isotropy, $T^i_0 = 0$, so

$$\frac{1}{c}\frac{d\left(-\rho c^2\right)}{dt} + \Gamma^\mu_{\mu 0}\left(-\rho c^2\right) - \Gamma^\lambda_{\mu 0} T^\mu_\lambda = 0.$$

$\Gamma^\lambda_{\mu 0}$ vanishes unless λ and μ are spatial indices and equal to each other. In this case, $\Gamma^i_{i0} = 3c^{-1}\dot{a}/a$. The continuity equation becomes

$$-c\dot{\rho} - 3c\frac{\dot{a}}{a}\rho - \frac{3}{c}\frac{\dot{a}}{a}P = 0.$$

$$\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + \frac{P}{c^2}\right) = 0 \tag{10.19}$$

This equation describes "energy conservation" in the cosmological context.

Define a constant **equation of state**,

$$w = \frac{P}{\rho c^2} \tag{10.20}$$

Most cosmological fluids can be parameterized in terms of a constant equation of state. The energy continuity equation is then

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} \quad \Longrightarrow \quad \rho \propto a^{-3(1+w)}.$$

10.3.2 Matter and Radiation

In this part, **matter** denotes a fluid with pressure much smaller than its energy density, $|P| \ll \rho c^2$. It is a gas of nonrelativistic particles, i.e. the energy density is dominated by their rest mass,

$$\rho \propto a^{-3}, \quad [w = 0].$$

This reflects as the energy within a region stays the same, the region increases as $V \propto a^3$.

Radiation denotes anything with the pressure is one third of the energy density, $P = \frac{1}{3}\rho c^2$. In this case, the gas is relativistic, i.e. the energy density is dominated by the kinetic energy,

$$\rho \propto a^{-4}, \quad \left[w = \frac{1}{3}\right].$$

This means the dilusion includes not only volume expansion, $V \propto a^3$, but also energy redshift, $E \propto a^{-1}$.

- **Light particles**
  At early universe, all particles of the Standard Model acted as radiation because of the high temperature. As the temperature

dropped, many particles started to behave like matter.

- **Photons**
  Massless particles are always relativistic. During Big Bang nucleosynthesis, with neutrinoes, they were the dominant energy density. Today, these photons appear as cosmic microwave background.

- **Neutrinos**
  Because of their extremely small mass, neutrinoes behave like radiation for most of the history of universe. Recently, they act more like matter.

- **Graviton**
  The energy desity of gravitons is predicted to be very small, so they have a negligible effect on the expansion of the universe.

10.3.3 Dark Energy

The universe today is dominated by **dark energy** with negative pressure, $P = -\rho c^2$. They have constant energy density,

$$\rho \propto a^0, \quad [w = -1]$$

In this way, energy has to be created, but still it does not violate the conservation of energy according to 10.19.

- **Vacuum energy**
  As the universe expands, more space is created and the energy for empty space increases proportionally to volume. This leads to an energy-momentum tensor
  $$T^{\text{vac}}_{\mu\nu} = -\rho_{\text{vac}} c^2 g_{\mu\nu}.$$
  This gives the pressure of vacuum $P_{\text{vac}} = -\rho_{\text{vac}} c^2$

- **Cosmological constant**
  We can add the term $\Lambda g_{\mu\nu}$ to the LHS of the Einstein equation without changing the conservation of the energy-momentum tensor, $\nabla^\mu T_{\mu\nu} = 0$. Therefore, we can rewrite the Einstein equation as
  $$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}.$$
  We can move this extra term to the RHS as contribution to the energy-momentum tensor
  $$T^{(\Lambda)}_{\mu\nu} = -\frac{\Lambda c^4}{8\pi G} g_{\mu\nu} \equiv -\rho_\Lambda c^2 g_{\mu\nu}.$$

10.3.4 Spacetime Curvature

Now is to understand how matter and energy satisfy the dynamics of the spacetime. Compute the Einstein tensor on the LHS of the Einstein equation,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}, \tag{10.21}$$

where $R_{\mu\nu}$ is the **Ricci tensor** and $R = R^\mu_\mu = g^{\mu\nu} R_{\mu\nu}$ is its trace, the **Ricci scalar**. The Ricci tensor can be written in terms of Christoffel symbols,

$$R_{\mu\nu} = \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu \Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\lambda\rho} \Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\mu\lambda} \Gamma^\lambda_{\nu\rho}. \tag{10.22}$$

Because of the isotropy of the Robertson-Walker metric, the three vectors $R_{i0} = R_{0i} = 0$. The non-vanishing components of the Ricci tensor are

$$R_{00} = -\frac{3}{c^2} \frac{\ddot{a}}{a},$$

$$R_{ij} = \frac{1}{c^2} \left[\frac{\ddot{a}}{a} + 2 \left(\frac{\dot{a}}{a} \right)^2 + 2 \frac{kc^2}{a^2 R_0^2} \right] g_{ij}.$$

The Ricci scalar is

$$\begin{aligned} R = g^{\mu\nu} R_{\mu\nu} &= -R_{00} + \frac{1}{a^2} R_{ii} \\ &= \frac{6}{c^2}\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2 R_0^2}\right]. \end{aligned} \tag{10.23}$$

The nonzero components of the Einstein tensor, $G^\mu_\nu = g^{\mu\lambda} G_{\lambda\nu}$, are

$$\begin{aligned} G^0_0 &= -\frac{3}{c^2}\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2 R_0^2}\right], \\ G^i_j &= -\frac{1}{c^2}\left[2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{kc^2}{a^2 R_0^2}\right]\delta^i_j. \end{aligned}$$

10.3.5 Friedmann Equations

Now is to evaluate the Einstein equation. Setting

$$G^0_0 = \left(\frac{8\pi G}{c^4}\right) T^0_0 = -\frac{8\pi G}{c^2}\rho,$$

we get the **first Friedmann equation**:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2 R_0^2}, \tag{10.24}$$

where ρ is the sum of all contributions to the energy density. It will be separated into several energy density: The spatial part

