Modern physics

An overview

John O'Neall

Note to the reader (assuming there is one...).

This document is just an overview of the subject, composed essentially of notes I have made to remind myself of the math behind the main principles of modern physics. I should say, the mathematical and physical concepts which *are* the main principles of modern physics. No attempt to teach anything or to indicate any real calculations of cross sections or whatever. It presents, for me, the mathematics of physics, the *concepts* you need to know in order to understand the principles presented in simpler books, such as for laymen, no more.

*Note: The Minkowski **metric signature** used will be (+1,-1,-1,-1). Among authors cited, only Robinson uses the opposite. See Table 1.*

Table of Contents

1 Classical mechanics....5
1.1 Newton's laws of motion....5
1.2 Energy and conservation....5
1.3 Action, the Lagrangian and the Euler-Lagrange equations....6
1.4 Symmetries and conservation laws....7
1.5 Energy and the Hamiltonian....8
1.6 Poisson brackets....10
1.7 Résumé of classical mechanics....11
2 The laws of thermodynamics....11
2.1 Basics....12
2.2 Legendre transformations....13
2.3 Principle thermodynamic potentials....14
2.4 Spontaneous reactions....16
2.5 Note on entropy....16
3 Special relativity....18
3.1 The principles of Special Relativity....18
3.2 The spacetime diagram and the Lorentz transform....18
3.3 Four-velocity and four-momentum....22
3.4 Relativistic free-particle Lagrangian....24
3.5 Curious results of SR....24
3.6 The Lorentz transformation as a hyperbolic rotation....26
4 Electromagnetism....26
4.1 Maxwell's equations of electromagnetism....27
4.2 The vector and scalar potentials....28
4.3 Gauge transformations....28
4.4 EM four-potentials....29
4.5 Relativistic particle in EM field and the EM field tensor....30
5 Quantum mechanics....32
5.1 Vector and Hilbert spaces....32
5.2 Physical states, observables and operators....32
5.3 Principles of quantum mechanics....33
5.4 Unitarity and evolution in time....35
5.5 Simultaneous observables and uncertainty....38
5.6 Bases and representations....39
5.7 A useful example – the harmonic oscillator (SHO)....41
5.8 Angular momentum in QM....42
5.8.1. Angular momentum....42
5.8.2. Spin....43
5.8.3. Spin ½....44
5.9 Composite systems and entanglement....45
5.9.1. Density operator and entanglement....45
5.9.2. Locality and hidden variables....47
6 Symmetry, groups and quantum field theory....48
7 General relativity....49
7.1 Principles of General Relativity....49
7.2 The equivalence principle and curvature....49
7.3 Curvature and the mathematics of GR....51
7.3.1. Manifolds....52
7.3.2. Tensors....53
7.3.3. Vectors, one-forms (dual vectors) and basis vectors....53
7.3.4. Of curves, surfaces and normals....55
7.3.5. Higher-rank tensors....56
7.3.6. The metric....56

7.3.7. The covariant derivative....57
7.3.8. Parallel transport....59
7.3.9. Geodesics....61
7.3.10. The Riemann curvature tensor and friends....61
7.4 Back to physics....62
7.4.1. Four-velocity and four-momentum....62
7.4.2. The energy-momentum tensor....63
7.4.3. Einstein's gravitational field equations....64
7.4.4. A bit of philosophy....66
7.4.5. Consistency – the equations of motion....66
7.5 What GR tells us....67
7.6 The Schwarzschild metric, stars and black holes....68
7.6.1. Proper time and distance....69
7.6.2. A bit of history....71
7.6.3. Schwarzschild black holes....71
7.7 The cosmological principle....74
7.7.1. Robertson-Walker metric....74
7.7.2. The expanding universe....75
7.7.3. The Friedmann equations....76
7.8 Other models of the universe....77
7.8.1. Empty-universe model....77
7.8.2. Static Einstein model....77
7.8.3. De Sitter model....78
7.8.4. Radiation-only model....79
7.8.5. Einstein-de Sitter model....79
7.8.6. Critical density and density parameters....79
7.8.7. Lambda-CDM model....81
7.8.8. But....81
7.8.9. Gravitational waves....81
7.9 Annex: All you need to solve Einstein's equations....82
8 Bibliography....82

1 Classical mechanics

1.1 Newton's laws of motion

Around 1610, Galileo started working on laws of motion and gravity by measuring the coordinates of rolling or falling objects. Following Galileo, we can say classical mechanics began quantitatively with Newton's three laws of motion in 1687. Here are Susskind's versions of them.

1. Every object in a state of uniform motion tends to remain in that state of motion unless an external force is applied to it.
2. The relation between an object's mass m, its acceleration a, and the applied force F is F=ma.
3. For every action [force] there is an equal and opposite reaction.

These are somewhat brief and can do with some explaining.

The first law, known as the ***law of inertia***, says, in more detail:

> Every object persists in its state of rest or uniform motion in a straight line unless it is compelled to change that state by forces imposed on it.

This is important because if it is true for an observer, such as yourself, then you are in an inertial frame. (For more on that subject, see General Relativity.)

The second law may be expressed mathematically in terms of force F, mass m and acceleration a; or in terms of mass and coordinates q and their time derivatives; or in terms of mass, coordinates and momentum p = mv.

$$F = ma = m\frac{d^2q}{dt^2} = m\ddot{q} = \frac{\dot{p}}{m} \tag{1.1}$$

The third law says that if a charge exerts an electrostatic force on a second charge, then the second charge exerts the negative of the same force on the first. This is often stated in terms of action, where one really means force (and certainly not the action of the principle of extremized Action). One then says the oft-heard statement that action and reaction are equal and opposite.

1.2 Energy and conservation

A ***conservative force*** is one for which the work (integral of force times distance) going from one point to another is independent of the path taken. The important point is that every conservative force can be obtained by differentiating a potential energy V(q).

$$F- = -\frac{dV}{dq}$$

The kinetic energy of a particle is defined as

$$T = \frac{1}{2}mv^2 = \frac{1}{2}m\dot{q}^2. \tag{1.2}$$

Using

$$\frac{dV}{dt} = \frac{dV}{dq}\frac{dq}{dt} = \frac{dV}{dq}v$$

the rate of change of the total energy E = T + V is

$$\frac{dE}{dt} = \dot{E} = \dot{T} + \dot{V} = mva + \frac{dV}{dq}v = m\dot{q}\ddot{q} + \frac{dV}{dq}\dot{q}$$

so

$$\frac{dE}{dt} = \dot{q}\Big(m\ddot{q} + \frac{dV}{dq}\Big).$$

Newton's second law says that the bit in parentheses is zero and so energy is ***conserved*** over time. It follows from the force's being conservative and so, expressed as the negative derivative of a potential. Later, we we will see that it is due to symmetry under a time translation of the system studied.

1.3 Action, the Lagrangian and the Euler-Lagrange equations

Next to come was a method of attacking mechanics problems not through coordinates, but through energies. These ideas were developed during the mid-eighteenth century, principally by Euler and Lagrange.

The ***action*** is defined as

$$A = \int_{t_0}^{t_1} L dt \tag{1.3}$$

where the ***Lagrangian*** is most simply (at least in classical physics) defined as

$$\boxed{L = T - V}, \tag{1.4}$$

the difference of the kinetic and potential energies of the system in question. The ***principle of stationary*** (often, least) ***action*** then says:

$$\delta A = \delta \int_{t_0}^{t_1} L(q_i, \dot{q}_i) dt = 0 \tag{1.5}$$

The use of incremental differences and the product rule then allow derivation of the ***Euler-Lagrange equations***

$$\boxed{\frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{q}_i}\Big) = \frac{\partial L}{\partial q_i}} \tag{1.6}$$

Susskind calls this "...all of classical physics in a nutshell."[1]

From Newton's second law (1.1) the simple formula for kinetic energy (1.2), we can identify

$$\boxed{p_i = \frac{\partial L}{\partial \dot{q}_i}} \tag{1.7}$$

as the ***conjugate momentum*** to q_i, or ***canonical momentum conjugate to the coordinate*** q_i, so that the Euler-Lagrange equation becomes

$$\frac{dp_i}{dt} = \frac{\partial L}{\partial q_i}. \tag{1.8}$$

In the case of kinetic energy depending only on the squares of the velocities and potential energy on the coordinates, this is easily seen to be ***Newton's second law***.

1 Susskind, 123.

1.4 Symmetries and conservation laws

Susskind defines a ***symmetry*** as "... an ***active*** coordinate transformation that does not change the value of the Lagrangian." By "active", he means the particle in question is, for instance, translated from one place to another, rather than just changing the coordinates. The particle moves, not the Earth around it and therefore not any potential energy not due to the particle. Repeat: The potential energy V(q) seen by the particle may change. This is an active transformation, as opposed to a passive one, where one simply changes the coordinates. It is assumed to happen at all times, meaning it has no dependence on time.

Klauber says that symmetry is "… the propensity for non-change with change, for invariance under transformation." It is "… a relationship between the whole and the parts in which the whole can exhibit changelessness while the component parts change."[2]

Consider an arbitrary translation, a transformation depending on an infinitesimal quantity δ and such that the transformation depends on all the coordinates.

$$\delta q_i = f_i(q)\delta \tag{1.9}$$

The existence of δ assures that the coordinate differences are infinitesimal. Otherwise, it is a bookkeeping device. The $f_i(q)$ do not depend on time, so velocities do not change. Differentiation says that

$$\delta L = \sum_i \Big(\frac{\partial L}{\partial \dot{q}_i} \delta \dot{q}_i + \frac{\partial L}{\partial q_i} \delta q_i \Big). \tag{1.10}$$

(All this business of infinitesimal variations forms the subject matter of the ***calculus of variations***.) Assuming that the equations of motion are satisfied, we can insert the momentum canonical to q_i from (1.7) and use the Euler-Lagrange equation (1.8), in which case this becomes

$$\delta L = \sum_i \Big(p_i \delta \dot{q}_i + \dot{p}_i \delta q_i \Big) = \frac{d}{dt} \sum_i p_i \delta q_i = \frac{d}{dt} \sum_i p_i f_i(q) \delta$$

If the transformation (1.9) represents a symmetry, then this δL is zero and the quantity

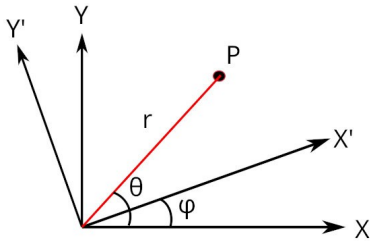
$$Q = \sum_i p_i f_i(q) \tag{1.11}$$

is ***conserved*** over time. This is an example of ***Noether's theorem*** (published in 1918), which itself is much more general. Note these important assumptions/conditions:

- We are in a context of least action, so $\delta L = 0$..
- The $f_i(q)$ are time-independent.
- The transformation is continuous, made up of infinitesimal transformations, so that the first-order changes in the Lagrangian are zero. Reflection would be a discrete transformation and so not an example of this type of symmetry.

Consider simple 2d rotation of angle ϕ (equivalent to the x-y plane under a rotation about the z-axis).

2 Klauber, 162.



The original coordinates are

$$x = rcos\theta,\ y = rsin\theta.$$

After rotation by ϕ,

$$x' = rcos(\theta - \phi) = r(cos\phi cos\theta + sin\phi sin\theta) = xcos\phi + ysin\phi$$

$$y' = rsin(\theta - \phi) = r(sin\theta cos\phi - cos\theta sin\phi) = ycos\phi - xsin\phi,$$

or in matrix notation

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} cos\phi & sin\phi \\ -sin\phi & cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{1.12}$$

The signs of the sine functions are inverted if the rotation angle ϕ is greater than the original polar coordinate θ.

Now, consider this example.

$$x \to x + y\delta$$

$$y \to y - x\delta$$

This is equivalent to an infinitesimal rotation through angle δ, as can be seen from (1.12)

$$x \to x\ cos\delta + y\ sin\delta \approx x + y\delta$$

$$y \to -x\ sin\delta + y\ cos\delta \approx -x\delta + y.$$

The Lagrangian is unchanged to first order in δ so, from (1.11), the following quantity is conserved.

$$Q = p_x y - p_y x$$

which is the (negative) angular momentum of this 2-d particle. So rotational symmetry means angular momentum is conserved. We will look more at this subject when we consider quantum field theory (QFT), in another document.

1.5 Energy and the Hamiltonian

The next step in the use of energy was taken by William Rowan Hamilton around 1830.

By the same method used to obtain equation (1.11) from equation (1.10), one can derive

$$\frac{dL}{dt} = \frac{d}{dt}\sum_i (p_i \dot{q}_i) + \frac{\partial L}{\partial t} \tag{1.13}$$

which says among other things that the Lagrangian is not conserved over time.

A different way of approaching the subject is to define the ***Hamiltonian*** by

$$\boxed{H = \sum_i (p_i \dot{q}_i) - L}. \tag{1.14}$$

From section (2.2), one can see this is a Legendre transform of the Lagrangian with respect to $\dot{q}_i$, where we have taken the negative of the transform in order to have positive energy. Then from (1.13) we have

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t}. \tag{1.15}$$

So H varies in time only if the Lagrangian has a specific time dependence. Equivalently, if a system (Lagrangian) is invariant under time translation, then H is conserved.

In the simple case where the Lagrangian is

$$L = \frac{m}{2}\dot{q}^2 - V(q), \tag{1.16}$$

with canonical momentum $p = m\dot{q}$, the Hamiltonian is

$$H = (m\dot{q})\dot{q} - \frac{m}{2}\dot{q}^2 + V(q) = \frac{m}{2}\dot{q}^2 + V(q) = T + V,$$

the total energy, the sum of the kinetic and potential energies. This is only true if the Lagrangian has the simple form in equation (1.16). Nevertheless, the Hamiltonian is always the energy. It is also the basis for the Hamiltonian formulation of mechanics – and, especially – non-relativistic quantum mechanics.

The Lagrangian formulation relies on a notion of configuration space, here represented by the set of position coordinates q. The Hamiltonian formulation focuses on ***phase space***, including the q and p variables so that a state of the system represents a specific point in the space. Any time dependence comes through them.

$$H = H(q_i(t), p_i(t)), \tag{1.17}$$

so the simple formula becomes

$$H = \frac{p^2}{2m} + V(q).$$

Then

$$F = -\frac{dV}{dx} = ma = \dot{p}$$

implies that

$$\boxed{\dot{p}_i = -\frac{\partial H}{\partial q_i}} \tag{1.18}$$

and

$$\boxed{\dot{q}_i = \frac{\partial H}{\partial p_i}}, \tag{1.19}$$

as will now be proven more rigorously. These last two are ***Hamilton's equations***. More rigorously, without supposing the simple form (1.16) for the Lagrangian, expand equation (1.14) to

$$\delta H = \sum_i \left(p_i \delta\dot{q}_i + \dot{q}_i \delta p_i - \frac{\partial L}{\partial q_i}\delta q_i - \frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i \right)$$

where by equation (1.8) the first and last terms cancel. Comparing the result to a standard derivative

$$\delta H(q,p) = \sum_i \Big(\frac{\partial H}{\partial p_i}\delta p_i + \frac{\partial H}{\partial q_i}\delta q_i \Big)$$

and using equation (1.8) leads to Hamilton's equations.

1.6 Poisson brackets

For a function F(q,p) in phase space, q and p are generally functions of t, so

$$\dot{F} = \sum_i \Big(\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial p_i}\dot{p}_i \Big) \tag{1.20}$$

From Hamilton's equations (1.18) and (1.19), this becomes

$$\dot{F} = \sum_i \Big(\frac{\partial F}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q_i} \Big) := \{F, H\} \tag{1.21}$$

which defines the ***Poisson brackets***, or PBs. It also says that the time derivative of any object is the PB of the object and the Hamltonian. If F depends directly on t, then

$$\dot{F} = \{F, H\} + \frac{\partial F}{\partial t}. \tag{1.22}$$

PBs are antisymmetric

$$\{F, H\} = -\{H, F\}$$

and linear and can be combined through

$$\{(AB), C\} = B\{A, C\} + A\{B, C\}.$$

Specifically,

$$\{q_i, q_j\} = 0 \tag{1.23}$$

and

$$\{p_i, p_j\} = 0 \tag{1.24}$$

for all I and j, and

$$\{q_i, p_j\} = \delta_{ij} \tag{1.25}$$

Since

$$\dot{F} = \{F, H\} \tag{1.26}$$

Hamilton's equations in Poisson-bracket form are

$$\dot{q}_k = \{q_k, H\} \tag{1.27}$$

and

$$\dot{p}_k = \{p_k, H\} \tag{1.28}$$

One can use power series and linearity, along with

$$\{q^n, p\} = \frac{\partial |q^n|}{\partial q} \tag{1.29}$$

to show that for a continuous, smooth function, F

$$\{F(q,p), p_i\} = \frac{\partial F(q,p)}{\partial q_i} \tag{1.30}$$

and

$$\{F(q,p), q_i\} = -\frac{\partial F(q,p)}{\partial p_i} \tag{1.31}$$

PBs constitute a link with quantum mechanics, where a look forward to equation (5.18) shows that the expectation value of an operator L is given by

$$\frac{d\langle L\rangle}{dt} = -\frac{i}{\hbar}\langle [L,H]\rangle$$

from which we will derive a correspondence between PBs and QM commutators:

$$[F,H] \Leftrightarrow i\hbar\{F,H\}. \tag{1.32}$$

1.7 Résumé of classical mechanics

It is time for a short résumé:

- The principle of extremized action (1.5) led to the definition of the momentum conjugate to a coordinate (1.7) and to the Euler-Lagrange equations (1.6).
- Definition of the Hamiltonian (1.14) plus simple differentiation, conjugate momentum and the Euler-Lagrange equation led to Hamilton's equations (1.18) and (1.19).
- Hamilton's equations plus differentiation led to the definition of Poisson brackets (1.21) and then to the time derivative (1.26).

That's it!

2 The laws[3] of thermodynamics

Thermodynamics is the science of energy transfer in the form of heat. Usually, we only talk about ***equilibrium thermodynamics***, meaning that changes in the observed system are slow and continuous. The system passes from one equilibrium state to another infinitesimally close equilibrium state, without going first through a state of non-equilibrium. (Non-equilibrium thermodynamics is a whole other subject.) There are four laws, usually given in the following order.

1. The energy of the universe does not change; it is always conserved.
2. In a physical process, the entropy of the universe always increases.
3. The entropy of a system at a temperature of absolute zero[4] is zero.

The last one is the zeroth:

- If A is a system in thermal equilibrium with the system B, and B is in thermal equilibrium with the system C, then the systems A and C are also in thermal equilibrium.

3 In English, one always refers to these as laws – physical laws – whereas the French, ever alert to a nuance, refer to them as principles.

4 Absolute zero is 0 Kelvin, or -273.15° Celsius.

2.1 Basics

Thermodynamics is the general branch of physics which describes static states of macroscopic systems. By "static", we mean equilibrium states. It is mainly concerned with forms of energy, especially heat. The word heat indicates a transfer of energy, not just energy *per se*. Only differences of energy are measured, and for any state, its internal energy U is defined relative to some standard or base state, a fiducial state,

Ignoring electromagnetic or atomic or nuclear effects, the internal energy of a system can be written as

$$U(V, N_1, N_2, \ldots) \tag{2.1}$$

where V is its volume and the N_i are the mole numbers of its constituent chemical components. All these parameters, including the energy U, are ***extensive parameters***, meaning that if two subsystems are combined, the resultant parameters are the sums of those of the subsystems.

The energy of a system can be increased mechanically or by adding heat energy.

$$dU = dQ + dW_M \tag{2.2}$$

with

$$dW_M = -PdV \tag{2.3}$$

where P is the pressure and V is the volume of the system. The term -pdV is the work done on the system by a volume change; its sign is chosen so as to be positive if the energy of the system is increased. Q is the thermal energy transferred to the system.

We then define the ***entropy*** S of a system as an extensive parameter defined for equilibrium states. "The values assumed by the extensive parameters in the absence of an internal constraint are those that maximize the entropy over the manifold (continuous set) of constrained equilibrium states."[5] So S must be added to the equation of state.

$$U = U(S, V, N_1, \ldots N_r). \tag{2.4}$$

The entropy is defined to vanish at absolute zero:

$$\left(\frac{\partial U}{\partial S}\right)_{V,N_1,\ldots,N_r} = 0, \text{ for T=0.}^{6} \tag{2.5}$$

The following ***intensive parameters*** (non-addable) are defined:[7]

$$\left(\frac{\partial U}{\partial S}\right)_{V,N_1,\ldots,N_r} := T \tag{2.6}$$

$$-\left(\frac{\partial U}{\partial V}\right)_{S,N_1,\ldots,N_r} := P \tag{2.7}$$

$$\left(\frac{\partial U}{\partial N_j}\right)_{V,N_1,\ldots,N_k,\ldots} := \mu_j \tag{2.8}$$

where T is the temperature, P the pressure and μ_j the chemical potential of the j^{th} component.

An equilibrium state is then described by one of several principles. From Callan, we have two equivalent principles of thermodynamics:[8]

5 Callan, 27.
6 Callan, 30. I do NOT understand this one. *Peu importe*?
7 Callan, 35.
8 Callan, 133.

1. “**Entropy maximum principle**. The equilibrium value of any unconstrained internal parameter is such as to maximize the entropy for the given value of the total internal energy.”
2. **“Energy minimum principle**. The equilibrium value of any unconstrained internal parameter is such as to minimize the energy for the given value of the total entropy.”

These two principles are equivalent.

The basic formula for the internal energy U as a function of entropy S, temperature T, pressure p and volume V (and the chemical potentials, N_i) is

$$U = TS - pV + \mu N \tag{2.9}$$

where μN stands for the sum over different chemical potentials

$$\sum_i \mu_i N_i$$

representing the quasi-static chemical work and

$$dU = TdS - pdV + \mu dN \tag{2.10}$$

The independent variables here are the extensive (additive) ones, S, V and N.

Experimental conditions vary. They may occur, like most chemical experiments, at constant (atmospheric) pressure. Others may be restricted to a container at constant volume or may be in contact with heat sinks and sources to maintain constant temperature. Depending on the problem at hand, It may be simpler to solve for the equilibrium state if the formula had different independent variables, among the intensive ones. That is what Legendre transformations allow us to do.

2.2 Legendre transformations

Given a function $f(x, y)$ of two independent variables x and y, we might like to look at it in terms of other independent variables, such as derivatives. Since

$$df(x, y) = \frac{df}{dx}dx + \frac{df}{dy}dy := udx + wdy,$$

we can define a new function, the ***Legendre transform***,

$$g = f - ux = f - \frac{\partial f}{\partial x}x.$$

Then, using differentiation by parts,

$$dg = d(f - ux) = udx + wdy - udx - xdu = wdy - xdu,$$

and we see that g is function of independent variables x and w.

> In order to eliminate x as a variable, we subtract x multiplied by the derivative of the original function with respect to x.

Consider the Lagrangian,

$$L = L(q, \dot{q}).$$

Doing just a reversal of the sign, the Legendre transform is then

$$H = \frac{\partial L}{\partial \dot{q}}\dot{q} - L = p\dot{q} - L, \tag{2.11}$$

where p is the canonical momentum and H is the Hamiltonian, which we now see to be a simple Legendre transform of the Lagrangian.

This can be visualized graphically by considering a function

$$Y = Y(X) \tag{2.12}$$

where

$$P = \frac{dY}{dX} \tag{2.13}$$

is the slope of the curve at a given point. We can pass from a point representation of the curve Y in terms of X, as in (2.12), to one in which the curve is represented as the envelope of tangent lines (2.13). This will effectively change the independent variable from X to P. Just knowing the slope is not sufficient to fix the position of a line, so in order to do this, we can use the X-intercept ψ on the Y axis of a tangent of constant slope at X. Then the curve may be represented by the set of slopes and intercepts just as well as by the formula Y(X), without any loss of information.

The slope is given by

$$P = \frac{Y - \psi}{X - 0}$$

which can be arranged to give the **Legendre transformation**

$$\psi = Y - \frac{dY}{dX}X = Y - PX.$$

Then

$$d\psi = dY - PdX - XdP = -XdP,$$

since $dY = PdX$. Then

$$-X = \frac{d\psi}{dP}$$

shows that the inverse transformation is just

$$Y = XP + \psi.$$

In general, for

$$Y = Y(X_o, X_2, \ldots, X_i)$$

the partial slope of this hypersurface is given by

$$P_k = \frac{\partial Y}{\partial X_k}$$

and the Legendre transformation is

$$\psi = Y - \sum_k P_k X_k$$

This is not only useful for thermodynamic potentials but also for switching between the Lagrangian and the Hamiltonian in mechanics, as we have seen (2.11).