ρ_r for contribution from radiation	
ρ_γ	photons
ρ_ν	neutrinos
ρ_m for contribution from matter	
ρ_c	cold dark matter
ρ_b	baryons
ρ_Λ for contribution from vacuum energy	

of the Einstein equation, $G^i_j = (8\pi G/c^4) T^i_j$, leads to the **second Friedmann equation** by taking the time derivative of the first Friedmann equation and use the continuity equation 10.19 for $\dot{\rho}$,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right), \tag{10.25}$$

The first Friedmann equation can be written in terms of the Hubble parameter,

$$H^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2 R_0^2}. \tag{10.26}$$

A flat universe with $k = 0$ have **critical density** today

$$\begin{aligned} \rho_{\text{crit},0} = \frac{3H_0^2}{8\pi G} &= 1.9 \times 10^{-29} h^2 \,\text{gram}\,\text{cm}^{-3} \\ &= 2.8 \times 10^{11} h^2 M_\odot \text{Mpc}^{-3} \\ &= 1.1 \times 10^{-5} h^2 \text{protons}\,\text{cm}^{-3}. \end{aligned}$$

For convenience, we can measure all densities relative to the critical density today with dimensionless density parameters

$$\Omega_{i,0} = \frac{\rho_{i,0}}{\rho_{\text{crit},0}}, \quad i = r, m, \Lambda, \ldots$$

The subscript “0” is often dropped in $\Omega_{i,0}$ into Ω_i. The Friedmann equation 10.26 can then be written as

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-3} + \Omega_\Lambda, \tag{10.27}$$

where Ω_k is the curvature density parameter, $\Omega = -kc^2/(R_0 H_0)^2$.

At present time t_0, the scaling factor $a(t_0) = 1$ will lead to the constraint

$$1 = \Omega_r + \Omega_m + \Omega_\Lambda + \Omega_k.$$

Define $\Omega_0 = \Omega_r + \Omega_m + \Omega_\Lambda$, we can write the curvature parameter as

$$\Omega_k = -\frac{kc^2}{(R_0 H_0)^2} = 1 - \Omega_0.$$

10.3.6 Exact Solutions

In this section we will present some special cases for which exact solutions are possible.

Consider a flat universe ($k = 0$) with a single fluid component. In fact, most of the the history of the universe was dominated by a single component (first radiation, then matter, then vacuum energy). Using the equation of state w_i, the Friedmann equation 10.27 can be written as

$$\frac{\dot{a}}{a} \approx H_0 \sqrt{\Omega_i} a^{-\frac{3}{2}(1+w_i)}.$$

Integrating this equation, we get

$$a(t) \propto \begin{cases} t^{2/3(1+w_i)} & w_i \neq -1 \\ e^{H_0 \sqrt{\Omega_\Lambda} t} & w_i = -1 \end{cases}$$

A few comments to this result:

- The solution for a pure matter universe is known as the **Einstein-de Sitter universe**:

  $$a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3}}.$$

  It is the standard cosmological approximation, especially to the long matter-dominated period in the history of our universe. This equation follows the relation between the Hubble constant and the age of the universe,

  $$H_0 = \frac{2}{3}\frac{1}{t_0}.$$

  Input the value of the Hubble constant, $H_0 \approx 70\,\mathrm{km\,s^{-1}Mpc^{-1}}$, we can calculate the age of the universe,

  $$t_0 = \frac{2}{3} H_0^{-1} \approx 9\,\text{billion yrs}.$$

  This shows the “age problem”, the age of the pure matter universe is shorter than that of the oldest stars.

- The result apply to a universe without any matter and only a curvature term. The a^{-2} scaling implies $w_k = -1/3$ and the Friedmann equation only have a solution for $k = -1$. We will then get

  $$a(t) = \frac{t}{t_0},$$

  which is called the **Milne universe**.

- The solution for $w_\Lambda = -1$ is known as **de Sitter space**. It is a good approximation to the evolution of our universe in the far past and the far future.

Except the de Sitter solution, the other solutions all have a singularity at $t = 0$, when the scale factor becomes zero and the density diverses as $\rho \approx t^{-2}$. Consider the second Friedmann equation

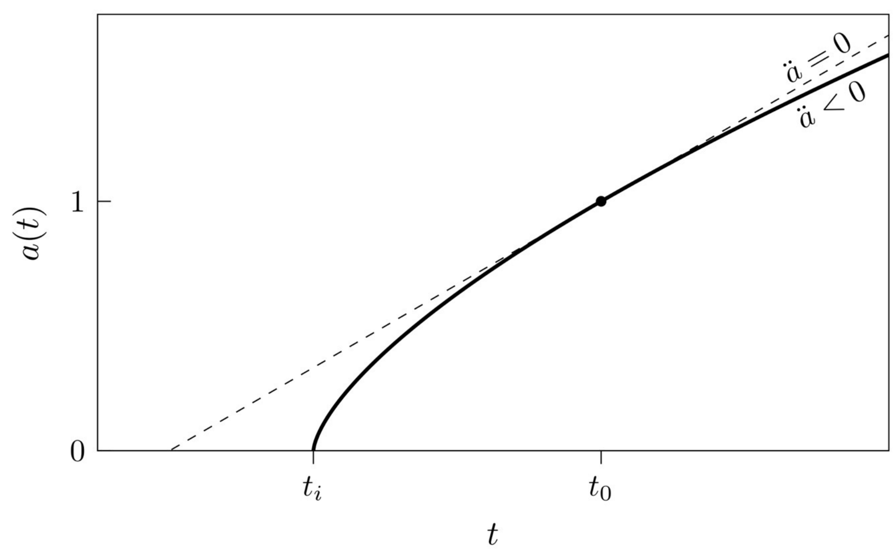
$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right).$$

As long as the matter obeys the **strong energy condition**, $\rho c^2 + 3P \geq 0$ so that $\ddot{a} < 0$, the second Friedmann equation implies a singularity in the past. The strong energy condition holds because the gravity is attractive so it slows down the expansion.