2.3 Principle thermodynamic potentials

The most used thermodynamic potentials are indicated in the following table and are all Legendre transforms of the internal energy U as shown.

We can identify each term with a type of energy:

- TS with heat energy,
- pV with work done on the system, due to volume-change energy (as when a piston is pushed) and
- the sum $\sum_i \mu_i N_i$ as the energy of chemical bonding.

We will ignore such things as electromagnetic or atomic energy.

Name	Symbol	Formula	Natural variables	Variable change	Physical interpretation (infinitesimal, quasi-equilibrium changes)
Internal energy	U	$TS - pV + \sum_i \mu_i N_i$ $dU = TdS - pdV$	S, V, {N_i}	-	Total internal energy of system
Helmholtz free energy	F (or A)	$F = U - TS$ $dF = -SdT - pdV...$	T, V, {N_i}	S → T	Max usable energy at constant T & V after dispersal of energy (heat) measured by entropy ("entropy tax")
Enthalpy	H	$U + pV$ $dH = TdS + Vdp...$	S, p, {N_i}	V → p	Total heat energy available in system at constant p after V change
Gibbs free energy	G	$U + pV - TS$ $dG = -SdT + Vdp...$	T, p, {N_i}	S → T and V → p	Max energy available at constant T & p after V change and "S tax"

The above-cited energy-minimum principle applies to all these potentials under specific circumstances.

> **Helmholtz Potential Minimum Principle**. "The equilibrium value of any unconstrained internal parameter in a system in diathermal contact with a heat reservoir minimizes the Helmholtz potential over the manifold of states for which T = T^r [at constant T]."[9]
>
> **Enthalpy Minimum Principle**. "The equilibrium value of any unconstrained internal parameter in a system in contact with a pressure reservoir minimizes the enthalpy over the manifold of states of constant pressure (equal to that of the pressure reservoir)."[10]
>
> **Gibbs Potential Minimum Principle**. "The equilibrium value of any unconstrained internal parameter in a system in contact with a thermal and a pressure reservoir minimizes the Gibbs potential at constant temperature and pressure (equal to those of the respective reservoirs)."[11]

We can attribute physical meanings to these potentials.

In the case of ***enthalpy***, a system at constant P can expand or contract, therefore losing energy to move around molecules. This volume-change energy is exactly what is subtracted from U in order to derive H. So enthalpy is the energy released as heat by the system at constant P. Or, in the other direction,

> ...heat added to a system at constant pressure and at constant values of all the remaining extensive parameters (other than S and V) appears as an increase in the enthalpy.[12]

As for the ***Helmholtz free energy***,

> … the work delivered in a reversible process, by a system in contact with a thermal reservoir [so at constant T], is equal to the decrease in the Helmholtz potential of the system.

9 Callan, 155.
10 Callan, 156.
11 This and the following definitions are all from Callan, 157, 161, 159, 167.
12 Callan, 161.

The Helmholtz free energy is thus the available work at constant temperature and volume. The term TS represents the “entropy tax” which the system must pay in order for the total entropy of the universe to remain zero.

The ***Gibbs free energy*** is useful when T and P are constant, which is the usual case of chemical reactions open to the atmosphere, which acts as a reservoir of both T and P. Gibbs free energy is thus much beloved of chemists. As Callan points out, it may also be true “… in a small subsystem of a larger system that acts as both a thermal and a pressure reservoir (as in the fermentation of a grape in a large wine vat).”[13]

2.4 Spontaneous reactions

Suppose there is only one thermodynamic system in the universe and that its T and V are constant. Then the sum of the entropy changes of the system and of its surroundings is the entropy change of the universe.

$$\Delta S_{univ} = \Delta S_{surr} + \Delta S_{sys} \tag{2.14}$$

If the energy of the system changes by an amount ΔU, then the surroundings will change by -ΔU, and

$$\Delta S_{surr} = \frac{-\Delta U}{T} \tag{2.15}$$

so

$$\Delta S_{univ} = \Delta S_{sys} - \frac{\Delta U}{T}$$

which means

$$\Delta U_{univ} = T\Delta S_{univ} = -(\Delta U - T\Delta S_{sys}) = -\Delta F_{sys}.$$

So the entropy maximum principle says that for constant T and V, the change in the Helmholtz free energy must be negative, in order that

$$\Delta S_{univ} = \frac{\Delta U_{univ}}{T} > 0.$$

Similarly, if only a spontaneous chemical reaction occurs at constant T and P, then a change in heat energy (enthalpy) of the system means the surroundings change by the negative of that., i.e.,$-\Delta H_{sys}$. So equation (1) gives

$$\Delta S_{univ} = -\frac{-\Delta H_{sys}}{T} + \Delta S_{sys}$$

The only change in the universe is due to the dispersal of energy in a quantity we may call $-\Delta G$. Therefore

$$\Delta S_{univ} = -\frac{\Delta G}{T} = -\frac{-\Delta H_{sys}}{T} + \Delta S_{sys}$$

which may be rewritten as

$$\Delta G = \Delta H_{sys} - T\Delta S_{sys} \tag{2.16}$$

which is the ***Gibbs free energy***.

2.5 Note on entropy

Entropy is generally stated to be related to the number of states available to a system in statistical mechanics by the Boltzmann entropy formula S = k ln(W), where k is the Boltzmann constant and W (for

13 Callan, 167,

Wahrscheinlichkeit) is the probability of the system, essentially the number of indistinguishable arrangements of its molecules. This model only works when chemical reactions do not take place, which they don't in the equilibrium thermodynamics we usually consider. Consider adding these two types of molecules:

$2H_2 + O_2 \rightarrow \quad 2H_2O$ + lots of energy.

The final molecules display less randomness than the first. So if we want to go beyond equilibrium thermodynamics, a better definition is Lambert's:

> "*Entropy change is the measure of how more widely a specific quantity of molecular energy is dispersed in a process, whether isothermal gas expansion, gas or liquid mixing, reversible heating and phase change or chemical reactions, as shown by the Gibbs free energy equation/T*." (The last means G/T.)[14]

This definition goes along with his statement of the second law as:

> "*Energy of all types changes from being localized to becoming more spread out, dispersed in space if that energy is not constrained from doing so*."

For more on this subject, see Lambert's web site.

14 Entropy sites, a guide. http://entropysite.oxy.edu/#intro.

3 Special relativity

3.1 The principles of Special Relativity

Special relativity (SR) is based on two principles:

1. The principle of relativity states that the laws of nature should be the same for all observers in inertial frames (defined below).
2. All such observers, upon measuring the speed of light in a vacuum, will find the same result, c = 299,792,458 m/sec. The speed of light is a *universal* constant.

The first requirement is necessary in order for physics to be coherent. It means that observers in inertial systems use the same equations. Rather than going on incessantly repeating "in an inertial reference frame", let's get it done with once and for all by stating:

> SR considers only observers in ***inertial*** reference frames, those in which Newton's first law holds. Different inertial frames move with constant (unaccelerated) velocity relative to one another.

This does not mean that SR is incapable of handling acceleration. The first requirement may be stated differently:

> No experiment can measure the absolute velocity of an observer; the results of any experiment performed by an observer do not depend on his speed relative to other observers who are not involved in the experiment.[15]

The second requirement is a result of rather astounding experimental results. There is no known reason for it, that's just how it is. Because of the constant speed of light, which is of course the spatial distance it travels in a given time interval, SR links space and time into a more global entity, ***spacetime***. In this way, it prevents us from considering space and time as being two separate things and explains how they are related. More on that in a moment.

3.2 The spacetime diagram and the Lorentz transform

We use the notion of a ***four-vector*** in tensor notation:[16]

X = (t, x, y, z) := (x^0, x^1, x^2, x^3)

or

$$X^\mu = (x^0, x^1, x^2, x^3).$$

We also use standard relativist units, where c = 1, so that

c = 3x10^8 m/sec = 1

meaning that time is measured such that

1 sec = 3x10^8 m.

In SR, we speak of spacetime as the total of space and time, which while different are no longer distinct. An ***event*** is something which happens instantaneously at a specific value of time and space, i.e., t and $\vec{x}$. An event may be simply a set of values of the four spacetime variables. An inertial observer is just a spacetime coordinate system which records the coordinates of any event. His space is Euclidean (in SR), distances he

15 Schutz, 1.
16 More on that in my GR overview.

measures between given fixed (to him) points are independent of time and his clocks are synchronized and run together.

The essential geometry of SR is shown in Figure 1. For simplicity, we ignore y and z. The black axes show the x and t variables of observer O; the blue, of observer O', whom we suppose is moving with velocity v along the x-direction. The path of the origin of O' as seen by O is the blue line labeled t'. This is because the origin of O' is always at $x' = 0$ and so this path is his time axis. From the diagram, the tangent of the upper angle, ϕ, is given by

$$tan\phi = \frac{dx}{dt} = v.$$

Any light wave on such a diagram must move with constant velocity, c = 1, and so is always represented by a 45° line (since we chose c=1). Imagine a light wave in the O' system starting at x=0 and t = -a, so it will move upward at an angle of 45°. Suppose there is a mirror to reflect it at the x-axis where x=a also (45°, remember) so it will move upwards, always at 45° and reach x=0 again at t=+a. Transfer this to the system of Figure 1 with the rays still at 45° and passing through t'=-a (arbitrary) and t'=+a, with both lines' slopes and intercepts on x' and t' well determined: Their point of intersection will define a point on the x' axis. Since we assume the particle started at x=t=0, we have the two points needed to define a straight line, the blue line making an equal angle ϕ with the x-axis of observer O and which is the x' axis.[17]

A point P in space is therefore represented in O's reference frame by the intersections of the fine-dotted lines with the x and t axes. They are obviously normal to t and x. The coordinates of O' are a bit more complicated: Since lines parallel to x' are lines of constant t', the upper dashed line intersects the t' axis at the value t' of P in the O' frame. Similarly, the lower dashed line intersects the x' axis at the value of x' for the point P.

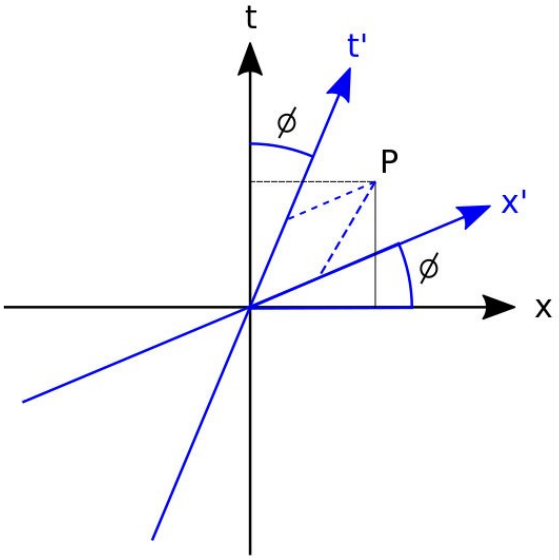


Figure 1: Geometry of SR – the spacetime diagram (after Collier)

We should note two things about this diagram. First, since the x-axis is orthogonal to the t-axis, events on it, except at the origin, are space-like separated. Secondly, imagine an event between the x and x' axes. Assuming it is within the light cone, it is in the future for x but in the past for x', so obviously not at the same time for both. Simultaneity, as we will see further, can no longer be defined.

The fact that the speed of light is constant for all observers leads (by means of several sorts of derivations) to the ***Lorentz transformation***s between the reference frames of the two observers:

17 Schutz, 6-8.

$$t' = \frac{t - vx}{\sqrt{1 - v^2}} = \gamma(t - vx) \tag{3.1}$$

$$x' = \frac{x - vt}{\sqrt{1 - v^2}} = \gamma(x - vt)$$

$$y' = y$$

$$z' = z,$$

This is often expressed as a transformation matrix:

$$(\Lambda^{\bar{\beta}}{}_{\alpha}) = \begin{pmatrix} \gamma & -v\gamma & 0 & 0 \\ -v\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.2},$$

where we use a bar instead of a prime to designate the second system, and

$$\gamma = (1 - v^2)^{-1/2} \tag{3.3}$$

More precisely, this is a Lorentz ***boost***, where one frame moves with a constant velocity relative to another. It may also differ by spatial rotation. This point of view leads to the idea of the Lorentz transformation as a rotation including time, as explained in Section 3.6. Whereas a simple rotation changes and mixes up the space coordinates, so does the boost change and mix up the space and time coordinates.

The Lorentz transformations (3.1) are easily derived from the two principles and the requirement of inertial frames.[18] Consider a reference frame (RF) $S'(x', y', z', t'')$ which is moving with velocity v along the x-axis frame $S(x, y, z, t)$. The absence of acceleration says that the transformation requires a linear transformation, so we can write

$$x' = b_1 x + b_2 t \quad \text{and} \quad t\prime = a_1 x + b_2 t,$$

since we know Lorentz transformations mix time and space. Considering the movement of the origin of S' and requiring the speed of light to be a constant

$$c = \frac{x}{t} = \frac{x'}{t'}$$

leads easily to equations (3.1).

Then the transformation formula for a four-vector is

$$A^{\bar{\beta}} = \Lambda^{\bar{\beta}}{}_{\alpha} A^{\alpha} \tag{3.4},$$

with $A^{\bar{\beta}}$ and A^{α} expressed as a column vectors. A four-vector is a set of coordinates (displacements) which transform by the Lorentz transformation.[19]

Using the Lorentz transform, one can derive somewhat laboriously all the following.

The defined ***interval*** between two events separated by the displacements $(\Delta t, \Delta \vec{x})$

$$\Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 = \Delta t^2 - \Delta \vec{x}^2 \tag{3.5}$$

is an invariant quantity across inertial systems. It can be expressed in terms of a metric tensor which is diagonal and has eigenvalues (1,-1,-1,-1) as

18 See, for instance, Collier, 128-130.

19 In GR, a vector is a set of components and *not* a thing which points from one point to another, since that would be impossible in a curved space.

$$\Delta s^2 = \eta_{\mu\nu}\Delta x^\mu \Delta x^\nu.$$

A vector space of this kind is called ***Minkowski space***. Look out, though, some authors (usually GR people) use the inverse of this metric, which can multiply one side of some equations by -1. See Table 1.

The ***proper time*** is also an invariant:

$$\Delta\tau^2 = \Delta s^2 = \Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2 = \Delta t^2 - \Delta\vec{x}^2 \quad (3.6).$$

(With the opposite metric, $\Delta\tau^2 = -\Delta s^2$!) It is clear from the definition that the proper time is the time measured by a clock at rest with respect to the particle or event, since then $\Delta\tau^2 = \Delta t^2$,

Author(s)	metric signature	most-used γ matrix basis
Barr et al.	+---	both
Blundell & Lancaster	+---	chiral
Carroll	-+++	
Griffiths	+---	Dirac
Jeevanjee	+++- (!)	NA
Klauber	+---	Dirac
Maggiore	+---	chiral
Peskin & Schroeder	+---	chiral
Robinson	-+++	chiral
Schutz	-+++	
Schwartz	-+++	chiral
Schwichtenberg	+---	chiral
Susskind[20]	-+++	NA
Zee	?	both

Table 1. Comparison of metric signatures and γ bases by author.

There exists in any frame a set of orthonormal ***basis vectors***

$$\begin{pmatrix}\vec{e}_0\\ \vec{e}_1\\ \vec{e}_2\\ \vec{e}_3\end{pmatrix} = \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} \quad (3.7)$$

so that a four-vector may be expressed as

$$\vec{A} = A^\alpha \vec{e}_\alpha \quad (3.8)$$

where $\alpha = 0, 1, 2, 3$ and is summed over. Such a four-vector transforms in spacetime by Lorentz transforms. Although it represents the same object, its components are different in different frames. The equality

$$\vec{A} = A^\alpha \vec{e}_\alpha = A^{\bar{\beta}} \vec{e}_{\bar{\beta}}$$

can be used to show that the basis vectors transform as

$$\vec{e}_\alpha = \Lambda^{\bar{\beta}}{}_\alpha \vec{e}_{\bar{\beta}} \quad (3.9).$$

Comparing this to

$$A^\alpha = \Lambda^\alpha{}_{\bar{\beta}} A^{\bar{\beta}}$$

20 Susskind claims (-1,1,1,1) is generally used in GR and he seems to be right. It is also used by Carroll, Schutz and Hartle, but not by Lambourne or Collier.

shows that the transformation matrix for basis vectors is the inverse of that for vectors. Such objects are not vectors, but are called ***one-forms***, or ***dual vectors***.[21]

We can diagram the relation of space and time as in Figure 2.

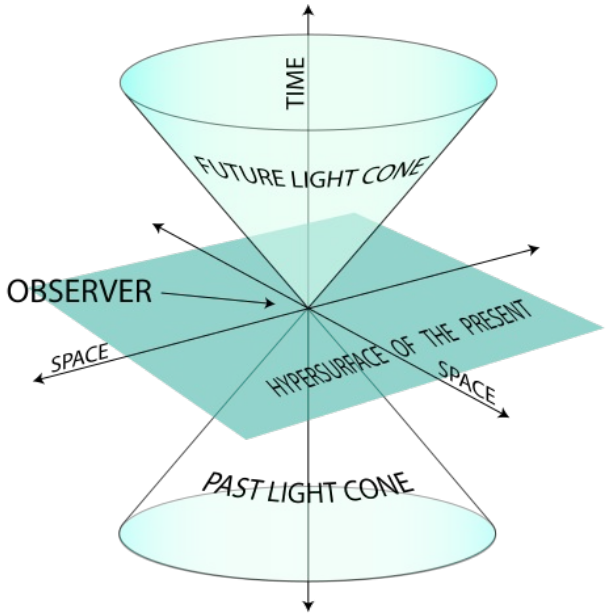


Figure 2: Light cones (by K. Aainsqatsi via Wikimedia Commons[22])

Light rays satisfy $\Delta s^2 = \Delta\tau^2 = 0$ and so move at 45° angles to the t-axis and x-axis (and y-axis) on a spacetime diagram, extended to represent two spatial dimensions. Inside the cone along the t-axis, $\Delta\tau^2 > 0$, so an event at the origin has time to influence any point within the cone without going faster than the speed of life. The interior of the cone, for positive or negative t, is therefore referred to as ***timelike***. Outside the cones, this is not possible, so these regions are called ***space-like***.. The cone for t>0 is the ***future*** relative to the origin, the cone for t<0, the ***past***.

3.3 Four-velocity and four-momentum

The ***four-velocity***, $\vec{U}$, of a particle may be defined as a vector tangent to its ***world line***, the path it traces in space as time passes, and one time unit long in its rest frame. Observer O sees the origin of O' move along its world line in his frame. In the frame of O', this origin is at $\vec{x} = 0$, so its only non-zero component is along the basis vector $\vec{e}_0$. So the particle's four-velocity is $\vec{e}_0$, the unit vector along its time axis in its inertial rest frame, O'.

From another point of view, we want to define four-velocity as a derivative, as in classical physics. But instead of dt, an invariant quantity, $d\tau$, is used, so the components of the four-velocity $\vec{U}$ are given by

$$U^\alpha = \frac{dx^\alpha}{d\tau} \tag{3.10}$$

which reduces to $\vec{e}_0$ in the particle's rest (proper) frame. Look at the time component of $\vec{U}$.

$$U^0 = \frac{dt}{d\tau} = \frac{1}{d\tau/dt}.$$

21 See the GR overview.
22 https://commons.wikimedia.org/wiki/File:World_line.svg.

But

$$\frac{d\tau}{dt} = \frac{\sqrt{dt^2 - d\vec{x}^2}}{dt} = \sqrt{1 - v^2} := \frac{1}{\gamma}.$$

Then the spatial components are

$$U^i = \frac{dx^i}{d\tau} = \frac{dt}{d\tau}\frac{dx^i}{dt} = \gamma V^i,$$

where V^I is the ordinary velocity. So finally

$$\vec{U} = (\gamma, \gamma\vec{V}). \qquad (3.11).$$

Note that

$$U^2 = \vec{U} \cdot \vec{U} = \gamma^2(1 - v^2) = 1 \qquad (3.12),$$

by the definition of γ. This is also an invariant.

The ***four-momentum***, $\vec{p}$, is defined, similarly to classical mechanics, by

$$\vec{p} = m\vec{U}, \qquad (3.13)$$

where m is the particle's mass in its own rest frame. So

$$\vec{p} = m\gamma(1, \vec{V}) \qquad (3.14)$$

and

$$\vec{p} \cdot \vec{p} = m^2\vec{U} \cdot \vec{U} = m^2$$

so $p^0 = m\gamma$ is the particle's energy and $\gamma\vec{V}$ its three-momentum. Also,

$$\vec{p} \cdot \vec{p} = E^2 - p^2 = m^2,$$

where p^2 refers to the three-momentum, and so

$$E^2 = m^2 + p^2. \qquad (3.15)$$

Such a relation of of energy and momentum is called a ***dispersion relation***. This formula is good even for massless particles like photons. SR *unites* energy and momentum, the former being the time-like component of the latter.

Note that when v << c, the energy

$$E = p^0 = m(1 - v^2)^{-1/2} \approx m + \frac{1}{2}mv^2,$$

the rest-mass energy plus the kinetic energy of classical mechanics.

If particle p' moves in the frame of p with velocity v along the x-axis, and a particle is moving with velocity W along the x'-axis of p' , then its velocity in p's frame is given by the ***Einstein law of composition of velocities***:

$$w = \frac{u + v}{1 + uv} \qquad (3.16)$$

which can never exceed the speed of light (1, in this system).

3.4 Relativistic free-particle Lagrangian

In order to have Lorentz-invariant equations of motion via the Euler-Lagrange equations, we must start with a Lorentz-invariant scalar Lagrangian. Consider the case of a free particle. Later, we will study a free particle in an EM field.

The only thing invariant when a particle moves from one position to the next is its own proper time.[23] So a good candidate for the Lagrangian is the proper time increment, $d\tau$. Looking ahead, we will multiply this by the negative of the mass of the particle to give the Action

$$Action = -m\int_a^b d\tau = -m\int_a^b dt\sqrt{1-v^2} = -m\int_a^b dt\gamma^{-1}$$

so that the Lagrangian is.

$$\mathcal{L} = -m\sqrt{1-v^2}. \tag{3.17}$$

Then, using $v^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2$, the conjugate momentum is

$$p^i = m\frac{dx^i}{d\tau} = mU^i,$$

the mass multiplied by the spatial part of the 4-momentum. From (1.14) with (3.14), the Hamiltonian is

$$H = \sum_i \dot{x}^i p^i - \mathcal{L} = \frac{m}{\sqrt{1-v^2}} = m\gamma,$$

which is $mU^0 = P^0$, in agreement with (3.11).

3.5 Curious results of SR

We already saw in Figure 1 that some events may be in the past for one observer while in the future for another. Consider now the spacetime diagram in Figure 3 with an invariant hyperbola, Δs^2, about the t-axis, intersecting it at t=A.

All points on the hyperbola[24] have the same (invariant) interval or proper time, so point A is at 1 on the t-axis and point B is at 1 on the t'-axis. The red dotted line is tangent to the curve at B and parallel to the x' axis and is therefore a line of simultaneity for O' at constant t'=1. The horizontal dotted line through A and C is a similar line of simultaneity for O at constant t=1. What can this tell us?

23 Susskind and Friedman, 96.

24 It's really a parabola, since I could not figure out how to make a hyperbola in Inkscape. Hope nobody notices.

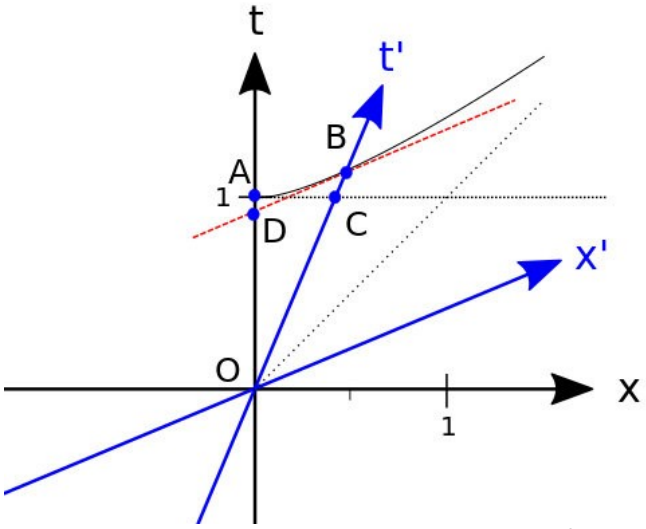


Figure 3: Time dilation (after Collier[25])

O' sees point D lying on its line t'=1. But O sees D at t<1. So O thinks the clocks of O' are running slow. Inversely, O' sees event C as occurring at time t'<1. But O sees it at t=1. So O' thinks the clocks of O are running slow.