At time $t = t_0$, we have two conditions

$$a(t_0) = 1,$$
$$\dot{a}(t_0) = H_0.$$

Assume the universe is expanding at t_0, $H_0 > 0$. Consider the case where $\ddot{a} = 0$. The solution is $a(t) = 1 + H_0(t - t_0)$ with a singularity at $t_0 - H_0^{-1}$. If $\ddot{a} < 0$, the solution will have the same tangent at t_0, but smaller $a(t)$ at $t < t_0$, thus a singularity at some $t_i > t_0 - H_0^{-1}$.



We can also find solutions for two fluid components. This is related to the early universe (a mixture of matter and radiation) and late universe (a mixture of matter and dark energy). We will work in conformal time where the primes denote derivatives with respect to η. In this way, $\dot{a} = a'/a$ and $\ddot{a} = a''/a^2 - (a')^2/a^3$. The two Friedmann equations then become

$$(a')^2 + \frac{kc^2}{R_0^2}a^2 = \frac{8\pi G}{3}\rho a^4 \tag{10.28}$$

$$a'' + \frac{kc^2}{R_0^2}a = \frac{4\pi G}{3}\left(\rho - \frac{3P}{c^2}\right)a^3 \tag{10.29}$$

A few examples:

- **Matter and Radiation**
  The total density can be written as

$$\rho = \rho_m + \rho_r = \frac{\rho_{\text{eq}}}{2}\left[\left(\frac{a_{\text{eq}}}{a}\right)^3 + \left(\frac{a_{\text{eq}}}{a}\right)^4\right],$$

  where "eq" denotes quantites evaluated at matter-radiation equality (approximately 50000 years after Big Bang). Since $\rho_r c^2 - 3P_r = 0$, radiation does not contribute as a source term in 10.29. Also, $\rho_m a^3 = \text{const.} = \frac{1}{2}\rho_{\text{eq}} a_{\text{eq}}^3$ and $k = 0$, so we write 10.29 as

$$a'' = \frac{2\pi G}{3}\rho_{\text{eq}} a_{\text{eq}}^3.$$

  The solution is

$$a(\eta) = \frac{\pi G}{3}\rho_{\text{eq}} a_{\text{eq}}^3 \eta^2 + C\eta + D.$$

The inital condition $a(\eta = 0) = 0$ gives $D = 0$. Subsituting the solution and the density equation to 10.28,

$$C = \left(\frac{4\pi G}{3}\rho_{\text{eq}}a_{\text{eq}}^4\right)^{\frac{1}{2}}.$$

The solution then can be written as

$$a(\eta) = a_{eq}\left[\left(\frac{\eta}{\eta_*}\right)^2 + 2\left(\frac{\eta}{\eta_*}\right)\right], \tag{10.30}$$

where

$$\eta_* = \left(\frac{\pi G}{3}\rho_{\text{eq}}a_{\text{eq}}^2\right)^{-\frac{1}{2}} = \frac{\eta_{\text{eq}}}{\sqrt{2}-1}.$$

For $\eta \ll \eta_{\text{eq}}$, radiation-dominated; $\eta \gg \eta_{\text{eq}}$, matter-dominated.

- **Matter and Curvature**

  Without dark energy, a positively-curved universe would end in a Big Crunch; a negatively-curved would expand forever. Define the conformal Hubble parameter as $\mathcal{H} = a'/a$, the two Friedmann equations are

$$\mathcal{H}^2 + \frac{kc^2}{R_0^2} = \frac{8\pi G}{3}\rho_m a^2, \tag{10.31}$$

$$\frac{d\mathcal{H}}{d\eta} = -\frac{4\pi G}{3}\rho_m a^2. \tag{10.32}$$

  Combining the two equations,

$$2\frac{d\mathcal{H}}{d\eta} + \mathcal{H}^2 + \frac{kc^2}{R_0^2} = 0.$$

  Define $\theta = c\eta/R_0$ and rescaled Hubble parameter $\widetilde{\mathcal{H}} = R_0\mathcal{H}/c$, this becomes

$$2\frac{d\widetilde{\mathcal{H}}}{d\theta} + \widetilde{\mathcal{H}}^2 + k = 0.$$

  The solutions are

$$\widetilde{\mathcal{H}}(\theta) = \frac{1}{a}\frac{da}{d\theta} = \begin{cases} \cot(\theta/2) & k = +1 \\ 2/\theta & k = 0 \\ \coth(\theta/2) & k = -1 \end{cases}.$$

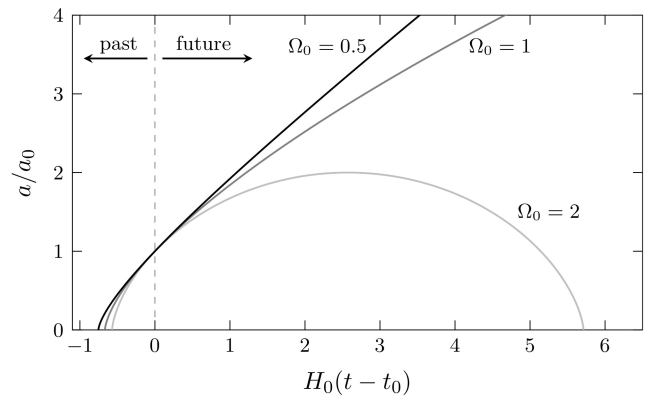
  The scale factor solutions are

$$a(\theta) = A\begin{cases} \sin^2(\theta/2) & k = +1 \\ \theta^2 & k = 0 \\ \sinh^2(\theta/2) & k = -1 \end{cases},$$

  and the physical time solutions are

$$t = \frac{R_0}{c}\int a(\theta)\,d\theta = \frac{R_0 A}{2c}\begin{cases} \theta - \sin(\theta) & k = +1 \\ \theta^3 & k = 0 \\ \sinh(\theta) - \theta & k = -1 \end{cases}.$$

  From these solutions, we can plot the fate of the universe.



- **Cosmological Constant and Curvature**
  The Friedmann equation is

$$H^2 = \frac{\Lambda c^2}{3} - \frac{kc^2}{a^2 R_0^2}.$$

When $\Lambda > 0$,

$$a(t) = A \begin{cases} \cosh(\alpha t) & k = +1 \\ \exp(\alpha t) & k = 0 \\ \sinh(\alpha t) & k = -1 \end{cases},$$

where $\alpha = \sqrt{\Lambda c^2/3}$. The normalization $A = \sqrt{3/\Lambda}/R_0$ for $k \pm 1$ and is arbitrary for $k = 0$. The $k = +1$ solution does not have a singularity. The singularity exists at $t = -\infty$ and $t = 0$ for $k = 0$ and $k = -1$, respectively. For $\Lambda < 0$, the only solution is for $k = -1$,

$$a(t) = A \sin(\alpha t),$$

where $\alpha = \sqrt{\Lambda c^2/3}$. This solution is known as the **anti-de Sitter space**.

- **Matter and Cosmological Constant**
  This is a good approximation to our universe today. The Friedmann equation is

$$\frac{H^2}{H_0^2} = \frac{\Omega_m}{a^3} + \Omega_\Lambda,$$

where $\Omega_m + \Omega_\Lambda = 1$. It has the solution

$$a(t) = \left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{\frac{1}{3}} \sinh^{\frac{2}{3}}\left(\frac{3}{2}\alpha t\right),$$

where $\alpha = \sqrt{\Omega_\Lambda} H_0$. With $a(t_0) = 1$, we can calculate the age of the universe,

$$t_0 = \frac{2}{3\sqrt{\Omega_\Lambda} H_0} \sinh^{-1}\left(\sqrt{\frac{\Omega_\Lambda}{\Omega_m}}\right).$$

The observed cosmological parameters are $\Omega_m = 0.32$ and $\Omega_\Lambda = 0.68$, so

$$t_0 = 0.96 H_0^{-1} \approx 14 \text{ billion yrs.}$$

There appears the **coincidence problem**: why dark energy came to dominate so close to the present time. While matter dominated in the past, the dark energy will dominate in the future.

A Proofs and Derivations

A.1 Lorentz Transformation

To derive the Lorentz transformation, we need to use both postulates of special relativity:

1. The laws of physics take the same form in all inertial reference frames. All inertial frames are fundamentally indistinguishable based on experiments in that frame only.
2. The speed of light (and other massless particles) c is the maximum speed in the universe and has the same value in all inertial reference frames.

Consider two frames S and S', moving at a velocity v relative to S, with parallel coordinate axes. The x-axis and x'-axis are aligned. When origins of S and S' coincide (call them $\mathcal{O}$ and $\mathcal{O}'$ respectively), the time in both frames is set to zero, $t = t' = 0$. Since the x-axes are aligned, we will ignore y- and z-components here because they are not interesting. Suppose the Lorentz transformation from inertial frame S to another inertial frame S' takes the form

$$t' = f(x,t), \quad x' = g(x,t),$$

and f and g are linearly independent. Consider a free particle moving with constant velocity $v_x = dx/dt$ in S frame. By postulate 1, this particle observed in S' must also move at a constant velocity $v'_x = dx'/dt'$, which means

$$v'_x = \frac{dx'}{dt'} = \frac{dx'/dt}{dt'/dt} = \frac{\frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial t}}{\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial t}} = \frac{\frac{\partial g}{\partial x}v_x + \frac{\partial g}{\partial t}}{\frac{\partial f}{\partial x}v_x + \frac{\partial f}{\partial t}}.$$

Since v_x and v'_x can be arbitrary constants, all partial derivatives must also be constants. This means the Lorentz transformation must be linear:

$$\begin{aligned} t' &= a_s + a_0 t + a_1 x, \\ x' &= b_s + b_0 t + b_1 x. \end{aligned}$$

The goal is to find the constants a_s, a_0, a_1, b_s, b_0, and b_1. When $\mathcal{O}$ and $\mathcal{O}'$ coincide, we have $x = x' = 0$ at $t' = t = 0$. This synchronization event sets $a_s = b_s = 0$. Now, the frame S' is moving at a velocity v relative to S, then the position of $\mathcal{O}'$ in S is $x = vt$. In S', its origin $\mathcal{O}'$ is always at $x' = 0$, so

$$0 = x' = b_0 t + b_1(vt) \quad \Longrightarrow \quad b_0 = -b_1 v \quad \Longrightarrow \quad x' = b_1(x - vt).$$

Moreover, by symmetry, S is moving at a velocity of $-v$ relative to S'. If the transformation works for all frames (by postulate 1), then we should obtain

$$x = b_1(x' + vt').$$

The transformation equations now look like

$$\begin{aligned} t' &= a_0 t + a_1 x, \\ x' &= b_1(x - vt), \\ x &= b_1(x' + vt'). \end{aligned}$$

It's time to use postulate 2. Suppose a light pulse is emitted at $(t,x) = (0,0)$ and $(t',x') = (0,0)$, i.e., at synchronization. By the constancy of the speed of light, the event $(t,x) = (t,ct)$ in S should have coordinates $(t',x') = (t',ct')$ in S'. Thus,

$$\begin{aligned} x' &= ct' = b_1(x - vt) = b_1(ct - vt) = b_1(1-\beta)ct, \\ x &= ct = b_1(x' + vt') = b_1(ct' + vt') = b_1(1+\beta)ct', \end{aligned}$$

where $\beta \equiv v/c$. Substituting ct into ct', we have

$$ct' = b_1^2(1-\beta^2)ct' \quad \Longrightarrow \quad b_1 = \pm\frac{1}{\sqrt{1-\beta^2}}.$$

We will keep the positive solution as the negative solution does not satisfy $x = x'$ and $t = t'$ at $v = 0$. This b_1 is the Lorentz factor,

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}}.$$

Using $x' = \gamma(x - vt)$ and $x = \gamma(x' + vt')$ to solve for t' and we will get

$$t' = \gamma t - \gamma\beta\,(x/c)\,.$$

The Lorentz transformation is

$$\begin{aligned} ct' &= \gamma(ct - \beta x), \\ x' &= \gamma(x - \beta ct), \\ y' &= y, \\ z' &= z. \end{aligned}$$