Using the Lorentz transform, one can calculate the quantity of this ***time dilation*** as

$$(\Delta t)_O = \gamma(\Delta t'_O) < \Delta t'_O.$$

If the diagram of Figure 3 is rotated 90° clockwise, similar considerations show that each observer thinks the other's distance is contracted in a direction along its trajectory. This is ***length contraction***.

Several so-called paradoxes, which are in fact incorrect ways of framing the situation, are well known.

Time dilation leads to the so-called ***twin paradox***: If one of two twins moves away at great speed, then reverses direction and returns, she will have aged much less than her earth-bound twin. The reason why the inverse is not also true (aside from the impossibility of it) is because she undergoes acceleration and so is not always in an inertial system. Further consideration falls into the realm of GR.

Length contraction leads to the problem of the pole in the barn and shows that simultaneity must be abandoned. Leonard Susskind has updated this problem to a stretch limousine and a garage for a VW beetle. As seen by an observer stationary relative to the garage, which has doors at both ends, the limo, if it moves at a speed close to that of light, will be contracted so that it might fit all into the garage at once. In particular, O will see the following sequence of events:

1. Limo front enters garage front door;
2. limo tail enters garage front door;
3. limo front leaves garage back door.

But the limo driver rather sees the barn as being contracted, so there is no way he can fit into it all at once. He sees the following sequence:

1. Limo front enters garage front door;
2. limo front leaves garage back door;

25 All the reasoning concerning this diagram comes from Collier, 123-4.

3. limo tail enters garage front door.

Note the reversal of the sequence of events 2 and 3. Time ordering and, so, simultaneity are out the door!

3.6 The Lorentz transformation as a hyperbolic rotation

The Lorentz transformation may be seen as a rotation involving the time coordinate. For a spatial rotation, a typical Lorentz transformation matrix might be the following:

$$(\Lambda^{\bar{\beta}}{}_{\alpha}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\theta & sin\theta & 0 \\ 0 & -sin\theta & cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

So a rotation including time may be written

$$(\Lambda^{\bar{\beta}}{}_{\alpha}) = \begin{pmatrix} cosh\phi & -sinh\phi & 0 & 0 \\ -sinh\phi & cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The $\bar{O}$ frame ($\bar{x} = 0$) is then moving with velocity v such that

$$\bar{x} = -tsinh\phi + xcosh\phi = 0$$

and

$$v = \frac{sinh\phi}{cosh\phi} = tanh\phi.$$

Then

$$1 - v^2 = 1 - \frac{sinh^2\phi}{cosh^2\phi} = \frac{1}{cosh^2\phi},$$

since

$$cosh^2\phi - sinh^2\phi = 1.$$

So

$$\gamma = \frac{1}{\sqrt{1 - v^2}} = cosh\phi,$$

$$sinh\phi = cosh\phi \cdot tanh\phi = \gamma v$$

and we get back to the usual form of the Lorentz transform.

4 Electromagnetism

For now, take these equations as determined by experiment, which they are. We will see later, in another document, how to derive them from symmetry considerations. The variables represent

- $\vec{E}$, the electric field vector
- $\vec{B}$, the magnetic field vector
- $\vec{j}$, the electric current density vector

- ρ, the electric charge density (scalar)

4.1 Maxwell's equations of electromagnetism

Maxwell\s equations in differential form are as follows.

I. $$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0} \tag{4.1}$$

II. $$\nabla \cdot \vec{B} = 0 \tag{4.2}$$

III. $$\nabla \times \vec{E} = -\frac{\partial B}{\partial t} \tag{4.3}$$

IV. $$\nabla \times \vec{B} = \mu_0 \Big(\vec{j} + \epsilon_0 \frac{\partial E}{\partial t} \Big) \tag{4.4}$$

In integral form:

I. $$\oint_{S=\partial V} \vec{E} \cdot d\vec{S} = \frac{1}{\epsilon_0} \int_V \rho d^3\vec{r} \tag{4.5}$$

II. $$\oint_{S=\partial V} \vec{B} \cdot d\vec{S} = 0 \tag{4.6}$$

III. $$\oint_{C=\partial S} \vec{E} \cdot d\vec{r} = \int_S \Big(- \frac{\partial \vec{B}}{\partial t} \Big) \cdot d\vec{S} \tag{4.7}$$

IV. $$\oint_{C=\partial S} \vec{B} \cdot d\vec{r} = \mu_0 \int_S \Big(\vec{j} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \Big) \cdot d\vec{S} \tag{4.8}$$

And in words:

I. Gauss's law for an electric field: The net electric flux through any hypothetical closed surface is proportional to the net electric charge within that surface.

II. Gauss's law for magnetic fields is similar, but with a different result: Because there are no magnetic monopoles, the net magnetic flux through a surface must be zero.

III. Faraday's law of induction says that a changing magnetic field generates an electric field. More precisely: The work (electromotive force) required to move a unit charge around a closed path is equal to the negative of the time rate of change of the magnetic flux enclosed by the path.

IV. Ampere + Maxwell. Ampere's law says that an electric current generates a magnetic field around it. This is the principal of induction. Maxwell added the complement of Faraday's law in the term stating that a changing electric field also generates such a magnetic field.

Classical electrodynamics is completed by the ***Lorentz force*** on a moving charge q:

V. $$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) \tag{4.9}$$

which includes Coulomb's law and the force due to current in a magnetic field.

The constants are:

$$\frac{1}{4\pi\epsilon_0} = 9 \cdot 10^9 N \cdot m^2 \cdot C^{-2},$$

$$\frac{\mu_0}{4\pi} = 10^{-7} N \cdot s^2 \cdot C^{-2},$$

where N = Newton, s = second and C = Coulomb.

4.2 The vector and scalar potentials

Define scalar potential V and vector potential $\vec{A}$ such that

$$\vec{B} = \nabla \times A \tag{4.10}$$

and

$$\vec{E} = \nabla V - \frac{\partial \vec{A}}{\partial t}. \tag{4.11}$$

For any such V and $\vec{A}$, Maxwell II and III are automatically satisfied. The vector potential is an *essential* concept because it is essential to writing a Lagrangian for the motion of a particle in a EM field.

4.3 Gauge transformations

Now consider a transformation of the potentials via an arbitrary function χ such that

$$\tilde{A} = \vec{A} + \nabla\chi \tag{4.12}$$

and

$$\tilde{V} = V - \frac{\partial \chi}{\partial t}. \tag{4.13}$$

The function χ is not only arbitrary, it is a function of position. This is what makes it a ***local gauge transformation***.[26] The potentials $\tilde{V}$ and $\tilde{A}$ are not observables, so can be changed at will as long as they lead to the same $\vec{E}$ and $\vec{B}$, which is guaranteed by the requirements of equations (4.12) and (4.13). Note that this is not about coordinate transformations but represents a redundancy in the description, different ways of describing the same physical situation. According to Susskind, "Gauge invariance is a feature of every known fundamental theory of physics."[27]

The function χ can be chosen according to the situation. We can limit the choice to those which satisfy a certain condition, known as a ***gauge condition***. In particular, we can pick those which satisfy the following condition, known as the ***Lorenz gauge***[28]:

$$\frac{\partial \tilde{A}_x}{\partial x} + \frac{\partial \tilde{A}_y}{\partial y} + \frac{\partial \tilde{A}_z}{\partial z} + \frac{1}{c^2}\frac{\partial \tilde{V}}{\partial t} = 0 \tag{4.14}$$

or, expressed differently,

$$\nabla \cdot \vec{A} = -\frac{1}{c^2}\frac{\partial V}{\partial t}. \tag{4.15}$$

Combining (4.12) through (4.15) with M I and M IV leads to the following equations:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = -\frac{\rho}{\epsilon_0} \tag{4.16}$$

26 A more precise definition of gauge theory is given in my notes on symmetry and QFT.
27 Susskind and Friedman, 267.
28 The gauge was first proposed by Ludvig Lorenz (no "t"), but credit is usually given to Hendrick A. Lorentz, he of the SR Lorentz transformation.

and

$$\frac{\partial^2 \vec{A}}{\partial x^2} + \frac{\partial^2 \vec{A}}{\partial y^2} + \frac{\partial^2 \vec{A}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \vec{A}}{\partial t^2} = -\mu_0 \vec{j} \tag{4.17}$$

which are inhomogeneous wave equations, or wave equations with sources (the right-hand sides). If the sources are absent, the equations describe waves propagating with velocity c through space. The essential and amazing point here is that the velocity of these waves is

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 3 \times 10^8 m/s, \tag{4.18}$$

which is the measured ***speed of light*** in a vacuum.

The wave is propagated such that the electric and magnetic vectors are orthogonal and the wave propagates along the direction $\vec{E} \times \vec{B}$, as in Figure 4.

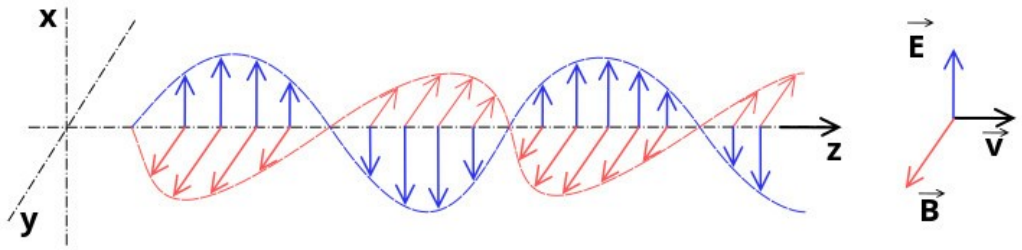


Figure 4: Electromagnetic wave propagation. from Supermanu via Wikimedia Commons.[29]

4.4 EM four-potentials

The EM vector potential can be expressed in covariant form as the four-potential

$$A_\mu = \left(\frac{V}{c}, \vec{A}\right) := (A_0, A_1, A_2, A_3) \tag{4.19}$$

and the four-current

$$J = \left(\rho c, \vec{j}\right). \tag{4.20}$$

We then also use

$$\frac{1}{\epsilon_0 c^2} = \mu_0. \tag{4.21}$$

Then the gauge transformation of equations (4.12) and (4.13) can be written in 4-vector form as

$$\tilde{A}_\mu = A_\mu + \frac{\partial \chi}{\partial x^\mu} \tag{4.22}$$

(for metric signature -1,1,1,1). The two wave equations (4.16) and (4.17) can be expressed as a single wave equation

$$\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} + \frac{\partial^2 A}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 A}{\partial t^2} = -\mu_0 J, \tag{4.23}$$

which can be expressed covariantly as[30]

29 Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Onde_electromagnetique.svg.
30 Using tensor subscript-superscript notation as in GR.

$$\partial_\mu \partial^\mu A^\nu = \mu_0 j^\nu \tag{4.24}$$

or as

$$\nabla \cdot \nabla A = \mu_0 J. \tag{4.25}$$

One can then write the ***Lorenz gauge*** (4.14) in 4-d notation as

$$\nabla \cdot A = 0 \quad \text{or} \qquad \partial_\nu A^\nu = 0. \tag{4.26}$$

Maxwell's equations in tensor form are then

$$\partial^\alpha \partial_\alpha A^\mu(x) - \partial^\mu(\partial_\nu A^\nu(x)) = j^\mu(x), \tag{4.27}$$

which in the Lorenz gauge is simply

$$\partial^\alpha \partial_\alpha A^\mu(x) = j^\mu(x). \tag{4.28}$$

The EM four-potential is indeed a four-vector. Just as Lorentz boosts mix time and space of a spacetime four-vector, so also do they mix by induction, MIII, electric and magnetic electric fields of A_μ.

4.5 Relativistic particle in EM field and the EM field tensor

Now add a field term to the free-particle Lagrangian (3.17), proportional to the electric charge (which we want in the ensuing Lorentz force) so the action looks like[31]

$$\mathcal{A} = -m \int_a^b \sqrt{1 - \dot{x}^2} dt + e \int_a^b dX^\mu A_\mu(t, x). \tag{4.29}$$

This is an incorrect expression for an action in classical mechanics, since the second term is integrated over space and not time. But the new term can be expressed as

$$e \int_a^b \frac{dX^\mu}{dt} A_\mu(t, x) dt$$

so the Lagrangian density for this term, the interaction term is

$$\mathcal{L}_{int} = e \frac{dX^\mu}{dt} A_\mu(t, x).$$

Using $\dfrac{dx^0}{dt} = 1$ gives

$$\mathcal{L} = -m\sqrt{1 - \dot{x}^2} + eA_0(t, x) + e\dot{X}^P A_p(t, x). \tag{4.30}$$

The Euler-Lagrange equations then lead (eventually) to a force

$$m \frac{d}{dt} \frac{\dot{X}_p}{\sqrt{1 - \dot{x}^2}} = e\Big(\frac{\partial A_0}{\partial X^p} - \frac{\partial A_p}{\partial t}\Big) + e\dot{X}^n \Big(\frac{\partial A_n}{\partial X^p} - \frac{\partial A_p}{\partial X^n}\Big). \tag{4.31}$$

This is beginning to look like the Lorentz force. In this case, from (4.9), it is clear that the first term between parentheses on the right must be the p-th component of $\vec{E}$. If we expand $\vec{v} \times \vec{B}$ in terms of $\nabla \times \vec{A}$ and compare it to the second term in parentheses, noting that terms for $n = p$ cancel out, this is equivalent to

$$m \frac{d}{dt} \frac{\dot{X}_p}{\sqrt{1 - \dot{x}^2}} = eE_p + e\dot{X}^n B_n.$$

31 Susskind, SR, 232ff.

Expressed properly in terms of four-vectors, this is indeed the relativistic Lorentz force:

$$m\frac{d^2X_\mu}{d\tau^2} = e\frac{dX^\nu}{d\tau}\Big(\frac{\partial A_\nu}{\partial X^\mu} - \frac{\partial A_\mu}{\partial X^\nu}\Big). \tag{4.32}$$

Call the expression in parentheses

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{4.33}$$

The correspondence between the derivative differences in the last two equations defines the elements in the ***electromagnetic field tensor***

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}, \tag{4.34}$$

as can be seen, for example, from

$$F_{10} = \partial_1 A_0 - \partial_0 A_1 = \partial_x V - \partial_t A_x = E_x$$

or

$$F_{12} = \partial_1 A_2 - \partial_2 A_1 = \partial_x A_y - \partial_y A_x = (\nabla \times \vec{A})_z = -B_z.$$

The EM field tensor is important because it shows how to transform EM fields between frames related by Lorentz transformations.

In terms of the four-velocity

$$\vec{U} = (\gamma, \gamma\vec{V}) \tag{4.35}$$

and the EM field tensor, the Lorentz four-force is

$$F^\mu = qF^{\mu\nu}U_\nu, \tag{4.36}$$

as can be verified from equations (4.34) and (4.35). E.g.,

$$F^2 = qF^{2\nu}U_\nu = q\gamma(E_y + B_z v_x - v_z B_x) = q\gamma[E_2 + (\vec{v} \times \vec{B})_2],$$

the y-th component of a four-force.

More on Maxwell's equations in my notes on QFT.

5 Quantum mechanics

What follows is an overview of QM. For more details, see my QM notes.

5.1 Vector and Hilbert spaces

In QM, the space of states of a physical system is a mathematical ***Hilbert space***, an abstract construction of a type of space called a ***vector space***. A Hilbert space is a complete vector space possessing the structure of an inner product.[32] It may have any number of dimensions, often an infinite number.

A vector space must obey a number of rules. The sum of any two vectors is a vector, and addition is commutative and associative. There exists a zero vector and, for every vector, an inverse vector. The product of a scalar and a vector is a vector and therefore so is a ***linear combination*** of vectors, a sum of such products. Scalar addition is associative and scalar multiplication is distributive and associative.

A vector is ***linearly independent*** of a set of vectors if it can not be represented as a linear combination of vectors in the set. Each vector in a set of linearly independent vectors is independent of all the other members of the set. If every vector in the space can be expressed as a linear combination of vectors in the set, the set Is said to ***span*** the state. Such a set of linearly independent vectors which span the space is called a ***basis***. (Think of $\hat{\imath}, \hat{\jmath}, \hat{k}$ in 3d Euclidean space.) Since they are linearly independent their number is minimal and represents the dimension of the space.

In physics, the basis vectors of such a state may be discrete or continuous. In the latter case, we are interested in functions which are square-integrable over a specified interval, a to b, a space mathematicians call L^2([a,b]).[33]

5.2 Physical states, observables and operators

The ***state*** of a system is a represented by its state vector, a unit (normalized) vector in the vector space of states. A physical ***observable*** is described by a linear ***operator***.

Using Dirac's bra-ket notation, a state vector is a ***ket*** represented by $|A\rangle$. A bra $\langle B|$ is a member of the ***dual space*** of complex conjugates of the bras.

Vectors need not be intervals in the vector space. As soon as a set of functions obeys the above requirements for a vector space, it may be considered a set of vectors. This is true for the functions of physics.

Like good vectors in any Hilbert space, bras and kets can form an inner product denoted by

$$\langle A|B\rangle \text{ with } \langle B|A\rangle = \langle A|B\rangle^*,$$

its inverted complex conjugate. The Dirac notation is conveniently the same whether the basis be discrete or continuous. In the discrete case, a ket vector may be represented as a sum of basis vectors, in which case

$$|A\rangle = \sum_i a_i|\lambda_i\rangle \text{ and } |B\rangle = \sum_i b_i|\lambda_i\rangle$$

for complex a_i and b_i, assuming the $|\lambda_i\rangle$ are orthonormal basis vectors. Then the inner product is given by

32 Or, a Hilbert space is an inner-product space which is complete. Jeevanjee, 37.
33 Jeevanjee, 14.

$$\langle A|B\rangle = \sum_i {a_i}^* b_i$$

In the discrete case, it is often convenient to represent the bra and the ket as matrices.

When an observable is continuous, say x, the state is itself a function of continuous variables. The bra $\langle\psi|$ corresponding to a ket $|\psi\rangle$ represented by the function $\psi(x)$ is the complex conjugate $\psi^*(x)$. In these equations, $\psi(x)$is a complex-valued function of an independent real variable x. Then

$$\sum_i \rightarrow \int_i dx$$

and the inner product is defined by

$$\langle\Psi|\Phi\rangle = \int_{-\infty}^{\infty} \psi^*(x)\phi(x)dx.$$

Normalization of probability then requires that

$$\langle\Psi|\Psi\rangle = \int_{-\infty}^{\infty} \psi^*(x)\psi(x)dx \ = 1.$$

Although this integral may be infinite (and so not normalizable), it may nevertheless be useful over a limited range (a,b) of x, so the space is L^2([a,b]). This means integrating from a to b rather than from $-\infty$ to ∞ in the above equation.

The important operators in QM are Hermitian operators. The ***Hermitian conjugat***e $T^\dagger$ of a linear operator (transformation) T is defined by

$$\langle\phi|T\psi\rangle = \langle T^\dagger\phi|\psi\rangle \qquad (5.1)$$

which in discrete matrix form is the complex conjugate of the transposed (inverted) matrix

$$T^\dagger_{ij} = T^*_{ji}.$$

5.3 Principles of quantum mechanics

Several principles govern observables and basis vectors.[34]

Principle 1. A physical observable is represented by a linear Hermitian operator L. where the Hermitian conjugate is defined above.

Principle 2. The possible results of measurements of observables are eigenvalues of the Hermitian operator representing the observable:

$$L|\lambda\rangle = \lambda|\lambda\rangle \qquad (5.2)$$

where λ is the ***eigenvalue*** of the eigenvector ket $|\lambda\rangle$. The operator transforms the vector into a multiple of itself. A ***Hermitian operator*** is one which is equal to its Hermitian conjugate

$$L^\dagger = L,$$

which guarantees that its eigenvalues are positive and ***real***. The set of all eigenvalues of an operator is called its ***spectrum***.

34 Principles adapted from Susskind and Friedman, 69-74.

Principle 3a. Unambiguously distinguishable states are represented by ***orthogonal*** vectors. i.e., for two eigenvalues, $\lambda_1 \neq \lambda_2$, the corresponding eigenvectors are orthogonal:

$$\langle \lambda_1 | \lambda_2 \rangle = \langle \lambda_2 | \lambda_1 \rangle = 0 \tag{5.3}$$

Principle 3b. The eigenvectors of an operator are a ***complete*** set. This means that any vector can be expressed as a linear sum of eigenvectors.

$$|A\rangle = \sum_i \alpha_i |\lambda_i\rangle \tag{5.4}$$

where the complex coefficients are inner products

$$\alpha_i = \langle \lambda_i | A \rangle \tag{5.5}$$

In other words, the eigenvectors of a Hermitian operator may be chosen to form an ***orthonormal*** basis. The set of inner-product coefficients α_i is the ***wave function*** of the system in the basis defined by the observable $\mathbf{L}$.[35] More on that in section 5.6. The number of eigenvectors in the basis is the dimension of the Hilbert space, or at least that part of it corresponding to this observable. (Think of a fermion's angular momentum in 3-d space and its spin, in 2-d spinor space.) So the dimension is greater than or equal to the number of eigenvalues. If the eigenvalues are continuous, the dimension is infinite and this may apply to many or all observables.

Principle 3c. Two orthogonal eigenvectors may have the same eigenvalue, in which case they are referred to as ***degenerate*** states. Even then, it is possible to construct a set of orthonormal basis vectors.

Principle 4. For a normalized state vector $|A\rangle$ and observable $\mathbf{L}$, the probability of observing value λ_i by measurement is given by the ***Born rule***

$$P(\lambda_i) = \langle A | \lambda_i \rangle \langle \lambda_i | A \rangle = |\langle A | \lambda_i \rangle|^2 \tag{5.6}$$

where

$$L|\lambda_i\rangle = \lambda_i |\lambda_i\rangle$$

and

$$\sum_i P(\lambda_i) = 1$$

for total probability (normalization).

With the Dirac notation, we can define the expectation value of a Hermitian operator $\mathbf{L}$ by

$$\langle L \rangle = \langle \psi | L | \psi \rangle = \langle \psi | L\psi \rangle = \langle \psi L | \psi \rangle \tag{5.7}$$

since L is Hermitian. More generally, if a basis of orthonormal state vectors is represented by $|a, b, c, ...\rangle$ for eigenvalues (a,b,c,...), then any state vector may be expanded as

$$|\Psi\rangle = \sum_{a,b,c,\ldots} \psi(a, b, c, ...) |a, b, c, ...\rangle \tag{5.8}$$

where the *set of coefficients* $\psi(a, b, c, ...)$ is called the ***wave function*** of the system in the basis defined by the observables, A, B, C, ….[36] We can expand $|\Psi\rangle$in terms of another set of basis vectors corresponding to different observables and the wave function set will be different even though they describe the same system

35 Susskind and Friedman, QM, 134. Some authors confuse wave function and state vector. Grifffiths, 114, restricts the definition to the position representation.

36 Susskind's notation. Griffiths uses a different definition of the wave function, but this one keeps it separate from the state vector.

state. The two sets of basis vectors correspond to different ***representations*** of the system, for the two different observables.

After measurement of an observable in such a ***superposition*** of states, the wave function is no longer a superposition and a second measurement of the state gives the same result as the first, which is logical ... in a way. The standard (Copenhagen) interpretation says that the wave function has “collapsed” to a single state. But what does “collapse” mean? This is the so-called ***measurement problem*** of QM.

There are many proposals for interpreting what happens when measurement takes place. The ***instrumentalist*** approach, for instance, interprets the wave function as merely an instrument to calculate probabilities, without having any real meaning in itself. This is opposed to the ***realist*** approach, which interprets the wave functions as a thing, a new physical entity or field. There exist nuances of both these ideas. And we haven't mentioned many worlds or pilot waves or other suggestions.

Consider this though. A superposition is not a thing in itself, it is always a superposition of some thing or things. If we have, for instance

$$|\Psi_3\rangle = \tfrac{1}{\sqrt{2}}(|\Psi_1\rangle + |\Psi_2\rangle) \qquad \text{and}$$

$$|\Psi_4\rangle = \tfrac{1}{\sqrt{2}}(|\Psi_1\rangle - |\Psi_2\rangle),$$

then we can also write

$$|\Psi_1\rangle = \tfrac{1}{\sqrt{2}}(|\Psi_3\rangle + |\Psi_4\rangle) \qquad \text{and}$$

$$|\Psi_3\rangle = \tfrac{1}{\sqrt{2}}(|\Psi_3\rangle - |\Psi_4\rangle).$$

Would you really accept all four wave functions as representing real entities?[37]

Meanwhile, back to our wave function (5.3). A method for solving such an equation is to first rewrite it in matrix form (if possible) as

$$(L - \lambda\mathbf{I})|\lambda\rangle = 0$$

where $\mathbf{I}$ is the identity matrix. If $(\mathbf{L} - \lambda_i\mathbf{I})$ has an inverse, we can let It operate on both sides of the above equation and show that $|\lambda\rangle$ must be zero, which is not an interesting case. So it must not have an inverse and therefore is singular and has determinant zero.

$$det(L - \lambda\mathbf{I}) = \begin{vmatrix} (L_{11} - \lambda) & L_{12} & ... & L_{1n} \\ L_{21} & (L_{22} - \lambda) & ... & L_{2n} \\ ... & & & \\ L_{n1} & L_{n2} & ... & (L_{nn} - \lambda) \end{vmatrix} = 0$$

This gives an equation, the ***characteristic equation***, which (hopefully) may be solved for the eigenvalues. Then, putting the eigenvalues back into equation (5.2), one can solve for the eigenvectors.[38]

5.4 Unitarity and evolution in time

Susskind calls the minus-first law the statement that information is never lost, so two identical isolated systems which begin in different states remain in different states: Distinctions are conserved. The QM version of this is ***unitarity***, the fifth principle.

37 Thanks to Sabine Hossenfelder for this idea. Understanding Quantum Mechanics #2: Superposition and entanglement. https://www.youtube.com/watch?v=j6Mw3_tOcNI

38 Griffiths, 476-7.

Principle 5. The evolution of state-vectors with time is ***unitary***.

Then distinctions are conserved, which means that eigenvectors remain orthogonal over time. If

$$\langle\Psi(0)|\Phi(0)\rangle = 0 \tag{5.9}$$

and the ***time-development operator*** $\mathbf{U}$ is defined by

$$|\Psi(t)\rangle = \mathbf{U}(t)|\Psi(0)\rangle \tag{5.10}$$

then orthogonality leads to U being ***unitary***, meaning that

$$\mathbf{U}^\dagger\mathbf{U} = \mathbf{I} \tag{5.11}$$

the identity, i.e.

$$\langle i|\mathbf{U}^\dagger(t)\mathbf{U}(t)|j\rangle = \delta_{ij}, \tag{5.12}$$

which is the defining property of a unitary group U(n), (In group-speak, we say that a unitary group conserves complex inner products. For much more, see my notes on symmetry and QFT.) Then time evolution is unitary. For small time ϵ, we can write[39]

$$\mathbf{U}(t) = \mathbf{I} - i\epsilon\mathbf{H} \tag{5.13}$$

which leads through (5.11) to the result that $\mathbf{H}$ must be Hermitian

$$\mathbf{H}^\dagger = \mathbf{H}. \tag{5.14}$$

Then (tossing in $\hbar$ to correct units) equation (5.13), we can rearrange

$$|\Psi(\epsilon)\rangle = |\Psi(0)\rangle - i\epsilon\mathbf{H}|\Psi(0)\rangle$$

and divide by ϵ, which leads to[40]

$$\boxed{i\hbar\frac{\partial|\Psi\rangle}{\partial t} = \mathbf{H}|\Psi\rangle} \tag{5.15}$$

which is the ***generalized*** (***time-dependent***) ***Schrödinger equation***. The presence of $\hbar$ gives $\mathbf{H}$ the units of energy: It is the ***quantum Hamiltonian***.

The Hamiltonian represents energy, so the eigen-equation for the energy must be

$$\mathbf{H}|E_j\rangle = E_j|E_j\rangle. \tag{5.16}$$

We can expand the definition of the Hamiltonian by using the standard operator for the momentum, $i\hbar\nabla$ and taking the energy operator, $i\hbar\partial_t$, from (5.15) in the classical equation

$$E = \frac{p^2}{2m} + V.$$

Then what do we do with this? The answer is, we'll let the result act on something we will call a wave function. This gives us the iconic Schrödinger equation:

$$\boxed{i\hbar\frac{\partial\Psi(x)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi(x) + V(x)\Psi(x)}. \tag{5.17}$$

For an operator $\mathbf{L}$, one can show from the definition of the expectation value (5.7) and the Schrödinger equation that

39 In group-theory parlance, H is the ***generator*** of a unitary group, in this case U(3).
40 Susskind and Friedman, QM, 111-112.

$$\frac{d\langle L\rangle}{dt} = -\frac{i}{\hbar}\langle [L, H]\rangle \tag{5.18}$$

where the ***commutator***

$$[L, M] = LM - ML \tag{5.19}$$

Is in general not equal to zero.

This is the core of quantum mechanics, that the operators for observables, in general, do not commute.[41] This fact leads, as we shall see very soon, to the Uncertainty Principle. The simplest way to determine the result of commutation of two operators is to apply them to a dummy wave function, such as follows.

$$[\hat{x}, \hat{p}] = \left[x(-i\hbar)\frac{d}{dx}(f) - (-i\hbar)\frac{d}{dx}(xf)\right] = i\hbar f(x),$$

which demonstrates the ***canonical commutation relation*** of QM:

$$[\hat{x}, \hat{p}] = i\hbar. \tag{5.20}$$

In more than one dimension, this equation can be generalized as

$$\boxed{[x_i, p_j] = i\hbar\delta_{ij}}. \tag{5.21}$$

We have already seen that, in classical mechanics, the time derivative can be expressed in terms of Poisson brackets as

$$\dot{F} = \{F, H\} \tag{1.26}$$

so that the correspondence between classical and quantum mechanics is

$$[F, G] \leftrightarrow i\hbar\{F, G\} \tag{5.22}$$

where the smallness of $\hbar$ makes the term negligible in the classical limit.

So if the commutator of an observable with the Hamiltonian is zero, the quantity does not change with time and is ***conserved***. The simplest case

$$[H, H] = 0$$

assures ***conservation of energy***. Possible energy states are represented by the eigenvalues of the Hamiltonian as seen in (5.16).

Since the eigenvectors form a complete set of basis vectors,

$$|\Psi\rangle = \sum_i \alpha_i |E_i\rangle \tag{5.23}$$

and if the Hamiltonian does not depend explicitly on time, then

$$\alpha_j(t) = \alpha_j(0)e^{-\frac{i}{\hbar}E_j t}. \tag{5.24}$$

Since $\alpha_i(0) = \langle E_i|\Psi(0)\rangle$, then

41 Sabine Hossenfelder, Understanding quantum mechanics #1, https://backreaction.blogspot.com/2020/04/understanding-quantum-mechanics-1-its.html?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+blogspot%2Fermku+%28Backreaction%29

$$|\Psi(t)\rangle = \sum_i |E_i\rangle\langle E_i|\Psi(0)\rangle e^{-\frac{i}{\hbar}E_j t} \tag{5.25}$$

The concept of conservation applies also to the probability, which must satisfy a continuity equation. This says that change of probability in a region must be accompanied by a flow of probability into or out of the region. The continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \vec{j} = 0. \tag{5.26}$$

Here, ρ is the density and $\vec{j}$ the current, be it of probability or some physical field such as the EM field. By multiplying the Schrödinger equation by the complex conjugate of the wave function, doing the same thing with the complex conjugate of both, subtracting and doing some differentiation by parts[42], one can show that the NRQM ***probability density and current*** are given by

$$\rho = \Psi^\dagger \Psi$$

and

$$\vec{j} = \frac{\hbar}{2iM}\{\Psi^\dagger(\nabla\Psi) - (\nabla\Psi)\Psi\} := -i\frac{\hbar}{2M}\Psi^\dagger \overleftrightarrow{\nabla} \Psi, \tag{5.27}$$

where M is the mass of the body concerned in the Schrödinger equation.

5.5 Simultaneous observables and uncertainty

It is clear from

$$\mathbf{LM}|\lambda,\mu\rangle = \lambda\mu|\lambda,\mu\rangle$$

that if $|\lambda,\mu\rangle$ is a simultaneous eigenvector basis for the two operators $\mathbf{L}$ and $\mathbf{M}$, then the result is independent of order and the two operators must commute:

$$[\mathbf{L},\mathbf{M}]|\lambda,\mu\rangle = 0.$$

In classical mechanics, from (1.25),

$$\{x,p\} = 1$$

and in quantum mechanics

$$[x,p] = i\hbar. \tag{5.28}$$

The case of non-commutation of two operators A and B plus some geometry leads to the ***generalized uncertainty principle***:

$$\Delta A \Delta B \geq \frac{1}{2}\langle\Psi|[A,B]|\Psi\rangle. \tag{5.29}$$

This result is not a separate postulate, but a consequence of the statistical interpretation of QM. The original ***Ungenauigkeitsprinzip*** would have been better and more accurately translated as “inexactitude principle” or “imprecision principle” and could have avoided overly hasty interpretations, but... The result can also be shown to be

$$\sigma_a^2 \sigma_b^2 \geq \left(\frac{1}{2i}\langle[A,B]\rangle\right)^2 \tag{5.30}$$

where the standard deviation of an observable is given by

42 Klauber, 45.

$$\sigma^2 = \langle (A - \langle A \rangle)\Psi | (A - \langle A \rangle)\Psi \rangle. \tag{5.31}$$

This quickly leads to the usual version concerning position and momentum

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

and can be shown to be true also for energy and time

$$\Delta E \Delta t \geq \frac{\hbar}{2}.$$

This is a QM result and has nothing to do with SR four-vectors.

“No experiment can ever beat this limitation.”[43]

In my notes on symmetry groups, it is shown that the QM operators for observables are the generators for the corresponding symmetries. Therefore, a measurement of momentum is equivalent to the action of a translation generator.[44] So a measurement of momentum moves the system some, thus making the position less certain.

5.6 Bases and representations

Define two linear operators

$$\mathbf{X}\psi(x) = x\psi(x) \tag{5.32}$$

and

$$-i\hbar\mathbf{D}\psi(x) = -i\hbar\frac{d\psi(x)}{dx} := \mathbf{P}\psi(x) \tag{5.33}$$

which is the definition of the ***momentum operator***, the factor $-i\hbar$ being necessary to ensure that **P** be Hermitian and the units correct. The momentum eigen-equation then is

$$-i\hbar\frac{d\psi(x)}{dx} = p\psi(x) \tag{5.34}$$

of which “normalized” solutions are of the form

$$\psi_p(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{ipx}{\hbar}}. \tag{5.35}$$

What is meant in this case by “normalized” is

$$\langle \psi_{p'} | \psi_p \rangle = \delta(p - p'),$$

which is referred to as ***Dirac normalization***. In fact, the eigenfunctions of a hermitian operator whose spectrum is continuous like this are not normalizable (Think uncertainty principle.), but a wave packet around this momentum value is. Fortunately, eigenfunctions with real eigenvalues are Dirac orthonormalizable and complete.

Equation (5.35) is the equation in the x basis of a wave of wavelength

$$\lambda = \frac{2\pi\hbar}{p}, \tag{5.36}$$

43 Susskind, 270.
44 Schwichtenerg, PS, 197.

which is the ***De Broglie formula***. Equations (5.32) and (5.35) are in the ***x basis*** or ***position representation***. So, anticipating equation (5.42), we could also write

$$\psi_p(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{ipx}{\hbar}} = \langle x|p\rangle. \tag{5.37}$$

In the position representation, the probability for a given value of x (although we should multiply by dx and say for a given increment of x) is

$$P(x) = \psi^*(x)\psi(x) \tag{5.38}$$

and for a given value of momentum

$$P(p) = {\psi_p}^*(x)\psi_p(x).$$

But we can also say that

$$P(p) = \widetilde{\psi}^*(p)\widetilde{\psi}(p) \tag{5.39}$$

where $\widetilde{\psi}(p)$ is the wave function in the ***momentum representation***.

Slipping back into Dirac notation, define the identity operator in both bases

$$\mathbf{I} = \int dx|x\rangle\langle x| \tag{5.40}$$

and

$$\mathbf{I} = \int dp|p\rangle\langle p|. \tag{5.41}$$

Supposing a state vector $|\Psi\rangle$, the wave function in the position-representation is

$$\psi(x) = \langle x|\Psi\rangle \tag{5.42}$$

and in the momentum representation

$$\tilde{\psi}(p) = \langle p|\Psi\rangle. \tag{5.43}$$

Note the order here: *The **wave function** is the **projection** of the **state vector onto** the **eigenvectors** of the **representation**.* This is just the content of equations (5.42) and (5.43).

Using the identity operator in the x basis

$$\tilde{\psi}(p) = \int dx\langle p|x\rangle\langle x|\Psi\rangle$$

so from the complex conjugate of the momentum eigenvector (5.37) and the wave function in the p basis (5.42),

$$\widetilde{\psi}(p) = \frac{1}{\sqrt{2\pi}} \int dx e^{\frac{-ipx}{\hbar}} \psi(x). \tag{5.44}$$

Starting from the x-basis wave function (5.42) and using the identity operator (5.41) leads to the inverse

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int dp e^{\frac{ipx}{\hbar}} \widetilde{\psi}(p) \tag{5.45}$$

The two representations, position (5.45) and momentum (5.44), are ***Fourier transforms*** of each other. They are also the means for transforming a wave function from the position representation to the momentum representation or vice versa.

5.7 A useful example – the harmonic oscillator (SHO)

A SHO is defined by a force proportional to the displacement of a mass. So for a classical 1-d SHO

$$F = -kx,$$

and one solution is of the form

$$x(t) = Acos(\omega t - \phi).$$

The kinetic and potential energies are then

$$K(t) = \frac{1}{2}mv^2(t) = \frac{1}{2}kA^2sin^2(\omega t - \phi) \tag{5.46a}$$

and

$$U(t) = \frac{1}{2}kx^2(t) = \frac{1}{2}kA^2cos^2(\omega t - \phi), \tag{5.46b}$$

both of which depend on the square of the amplitude, A. The total energy is obviously

$$E = K + U = \frac{1}{2}kA^2. \tag{5.47}$$

In QM, a SHO must satisfy a Schrödinger equation of the form

$$\Big(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \frac{1}{2}Kx^2\Big)\psi = E\psi. \tag{5.48}$$

.The solutions involve Hermite polynomials of degree n and the energy of each state is

$$E_n = \Big(n + \frac{1}{2}\Big)\hbar\omega, \tag{5.49}$$

where $\omega = \sqrt{K/m}$. However, there is a more interesting way to solve for the energy eigenvalues.

Rewrite the Schrödinger equation in terms of momentum and position operators.

$$\frac{1}{2m}\Big(\hat{p}^2 + m\omega x^2\Big)\psi = E\psi.$$

It would be nice to factor this into something like $(\hat{p} + im\omega\hat{x})(\hat{p} - im\omega\hat{x})$, but that would not work because in QM the momentum and position operators do not commute, as we saw in (5.20). Instead, define two *operators*,

$$\hat{a} = \sqrt{\frac{m\omega}{2\hbar}}\Big(\hat{x} + \frac{i}{m\omega}\hat{p}\Big) \tag{5.50}$$

and its *complex (Hermitian) conjugate*

$$\hat{a}^\dagger = \sqrt{\frac{m\omega}{2\hbar}}\Big(\hat{x} - \frac{i}{m\omega}\hat{p}\Big) \tag{5.51}$$

in terms of operators x and $\hat{p}$.[45] Since neither one is Hermitian, they are not operators for observables (what we will call Cartan operators in group theory). Note that

$$[\hat{a}, \hat{a}^\dagger] = 1. \tag{5.52}$$

Then the Hamiltonian can be written

$$\hat{H} = \hbar\omega\Big(\hat{a}^\dagger\hat{a} + \frac{1}{2}\Big), \tag{5.53}$$

where where comparison with (5.49) shows that the ***number operator***

$$\hat{n} = \hat{a}^\dagger\hat{a} \tag{5.54}$$

returns the number of entire units $\hbar\omega$ of energy in a given state:

$$\hat{n}|n\rangle = n|n\rangle. \tag{5.55}$$

Then $|n\rangle$ is also an ***eigenstate*** of the Hamiltonian, which may be written as

$$\hat{H} = \hbar\omega\Big(\hat{n} + \frac{1}{2}\Big) \tag{5.56}$$

so that

$$\hat{H}|n\rangle = \Big(n + \frac{1}{2}\Big)\hbar\omega|n\rangle, \tag{5.57}$$

The eigenvalues are equally spaced at intervals of $\hbar\omega$. Since

$$\hat{n}\hat{a}^\dagger|n\rangle = \hat{a}^\dagger\hat{a}\hat{a}^\dagger|n\rangle = \hat{a}^\dagger(1 + \hat{a}^\dagger\hat{a})|n\rangle = (n+1)\hat{a}^\dagger|n\rangle, \tag{5.58}$$

it turns out that $\hat{a}^\dagger$ has the effect of increasing the energy of state $|n\rangle$ by one unit of $\hbar\omega$; it is called a ***raising operator***.

Similarly,

$$\hat{n}\hat{a}|n\rangle = (n-1)|n\rangle, \tag{5.59}$$

so $\hat{a}$ is considered a ***lowering operator*** and the pair, raising and lowering, are called ***ladder operators***. Energy is positive, so there must be a ground state to which application of $\hat{a}$ gives a zero or negative result, and this can easily be used to find the wave function of that state.[46] Plugging that back into the Schrödinger equation yields the result for the ground-state energy

$$E_0 = \frac{1}{2}\hbar\omega.$$

Since application of the raising operator to the ground state $|0\rangle$ can create any energy state, we can write the normalized state

$$|n\rangle = \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}}|0\rangle. \tag{5.60}$$

Remember, all this follows from the (classical) form of the Hamiltonian and the QM commutation relations for the position and momentum operators.

45 For details, see Griffiths and Schroeter, 40-46 or Lancaster and Blundell, 19-23.
46 Griffiths and Schroeter, 43-44.

5.8 Angular momentum in QM

5.8.1. Angular momentum

Start with CM definition of angular momentum

$$\vec{L} = \vec{r} \times \vec{p}$$

and use the canonical commutation relations for position and momentum operators:

$$[r_i, p_j] = -[p_i, r_j] = i\hbar_{ij}, \quad [r_i, r_j] = [p_i, p_j] = 0. \tag{5.61}$$

The result (cyclic in I,j,k or x,y,z) is

$$[L_i, L_j] = i\hbar L_k \tag{5.62}$$

and

$$[L^2, \vec{L}] = 0, \tag{5.63}$$

where

$$L^2 = L_x^2 + L_y^2 + L_z^2.$$

The commutation relations of equation (5.61) are the basis of angular momentum (and spin) in QM; everything else concerning angular momentum follows from them. The same equations show that although the three components of L are incompatible as observables, one of them, say L_z, which would be called a Cartan operator in group theory, is compatible (commutes) with L^2, so the two may have simultaneous eigenvalues. We will suppose them to be as follows:

$$L^2 = \lambda f \quad \text{and} \quad L_z f = \mu f.$$

Define operators

$$L_\pm = L_x \pm iL_y. \tag{5.64}$$

Because of the commutation relations,

$$[L_z, L_\pm] = \pm\hbar L_\pm,$$

it turns out $L_\pm$ is a ladder (raising/lowering) operator, since

$$L_z(L_\pm f) = (\mu \pm \hbar)(L_\pm f). \tag{5.65}$$

The raising operator raises the L_z value of the eigenstate by $\hbar$. The raised or lowered eigenstate has the same value of L^2. Since L_z has to be less than or equal to L, taking its maximum value to be l (the negative of which turns out to be its minimum), we find that

$$L_z f_t = \hbar l f_t,$$

where f_t is the "top" eigenfunction, as for the case of a SHO. This in turn leads to

$$L^2 f_l^m = \hbar^2 l(l+1) f_l^m \tag{5.66}$$

and

$$L_z f_l^m = \hbar m f_l^m, \tag{5.67}$$

where l = 0, ½, 1, 3/2, ... and $m = -l, -l+1, ..., l-1, l$.

These values follow from the requirement that m take on integrally-spaced values from $-l$ to l, so $l = N/2$ and must therefore be an integer or half integer.

Note that the maximum measurable (observable) value of the z component of L is less than the magnitude of L!

5.8.2. Spin

Particles have spin which look mathematically like intrinsic angular momentum, even though the particles are mere points without extension and so have no real angular momentum. In fact, spin exists not in 4-d spacetime but in another, intrinsic space. Nevertheless, starting with the commutation relations, all the QM methods for angular momentum apply to the spin S.

The corollaries of equations (5.62), (5.66) and (5.67) for spin are[47]

$$[S_i, S_j] = i\hbar S_k \tag{5.68}$$

$$S^2|sm\rangle = \hbar^2 s(s+1)|sm\rangle; \;\; S_z|sm\rangle = \hbar m|sm\rangle, \tag{5.69}$$

where

$$s = 0, \frac{1}{2}, 1, \frac{3}{2}, ... \;\; \text{and} \;\; m = -s, -s+1....s-1, s. \tag{5.70}$$

Every elementary particle has an immutable spin with one of these values. As for angular momentum, there are raising and lowering operators for spin

$$S_\pm = S_x \pm iS_y$$

such that

$$S_\pm|sm\rangle = \hbar\sqrt{s(s+1) - m(m\pm 1)}|s(m\pm 1)\rangle.$$

Much more on spin in my document on symmetry and gauge thoery.

5.8.3. Spin ½

Let's consider the case of ***fermions***, having spin ½. In this case S_z can only have two values, $\pm\frac{1}{2}$ and it is convenient to write eigenvectors and operators as matrices. We define S_z =+½ as up, so the two eigenvectors for spin up and spin down are

$$\chi+ = \begin{pmatrix}1\\0\end{pmatrix} \qquad \text{and} \quad \chi- = \begin{pmatrix}0\\1\end{pmatrix} \tag{5.71}$$

and a general state is a sum of these two orthogonal eigenstates:

$$\chi = \begin{pmatrix}a\\b\end{pmatrix} = a\chi^+ + b\chi^-.$$

Note that χ is a set of components in a particular basis, which make up a ***spinor*** in spinor space. A ket is a vector in Hilbert space.

It is important to understand that spinors are not scalars or vectors for the reason that they do not transform like either of these. While equations with spinors must remain invariant under Lorentz transformations, the equations for the transformations are different from those for scalars or vectors. Whereas scalars remain the same under coordinate transformations and vectors (four-vectors) are transformed by multiplication by 4x4 matrices, spinors are transformed by 2x2 matrices.

47 All this from Griffiths, QM, 165-170.

Using these eigenfunctions and the commutation relations, we can show that the operators of the spin ½ components are three ***Pauli spin matrices***:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{5.72}$$

There is much more on spin-½ in the discussion of quantum field theory in my notes on symmetry and field theory.

5.9 Composite systems and entanglement

We've been considering single systems with their own wave functions and eigenvalues – what we can call ***pure states***. Then the expectation value of an observable A is

$$\langle A \rangle = \langle \Psi | \hat{A} | \Psi \rangle.$$

A state vector for two different, independent measurements or subsystems may be represented sometimes as a product of the state vectors for each subsystem. For a product state, there exist two sets of normalization equations. For instance, in a product of two independent states each with two eigenvalues, each is represented by two complex components and so four variables, but normalization reduces each by one and a negligible phase factor eliminates another, leaving only four independent components (degrees of freedom) for the product state.

But in general, composite systems are not product states. For a state vector for the same subsystems which is not a product state, there is only one normalization requirement and one phase factor, so there remain six independent parameters, making this a more complicated state. Such a state Is said to be ***entangled***.

5.9.1. Density operator and entanglement

Let's define a new tool, the ***density operator***:

$$\hat{\rho} \equiv |\Psi\rangle\langle\Psi|, \tag{5.73}$$

which we recognize as a projection operator. It projects an operator A onto the direction of $|\psi\rangle$:

$$\hat{\rho}|A\rangle = |\psi\rangle\langle\psi| \; |A\rangle = \langle\psi|A\rangle|\psi\rangle$$

In an orthonormal basis $\{|e_j\rangle\}$, its matrix elements are

$$\rho_{ij} = \langle e_i|\hat{\rho}|e_j\rangle = \langle e_i|\Psi\rangle\langle\Psi|e_j\rangle,$$

which constitute the ***density matrix***.

Case 1. Consider a standard case of spin up or down along the z-axis but write the density matrix for spin up along the x-axis. For this case[48]

$$\Psi = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

so brute-force insertion of this into (5.73) gives

$$\rho = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Obviously, ρ is Hermitian, has trace 1 and is idempotent:

$$\rho^2 = \rho.$$

48 Griffitys, QM, 169.

These are properties of density matrices for all pure states. We will see that the last one is not true for entangled states.

The density matrix allows us to calculate the expectation value of an operator:

$$\langle A\rangle = Tr|\psi\rangle\langle\psi|A = Tr(\rho L), \tag{5.74}$$

because

$$Tr|\psi\rangle\langle\psi|A = \sum_i \langle i|\psi\rangle\langle\psi|A|i\rangle = \sum_i \langle\psi|A|i\rangle\langle i|\psi\rangle$$

and

$$\sum_i |i\rangle\langle i| = 1.$$

What about the case where the system might be in any of a number of states, but we ignore which one? We only know the probability of each state. In this case, the definition of the density operator is expanded as a weighted sum of density operators

$$\rho = \sum_k p_k|\psi_k\rangle\langle\psi_k|, \tag{5.75}$$

with p_k being the probability of state ψ_k. A density matrix like this, which is a sum of projection operators, represents a ***mixed state***, one in which multiple states are possible.[49] A single projection operator represents a ***pure state***. Starting from (5.75), the same method of calculation used above leads again to equation (5.74). So knowing the density matrix, i.e., the states and their probabilities, we can calculate expectation values of operators in pure or mixed states.