The line element $ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2$ is invariant under Lorentz transformation:

$$\begin{aligned} -c^2dt'^2 + dx'^2 + dy'^2 + dz'^2 &= -\gamma^2(c\,dt - \beta\,dx)^2 + \gamma^2(dx - \beta c\,dt)^2 + dy^2 + dz^2 \\ &= \gamma^2(-c^2dt^2 + 2\beta c\,dt\,dx - \beta^2dx^2 + dx^2 - 2\beta c\,dt\,dx + \beta^2c^2dt^2) + dy^2 + dz^2 \\ &= \gamma^2c^2(\beta^2 - 1)\,dt^2 + \gamma^2(1-\beta^2)\,dx^2 + dy^2 + dz^2 \\ &= -dt^2 + dx^2 + dy^2 + dz^2. \end{aligned}$$

A.2 Geodesic Equation

The Lagrangian and Euler-Lagrange equation are given by

$$\mathcal{L} = \sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}} = \frac{d\tau}{d\lambda} \qquad \text{and} \qquad \frac{\partial\mathcal{L}}{\partial x^\alpha} - \frac{d}{d\lambda}\frac{\partial\mathcal{L}}{\partial(dx^\alpha/d\lambda)} = 0.$$

We want the derivatives,

$$\frac{\partial\mathcal{L}}{\partial x^\alpha} = \frac{1}{2\mathcal{L}}\left(-\frac{\partial g_{\mu\nu}}{\partial x^\alpha}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\right) = -\frac{1}{2}\frac{d\lambda}{d\tau}\left(\frac{\partial g_{\mu\nu}}{\partial x^\alpha}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\tau}\right) = -\frac{1}{2}\partial_\alpha g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\tau}.$$

$$\frac{\partial\mathcal{L}}{\partial(dx^\alpha/d\lambda)} = \frac{1}{2\mathcal{L}}\left(-2g_{\alpha\nu}\frac{dx^\nu}{d\lambda}\right) = -\frac{d\lambda}{d\tau}g_{\alpha\nu}\frac{dx^\nu}{d\lambda} = -g_{\alpha\nu}\frac{dx^\nu}{d\tau}.$$

Note the factor of 2 in the first equality comes from

$$\left[-g_{\mu\nu}\frac{\partial}{\partial(dx^\alpha/d\lambda)}\left(\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\right)\right] = -g_{\mu\nu}\delta^\mu{}_\alpha\frac{dx^\nu}{d\lambda} - g_{\mu\nu}\delta^\nu{}_\alpha\frac{dx^\mu}{d\lambda} = -g_{\nu\alpha}\frac{dx^\nu}{d\lambda} - g_{\mu\alpha}\frac{dx^\mu}{d\lambda} = -2g_{\nu\alpha}\frac{dx^\nu}{d\lambda}.$$

Multiply $d\lambda/d\tau$ on both sides of the Euler-Lagrange equation

$$\frac{d\lambda}{d\tau}\frac{\partial\mathcal{L}}{\partial x^\alpha} - \frac{d}{d\tau}\frac{\partial\mathcal{L}}{\partial(dx^\alpha/d\lambda)} = 0.$$

Substituting the derivatives,

$$-\frac{1}{2}\partial_\alpha g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} - \frac{d}{d\tau}\left(-g_{\alpha\nu}\frac{dx^\nu}{d\tau}\right) = 0.$$

Expanding the τ-derivative using the product rule and chain rule,

$$-\frac{1}{2}\partial_\alpha g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} + \partial_\mu g_{\alpha\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} + g_{\alpha\nu}\frac{d^2x^\nu}{d\tau^2} = 0.$$

Multiplying both sides by $g^{\alpha\beta}$, the inverse of $g_{\alpha\beta}$,

$$-\frac{1}{2}g^{\alpha\beta}\partial_\alpha g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} + g^{\alpha\beta}\partial_\mu g_{\alpha\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} + \delta^\beta{}_\nu\frac{d^2x^\nu}{d\tau^2} = 0.$$

The second term can be split into two parts,

$$\partial_\mu g_{\alpha\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = \frac{1}{2}\partial_\mu g_{\alpha\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} + \frac{1}{2}\partial_\nu g_{\alpha\mu}\frac{dx^\nu}{d\tau}\frac{dx^\mu}{d\tau} = \frac{1}{2}(\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\alpha\mu})\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau},$$

where we interchanged the dummy indices μ and ν. Hence

$$-\frac{1}{2}g^{\alpha\beta}\partial_\alpha g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} + \frac{1}{2}(\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\alpha\mu})\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} + \delta^\beta{}_\nu\frac{d^2x^\nu}{d\tau^2} = 0.$$

Rearranging some terms, we get the geodesic equation,

$$\frac{d^2x^\beta}{d\tau^2} + \frac{1}{2}g^{\alpha\beta}\left(\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\alpha\mu} - \partial_\alpha g_{\mu\nu}\right)\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0,$$

or equivalently,

$$\boxed{\frac{d^2x^\alpha}{d\tau^2} + \Gamma^\alpha{}_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0, \qquad \text{where} \qquad \Gamma^\alpha{}_{\mu\nu} = \frac{1}{2}g^{\alpha\beta}\left(\frac{\partial g_{\beta\nu}}{\partial x^\mu} + \frac{\partial g_{\beta\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\beta}\right).}$$

A.3 The Periapsis Precession

A.3.1 Direct Integration

The periapsis precession integral is

$$\Delta\phi = 2L\int_{r_1}^{r_2}\frac{dr}{r^2}\left(c^2e^2 - c^2 + \frac{2GM}{r} - \frac{L^2}{r^2} + \frac{2GML^2}{r^3c^2}\right)^{-1/2}.$$