Case 2. An example is an electron in either a spin up or spin down orientation along the z-axis, with equal probability.[50] Now $p_1 = p_2 = \frac{1}{2}$ and the density matrix is

$$\begin{aligned}\rho = \sum_k p_k|\Psi_k\rangle\langle\Psi_k| &= \frac{1}{2}\begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix} + \frac{1}{2}\begin{pmatrix}0\\1\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix}\\ &= \begin{pmatrix}1/2 & 0\\ 0 & 1/2\end{pmatrix}\end{aligned} \tag{5.76}$$

and

$$\rho^2 = \begin{pmatrix}1/4 & 0\\ 0 & 1/4\end{pmatrix} \neq \rho,$$

so this state is not pure.

Case 3. As another example, suppose that André and Béatrice (hereafter referred to as Andy and Bea) are studying the decay of a π^0 into an electron and a positron, each of which has a spin which can be either up or down along some axis. each represented by an operator (σ_A or σ_B). The π^0 is the singlet state,

$$|sing\rangle = \frac{1}{\sqrt{2}}(|ud\rangle - |du\rangle),$$

where each pair ud (up-down) or du represents the spin of the electron and the positron. We find

$$\langle\sigma_{Az}\rangle = \langle\sigma_{Bz}\rangle = 0, \tag{5.77}$$

49 Susskind and Friedman,
50 Griffiths, QM, 457.

but[51]

$$\langle \sigma_{Az}\sigma_{Bz}\rangle = -1. \tag{5.78}$$

Result (5.77) means nothing is known about what a measurement of σ_{Az} or σ_{Bz} will give, each being equally likely to return +1 or -1. But because of (5.78), once one σ_z is measured, the other is known. This state is ***maximally entangled***, which means that *although the state is a complete description of the system taken as a whole, nothing is known about either subsystem by itself*. By nothing, we mean that only probabilities for different eigenvalues can be calculated. However, ignorance of individual subsystem states is accompanied by correlation between measurements of the individual subsystems.

It is not possible to isolate the wave functions of one subsystem of an entangled system. "One cannot really speak of the 'the state' of either particle separately."[52] Only in a pure, unmixed state can Bea study her own subsystem without learning anything about Andy's.

The density matrix for the singlet state is

$$\rho = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix},$$

showing the 50/50 mixture of states. This makes it clear that

$$\rho^2 = \begin{pmatrix} 1/4 & 0 \\ 0 & 1/4 \end{pmatrix} \neq \rho.$$

It is the same as that for a particle of case (5.76), which represented our *ignorance* of the particle's state, only here it represents the physical *impossibility* of predicting a measured value for either spin direction.

Even if a subsystem changes, it must conserve distinctions (i.e., complex scalar products). Therefore it will change by a unitary transformation U such that

$$\mathbf{U}^\dagger\mathbf{U} = \mathbf{I},$$

and Andy's density matrix is unchanged under a unitary change of Bea's subsystem:

$$\rho^A_{aa\prime} = \sum_{bb\prime\prime b\prime} \psi^*(a\prime b\prime\prime)(U^\dagger_{b\prime\prime b}U_{bb\prime})\psi(ab\prime) = \sum_{b\prime} \psi^*(a\prime b\prime)\psi(ab\prime) = \rho^A_{aa\prime}.$$

This means that Bea cannot influence Andy's statistical results, meaning that "... no influence can propagate faster than the speed of light"[53], which is the ***principle of locality.*** The evolution of Bea's system, even if it is entangled with Andy's, has no influence on Andy's statistical predictions. This is all about expectation values, in the spirit of quantum mechanics. Adding the least little bit of non-unitary evolution to Bea's subsystem would mean that she could influence Andy's subsystem *faster than the speed of light*.

According to (5.18), since $\hat{\rho}$ is an operator,

$$i\hbar\frac{d\hat{\rho}}{dt} = [\hat{H}, \hat{\rho}],$$

if all $\frac{dp_k}{dt} = 0$.

5.9.2. Locality and hidden variables

We have defined ***locality*** as the principle that "... no influence can propagate faster than the speed of light." But in the singlet state for, for instance, the decay of a π^0 into an electron and a positron, we know

51 Susskind and Friedman, QM, 177.
52 Grifffiths, QM, 448.
53 Griffiths, QM, 447.

that if we measure the spin of one particle, then we know in advance of its measurement the spin of the other. But how can this information be passed from one measurement to the other and not disobey locality, i.e., by violating the speed limit c? It was suggested that QM, albeit correct, is incomplete, that there are so-called “hidden” variables which are unmeasured because unknown, and that these could at the moment of the decay determine the spin directions of the two particles. The challenge was taken up by John Bell.[54]

Bell considered a slightly different hypothetical experiment. Let the π^0 decay as before, but

(1) rotate the axes of spin measurement independently, so they are not always aligned;

(2) consider not the individual spin directions +1 or -1, but the average value of their product.

Because of these arbitrary and independent rotations, the product of the spin directions along the different directions is not necessarily -1. Call these directions $\mathbf{a}$ and $\mathbf{b}$. For given directions $\mathbf{a}$ and $\mathbf{b}$, the average of the spin product will be called $P(\mathbf{a}, \mathbf{b})$. For arbitrary orientations, QM says

$$P(\mathbf{a}, \mathbf{b}) = -\mathbf{a} \cdot \mathbf{b}.$$

Now let’s add locality in the form of a hidden variable λ, which is the argument to functions $A(\mathbf{a}, \lambda)$ and $B(\mathbf{b}, \lambda)$ which will fix in advance the result of the spin measurements. Evidently,

$$A(\mathbf{a}, \lambda) = \pm 1 \qquad \text{and} \quad B(\mathbf{b}, \lambda) = \pm 1.$$

Perfect alignment of the detectors forces

$$A(\mathbf{a}, \lambda) = -B(\mathbf{b}, \lambda).$$

Given any arbitrary probability density ρ, the average of the product of the measurements is

$$P(\mathbf{a}, \mathbf{ab}) = \int \rho(\lambda) A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda) d\lambda.$$

Using only these equations and some simple algebra, one can imagine a third unit vector $\mathbf{c}$ and calculate the ***Bell inequality***

$$|P(\mathbf{a}, \mathbf{b}) - P(\mathbf{a}, \mathbf{c})| \leq 1 + P(\mathbf{b}, \mathbf{c}),$$

which must therefore hold for any local hidden-variable theory. But it doesn’t. For instance, for the three vectors in a plane, $\mathbf{a}$ and $\mathbf{b}$ perpendicular to each other with $\mathbf{c}$ at a 45° angle between them,

$$P(\mathbf{a}, \mathbf{b}) = 0, \qquad P(\mathbf{a}, \mathbf{c}) = P(\mathbf{b}, \mathbf{c}) = -0.707,$$

which is clearly inconsistent with Bell’s equality. If QM is correct, then no hidden-variable theory can explain the apparent non-locality of the π^0 decay or restore locality. This result has been tested by many experiments and even received a Nobel Prize.

Entanglement is real and is not due to a local hidden-variable theory. Perhaps a worse blow to locality is the supposed instantaneous “collapse” of the wave function in the Copenhagen

6 Symmetry, groups and quantum field theory

See my notes on this subject in a separate file.

54 Following from Girffiths, QM, 449-452.

7 General relativity

GR is a classical theory in the sense that it does not take into account the tenets of QM.

7.1 Principles of General Relativity

A restatement of Special Relativity (SR) from section 3.1 begins by pointing out that it is based on two principles:

1. The ***principle of relativity*** states that the laws of nature should be the same for all observers in inertial frames (defined below).
2. All such observers, upon measuring the speed of light in a vacuum, must find the same result, c = 299,792,458 m/sec.[55]

The first requirement is necessary in order for physics to be coherent. It means that observers in inertial systems use the same equations. Rather than going on incessantly repeating "in an inertial reference frame", let's get it done with once and for all by stating:

> SR considers only observers in ***inertial*** reference frames, those which move with constant, unaccelerated velocity relative to one another.

However, this inertial observer can perfectly well observe objects which are accelerating in his frame. We will see what GR has to say about inertial frames at the end of the next section.

General Relativity (GR) builds on top of SR to take into account three additional requirements:

1. The ***principle of covariance*** is an extension of the first requirement of SR, requiring that physical equations be true in all coordinate systems. GR drops the requirement of inertial systems. In order to do this, new mathematical objects are required – tensors.
2. The ***principle of consistency*** states that the equations of GR must give the same result as the classical (Newtonian) equations when applied in the limiting case, e.g., in inertial frames.
3. The ***principle of equivalence*** exists in two or three different versions, but the strongest one is the most interesting, ***Einstein's equivalence principle***. This states that in a very small frame of reference, no experiment can distinguish between objects in a gravitational field or objects subject to uniform acceleration.

The principle of equivalence leads to new physics and a new understanding of the universe.

7.2 The equivalence principle and curvature

Imagine four cases:

1. An unaccelerated rocket is coasting along in space with its motors off, far from any matter which might give rise to a gravitational field, and with all the portholes covered so that an astronaut can see nothing outside. If the astronaut holds a wrench and opens her hand, the wrench stays next to her hand. In fact, if she drops two wrenches of different weights, they will both undergo the same acceleration, exactly as under the influence of gravity. Both will remain wherever she let go of them.

55 Or 1.0792528488 km/hr.

2. A person in a very high elevator, or an elevator in a very deep mine shaft, also has no view of the exterior. If the elevator cable breaks so that the elevator is falling freely and the person lets go of a wrench, the wrench will remain by his hand.
3. Now consider the elevator as sitting still on the surface of the Earth. If the person in the stationary elevator lets go of the wrench, it is in his interest first to assure that his foot is not just below it, as the wrench will fall to the floor.
4. Meanwhile, back in the rocket, the motors are turned on so as to give an acceleration of 1g to the rocket. If the astronaut releases her wrench, it too will drop to the "floor" of the rocket. This is entirely due to the acceleration of the rocket, not due to gravity. In fact, there is nothing you can do, no experiment possible, to determine whether you are accelerating or sitting in a gravitational field.

If you don't like the idea of falling elevators (which was Einstein's example), think of the reduced-gravity aircraft (aka the "vomit comet") used for habituating astronauts to near-weightless conditions.

The first two cases illustrate the impossibility of distinguishing between free fall and an absence of a gravitational field; the last two, of distinguishing between a gravitational field and uniform acceleration. GR really starts with the recognition that uniform acceleration is indistinguishable from a gravitational field. Einstein's equivalence principle is a precise statement of this idea. It should be understood as applying to small regions of spacetime, since in a large region above, say, the Earth, moving from one point to another will experience different values (directions) of the field.

The above four cases depend on one more very important – indeed, essential – point. Newton's second law of motion says that the acceleration of a body is proportional to the force on it, the proportionality constant being the object's inertial mass, m_I.

$$F = m_I a$$

Newton's law of gravitation says that the gravitational force on the same object is proportional to the products of the object's gravitation mass, m_G, and the mass, M, of the source of the gravitational field.

$$F = G\, m_G M/r^2.$$

Experiments beginning with Galileo show to a very high degree of accuracy that the acceleration due to gravity is independent of the mass of the object. Equating the force and its effect and dividing by m_I gives

$$a = G\, m_G M/m_I r^{2.}$$

Galileo's observation means that the m_I and m_G must be equal, leading to the following important requirement:

The inertial mass of an object is the same as its gravitational mass: $m_I = m_{G.}$

All of GR depends on this equality. making gravity accelerate all objects the same way, regardless of their mass or composition.[56] Otherwise, the person in the elevator or rocket could distinguish between the cases of gravity and of acceleration.

But how can gravity know how to pull on such different objects so as to impart the same acceleration to both?[57] It seems that gravity has nothing to do with the object, but is simply a property of space: In that case, what else can it be but curvature?

Consider race cars. At each end of a track, the roadbed is steeply banked to keep the cars from skidding off the road (as are highways, but less so, because motorists are not supposed to drive that fast). The angle and

56 We can ignore friction, since experiments have shown that feathers and anvils fall at the same rate in a vacuum.

57 I lifted this question from Russell Stannard. *Relativity: A very short introduction*. Oxford: Oxford University Press, 2008. Print.

curvature of the track are partly responsible for the cars' remaining on the track. Curving, even at constant linear speed, is a form of acceleration.

So curvature – curvature of space – can explain the force of gravity in a far "simpler" way than having gravity do something or other depending on the mass and composition of objects – and that across a distance. In that case, gravity is not even a "force", but simply the curvature of space.[58] As objects not submitted to other forces try to follow a straight-line path in a curved space, they naturally act as if they were accelerated – by gravity! Like the car on the race track.

But what the dickens would a straight line be in a curved space? Well, that is the matter of GR, the theory of gravity. In order to do handle this problem, GR must employ the mathematics of curved spaces called ***Riemann spaces***.

Gravity is present everywhere, maybe not much at great distances from stars and planets, but certainly here on the surface of the earth. So the only way to have an inertial reference frame, relative to which a particle is not accelerating, is to use a frame which is freely falling with the particle, as in our first two cases. And the fact that all particles or objects are accelerated the same way in a gravitational field means that the one freely falling reference frame is good for all objects in that frame, as long as there are no other forces present which distinguish one object from another. The conclusion of this reasoning is:

> An ***inertial reference frame*** is one which objects and observers are falling freely.

Now we must see how to take curvature into account mathematically.

7.3 Curvature and the mathematics of GR

First, a word on philosophy.[59] Coulomb's law suffered from the problem of "action at a distance", in violation of SR. Maxwell's equations replaced the static force by a field compatible with SR and resolved the problem for EM.

Newton's law of gravitation suffers from the same violation of SR as Coulomb's law. It also depends on an ill-defined notion of absolute space. Just as SR did for EM, GR replaces gravity by a field (and a force equation) and solves Newton's problem. It does so by *defining space and time as manifestations of a real, dynamical field, which is the gravitational field*. The metric $g_{\mu\nu}$ not only describes the variation of the field, it *is* the gravitational field.

The mathematical formulation of GR continues as follows:

- Space cannot vary too wildly or we could not calculate in it. So we assume that over a very small region of space we can use standard methods of differentiation. This assumption leads to the notion of ***manifolds***.
- In order to study curved space, we need to manipulate objects which enable us to write equations that satisfy the requirement that they be valid in all reference frames. Such objects are called ***tensors***.
- The object used to measure the space, the ***metric***, must be modified to become a function of spacetime, not just the constant one of Cartesian coordinates. It must reduce to the Cartesian metric in classical physics.
- In order to have valid differential tensor equations, the classical derivative must be replaced by a tensor called the ***covariant derivative***. In the limit of ordinary physics, this should – and does –

58 Some far-out current hypotheses like string theory suppose more (invisible to us) dimensions to spacetime. The curvature of these spaces could explain the other three forces, weak, strong and electromagnetic.

59 Thanks to Rovelli, 14-15.

reduce to the classical derivative. This allows us to calculate the change in a tensor over an infinitesimal distance of smooth, differentiable curved space.

- The tensor equations of motion must reduce to their classical versions under the conditions of classical physics (the principle of consistency).
- And, of course, the theory must agree with observation – hopefully new ones or ones which did not yet have an adequate explanation. This is indeed the case.

An example of the last requirement: SR says that fast-moving objects have slower clocks relative to those of a stationary observer. We shall see that GR says that objects in a stronger gravitational field (a more curved space) have clocks which slow down. So objects like GPS satellites, far above the earth, appear to have slower clocks than those on the surface of the earth because they are moving (SR), but faster because they are farther from the source (Earth) of the gravitational field (GR). Both these opposite but unequal effects must be taken into account by GPS software.

7.3.1. Manifolds

We must be careful now. The curvature of space can change as we move from one point to another. It can change differently around every point and in different directions from each point. In a "flat" Euclidean space we calculate the net effect of two vectors by translating the origin of one to the end of the other. But in a general curved space, there is no direct way to do this, since the vector may change as the coordinates do. Instead, we require that a curved space be such that, for an infinitesimally small distance around a point, the space behaves like an n-dimensional Euclidean space. Then we can translate by a succession of infinitesimal displacements. The totality of all these small regions constitutes a ***manifold***.[60] Note that a manifold has a well-defined topology and geometry and so should be thought of intrinsically, not extrinsically.

A manifold must obey the following conditions:

- In the immediate neighborhood of a point, the space is flat, meaning the first derivative (but not necessarily the second) *of a curve* in the space with respect to displacement in any spatial direction must be zero.
- From one point to another point infinitesimally separated from it, the space is smooth, allowing us to calculate the value of a function from one point to the next. At the next point, although the same two requirements must be satisfied, the shape of the space as well as the coordinates will have changed.

A tensor called the ***metric tensor*** may describe the distance between two neighboring points in a differentiable manifold. If the metric exists and is symmetric, the manifold is a ***Riemannian manifold***. This is the type of manifold used in GR.[61]

In order to pass from a vector or tensor field at one point P to the same field at an infinitesimal distance from P, we must employ a mathematical device called a ***connection***. In GR, this may be of different sorts.[62] We will consider only the covariant derivative. We can understand why we must use curved spaces by realizing that the surface of the Earth is such a space, in two dimensions. Measurement of the sum of the angles of a triangle on such a surface always gives a result greater than 180°, so the space is non-Euclidean. A saddle-shaped surface will give a sum of the angles of a triangle on the surface less than 180°. Also, only on a flat

60 Schutz 2009, 142: "This is the way to think of a manifold: it is a space with coordinates, that locally looks Euclidean but that globally can warp, bend, and do almost anything (as long as it stays continuous.)" On p 144, he adds: "The differentiable manifold itself is 'primitive': an amorphous collection of points, arranged locally like the points of Euclidean space, but not having any distance relation or shape specified. Giving the metric **g** gives it a specific shape..."

61 I don't know what the justification for this is, but it works.

62 In differential geometry or topology, one speaks of ***affine*** connections, of which a special case is the Levi-Civita connection, which in the case of manifolds reduces to the covariant derivative.

surface will a rectangular path, defined by four right angles and opposite sides of equal length, end where it started. And the surface of a sphere is equal to 4π times the square of its radius only in a flat, Euclidean space. So it is possible, although oft-times difficult to measure the geometry of a space.

7.3.2. Tensors

A ***tensor*** is a geometric object which makes a linear map of other objects, vectors or one-forms or a mixture (product) of the two, into a scalar. A tensor of type $\binom{0}{n}$ makes a linear mapping of n vectors into a scalar. A popular image shows something like a vending machine with n input slots for vectors: Put in the appropriate number of vectors and out pops a scalar!

Similarly, a tensor of type $\binom{m}{0}$ does a linear mapping of m one-forms into a scalar.

We start with single-rank tensors, with a single index (set of components), and define ***vectors***; denoted by one upper index, and ***one-forms*** (or ***dual vectors***), denoted by one lower index.[63] A one-form linearly maps a vector into a scalar and a vector linearly maps a one-form into a scalar (by a scalar product), so each is a tensor. The two are distinguished by the way they transform under a coordinate transformation, as follows in the table.

$\binom{1}{0}$ Vector	$V'^{\alpha} = \frac{\partial x'^{\alpha}}{\partial x^{\beta}} V^{\beta} \equiv \Lambda^{\alpha}{}_{\beta} V^{\beta}$ (7.1)
$\binom{0}{1}$ One-form (or dual vector)	$V'_{\alpha} = \frac{\partial x^{\beta}}{\partial x'^{\alpha}} V_{\beta} \equiv \Lambda^{\beta}{}_{\alpha} V_{\beta}$ (7.2)

Table 2: Vector and one-form transformation properties

We shall soon see that the transformation matrix for one-forms is the inverse of that for vectors.

One of the rules of tensor manipulation stipulates that a given index can exist only once or twice. If it exists twice, there must be one upper and one lower index and they are summed over. This is called the ***Einstein summation convention***.

The set of all possible vectors at a point p in spacetime is called the ***tangent space*** at p and constitutes a ***vector space***. (Think of the set of all the vectors attached to a point on a curved 2-d space embedded in a higher dimension.) Carroll stresses the importance of considering these objects as being located at a single point rather than extending from one point to another.[64] Similarly, the space of all linear maps from the vector space at p to real numbers constitutes the ***dual vector space***, also called the ***cotangent space***.[65] So the cotangent space is the space of all one-forms at a point.

7.3.3. Vectors, one-forms (dual vectors) and basis vectors

A vector may be represented as sums of components which are the coefficients of ***basis vectors*** $\vec{e}_{\beta}$[66]

$$\vec{A} = A^{\alpha}\vec{e}_{\alpha} = A^{\bar{\alpha}}\vec{e}_{\bar{\alpha}}. \tag{7.3}$$

Then, for the Lorentz transformation Λ,

$$A^{\bar{\alpha}} = \Lambda^{\bar{\alpha}}{}_{\beta} A^{\beta}$$

63 In days of yore, when I studied physics, vectors were called ***contravariant*** tensors and one-forms, ***covariant*** tensors.
64 Carroll (2913), 17.
65 Carroll (2013), 18.
66 These and following examples from Schutz, 37-38.

and some creative diddling of summed indexes leads to the transformation rule for the basis vectors[67]:

$$\vec{e}_\alpha = \Lambda^\beta{}_\alpha \vec{e}_\beta \tag{7.4}$$

So, by (7.2), basis vectors transform like one-forms.

A one-form maps a vector into a real number (scalar) and so can be represented as a scalar function of a vector:

$$\tilde{p}(\vec{A}) = \text{real number,}$$

where a tilde over a letter is used to designate a one-form. The ***components*** of a one-form are the result of mapping of the basis vectors:

$$p_\alpha := \tilde{p}(\vec{e}_\alpha)$$

which leads to

$$\tilde{p}(\vec{A}) = A^\alpha p_\alpha$$

and

$$p_{\bar{\beta}} = \Lambda^\alpha{}_{\bar{\beta}} p_\alpha$$

so components of a one-form transform like basis vectors. This inverse transformation leads to the frame invariance of

$$A^{\bar{\alpha}} p_{\bar{\alpha}} = A^\beta p_\beta.$$

Analogously to $\vec{A} = A^\alpha \vec{e}_\alpha$ for vectors, one-forms are defined in terms of basis one-forms ω by

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

and these formulae lead to the basis vectors and basis one-forms satisfying

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

so this equation defines the βth (vector) component of the αth basis one-form(!). $\vec{\delta}^\alpha{}_\beta$ is the Kronecker delta function. This equation may be taken as the definition of the relation between basis vectors and one-forms.

The ***tangent*** to a curve parametrized by λ is a vector

$$t^\beta = \frac{dx^\beta}{d\lambda} = \Big(\frac{dx^0}{d\lambda}, \frac{dx^1}{d\lambda}, \frac{dx^2}{d\lambda}, \frac{dx^3}{d\lambda}\Big) \tag{7.5}$$

which undergoes a coordinate transformation as follows

$$t'^\alpha = \frac{dx'^\alpha}{d\lambda} = \frac{dx'^\alpha}{dx^\beta}\frac{dx^\beta}{d\lambda} = \Lambda^\alpha{}_\beta t^\beta$$

like a vector should, hence its being written as t^α or $\vec{t}$.

The 4-dimensional ***gradient*** of a scalar field Φ is

$$\tilde{d}\phi = \Big(\frac{\partial\phi}{\partial t}, \frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}\Big) \tag{7.6}$$

which may be represented more generally as

67 Examples from Schutz, 37.

$$\tilde{d}\phi = \left(\frac{\partial\phi}{\partial x^0}, \frac{\partial\phi}{\partial x^1}, \frac{\partial\phi}{\partial x^2}, \frac{\partial\phi}{\partial x^3}\right)$$

and is a one-form. This is seen by using the product (Leibniz) rule, for instance,

$$V_{\bar{\mu}} = \frac{\partial\phi}{\partial x^{\bar{\mu}}} = \frac{\partial x^\nu}{\partial x^{\bar{\mu}}}\frac{\partial\phi}{\partial x^\nu} = \Lambda^\nu{}_{\bar{\mu}} V_\nu$$

like a good one-form.

To note:

- A ***basis vector*** is tangent to a coordinate curve, along which only one coordinate changes.
- A ***basis one-form*** is the gradient of a coordinate surface, where one coordinate is constant.
- Since the *bases* change with the coordinates (unlike Cartesian coordinates), they are called ***coordinate bases***. They are *not* orthonormal.

Since tensor equations are independent of the coordinates (of the bases), they are true in all coordinate systems if they are true in one. This of course is why they are useful in GR.

7.3.4. Of curves, surfaces and normals

A good example of the interplay of vectors and one-forms is the change of a scalar field along a path parametrized by a variable λ, the ***directional derivative***:

$$\frac{d\phi}{d\lambda} = \frac{\partial\phi}{\partial x^\mu}\frac{dx^\mu}{d\lambda} = \tilde{d}\phi_\mu t^\mu \tag{7.7}$$

where $\tilde{d}\phi$ is a one-form, the gradient, mapping a vector, the tangent to the curve, into a scalar which represents the rate of change of ϕ along the direction defined by the vector $\vec{V}$. One can also say that a vector is a tangent to a curve and is the function which returns $d\phi/d\lambda$ when It takes $\tilde{d}\phi$ as an argument.[68]

Some (unfortunately for those with weak eyes) oft-seen notation:

$$\frac{\partial\phi}{\partial x^\mu} = \delta_\mu\phi = \phi_{,\mu} \tag{7.8}$$

Don't miss that comma in the last subscript.

A one-form is considered ***normal*** to a surface if it maps all vectors tangent to the surface to zero,

$$\tilde{p}(\vec{t}) = 0$$

for all tangent vectors $\vec{t}$.

For a function f(x), assumed "well behaved", the gradient

$$\nabla f(x^\alpha)$$

is the steepest slope at a point. An isosurface

$$f(x^\alpha) = constant$$

defines a set of variables, a level set, of members on the surface. Then the gradient of the function at each point on the surface is normal to the surface. (Think of hills and contour lines.) Also,

$$\nabla f(x^\alpha)\cdot\hat{e}$$

68 Schutz, 122,

is the directional derivative along $\hat{e}$.

7.3.5. Higher-rank tensors

Higher-rank tensors can be formed by ***outer products*** of simple tensors:

$$X = V \otimes W \tag{7.9}$$

where $X^{\alpha\beta} = V^{\alpha}W^{\beta}$. In general, a tensor $X = V \otimes W$ transforms like this:

$$T^{\mu'_1..\mu'_j}{}_{\nu'_1...\nu'_k} = \Lambda^{\mu'_1}{}_{\mu_1}...\Lambda^{\mu'_j}{}_{\mu_j}\Lambda^{\nu_1}{}_{\nu'_1}...\Lambda^{\nu_k}{}_{\nu'_k}T^{\mu_1..\mu_j}{}_{\nu_1...\nu_k}.$$

The rank is the total number of one-forms and vectors concerned. As an example, consider two one-forms $\tilde{p}$ and $\tilde{q}$. Then their outer product $\tilde{p} \otimes \tilde{q}$ is a rank two tensor of type $\binom{0}{2}$. Operating on vectors $\vec{A}$ and $\vec{B}$ gives

$$\tilde{p}(\vec{A})\tilde{q}(\vec{B})$$

which we we call $\tilde{f}(\vec{A}, \vec{B})$. The most general tensor can be represented as a sum of outer products. For our rank-2 tensor with components

$$f_{\alpha\beta} := \tilde{f}(\vec{e}_{\alpha}, \vec{e}_{\beta}),$$

we can define basis $\binom{0}{2}$ tensors

$$\tilde{\omega}^{\alpha\beta} = \tilde{\omega}^{\alpha} \otimes \tilde{\omega}^{\beta}$$

and show that

$$\tilde{f} = f_{\alpha\beta}\tilde{\omega}^{\alpha} \otimes \tilde{\omega}^{\beta}.$$

7.3.6. The metric

In SR, the square of the invariant infinitesimal distance is given by (3.6) in incremental form

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 \tag{7.10}$$

which can be understood as a generalization of Pythagoras's famous formula. This could be written as

$$ds^2 = \eta_{\mu\nu}dx^{\mu}dx^{\nu} \tag{7.11}$$

where the metric $\eta_{\mu\nu}$ is a diagonal matrix with eigenvalues (+1,-1,-1,-1)[69] in Minkowski space. SR also defines the square of the ***proper time*** as

$$d\tau^2 := ds^2 \tag{7.12}$$

(Remember, it's the negative with the opposite metric convention.) In GR, this is generalized to a ***metric*** $g_{\mu\nu}$, which generally is not diagonal and has non-constant components. It may be written as a matrix or a rank 2 one-form. It has an inverse and may be used to raise or lower indices, as shown now and summarized in Table 3.

In terms of tensors, the general ***metric tensor*** g is defined by the ***inner product*** of two vectors as

$$\mathbf{g}(\vec{A}, \vec{B}) := \vec{A} \cdot \vec{B} \tag{7.13}$$

which is symmetric by its definition. The metric is a tensor of type $\binom{0}{2}$ and converts two vectors into a scalar. Also,

69 Also known as the signature of the metric.

$$\mathbf{g}(\vec{A},) := \tilde{V}() \tag{7.14}$$

converts a single vector into a scalar and is therefore a one-form. So we can write

$$V_\alpha := \tilde{V}(\vec{e}_\alpha) = \vec{V} \cdot \vec{e}_\alpha = \vec{e}_\alpha \cdot (V^\beta \vec{e}_\beta) = (\vec{e}_\alpha \cdot \vec{e}_\beta) V^\beta = \eta_{\alpha\beta} V^\beta$$

illustrating how the metric can be used to lower indices in a tensor equation. The opposite is also true. (If we used the -1,1,1,1 metric, g would simply invert the sign of A^0 on lowering its indices,).

Note that equation (7.13) defines the inner product of two vectors. In its simplest form in Cartesian coordinates in Euclidean space,

$$\vec{A} \cdot \vec{B} = g_{\mu\nu} A^\mu B^\nu. \tag{7.15}$$

In GR, the use of a metric is possible because we restrict ourselves to Riemannian manifolds. In this case, the metric g exists and can also be used for raising and lowering of indices.

$$V_\alpha = g_{\alpha\beta} V^\beta.$$

Metric	$\eta_{\mu\nu}$
Inverse metric	$\eta^{\mu\nu}$
Incremental distance and proper time	$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = d\tau^2$
Lowering of vector index (map vector into one-form)	$V_\mu = \eta_{\mu\nu} T^\nu$
Raising of one-form index (map one-form into vector)	$V^\mu = \eta^{\mu\nu} T_\nu$

Table 3: Uses of the metric with tensors

SR in a nutshell: Spacetime has a Minkowski metric.[70]

7.3.7. The covariant derivative

In a curved space, any displacement takes us from one tangent space to another one with different basis vectors. Changes in a vector's components therefore have two sources: the vector field itself and changes in the basis vectors, due to the curvature of space. We need to take the variation due to curvature into account in any differentiation. Using differentiation by parts on

$$V = V^\alpha e_\alpha,$$

we get for the β-th component

$$\frac{\partial}{\partial x^\beta}(V^\alpha e_\alpha) = \frac{\partial V^\alpha}{\partial x^\beta} e_\alpha + V^\alpha \left(\frac{\partial e_\alpha}{\partial x^\beta} \right)$$

Defining

$$\frac{\partial e_\alpha}{\partial x^\beta} = \Gamma^\mu_{\alpha\beta} e_\mu \tag{7.16}$$

meaning that $\Gamma^\mu_{\alpha\beta}$ is the μth component of $\dfrac{\partial e_\alpha}{\partial x^\beta}$ (its component along the e_μ direction), we can diddle indexes a little to get the ***covariant derivative***

$$\nabla_\beta V^\alpha = \partial_\beta V^\alpha + V^\mu \Gamma^\alpha_{\mu\beta} \tag{7.17}$$

70 Carroll (2022), 220.

or, in yet another notation,

$$\frac{DV^\alpha}{Dx^\beta} = \frac{\partial V^\alpha}{\partial x^\beta} + \Gamma^\alpha_{\mu\beta} V^\mu$$

This can be understood roughly as

> covariant derivative = vector change + coordinate change.

We can show that

$$\nabla_\beta g_{\mu\nu} = 0$$

or

$$g_{\mu\nu;\beta} = 0,$$

where the *semicolon* (See it?) indicates a covariant derivative. Then use this plus some laborious index diddling to find the value of $\Gamma^\nu_{\alpha\beta}$

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\sigma\mu} \Big(\frac{\partial g_{\alpha\sigma}}{\partial x_\beta} + \frac{\partial g_{\beta\sigma}}{\partial x_\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x_\sigma} \Big) \tag{7.18}$$

Instead of equation (7.18), it is often easier just to use equation (7.16) to calculate the gammas.

The set of $\Gamma^\alpha_{\nu\beta}$ are called ***connection symbols*** or ***Christoffel symbols***[71]. They represent terms added to the ordinary partial derivatives in order to correct for the curvature of spacetime, thus allowing us to calculate the value of a tensor field at an infinitesimal distance from the current value. They "connect" the tangent spaces at the two points. They may be viewed as needed to represent the parallel transport of the later vector back to the origin of the first for comparison.

Note that even in flat, Euclidean space where the connection symbols are zero for Cartesian coordinates (x,y,z), this is not true for polar coordinates.

For a one-form, the covariant derivative is different by a sign:

$$\nabla_\beta V_\alpha = \partial_\beta V_\alpha - V_\mu \Gamma^\mu_{\alpha\beta}.$$

As already noted, one more, worse notation for the covariant derivative is to use a semi-colon:

$$V^\alpha_{;\beta} = V^\alpha_{,\beta} + V^\mu \Gamma^\alpha_{\mu\beta}. \tag{7.19}$$

Be sure to distinguish the comma and the semi-colon.

For higher rank tensors, to do a covariant derivative of a general tensor with respect to β, ∇_β, take $\partial_\beta T^{\mu\nu\ldots}$ and

> for each upper index, V^α, add a term $\Gamma^\alpha_{\gamma\beta}(V^\gamma)$
>
> for each lower index, V_α, subtract a term $\Gamma^\gamma_{\alpha\beta} V_\gamma$.

So for rank-2 tensors, one gets:

$$\nabla_\beta T^{\mu\nu} = \partial_\beta T^{\mu\nu} + \Gamma^\mu_{\alpha\beta} T^{\alpha\nu} + \Gamma^\nu_{\alpha\beta} T^{\mu\alpha},$$

$$\nabla_\beta T^\mu{}_\nu = \partial_\beta T^\mu{}_\nu + \Gamma^\mu_{\alpha\beta} T^\alpha{}_\nu - \Gamma^\alpha_{\nu\beta} T^\mu{}_\alpha,$$

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

71 Or sometimes, depending on the circumstances, Levi-Civita symbols. Or, because of their formula, X-awful symbols.

These will be needed for deriving the Riemann curvature tensor.

In fact, all these results about covariant derivatives and connection symbols can be derived in two-dimensional Euclidean space using polar coordinates, in which the basis vectors depend on position, as in curved space.[72]

7.3.8. Parallel transport[73]

In order to derive equations of motion, we must consider movement *along paths* in spacetime. Then we can derive a geodesic, an extremized path equivalent to a Euclidean straight line, the shortest distance between two points. Now such a straight line is also the path which follows its own tangent. What interests us are vector and one-form fields, so we must consider the movement of such objects along a path.

In order to understand tensors in curved spacetime, we can start with vectors in *Euclidean* space. In order to compare two such vectors, we move one from its own origin to the head or tail of another, keeping it parallel to itself, then add or subtract the two. Such movement of a vector is called ***parallel transport***. This is expressed mathematically by requiring that the vector maintain its length and direction at all points, which can be expressed by

$$\frac{d\vec{v}}{d\lambda} = \frac{d}{d\lambda}\left(v^\alpha \hat{e}_\alpha\right) = 0 \Rightarrow \frac{dv^\alpha}{d\lambda} = 0,$$

where the curve followed by the vector is expressed as a function of a variable λ,

$$\vec{x} = x^\alpha(\lambda) e_\alpha.$$

In this case, parallel transport works because both points (in fact, all points) are in the same tangent space, the set of all possible vectors at a point in spacetime.

In GR, the curvature of spacetime prevents us from just adding or subtracting the field's values at different places because the spacetime also changes. In another words, the tangent space at each point along the curve is different, so the vector moves from one tangent space to another. This fact is what led us to the covariant derivative.

Consider the 2-dimensional case in figure 5, where the vector is parallel transported around the triangle on the curved 2-d surface from A to N and back to A, such that it maintains its direction in the tangent plane to the surface at each point along the curve. Remember, this is a 2-d space and this requires any vector to be a tangent to the point in question of the manifold. As it moves around the curve, the tangent plane does not just tilt, it is a new tangent plane, the 2-d equivalent of a tangent space. The tangent plane at each point has a specific orientation. At the end of the circuit, the vector has returned to its original tangent plane, but no longer points in the same direction. This is due to the curvature of the sphere's surface along the path followed by the vector.

72 Schutz, 118-135.

73 I found the notion of parallel transport difficult to grasp. Thanks to the following video (https://www.youtube.com/watch?v=p1tfZD2Bm0w&feature=youtu.be) and to Physics Forum (www.physicsforums.com/threads/understanding-parallel-transfer.911852/#post-5743713 and https://www.physicsforums.com/threads/understanding-parallel-transport.915267/).

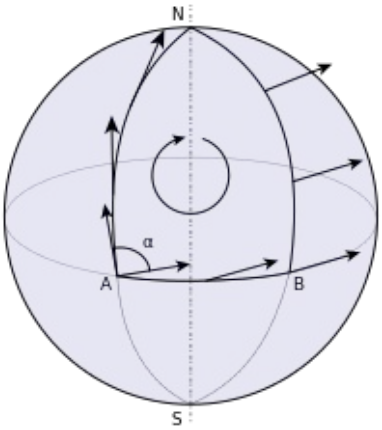


Figure 5: Parallel transport of a vector on a spherical surface, from Wikimedia Commons[74]

This shows clearly that there is no way to compare vectors in two different tangent spaces, so the idea of the relative velocity of, say, two galaxies has no meaning, since their velocity as we see it is not well defined.

> ***Expansion of the universe is better seen not as galaxies which are moving farther apart, but as change in the metric of the intervening space.***

There is a difficulty with Figure 5. We see that the 2-d tangent plane "tilts" in the extrinsic 3-d space in which we imagine it, in math speak, to be *embedded*. But we are unable (most of us, at least) of visualizing a 3-d space in a 4-d one somehow "external" to it. We are stuck with the 3-d intrinsic space. And the intrinsic geometry of curved spaces is just what Riemannian geometry was designed to handle.

In order to parallel transport a vector in curved space, we will employ the usual trick: We will require two vectors at infinitesimally close points to be parallel and of equal length. We must take into account the curvature of spacetime by using the covariant derivative. By analogy with the Euclidean case, we will define ***parallel transport*** of the vector along the path by requiring that the covariant derivative of the vector along the path vanish. Mathematically, we define the directional covariant derivative along the path as in (7.7), only we use the covariant derivative. The requirement that this be zero leads to the ***equation of parallel transport***:[75]

$$\frac{dV^\alpha}{d\lambda} + \Gamma^\alpha_{\nu\beta} V^\nu \frac{dx^\beta}{d\lambda} = 0. \tag{7.20}$$

Seen slightly differently, the derivative of a vector field along a path parametrized by λ is

$$\frac{d\vec{V}}{d\lambda} = \frac{\partial x^\mu}{\partial \lambda}\frac{\partial \vec{V}}{\partial x^\mu} = t^\mu \nabla_\mu V^\beta \hat{e}_\beta$$

where t^μ is the tangent to the path. So, using the covariant derivative, the condition for parallel transport of a vector becomes

$$t^\mu \nabla_\mu V^\beta = 0 \tag{7.21}$$

We need parallel transport because it will enable us to calculate a geodesic and identify the path followed by an object (say, a particle) on a pseudo-Riemannian manifold[76].

74 Parallel transport, https://commons.wikimedia.org/wiki/File:Parallel_Transport.svg.
75 Collier, 196- 7.
76 It's pseudo-Riemannian because the metric signature is not positive-definite, but ± 2 in our 4-d case.

7.3.9. Geodesics

In Euclidean geometry, the shortest path between two points is a straight line, which is a form of ***geodesic***, the extremized distance between two points. What interests us is that that this is the path followed by a moving particle subject to no external forces. One can state this more precisely: In Euclidean geometry, a particle follows a path in which the tangent to the path at a point is parallel to the tangent at the preceding point. In other words, the particle parallel-transports its own tangent vector, which explains the need for the preceding section.

We extend this idea into curved space by generalizing the statement that a ***geodesic*** is a path which parallel-transports its own tangent vector. Calling the tangent

$$t^\alpha = \frac{\partial x^\alpha}{\partial \lambda}$$

and using parallel transport, i.e., inserting t^α into (7.20), we see that a geodesic is represented by the ***covariant directional derivative***

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = \frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\rho\sigma}t^\rho t^\sigma = 0 \qquad (7.22)$$

which is therefore called the ***geodesic equation***. In this case, λ is an example of an ***affine*** parameter, a parametrization of a curve "such that the parametric equations for the curve satisfy the geodesic equation." Two affine parameters λ_1 and λ_2 are related by constants a and b such that $\lambda_2 = a\lambda_1 + b$. Or, if λ_1 is an affine parameter, then so is λ_2.[77]

7.3.10. The Riemann curvature tensor and friends

The Riemann curvature tensor is defined so as to measure the effect of curvature on a tensor which is parallel transported along different paths or along an infinitesimal loop. It is most easily derived by considering its definition in terms of non-commutating infinitesimal movements. We have seen in Figure 5 how parallel transfer of a vector around a closed path in curved space can lead to a different vector. We can consider an infinitesimal transport along a direction b, then along a direction c and compare this to the same transports in the opposite direction. This may be done by considering the commutator of the two operations:

$$[\nabla_c, \nabla_b]V_a = \nabla_c\nabla_b V_a - \nabla_b\nabla_c V_a \equiv R^d{}_{abc}V_d,$$

which defines the ***Riemann curvature tensor***:

This can be calculated by using the equations at the end of section 7.3.7 and, taking $\nabla_b V_a$ and $\nabla_c V_a$ to be rank-2 tensors. After some creative fiddling with indexes, one finds

$$R^d{}_{abc} = \partial_b(\Gamma^d_{ac}) - \partial_c(\Gamma^d_{ab}) + \Gamma^e_{ac}\Gamma^d_{eb} + {}^e_{ab}\,\Gamma^d_{ec}. \qquad (7.23)$$

It is based on the connection coefficients, which are products of the metric and partial derivatives of the metric.[78] So it's all made up from the metric. The metric knows all.

For a vector,

$$[\nabla_c, \nabla_b]V^a = \nabla_c\nabla_b V^a - \nabla_b\nabla_c V^a = R^a{}_{dbc}V^d$$

and this can be generalized to any type of tensor.

77 Planetmath.org, http://planetmath.org/affineparameter. Uh... I don't understand affine parameters.

78 Collier, 202, calls it "the Riemann curvature tensor, a glorious mixture of derivatives and products of connection coefficients."

Two other quantities used in GR are derived from the Riemann curvature tensor, the Ricci tensor and the Ricci scalar. The ***Ricci tensor*** is formed by contracting a pair of the Riemann tensor's indexes, either the first and last or the first and third. The result is the same give or take a plus or minus sign. Taking the first and third,

$$R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu} \tag{7.24}$$

The ***Ricci scalar*** is the trace of the Ricci tensor.

$$R = R^{\mu}{}_{\mu} = g^{\mu\nu} R_{\mu\nu} \tag{7.25}$$

Now we have defined or derived the quantities necessary to do GR. So let's get about it!

7.4 Back to physics

The principle of covariance must take into account curvature and for this, all this math in the form of tensors is needed, because:

> Tensor equations are independent of coordinates an so are the same in every reference frame.

Let's do some physics.

7.4.1. Four-velocity and four-momentum

Now we can have a slightly different view of what was discussed in section 3.3. The ***four-velocity*** $\vec{U}$ is defined to be a vector tangent to the ***world line*** of a particle, i.e., the path it follows through spacetime, and of length one unit of time in the particle's rest frame. This is equivalent to its time basis vector $\vec{e}_0$. If the particle is accelerated, the four-velocity is defined in an inertial frame which momentarily has the same velocity as the particle, the ***momentarily co-moving reference frame***, or ***MCRF***.[79] In any reference frame, it is the derivative of the components with respect to the proper time τ.

$$U^{\mu} = \frac{dx^{\mu}}{d\tau} = \Big(\frac{dt}{d\tau}, \frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau}\Big) \tag{7.26}$$

and using the chain rule, we find

$$U^{\mu} = \frac{dx^{\mu}}{d\tau} = \frac{dx^{\mu}}{dt}\frac{dt}{d\tau} = \gamma(1, \vec{v}). \tag{7.27}$$

Four momentum is simply the particle's four-velocity multiplied by its rest mass:

$$\vec{p} = m\vec{U} = (E, \vec{p}) = (\gamma m, \gamma m\vec{v})$$

where for small velocity

$$E := p^0 = m\gamma \approx m + \frac{1}{2}mv^2 \tag{7.28}$$

where c=1, as usual. Also,

$$\eta_{\mu\nu} p^{\mu} p^{\nu} = -m^2 = -E^2 + p^2$$

which gives us the dispersion relation of (3.15):

$$E^2 = m^2 + p^2 \tag{7.29}$$

None of this is really new, having been discussed in section (3.3).

79 Schutz (2016), 41.

7.4.2. The energy-momentum tensor

We know that gravity. i.e., the curvature of space, is brought about by energy and mass, which are equivalent in GR. So we must express mass and energy in the form of a tensor. Cosmologists do this by taking advantage of the ***Cosmological Principle***, which states that on very large (humongous) scales, the universe is homogeneous and isotropic. The universe is considered to be a continuum composed of ***elements*** which can be treated as points, each one of which possesses unique values for various properties such as energy and momentum. In other words, space is filled with ***fields***, each of which has a value at each point.

As a simplest case, cosmologists define ***dust*** as a collection of particles which are all at rest in some Lorentz frame, their mutually co-moving reference frame. In the MCRF, dust particles have a density of particles, or *number density*, n, which may vary from point to point. When they are moving relative to an observer, length contraction means that this observer sees a density γn. If the dust is moving with velocity $\vec{v}$, then the flux across a surface of constant $\vec{x}$ is $\gamma n v^{\vec{x}}$. We can thus construct the ***number-flux four-vector*** $\vec{N}$ from the four-momentum $\vec{U}$ as

$$\vec{N} = n\vec{U} = \gamma n(1, \vec{v})$$

In its MCRF, the ***energy density*** of dust is the number density times the energy of each particle, which is m in this frame.

$$\rho = nm \tag{7.30}$$

Seen from a moving frame, both n and m will change by a factor of γ, so that $\rho \to \gamma^2\rho$, This means the general energy density is not a vector, but a rank-2 tensor.

The ***energy-momentum tensor*** (or ***stress-energy tensor***) is defined such that its $\mu\nu$ component is the flux (rate of flow) of the μth component of four-momentum across a surface of constant x^ν. We can write it in any frame and the simplest is the MCRF where for dust it has only one term:

$$(T^{00})_{MCRF} = \rho = nm.$$

This suggests generalization of the tensor to

$$T = \vec{p} \otimes \vec{N}. \tag{7.31}$$

Then

$$T^{\alpha\beta} = mU^\alpha nU^\beta = \rho U^\alpha U^\beta, \tag{7.32}$$

which makes it clear that T is a rank-2 tensor (a vector) and is symmetric, $T^{\alpha\beta} = T^{\beta\alpha}$. Since the energy-momentum tensor component $T^{\alpha\beta}$ is the flux of α momentum across a surface of constant x^β, we see that:[80]

- T^{00} is the flux of 0-momentum, i.e., the energy, across no spatial surface, but only in time; it is therefore the energy density;
- T^{0i} is the 0-momentum (energy) flux across surface x^i, e.g., heat conduction;
- T^{i0} is the momentum density;
- T^{ij} is the flux of momentum I across surface x^j, called the ***stress***.

Remember symmetry:

80 Schutz, 92.

$T^{i0} = T^{0i} \Rightarrow$ i-th momentum density = energy flux across i-th surface.[81]

It is straightforward to write out the components of the tensor for dust using equation (7.32). For instance:

- $T^{00} = \rho U^0 U^0 = \gamma^2 \rho$,
- $T^{0i} = \rho U^0 U^i = \gamma^2 \rho v^i$.
- $T^{i0} = \rho U^i U^0 = T^{i0} = \gamma^2 \rho v^i$,
- $T^{ij} = \rho U^i U^j = \gamma^2 \rho v^i v^j$

A second special case, somewhat more general than dust, is that of a ***perfect fluid*** , which is composed of elements which have no sliding forces (viscosity) between them, so $T^{ij} = 0$ for $i \neq j$. It amounts to dust plus pressure, without heat conduction or viscosity in the MCRF. In this case, the energy-momentum tensor in the MCRF, where $U^i = 0$, is diagonal and its (0,0) element has the same value as that of dust, ρ. The spatial diagonal elements are momentum fluxes, i.e., changes in momentum per unit area, and changes in momentum are forces, according to Newton, who is correct in the MCRF. So the spatial diagonal elements all are equal to the pressure, p. In its generalized tensor form, then, the energy-momentum tensor for a perfect fluid in its MCRF is

$$(T^{\mu\nu}) = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \tag{7.33}$$

which can be written

$$T^{\mu\nu} = (\rho + p) U^\mu U^\nu + p g^{\mu\nu}, \tag{7.34}$$

since $U^0 = \gamma = 1$ and $U^i = 0$ in the MCRF.