We restore the G's and c's because then we can track the order of the effects. We want the lowest order ($1/c^2$) effects from the integral. Rearrange the integral,

$$\Delta\phi = 2L\int_{r_1}^{r_2}\frac{dr}{r^2}\left[c^2e^2 - \left(1-\frac{2GM}{rc^2}\right)\left(c^2+\frac{L^2}{r^2}\right)\right]^{-1/2} = 2L\int_{r_1}^{r_2}\frac{dr}{r^2}\left(1-\frac{2GM}{rc^2}\right)^{-1/2}\left[c^2e^2\left(1-\frac{2GM}{rc^2}\right)^{-1} - c^2 - \frac{L^2}{r^2}\right]^{-1/2}.$$

Now expand to leading order of c^2,

$$\left(1-\frac{2GM}{rc^2}\right)^{-1/2} \simeq 1 + \frac{GM}{rc^2}, \quad c^2e^2\left(1-\frac{2GM}{rc^2}\right)^{-1} \simeq c^2e^2\left(1+\frac{2GM}{rc^2}+\frac{4G^2M^2}{r^2c^4}\right).$$

Substituting them into $\Delta\phi$,

$$\begin{aligned}\Delta\phi &\simeq 2L\int_{r_1}^{r_2}\frac{dr}{r^2}\left(1+\frac{GM}{rc^2}\right)\left[c^2(e^2-1) + \frac{2GMe^2}{r} + \left(\frac{4G^2M^2e^2}{c^2L^2} - 1\right)\frac{L^2}{r^2}\right]^{-1/2} \\ &= 2L\int_{r_1}^{r_2}\frac{dr}{r^2}\left(1+\frac{GM}{rc^2}\right)\left[2E_{\text{Newt}} + \frac{2GM}{r}\left(1+\frac{2E_{\text{Newt}}}{c^2}\right) - \frac{L^2}{r^2}\left(1-\frac{4G^2M^2e^2}{c^2L^2}\right)\right]^{-1/2},\end{aligned}$$

where $E_{\text{Newt}} = c^2(e^2-1)/2$. In fact, if $e^2 = 1 + 2E_{\text{Newt}}/c^2$, the last term in the square bracket will be

$$-\frac{L^2}{r^2}\left(1-\frac{4G^2M^2e^2}{c^2L^2}\right) = -\frac{L^2}{r^2}\left(1-\frac{4G^2M^2}{c^2L^2} - \frac{8G^2M^2E_{\text{Newt}}}{c^4L^2}\right) \simeq -\frac{L^2}{r^2}\left(1-\frac{4G^2M^2}{c^2L^2}\right),$$

so we have no e's in all expressions. Making a change of variables $u = 1/r$, $du = -dr/r^2$,

$$\begin{aligned}\Delta\phi &\simeq 2L\int_{u_2}^{u_1}du\left(1+\frac{GMu}{c^2}\right)\left[2E_{\text{Newt}} + 2GMu\left(1+\frac{2E_{\text{Newt}}}{c^2}\right) - L^2u^2\left(1-\frac{4G^2M^2}{c^2L^2}\right)\right]^{-1/2} \\ &= \frac{2L}{L(1-4G^2M^2/c^2L^2)^{1/2}}\int_{u_2}^{u_1}du\left(1+\frac{GMu}{c^2}\right)\left[\frac{2E_{\text{Newt}}}{L^2(1-4G^2M^2/c^2L^2)} + \frac{2GMu(1+2E_{\text{Newt}}/c^2)}{L^2(1-4G^2M^2/c^2L^2)} - u^2\right]^{-1/2}.\end{aligned}$$

The terms in the square bracket is a mess, but recall that r_1 and r_2 are periapsis and apoapsis where $dr/d\tau = 0$, while $dr/d\tau$ is just the square bracket (to the power of $-1/2$). Thus, u_1 and u_2 are the roots of this quadratic expression, so it can be written as

$(u_1-u)(u-u_2)$. The integral looks a lot nicer now,

$$\Delta\phi \simeq 2\left(1+\frac{2G^2M^2}{c^2L^2}\right)\int_{u_2}^{u_1} du\,\frac{1+GMu/c^2}{\sqrt{(u_1-u)(u-u_2)}} = 2\left(1+\frac{2G^2M^2}{c^2L^2}\right)\int_{u_2}^{u_1}\frac{du}{\sqrt{(u_1-u)(u-u_2)}} + \frac{2GM}{c^2}\int_{u_2}^{u_1}\frac{u\,du}{\sqrt{(u_1-u)(u-u_2)}}.$$

Completing the square,

$$(u_1-u)(u-u_2) = -u_1u_2 + (u_1+u_2)u - u^2 = -u_1u_2 - \left(u-\frac{u_1+u_2}{2}\right)^2 + \frac{(u_1+u_2)^2}{4} = \frac{(u_1-u_2)^2}{4} - \left(u-\frac{u_1+u_2}{2}\right)^2.$$

Let

$$u-\frac{u_1+u_2}{2} = \frac{u_1-u_2}{2}\sin\theta, \quad du = \frac{u_1-u_2}{2}\cos\theta\,d\theta \quad\Longrightarrow\quad \left(u-\frac{u_1+u_2}{2}\right)^2 = \frac{(u_1-u_2)^2}{4}\sin^2\theta,$$

and

$$(u_1-u)(u-u_2) = \frac{(u_1-u_2)^2}{4} - \frac{(u_1-u_2)^2}{4}\sin^2\theta = \left(\frac{u_1-u_2}{2}\cos\theta\right)^2.$$

As a check if u covers $[u_2,u_1]$, we can find that $u=u_1$ when $\theta=\pi/2$ and $u=u_2$ when $\theta=\pi/2$. The integrals are now almost trivial because

$$\frac{du}{\sqrt{(u_1-u)(u-u_2)}} = \frac{[(u_1-u_2)/2]\cos\theta}{[(u_1-u_2)/2]\cos\theta}\,d\theta = d\theta.$$

The first integral becomes

$$\int_{u_2}^{u_1}\frac{du}{\sqrt{(u_1-u)(u-u_2)}} = \int_{-\pi/2}^{\pi/2} d\theta = \pi.$$