In general, the "… energy-momentum tensor summarizes everything we want to know about the mass, energy, momentum, pressure, stress and other energy-like features of a collection of matter (or radiation, or anything else)."[82] And gravity can be created by all these.

7.4.3. Einstein's gravitational field equations

Einstein spent almost ten years figuring out the equations of GR so that they would obey the three principles of covariance, consistency and equivalence. Thanks to him, we now can write the result simply as

$$\boxed{R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = -\kappa T_{\mu\nu}} \tag{7.35}$$

where

$$\kappa = 8\pi G.$$

Let's look at the pieces:

- $R_{\mu\nu}$ is the Ricci tensor and R is the Ricci scalar; both are functions of the curvature of spacetime via the connection coefficients and, therefore, the metric;
- $g_{\mu\nu}$ is the metric and describes the distance between two points, a function of the curvature of spacetime;

81 Collier, 218.
82 Carroll (2022), 223.

- G is Newton's gravitational constant, which is where gravity comes in;
- $T_{\mu\nu}$ is the energy-momentum tensor expressed as a double one-form or tensor of type $\binom{0}{2}$.

So the left-hand side as it is written represents the curvature of spacetime expressed as a sum of functions of the metric; the right-hand side, the gravity associated with energy and momentum in space. The equation(s) therefore express the relation between curvature and energy-momentum.

If we include the cosmological constant, we get:

$$\boxed{R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = -\kappa T_{\mu\nu}} \tag{7.36}$$

which can be rewritten as

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = -\kappa\Big(T_{\mu\nu} + \frac{\Lambda}{\kappa}g_{\mu\nu}\Big). \tag{7.37}$$

This formulation makes it clear that the cosmological constant may be seen as contributing to the energy-momentum of spacetime through the tensor

$$T^{(\Lambda)}_{\mu\nu} = \frac{\Lambda}{\kappa}g_{\mu\nu}.$$

In the MCRF, if we assume a perfect-fluid model for $T^{(\Lambda)}_{\mu\nu}$ as in equation (7.34), then from the matrix elements [ρ_Λ, p_Λ], we can read off the values of the momentum and pressure due to the cosmological constant and find that

$$\rho_\Lambda = \frac{\Lambda}{\kappa} = -p_\Lambda. \tag{7.38}$$

Surprise! This shows that a positive density, ρ_Λ, due to the cosmological constant gives a ***negative pressure***, p_Λ. The field equation can be split into parts due to matter (ordinary and dark) and the vacuum (the cosmological constant):

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = -\kappa\Big(T_{\mu\nu} + T^{(\Lambda)}_{\mu\nu}\Big) = -\kappa\Big(T_{\mu\nu} - \rho_\Lambda g_{\mu\nu}\Big). \tag{7.39}$$

The vacuum (cosmological constant) part of the right-hand side has opposite sign to the energy-momentum tensor, so if the former is the source of gravitational attraction, the latter must be a source of ***repulsion***, driving the universe to ***expand***. This is the basis of the de Sitter model, discussed in paragraph 7.8.3.

It is interesting to note that for Einstein, Λ is a constant of nature like Planck's constant. From the point of view of modern QFT, it is the energy associated with the vacuum expectation value (VEV) of a yet to be discovered quantum field.

Physicists sometimes use the ***Einstein tensor***, defined as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}. \tag{7.40}$$

Then equation (7.35) looks simpler:

$$G_{\mu\nu} = -\kappa T_{\mu\nu}. \tag{7.41}$$

As we shall see, it is possible to decompose the energy-momentum tensor further, into parts for matter, radiation and the vacuum (or cosmological constant, or dark energy).

In fact, Noether's theorem tells us that the energy-momentum tensor satisfies

$$\partial_\nu T^\nu_{\ \mu} = 0$$

so, by (7.41), this is also true for $G^\nu_{\ \mu}$ and the only quantity having this property is exactly the Einstein tensor![83]

Let's resume:

In GR, the connection and the metric are two conceptually different but not independent concepts. The connection allows us to take differentials on the manifold by taking into account the curvature of spacetime over an infinitesimal displacement. The metric is used to measure (infinitesimal) distances. Parallel transport of a vector along a closed curve and comparison of the difference between its initial and final states allows calculation of the Riemann tensor in terms of the connection coefficients, which can in turn be expressed in terms of derivatives of the metric. So the left side of Einstein's field equation of GR contains various derivatives of the metric and may be solved for the metric. (Good luck!)

7.4.4. A bit of philosophy

There is a double or mutual dependence in GR: spacetime, its curvature, depends on the energy (fields) within it; but the fields can only be expressed in terms of their extension or action in spacetime. In particular, the gravitational field does not exist within spacetime, it *is* spacetime. Flat spacetime can only exist where there is nothing, in which case it does not matter to us. Einstein himself said:

> *There is no such thing as an empty space, i.e., a space without field... Spacetime does not claim existence on its own, but only as a structural quality of the field.*[84]

This is a little like saying that space is only there where you can measure it. And to measure it, something must be there (a ruler).

7.4.5. Consistency – the equations of motion

According to the principle of consistency, these equations must reduce to those of Newton in the appropriate limit.

In an inertial frame where no gravity is present, where the connection coefficients are zero, the geodesic equation (7.22) reduces to

$$\frac{d^2x^\alpha}{d\tau^2} = 0$$

where the affine parameter λ has been chosen to be the proper time τ. In the non-relativistic limit, τ becomes t, time, and the equation says that the acceleration is zero, consistent with Newton's first law of motion.

The GR equivalent of Newton's second law

$$\vec{F} = m\vec{a} = \frac{d\vec{p}}{dt}$$

is easily shown to be

$$F^\mu = \frac{DP^\mu}{d\tau}$$

83 Schwichtenberg, PS, 246.

84 Einstein, *Relativity: The special and general theory (A popular exposition)*. Quoted by Al-Khalili, *The world according to physics,* 74.

and has the same form.

Newton's third law is complicated in GR. Usually, one assumes a large source of gravitation, say a star or a black hole, and considers the way a smaller particle moves along a geodesic in its gravitational field. Turning that around to consider the effect of the particle's tiny mass on the black hole's gravitational field is messy, and according to at least one authority, not addressed by Einstein's GR. So that subject is clearly beyond the scope of our study.

As for the law of gravitation, it suffices here to indicate that in a weak, static gravitational field, the force on a slowly moving particle can be shown to be equal to that of Newton's equation in the form of ***Poisson's equation***

$$\nabla^2\phi = \left(\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2}\right) = 4\pi G\rho$$

where ϕ is the gravitational potential field.[85]

7.5 What GR tells us

A general solution of the Einstein field equation is impossible. But several solutions have been found based on approximations or limiting cases. These are summarized in the following table.

Model or metric	Simplifying assumptions	Elements	Notes/validity
Schwarzschild	Spherically symmetric, static gravitational field due to single massive spherical body in otherwise empty space	Schwarzschild radius	Stars and black holes
Reissner-Nordström	Charged non-rotating body		
Kerr	Rotating black hole		See below
Robertson-Walker	Most general metric conforming to cosmological principle	Scale factor R(t) and curvature k = -1, 0, or 1	Metric of Friedmann equations
Perfect fluid		Dark matter and energy	$\rho_\Lambda > 0$ => $p_\Lambda < 0$ means dark energy
Plugging R-W metric into Einstein's field equation gives Friedmann equation(s), which lead in turn to several metrics.			
Empty universe ($\rho = 0$)	Empty space	Flat, static Minkowski space (k = 0) or negatively curved expanding universe (k = -1)	
Static Einstein universe	Static solution because of inclusion of cosmological constant	R(t) = constant, $\Lambda > 0$	k = +1, positive curvature, unstable
De Sitter universe	Only dark energy	$\rho_{m,0} = \rho_{r,0} = 0$	Recent universe (>9.8 Gy)
Radiation-only		k = 0, $\rho_{m,0} = \rho_\Lambda = 0$	Early universe (0-50 Ky)
Einstein-de Sitter universe	Mass only (dust)	$\rho_{r,0} = \Lambda = 0$	Mid-term universe; critical density Ω for k = 0

Table 4: Approximate solutions to Einstein gravitational field equations

85 Collier, 234.

We will look at them one by one.

7.6 The Schwarzschild metric, stars and black holes

The Schwarzschild metric, in polar coordinates, is as follows:

$$ds^2 = \Big(1 - \frac{2GM}{r}\Big)dt^2 - \Big(1 - \frac{2GM}{r}\Big)^{-1}dr^2 - r^2 d\Omega^2, \qquad (7.42)$$

where

$$d\Omega^2 = d\theta^2 + sin^2\theta d\phi^2 \qquad (7.43)$$

This metric assumes a spherically symmetric, static gravitational field in a vacuum outside a single massive spherical body which is the source of the field. The metric must be asymptotically flat, meaning that it must describe ordinary flat spacetime of SR at a large distance from the massive body. The object should not be rotating, but the solution can still be used for a slowly rotating object, so it is good outside a star like the sun or outside a black hole. It ignores the field due to the mass of the observer, generally a test particle. The parameter M is the mass within the radius r and need only be spherically symmetric. G is of course Newton's gravitational constant.

This metric is both ***stationary*** (invariant under translations in time) and ***static*** (invariant under time reversal, $t \to -t$), as it should be in order to be physically acceptable. It is also spherically symmetric, since $dt = dr = 0$ for a given value R of r gives

$$ds^2 = -R^2 d\theta^2 - R^2 sin^2\theta (d\phi)^2. \qquad (7.44)$$

Although this may look like it depends on θ, that is just because of the spherical coordinate system used. This shows that we must be careful in GR to avoid what look like physical results but which are merely artifacts of the coordinates chosen.

> "In general relativity, coordinates do not have immediate metrical significance... Intervals of time and distance must be measured by an observer who must make use of a frame of reference..."[86] Stated differently: "The coordinates don't have any meaning until the metric gives meaning to them, and the metric components make sense only with respect to some specified coordinates."[87]

Three frequently-cited "local" frames (local to an observer) are the frame of a freely falling observer, a frame at rest at some position and the frame of a distant observer. More on this in section 7.6.1.

Birkhoff's theorem, proven by George Birkhoff in 1923, states that as long as source of gravitation has isotropic effects, static or not, the vacuum solution to Einstein's field equation outside the source is the stationary Schwarzschild solution.[88] Even if the source is expanding, contracting or even pulsating radially, the Schwarzschild solution is correct outside the source. Any radial motion will have no effect outside the source.

The derivation starts by assuming a metric of the following form, which is the only one which satisfies the above conditions.[89]

$$ds^2 = U(r)dt^2 - V(r)dr^2 - W(r)r^2(d\theta^2 + sin^2\theta d\phi^2).$$

This resembles the Minkowski metric in spherical coordinates

$$ds^2 = dt^2 - dr^2 - r^2(d\theta^2 + sin^2\theta d\phi^2).$$

86 Lambourne, 155.
87 Carrol 2022, 242.
88 Lambourne, 154.
89 Collier, 239-245.

First, W is assumed to be 1, meaning that we can no longer take r to be the simple radial distance. At this point, we are just writing a solution in terms of some coordinates but we do not know yet just what they represent, aside from the fact that r^2 enters in such a way that a sphere at r will have area $4\pi r^2$ and circumference $2\pi r$. The metric matrix elements are then used to determine the Christoffel symbols and the Ricci tensor. Outside the mass M, the energy-momentum tensor must be zero. This fact is then used to determine the variables in the final Schwarzschild equation (7.42).

For very large values of r, the metric reduces to the metric for "flat" Minkowski space in polar coordinates, this being referred to as ***asymptotic flatness***. This is also true if there is no massive object, the case M=0. But the variables r and t represent real radial distance and clock time only when $M = 0$; otherwise, they are modified by the curvature of spacetime, and so are what we have called ***coordinate (bases) variables*** in section 3.3

The equation (7.42) indicates a real singularity at r=0, but the seeming singularity at r = 2GM, called the ***Schwarzschild radius***, is a coordinate singularity, due only to the choice of coordinate system. It can be eliminated, e.g., by converting to so-called ***Kruskal coordinates*** or ***Eddington-Finkelstein coordinates***, In either case, the singularity at the event horizon disappears. Nothing blows up at $r = 2Gm$.

It seems like cosmologists spend a lot of time searching for new coordinates which will better divulge the meaning of equations. Nevertheless, the ***Schwarzschild radius*** is defined as $R_S = 2GM$, remembering that we are using c=1, or that would be $R_S = 2GM/c^2$.

As examples, the Schwarzschild radius R_S measures about 9 mm for a body with the mass of the Earth and 3 Km for that of the Sun, well within the physical boundary of each one.

7.6.1. Proper time and distance

The meaning of the Schwarzschild coordinates remains arbitrary, simply distinguishing one event from another until we can relate them to intervals of space or time measured by observers. Using the Schwarzschild metric, one can derive the relation between the ***proper time*** of an event viewed by a distant observer, say at a nearly infinite distance, and that viewed by an observer at the event.[90]

$$d\tau_\infty = \left(1 - \frac{2GM}{r}\right)^{-1/2} d\tau \qquad (7.45)$$

To the distant observer, clocks will seem to run more slowly the closer they are to the earth's surface, which may be paraphrased roughly as "gravity makes time run slower." But this is not true: A clock in any system runs at 1 second per second, but the corresponding coordinate time is changed.This phenomenon of ***gravitational time dilation*** must be taken into account by GPS systems. This is not the same as the time dilation of SR, which is symmetric between two inertially moving observers.

Alternatively, one can say that *a body tends to move to where time runs more slowly.* That's where gravity is strongest, where the curvature of spacetime is greatest.

From the same equation (7.45), the frequency of light emitted by the owner of the clock in question will be seen by the distant observer as lower by

$$f_\infty = \left(1 - \frac{2GM}{r}\right)^{1/2} f_{em}$$

and so shifted towards the red (at least, if the emission frequency is higher than that of red, which is the case for most visible light). This is the ***gravitational redshift*** and has nothing to do with that of SR or a whatsoever velocity-related Doppler effect.

90 Collier, 248-249.

The geodesic equation of the Schwarzschild metric is messy to calculate. But once known, it can be used to predict both the the precession of the perihelion of Mercury, the experimental confirmation of which made Einstein world-famous; and the deflection of light in a gravitational field, predicting and explaining the phenomenon of ***gravitational lensing,*** which not only has been observed, but has become a standard tool of astronomers.

By analogy with the proper time, for which spatial distance is zero, we can define the ***proper distance***, $d\sigma$, as the distance with time fixed.[91]

$$d\sigma^2 = -ds^2 = \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2,$$

which for fixed r is

$$d\sigma^2 = ds^2 = r^2 d\Omega^2 = r^2 d\theta^2 + r^2 sin^2\theta d\phi^2.$$

This is just the line element for the surface of a sphere in 3-dimensional Euclidean space. The circumference of a circle in the $\theta = \frac{\pi}{2}$ plane is then $C = 2\pi r$. So for constant coordinate distance r and coordinate time t, the Schwarzschild metric defines a sphere in Euclidean space. "Effectively, for each constant value of coordinate time t, Schwarzschild spacetime can be thought of as a series of nested Euclidean spheres (think of a set of Russian dolls), each one representing a different value of r."[92] For fixed angles ($d\theta = d\phi = 0$), we have

$$d\sigma = \left(1 - \frac{2GM}{r}\right)^{-1/2} dr > dr.$$

Not only is $d\sigma$ greater than dr, it gets larger as we move inwards but approaches dr as we move outwards towards infinity.

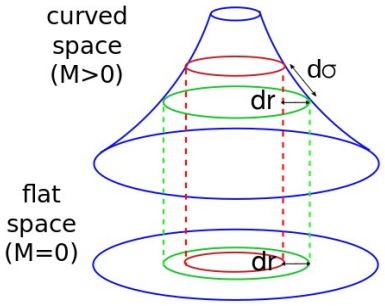


Figure 6: Embedding diagram of radial and proper distance in Schwarzschild spacetime, after Collier

This can be understood somewhat more intuitively by an ***embedding diagram***, showing how the difference between $d\sigma$ and dr entails a curved space (Figure 6).[93] The lower part shows the increase in size of the red circle as r is incremented by dr to the larger green circle in a flat Euclidean space. The upper part shows the same thing in curved space, where the curvature of space requires the proper distance $d\sigma$ to increase by an amount greater than dr. It also shows that $d\sigma$ approaches dr as both move outwards (downwards, in the diagram).

91 There is something here I do not understand. If dt=0, then the speed limit of SR would forbid any change in spatial coordinates.
92 Collier, 255.
93 Collier, 256.

7.6.2. A bit of history

This is about as good a time as any for a bit of history.

In 1905, Einstein published the special theory of relativity (SR). In 1907, he started thinking about free fall and acceleration, finally presenting his theory of general relativity to the Prussian Academy of Science in November 1915.

Karl Schwarzschild quickly found the metric named after him in December 1915, but unfortunately died of an autoimmune disease in 1916 while serving on the Russian front.

In 1939, Robert Oppenheimer and one of his students, Hartland Snyder, used relativity to show that someone falling into a black hole was never seen by a distant observer to arrive at the event horizon, although the observer himself definitely did get there in a finite (and for her all-too-short) period of time.

The term “black hole” was first used by theoretical physicist John Wheeler in 1967.

7.6.3. Schwarzschild black holes

According to NASA, “[b]lack holes are really just the evolutionary end points of massive stars.”[94] This is true, but deserves some amplification. A giant star of mass above 2-3 solar masses will eventually run out of nuclear fuel to “burn”. The reduced pressure due to the decreased burning is then not strong enough to counter the gravitational attraction of all that mass and the star will eventually implode and form a black hole. Black holes are well described by the Schwarzschild equation.[95]

It is an understatement to say that black holes have a number of interesting properties.

The gravitational field of a black hole is so strong that anything – even light – which falls into it, i.e., which gets within a distance of the Schwarzschild radius of the central singularity, will be drawn toward the singularity without any possibility of escape, as to do so would require a velocitiy greater than the speed of light. So the Schwarzschild radius describes a surface of no return, the ***event horizon***. Although the existence of the event horizon is a defining characteristic of black holes, it is not a rigid surface. Outside the event horizon, one could pass close to the black hole and then move away – if one calculates his orbit correctly. The black hole is not a cosmic vacuum cleaner, sucking up everything around it. But anything or anyone penetrating the event horizon is indeed doomed.

Since the black hole’s gravity keeps in everything which penetrates the event horizon, including light, the black hole is … black – and so invisible to observers. As light approaches it, a distant observer sees it as being increasingly gravitationally red-shifted until it also becomes black and disappears from view.

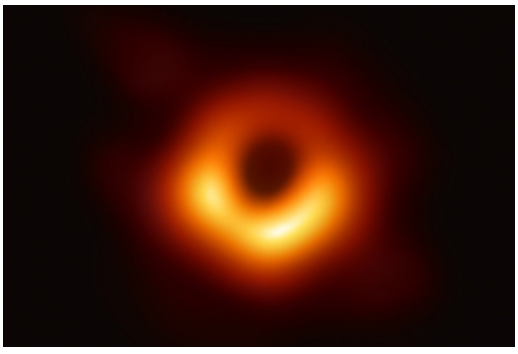


Figure 7. Photograph of a black hole by NASA.[96]

94 NASA chat. “Why do black holes suck? Or do they?” https://www.nasa.gov/connect/chat/black_hole_chat.html.
95 See my Natural Universe Part I for more details on the evolution of stars.
96 First image of a black hole. https://solarsystem.nasa.gov/resources/2319/first-image-of-a-black-hole/

In April 2019, astronomers using the globally distributed Event Horizon Telescope managed to obtain an image of a black hole (Figure 7). Or, rather, of the bright ring of light around it formed by gases accelerated by the BH's intense gravitational field. This particular BH is in the center of galaxy M87.

Interestingly, already near the end of the 18th century, John Michell and, slightly later, Pierre-Simon Laplace, conjectured that light, considered at the time to be composed of particles, would be attracted by gravity and so could only escape from a a planet's gravitational field if its speed, c, exceeded the escape velocity given by

$$c = \sqrt{\frac{2GM}{R}}$$

meaning that it could not escape if it were at a radius

$$R < \frac{2GM}{c^2}$$

which is – only coincidentally – the Schwarzschild radius. Neither of these men knew anything about black holes, but they theorized that very heavy stars would be dark because light could not escape from them. Laplace surmised that the largest objects in the universe might be invisible to us.

The gradient of the gravitational field within the event horizon is so strong that any poor astronaut who had the mishap of falling feet first "into" the hole, i.e., penetrating the event horizon, soon would feel a much stronger force on her feet than on her head so that, instants before before being torn apart, she would be stretched out, what cosmologists jocularly refer to as "spaghettification". Such forces due to the non-uniformity of a gravitational field are called ***tidal forces***.

It is possible to find approximate solutions to the equations both for distant observers and for the poor in-falling astronaut. Strangely enough, the distant observer of the doomed astronaut approaching the event horizon would "see" the astronaut's time slow down due to gravitational time dilation, as in equation (7.45). Every incremental distance of the falling astronaut would last longer, so that the observer would in fact never see the her fall in, i.e. disappear. This is also true if it is not an astronaut falling in, but the surface of a compacting star in the process of forming a black hole. One cannot see a black hole form because there is not enough time.

The unfortunate astronaut's clock, showing her proper time, does not slow down though, so she does indeed fall into the black hole, as does the surface of a collapsing star.

Incidentally, this is a good way to travel forward in time: Lurk for an hour or so close but not too close to a black hole event horizon, then go home to find centuries have gone by. 'S true.

Stranger yet, the Schwarzschild metric (7.42) shows clearly that inside the event horizon, the dt term becomes positive and the dr term, negative. In some sense, the coordinates of time and radial distance have exchanged rolls, so that falling in distance has become falling in time and vice versa. This can be seen in the behavior of the light cones in Figure 8, in Schwarzschild coordinates. It is clear that the light cones inside the event horizon all point towards the central singularity, so our in-falling astronaut will no choice but to be drawn to it.

However, in Eddington-Finkelstein coordinates, the light cones lean over as they approach the event horizon, where they become null (45°) and transition smoothly to leaning further so that they point towards the central singularity.

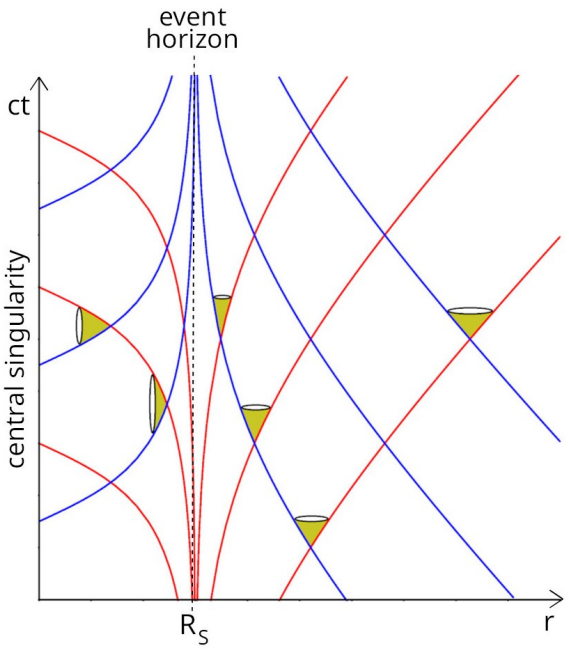


Figure 8. Schwarzschild null geodesics and light cones, after Collier.