The second integral is

$$\int_{u_2}^{u_1}\frac{u\,du}{\sqrt{(u_1-u)(u-u_2)}} = \int_{u_2}^{u_1} u\,d\theta = \int_{-\pi/2}^{\pi/2}\left(\frac{u_1+u_2}{2}+\frac{u_1-u_2}{2}\sin\theta\right)d\theta = \frac{1}{2}(u_1+u_2)\pi.$$

Substituting these integral into $\Delta\phi$,

$$\Delta\phi \simeq 2\pi\left(1+\frac{2G^2M^2}{c^2L^2}\right) + \frac{GM}{c^2}(u_1+u_2)\pi.$$

The second term is already in the order $1/c^2$, so we need the sum of the roots in the Newtonian limit of the quadratic equation

$$\frac{2E_{\text{Newt}}}{L^2} + \frac{2GMu}{L^2} - u^2 = 0.$$

The sum of roots of a quadratic in the form au^2+bu+c is just $-b/a$. In our case, $u_1+u_2 = 2GM/L^2$, so

$$\Delta\phi \simeq 2\pi\left(1+\frac{2G^2M^2}{c^2L^2}\right) + \frac{2G^2M^2}{L^2c^2}\pi = 2\pi + \frac{6G^2M^2}{c^2L^2}\pi.$$

In conclusion, the apsidal angle $\delta\phi = \Delta\phi - 2\pi$ is

$$\delta\phi = \frac{6\pi G^2M^2}{c^2L^2}.$$

A.3.2 Perturbation in Circular Orbit

To be written.

A.4 The Shapiro Delay Integral

The goal is to solve the integral

$$t(r,r_1) = \int_{r_1}^{r} dr'\frac{1}{b}\left(1-\frac{2M}{r'}\right)^{-1}\left[\frac{1}{b^2}-Q_{\text{eff}}(r')\right]^{-1/2}, \quad \text{where} \quad Q_{\text{eff}}(r) = \frac{1}{r^2}\left(1-\frac{2M}{r}\right).$$

We want to approximate each factor to first order in M/r or M/r_1, provided that $r_1 \gg 2M$. The first factor: impact parameter b is related to the radius of closest approach r_1 by

$$\frac{1}{b^2} = \frac{1}{r_1^2}\left(1 - \frac{2M}{r_1}\right) \quad \Longrightarrow \quad \frac{1}{b} \simeq \frac{1}{r_1}\left(1 - \frac{M}{r_1}\right).$$

The second factor:

$$\left(1 - \frac{2M}{r}\right)^{-1} \simeq 1 + \frac{2M}{r}.$$

The third factor:

$$\frac{1}{b^2} - Q_{\text{eff}}(r) \simeq \frac{1}{r_{\text{min}}^2}\left(1 - \frac{2M}{r_1}\right) - \frac{1}{r^2}\left(1 - \frac{2M}{r}\right) = \frac{r^2 - r_1^2}{r^2 r_1^2} - \frac{2M(r^3 - r_1^3)}{r^3 r_1^3}.$$

Then

$$\begin{aligned}
\left[\frac{1}{b^2} - Q_{\text{eff}}(r)\right]^{-1/2} &\simeq \frac{r r_1}{\sqrt{r^2 - r_1^2}}\left[1 - \frac{2M(r^3 - r_1^3)}{r r_1 (r^2 - r_1^2)}\right]^{-1/2} \\
&\simeq \frac{r r_1}{\sqrt{r^2 - r_1^2}}\left[1 + \frac{M(r^3 - r_1^3)}{r r_1 (r^2 - r_1^2)}\right] \\
&= \frac{r' r_1}{\sqrt{r^2 - r_1^2}} + \frac{M(r^3 - r_1^3)}{(r^2 - r_1^2)^{3/2}}.
\end{aligned}$$

Putting these terms into the integral, and keep terms to the first order in M/r or M/r_1,

$$\begin{aligned}
t(r, r_1) &\simeq \int_{r_1}^{r} \frac{1}{r_1}\left(1 - \frac{M}{r_1}\right)\left(1 + \frac{2M}{r'}\right)\left[\frac{r' r_1}{\sqrt{r'^2 - r_1^2}} + \frac{M(r'^3 - r_1^3)}{(r'^2 - r_1^2)^{3/2}}\right] dr' \\
&= \int_{r_1}^{r} dr' \left[\frac{r'}{\sqrt{r'^2 - r_1^2}}\left(1 - \frac{M}{r_1}\right) + \frac{2M}{\sqrt{r'^2 - r_1^2}} + \frac{M(r'^3 - r_1^3)}{r_1 (r'^2 - r_1^2)^{3/2}}\right].
\end{aligned}$$

The last term can be simplified as

$$\begin{aligned}
\frac{M(r'^3 - r_1^3)}{r_1 (r'^2 - r_1^2)^{3/2}} &= \frac{M}{r_1\sqrt{r'^2 - r_1^2}}\left(\frac{r'^3 - r_1^3}{r'^2 - r_1^2}\right) \\
&= \frac{M}{r_1\sqrt{r'^2 - r_1^2}}\left[\frac{r'(r'^2 - r_1^2) + r_1^2(r' - r_1)}{r'^2 - r_1^2}\right] \\
&= \frac{M r'}{r_1\sqrt{r'^2 - r_1^2}} + \frac{M r_1}{(r' + r_1)\sqrt{r'^2 - r_1^2}}.
\end{aligned}$$

Subtituting it back into the integral cancels one term:

$$t(r, r_1) \simeq \int_{r_1}^{r} dr' \left[\frac{r'}{\sqrt{r'^2 - r_1^2}}\left(1 - \cancel{\frac{M}{r_1}}\right) + \frac{2M}{\sqrt{r'^2 - r_1^2}} + \cancel{\frac{M r'}{r_1\sqrt{r'^2 - r_1^2}}} + \frac{M r_1}{(r' + r_1)\sqrt{r'^2 - r_1^2}}\right].$$

Now we can evaluate each term one by one: the first integral is

$$\int_{r_1}^{r} \frac{r'\,dr'}{\sqrt{r'^2 - r_1^2}} = \frac{1}{2}\int_{r_1}^{r} \frac{d(r'^2)}{\sqrt{r'^2 - r_1^2}} = \sqrt{r'^2 - r_1^2}\Big|_{r_1}^{r} = \sqrt{r^2 - r_1^2}.$$

The second integral is

$$\int_{r_1}^{r} dr' \frac{2M}{\sqrt{r'^2 - r_1^2}}.$$

Let $r' = r_1 \cosh\theta$, $dr' = r_1 \sinh\theta\, d\theta$. Using the identity $\cosh^2\theta - 1 = \sinh^2\theta$,

$$\int_{r_1}^{r} dr' \frac{2M}{\sqrt{r'^2 - r_1^2}} = 2M \int_{0}^{\cosh^{-1}(r/r_1)} \frac{r_1 \sinh\theta}{r_1 \sinh\theta}\, d\theta = 2M \cosh^{-1}\left(\frac{r}{r_1}\right).$$

The $\cosh^{-1}$ can be written in terms of a natural log,

$$\int_{r_1}^{r} dr' \frac{2M}{\sqrt{r'^2 - r_1^2}} = 2M \ln\left(\frac{r}{r_1} + \sqrt{\frac{r^2}{r_1^2} - 1}\right) = 2M \ln\left(\frac{r + \sqrt{r^2 - r_1^2}}{r_1}\right).$$

The third integral is

$$\int_{r_1}^{r} dr' \frac{Mr_1}{(r'+r_1)\sqrt{r'^2-r_1^2}}.$$

It turns out that the integrand can be written as a total derivative,

$$\frac{r_1}{\sqrt{r^2-r_1^2}(r+r_1)} = \frac{r_1}{\sqrt{r-r_1}(r+r_1)^{3/2}} = \frac{1}{2}\left(\frac{r+r_1}{r-r_1}\right)^{1/2}\left[\frac{r+r_1-(r-r_1)}{(r+r_1)^2}\right] = \frac{d}{dr}\left(\frac{r-r_1}{r+r_1}\right)^{1/2}.$$

Hence

$$\int_{r_1}^{r} dr' \frac{Mr_1}{(r'+r_1)\sqrt{r'^2-r_1^2}} = M\left(\frac{r-r_1}{r+r_1}\right)^{1/2}.$$

The Shapiro delay is

$$t(r,r_1) \simeq \sqrt{r^2-r_1^2} + 2M\ln\left(\frac{r+\sqrt{r^2-r_1^2}}{r_1}\right) + M\left(\frac{r-r_1}{r+r_1}\right)^{1/2}.$$

A.5 Black Holes

A.5.1 Kruskal-Szekeres Metric

The coordinates are defined as

$$u = \left(\frac{r}{2M}-1\right)^{1/2} e^{r/4M}\cosh\left(\frac{t}{4M}\right), \quad v = \left(\frac{r}{2M}-1\right)^{1/2} e^{r/4M}\sinh\left(\frac{t}{4M}\right),$$

for $r > 2M$ and

$$u = \left(1-\frac{r}{2M}\right)^{1/2} e^{r/4M}\sinh\left(\frac{t}{4M}\right), \quad v = \left(1-\frac{r}{2M}\right)^{1/2} e^{r/4M}\cosh\left(\frac{t}{4M}\right),$$

for $r > 2M$. The implicit definition of r and t in terms of u and v are

$$\left(\frac{r}{2M}-1\right)e^{r/2M} = u^2 - v^2, \quad \tanh\left(\frac{t}{4M}\right) = \begin{cases} v/u, & r > 2M, \\ u/v, & r < 2M. \end{cases}$$

Differentiating the r-equation on both sides:

$$\left[\frac{1}{2M} + \left(\frac{r}{2M}-1\right)\frac{1}{2M}\right] e^{r/2M}\, dr = 2u\,du - 2v\,dv,$$

$$dr = \frac{8M^2}{r} e^{-r/2M}(u\,du - v\,dv).$$

This works for all $r \in (0,\infty)$.

The t-equation has two cases, $r > 2M$ and $r < 2M$, but we will find that they produce the same result. For $r > 2M$,

$$\frac{dt}{4M\cosh^2(t/4M)} = \frac{u\,dv - v\,du}{u^2} = \frac{u\,dv - v\,du}{(r/2M)(1-2M/r)e^{r/2M}\cosh^2(t/4M)},$$

$$dt = \frac{8M^2}{r}e^{-r/2M}\left(1-\frac{2M}{r}\right)^{-1}(u\,dv - v\,du).$$

For $r < 2M$,

$$\frac{dt}{4M\cosh^2(t/4M)} = \frac{v\,du - u\,dv}{v^2} = \frac{v\,du - u\,dv}{(r/2M)(2M/r-1)e^{r/2M}\cosh^2(t/4M)},$$

$$dt = \frac{8M^2}{r}e^{-r/2M}\left(1-\frac{2M}{r}\right)^{-1}(u\,dv - v\,du),$$

which is the same as in $r > 2M$. We will focus on the r and t (u and v) part of the metric because the θ and ϕ part are the same as in Schwarzschild coordinates. Computing the line element,

$$
\begin{aligned}
-\left(1-\frac{2M}{r}\right)dt^2+\left(1-\frac{2M}{r}\right)^{-1}dr^2 &= \left(1-\frac{2M}{r}\right)^{-1}\frac{64M^4}{r^2}e^{-r/M}(-u^2dv^2-v^2du^2+2uv\,du\,dv+u^2du^2+v^2dv^2-2uv\,du\,dv)\\
&= \left(1-\frac{2M}{r}\right)^{-1}\frac{64M^4}{r^2}e^{-r/M}[(u^2-v^2)(-dv^2+du^2)]\\
&= \left(1-\frac{2M}{r}\right)^{-1}\frac{64M^4}{r^2}e^{-r/M}\frac{r}{2M}\left(1-\frac{2M}{r}\right)e^{r/2M}(-dv^2+du^2)\\
&= \frac{32M^3}{r}e^{-r/2M}(-dv^2+du^2).
\end{aligned}
$$

A.5.2 Unphysical Regions

To be written.