Cosmologists say that “black holes have no hair”, known as the ***no-hair theorem***. This statement refers to the fact that black holes have only three properties which can be measured by an external observer: mass, angular momentum and electric charge. Since all black holes necessarily possess mass, this leaves four kinds, with or without rotating or being electrically charged. Cosmologists have found several metrics to describe black holes with combinations of these properties:

Metric	Properties taken into account
Schwarzschild	Only mass
Kerr	Mass and angular momentum
Reissner-Nordström	Mass and electric charge
Kerr-Newmann	Mass, angular momentum and electric charge

Table 5: Black-hole metrics

A non-rotating black hole is spherical, but a rotating one will bulge out at its equator. The Kerr metric for a rotating black hole defines a region called the ***ergosphere*** where spacetime is ***dragged*** in the direction of rotation at a speed greater than the speed of light relative to the rest of the universe. This is due to an effect called the ***gravitomagnetism***. It is so named by analogy with the Lorentz force of electromagnetism because in the presence of such a field, the gravitational acceleration of a particle depends on its velocity.[97]

As matter is attracted toward a rotating black hole, it generally follows a path around the hole, forming a flat disk around the it, an ***accretion disk***. Such matter loses energy through heat and radiation, mainly as X-rays, and spirals into the black hole. The radiation follows paths curved by the black hole’s gravity. Observation of this radiation has identified many possible black holes.

Streams of matter also emerge in narrow ***jets*** with opening angles of <5°, thought to be emitted perpendicularly to the plane of the accretion disk and in both directions. Jets may be due to magnetic fields produced by the rotating disk and are an important subject of black-hole studies.[98]

97 Schutz (2003), 245.
98 For more on accretion disks and jets, see Abramowicz and Fragile, “Foundations of black hole accretion disk theory”, online at https://arxiv.org/pdf/1104.5499.pdf.

Figure 9: Jets from super-massive black hole in the center of the 4C+29.30 galaxy, from NASA.[99]

It is predicted that black holes also possess a finite temperature and entropy. If this were true, black holes would not be completely black. QM indeterminacy would form matter-antimatter pairs near the surface, using energy from the black hole itself. If one member of such a pair is closer and falls into the black hole, the other may escape, carrying off some of the energy. This is called ***Hawking radiation***, after its discoverer. If this happens, then black holes may eventually decrease in mass until they "fade" away over a very long period of time, the rate of Hawking radiation being tiny.[100] The effects of entropy and information in the case of black holes are widely debated, partly because there is no way to test the theories, the singularity in the center of the black hole being unattainable.[101]

7.7 The cosmological principle

Cosmologists have taken advantage of the cosmological principle to look for a metric describing a spacetime which is homogeneous and isotropic. Reminder: The cosmological principle states that on very large scales, the Universe is homogeneous and isotropic. You must go to really, really large scales for this to be true.

7.7.1. Robertson-Walker metric

The most fruitful method so far has been to idealize the case as follows. Spacetime is supposed to evolve in time so that at each value of time it consists of a hypersurface which is homogeneous and isotropic. This assumption leads to the ***Robertson-Walker metric***, shown to be the most general possible metric which describes such a spacetime. It can be written as follows:

$$ds^2 = -dt^2 + R^2(t)\Big[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\Big] \tag{7.46}$$

where

$$d\Omega^2 = d\theta^2 + sin^2\theta d\phi^2$$

The coordinates of this metric are rather special. The time coordinate t represents the ***cosmic time***, i.e., time measured by an observer whose motion is only due to the expansion of space, called a ***fundamental observer***, whose ***peculiar motion*** is zero. This would be the time seen by a galaxy, for example. The spatial coordinates, called ***co-moving coordinates***, are measured by a fundamental observer and are constant with time, all time dependence being found in the ***scale factor***, $R(t)$, which shows how big the space part is as it increases (or decreases) with time for an expanding (or contracting) universe. To repeat, the spatial

99 Black-hole powered jets plow into galaxy. https://www.nasa.gov/multimedia/imagegallery/image_feature_2510.html
100 Can you imagine something black fading?
101 See, for instance, Sabine Hossenfelder, "Why do physicists worry so much about the black hole information paradox?". http://backreaction.blogspot.fr/2017/04/dear-dr-b-why-do-physicist-worry-so.html

coordinates do not change with time: Each galaxy has a fixed set of spatial coordinates which lead to a temporally increasing metric only because of the scale factor. The parameter k is the ***curvature*** and is considered (i.e.,normalized) to take on the values -1, 0 or +1. Any other integer can be reduced to one of these three by appropriate normalization of the coordinate r,[102]

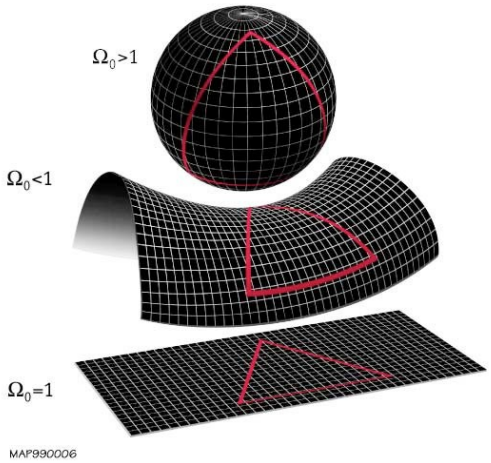


Figure 10. Spherical, hyperbolic and flat spacetime.[103]

The three permitted values of curvature lead to different types of universes:

- In the case of k=0, this reduces to ***flat*** Euclidean space ($\Omega_0 = 1$).
- For k=+1. It can be reduced to the metric for the surface of a three-sphere and is called the ***closed***, or ***spherical*** R-W metric ($\Omega_0 > 1$).
- For k=-1, it gives a hyperbolic, or ***open***, R-W metric ($\Omega_0 < 1$).

The density parameter Ω_0 will be explained in section 7.8.6.

7.7.2. The expanding universe

Consider only the radial part of the distance at a fixed time for constant θ and ϕ, The proper distance is

$$d\sigma = R(t)\Big[\frac{dr^2}{1-kr^2}\Big]^{1/2}.$$

This can be integrated for the three cases and, after some manipulation, all three give the same result:

$$\frac{d\sigma}{dt} = \Big(\frac{1}{R}\frac{dR}{dt}\Big)\sigma := H(t)\sigma \qquad (7.47)$$

where H(t) is the ***Hubble parameter***. Since the time derivative is a radial velocity, we get for t=0 (i.e., now)

$$v = H_0 d$$

which is ***Hubble's law***, telling us that the farther distant (d) a star, the faster it is receding from us – and from everything else. Among other things, this says that galaxies or stars farther than $d_H = c/H_0$ are receding from us faster than the speed of light. Not to worry, this is because of space expanding, nothing is traveling at superliminal speed in anybody's Lorentz rest frame.

102 Collier, 292-3; Schutz, 341-4.
103 From WMAP via NASA, http://map.gsfc.nasa.gov/universe/bb_concepts.html.

7.7.3. The Friedmann equations

Using a metric to calculate connection coefficients and then pushing all that into the various tensors is tedious, to say the least. So many cosmologists start with a predigested version of those results. The R-W metric can be used to calculate the connection coefficients, then the Riemann curvature tensor and the quantities derived from it, the Ricci tensor and scalar. They then assume that the Universe can be described as a perfect fluid and that its matter component can be treated as dust. Putting the energy-momentum tensor for all this into the Einstein equation (7.35) leads eventually to the Friedmann equations. The first, called the ***Friedmann equation***, is

$$\Big[\frac{1}{R}\frac{dR}{dt}\Big]^2 = \frac{8\pi G}{3}\rho - \frac{k}{R^2} \tag{7.48}$$

and the second, the ***Friedmann acceleration equation***, is:

$$\frac{1}{R}\frac{d^2R}{dt^2} = -\frac{4\pi G}{3}\big(\rho + 3p\big) \tag{7.49}$$

The sum $(\rho + 3p)$ of the pressure and 3-momentum densities from (7.33) on the right-hand side of this equation is referred to as the ***active gravitational mass***.[104] The quantity k is the curvature from the R-W metric of (7.46).

We can take the method of equation (7.39) farther and decompose the the energy-momentum tensor into three parts due to three different ***energy densities***, for matter, radiation and vacuum energy (the cosmological constant):

$$\rho(t) = \rho_m(t) + \rho_r(t) + \rho_\Lambda \tag{7.50}$$

where the last term indicates that the vacuum energy density is a constant across spacetime. As the energy scale factor increases, the matter and radiation energy densities decrease by a volume factor of R^3. In addition, since the energy of radiation is inversely proportional to its wavelength, which will increase with R, its density goes down by a further factor of R. One can show that the total energy density evolves as[105]

$$\rho(t) = \rho_{m,0}\Big(\frac{R_0}{R(t)}\Big)^3 + \rho_{r,0}\Big(\frac{R_0}{R(t)}\Big)^4 + \rho_\Lambda. \tag{7.51}$$

We will see in the following models that the different powers of $\frac{R_0}{R(t)}$ will be important in determining the rate of expansion. The pressure evolves as

$$p(t) = \frac{1}{3}\rho_{r,0}\Big(\frac{R_0}{R(t)}\Big)^4 - \rho_\Lambda, \tag{7.52}$$

assuming that matter is in the form of dust and we give or take a few factors of c (=1).

Defining a ***normalized scale factor***

$$a(t) = \frac{R(t)}{R_0} \tag{7.53}$$

allows the Friedmann equation to be written in another much-seen form:

$$\Big[\frac{1}{a}\frac{da}{dt}\Big]^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{7.54}$$

104 Schutz (2003), 242, says it is responsible for the gravitational redshift and the gravitoelectric field, that "part of the gravitational field that is most like the Newtonian gravitational acceleration.

105 See Collier, 300.

or

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}. \tag{7.55}$$

Similarly, if we assume pressure follows an equation of state

$$p = \omega\rho$$

(ω = 0, 1/3 or -1 in the case of dust, radiation or dark energy, respectively), then we can find the evolution of the pressure with time[106]

$$p(t) = \frac{\rho_{r,0}}{3}\left(\frac{R_0}{R(t)}\right)^4 - \rho_\Lambda. \tag{7.56}$$

7.8 Other models of the universe

The following models of the universe are all calculated using the Friedmann equations and therefore are based on the assumptions of the R-W metric and the energy-momentum tensor for a perfect-fluid. They are often called FRW models.

7.8.1. Empty-universe model

The simplest case is a completely empty universe:

$$\rho_m = \rho_r = \rho_\Lambda = 0$$

in which case the Friedmann equation becomes

$$\left[\frac{1}{R}\frac{dR}{dt}\right]^2 = -\frac{k}{R^2}$$

or

$$\frac{dR}{dt} = \sqrt{-k},$$

so k must be equal to 0 or -1. If k=0, the result is R constant and we have an empty, unchanging, quite boring space. Taking k=-1, integrating and dropping the constant factor gives $R = \pm t$ ($R = \pm ct$ if $c \neq 1$). In terms of the above normalized scale factor,

$$a(t) = \frac{t}{t_0}.$$

So an empty universe either remains static and empty or increases linearly with time.

7.8.2. Static Einstein model

Einstein did not like the idea of an expanding space and so finagled his equation to give a static one, adding in the cosmological constant, Λ. Beginning with that version of the Einstein equation (7.36), the Friedmann equations can be rederived to give

$$\left[\frac{1}{R}\frac{dR}{dt}\right]^2 = \frac{8\pi G}{3}\rho - \frac{k}{R^2} + \frac{\Lambda}{3}$$

and

106 Collier, 300.

$$\frac{1}{R}\frac{d^2R}{dt^2} = -\frac{4\pi G}{3}\Big(\rho + 3p\Big) + \frac{\Lambda}{3}.$$

In these equations, the vacuum energy is included in the Λ term, so $\rho = \rho_m + \rho_r$ only. The universe can only be static if the pressure p=0 and

$$\frac{dR}{dt} = \frac{d^2R}{dt^2} = 0.$$

The second equation above then says that

$$0 = -\frac{4\pi G}{3}\rho + \frac{\Lambda}{3},$$

so

$$4\pi G\rho = \Lambda$$

in agreement with equation (7.38). Adding these conditions to the first Friedmann equation above,

$$0 = \frac{8\pi G}{3}\rho - \frac{k}{R^2} + \frac{\Lambda}{3} = \frac{8\pi G}{3}\rho - \frac{k}{R^2} + \frac{4\pi G\rho}{3},$$

or

$$4\pi G\rho = \frac{k}{R^2}$$

so k must be positive and is taken to be k=+1, and the curvature of the static universe is positive. The problem with this solution, as Einstein realized, is that it is extremely unstable and therefore unlikely to be true.

Equation (7.38) gives the equivalence between Λ and ρ_Λ, from which the first equation here may be written as

$$\Big[\frac{1}{R}\frac{dR}{dt}\Big]^2 = \frac{8\pi G}{3}\big(\rho + \rho_\Lambda\big) - \frac{k}{R^2}$$

which is equivalent to the original Friedmann equation, now including ρ_Λ instead of Λ itself. This is how the cosmological constant leads to a static universe, by providing the effective energy density needed to combat expansion.

So it seems clear that neither of these two examples, the empty universe or the static one, are very useful descriptions of our universe.

7.8.3. De Sitter model

The de Sitter universe goes a step beyond the empty universe by adding vacuum energy only, leaving $\rho_{m,0} = \rho_{r,0} = 0$ as well as k=0. The Friedmann equation (7.48) then gives

$$\frac{dR}{dt} = \sqrt{\frac{8\pi G\rho_\Lambda}{3}}R$$

which can be resolved somewhat laboriously to give

$$R(t) = R_0 e^{\sqrt{\frac{8\pi G\rho_\Lambda}{3}}(t-t_0)}.$$

In this case the Hubble parameter is

$$H(t) = \frac{1}{R}\frac{dR}{dt} = \sqrt{\frac{8\pi G\rho_\Lambda}{3}}$$

so that

$$R(t) = R_0 e^{H_0(t-t_0)}. \tag{7.57}$$

Because of the small size of the Hubble constant (2.27x10^{-18}s^{-1}), this exponential curve is barely distinguishable from the t axis, but the universe is inexorably expanding. The expansion is entirely due to the positive cosmological constant.

7.8.4. Radiation-only model

As its name implies, this model assumes the initial matter energy density and the vacuum energy to be zero, as well as k: $k = 0 = \rho_\Lambda = \rho_{m,0}$. The Friedmann equation (7.48) gives

$$\frac{dR}{dt} = \sqrt{\frac{8\pi G}{3}\rho_{r,0}}\frac{R_0^2}{R} \tag{7.58}$$

This can be solved to give

$$R(t) = R_0\Big(\frac{t}{t_0}\Big)^{1/2}$$

or equivalently

$$a(t) = \Big(\frac{t}{t_0}\Big)^{1/2}. \tag{7.59}$$

7.8.5. Einstein-de Sitter model

This model assumes $k = 0 = \rho_\Lambda = \rho_{r,0}$, in other words, a universe of nothing but matter, assumed to behave like dust. So, one more time, the Friedmann equation (7.48) gives:

$$\frac{dR}{dt} = \sqrt{\frac{8\pi G}{3}\rho_{m,0}}\frac{R_0^{3/2}}{R^{1/2}}, \tag{7.60}$$

which describes an increasing scale factor, R(t), whose rate of increase decreases in time. After some time and labor, the equation renders

$$R(t) = R_0\Big(\frac{t}{t_0}\Big)^{2/3}$$

or equivalently

$$a(t) = \Big(\frac{t}{t_0}\Big)^{2/3}. \tag{7.61}$$

So this model describes a universe which expands forever, but at a continually decreasing rate, as seen from equation (7.60). This was considered the best model until the late 1990s, when it was discovered that the expansion is increasing.

7.8.6. Critical density and density parameters

The Friedmann equation can be written in terms of the Hubble parameter defined in equation (7.47) as follows:

$$\left(\frac{1}{R}\frac{dR}{dt}\right)^2 = H^2(t) = \frac{8\pi G}{3}\rho - \frac{k}{R^2} \tag{7.62}$$

where ρ represents the sum of all three energy densities. There are three cases to consider according to the values of k.

1. If the space is flat, then k=0 and

   $$\rho = \frac{3H^2(t)}{8\pi G} := \rho_c$$

   where ρ_c is the ***critical density*** necessary for a flat space.

2. If $\rho < \rho_c$, then equation (7.62) shows that k must be negative and therefore taken to be k=-1, so the curvature of space is negative. It also shows that R(t) will go on increasing forever.

3. If $\rho > \rho_c$, then k=1. In the context of the Einstein-de Sitter model, where $\rho = \rho_m$, we have seen that ρ_m, as a density, goes down as R^3, so that eventually $H^2(t)$ will become zero and the universe will collapse.[107]

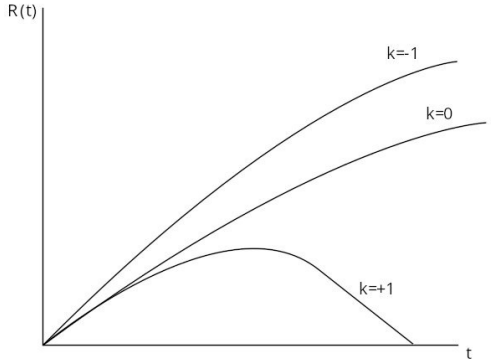


Figure 11: Effect of densities on expansion of space,.

Cosmologists also employ another set of ***density parameters*** based on the ratio of each energy density to the critical density

$$\Omega_x = \frac{\rho_x(t)}{\rho_c(t)} \tag{7.63}$$

for x = m, r or λ. The total density is the sum of the three

$$\Omega(t) = \Omega_m(t) + \Omega_r(t) + \Omega_\Lambda(t). \tag{7.64}$$

Inserting this into the Friedmann equation (7.62) and using (7.63) gives

$$\frac{k}{H^2(t)R^2(t)} = \Omega(t) - 1.$$

So we have the following correspondences:

$$k = -1 \Rightarrow \Omega(t) < 1$$

$$k = 0 \Rightarrow \Omega(t) = 1$$

$$k = +1 \Rightarrow \Omega(t) > 1.$$

107 Whatever does imaginary H(t) mean?

If one assumes no radiation and $\Omega_{m,0} + \Omega_{\Lambda,0} = \Omega_0 = 1$ (k=0), one can use the Friedmann equation to calculate the age of the universe to be 13.78 Gy, which is pretty damn close to the WMAP determination of 13.7 ± 0.13 Gy.[108]

7.8.7. Lambda-CDM model

The consensus is that the current Universe is best described by an ***accelerating model*** with k=0 and $\Lambda > 0$. Although the relative densities of matter, radiation and dark energy have changed over time, we are currently in a state dominated by dark energy. This model is known as the ΛCDM model, where CDM stands for Cold Dark Matter. The current best values of the cosmological parameters are:[109]

$$\Omega_{m,0} \approx 0.27,\ \Omega_{r,0} \approx 0,\ \Omega_{\Lambda,0} \approx 0.73,\ H_0 \approx 70.4 \text{ kms}^{-1}\text{Mpc}^{-1}.$$

We have broken the density into three parts due to matter, radiation and the vacuum. The two equations (7.51) and (7.52) show how the three forms of energy density and pressure vary with the expansion of the universe. Radiation energy density decreases faster than matter energy density, but both decrease faster than the vacuum energy density, which does not decrease at all. So consideration of these relative energy densities suggests that the universe has been through three periods:

1. An initial, short period of radiation dominance, estimated to measure on the order of 50,000 years, approximated by the radiation-only model, equation (7.59);
2. a longer period of matter dominance, estimated as lasting about 9.8×10^9 years, best represented by the Einstein-de Sitter matter-only model (EdS), equation (7.61);
3. a long (infinite?) period of dominance of vacuum energy, still running, approximated by the de Sitter model, equation (7.57).

Equation (7.56) implies that similar considerations hold for the pressure of radiation and the vacuum, with radiation pressure decreasing rapidly from its initial value.

7.8.8. But...[110]

All these results about the expanding universe and the models based on them could not be true.

Recent reanalysis of the supernovae data on expansion of the Universe finds no evidence for dark energy! What makes the difference is that the new analysis does not assume the validity of the Cosmological Principle over short distances.

Indeed, it is clear that the Cosmological Principle does not hold over short distances, as anyone can see by looking at the night sky. One mostly ignored paper calculated on the basis of the ΛCDM finds that deviations from a uniform distribution of matter should be less than 200-300 Mpc, or about 1 billion light years. However, exceptions to this requirement are well known. The Clowes-Campusano-Quasar group, a collection of thirty-four quasars extending over two billion light years is too large to be compatible the ΛCDM. So are the Great Wall (1.5 billion light years) and the Giant Arc (3 billion light years). And we on Earth live in a region of significantly lower density than the universal average called the “local hole”.

The next decade or two of cosmology may be quite interesting.

7.8.9. Gravitational waves

Beyond this document. They should exist but be very weak. LIGO and VIRGO have found them.

108 Collier, 316-7 and WMAP, https://map.gsfc.nasa.gov/resources/edactivity1.html.
109 Collins, 316.
110 New evidence against the standard model of cosmology. https://backreaction.blogspot.com/2021/09/new-evidence-against-standard-model-of.html.

7.9 Annex: All you need to solve Einstein's equations

Start with the equation itself, cosmological constant and all, equation (7.36):

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}.$$

Next, we need the Ricci tensor and scalar, from equations (7.24) and (7.25):

$$R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu}$$
$$R = R^{\mu}{}_{\mu} = g^{\mu\nu}R_{\mu\nu}$$

and that requires the Riemann curvature tensor's, from equation (7.23):

$$R^{d}{}_{abc} = \partial_b(\Gamma^{d}_{ac}) - \partial_c(\Gamma^{d}_{ab}) + \Gamma^{e}_{ac}\Gamma^{d}_{eb} + {}^{e}_{ab}\,\Gamma^{d}_{ec}$$

We also need the Christoffel or connection symbols, from equation (7.18):

$$\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2}g^{\sigma\mu}\Big(\frac{\partial g_{\alpha\sigma}}{\partial x_\beta} + \frac{\partial g_{\beta\sigma}}{\partial x_\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x_\sigma}\Big)$$

Now all you need is a metric, g, and an energy-momentum tensor (make it easy on yourself and use the perfect-fluid version) and you're ready to go!

8 Bibliography

Atkins, Peter, *The laws of thermodynamics: A very short introduction*. Oxford: Oxford University Press, 2010. Print.

Blundell, Katherine. *Black holes, a very short introduction*. Oxford: Oxford University Press, 2015. Print.

Blundell, Stephen J. and Lancaster, Tom. *Quantum field theory for the gifted amateur*. Oxford: Oxford University Press, 2014.

Callan, Herbert B. *Thermodynamics and an introduction to thermostatistics*. New York: John Wiley and Sons. 1985, 2005. Print.

Carroll, Sean M. *Spacetime and Geometry: An Introduction to General Relativity*. Harlow: Pearson, 2013. Print.

Carroll, Sean. *The biggest ideas in the Universe: space, time and motion*. Dutton, 2022. Print.

Collier, Peter. *A Most Incomprehensible Thing: Notes towards a Very Gentle Introduction to the Mathematics of Relativity*. Harlow: Incomprehensible, 2014. Print.

Griffiths, David. *Introduction to elementary particles*. Second, revised edition. Weinheim: Wiley-VCH, 2008. Print.

Griffiths, David J. and Schroeter, Darrell F. Introduction to quantum mechanis, 3rd edition. Cambridge: Cambridge University Press, 2018. Print.

Hartle, J. B. *Gravity: An Introduction to Einstein's General Relativity*. Harlow: Pearson Education, 2014. Print.

Jeevanjee, Nadir. An introduction to tensors and group theory for physicists. 2nd edition. Heidelberg: Springer, 2015. Print.

Klauber, Robert D. *Student friendly quantum field theory.* Fairfield: Sandtrove Press, 2013. Print.

Lambert, Frank L. The entropy site. http://entropysite.oxy.edu/.

Lambourne, Robert J. A. *Relativity, gravitation and cosmology*. Cambridge: Cambridge University Press, 2010. Print.

Rovelli, Carlo. *General relativity: The essentials*. Cambridge: Cambridge Unitarity Press, 2021. Print.

Schutz, Bernard F. *A First Course in General Relativity*. Cambridge: Cambridge UP, 2016. Print.

Schutz, Bernard F. *Gravity from the ground up*. Cambridge: Cambridge UP, 2003. Print.

Schwichtenberg, Jakob. *No-nonsense quantum field theory*. No-nonsense books, 2020. Print.

Schwichtenberg, Jakob. *Physics from finance.* No-nonsense books, 2020. Print.

Schwichtenberg, Jakob. *Physics from symmetry*, 2nd edition. Springer, 2018. Print.

Shankar, R. *Fundamentals of physics: Mechanics, relativity and thermodynamics*. New Haven: Yale University Press, 2014. Print.

Shankar, R. *Fundamentals of physics II: Electromagnetism, optics, and quantum mechanics*. New Haven: Yale University Press, 2016. Print.

Stannard, Russell. *Relativity: A very short introduction*. Oxford: Oxford University Press, 2010.

Susskind, Leonard and Hrabovsky, George. *The theoretical minimum: What you need to know to start doing physics*. New York: Basic Books, 2013. Print.

Susskind, Leonard and Friedman, Art. *Quantum mechanics, the theoretical minimum*. New York: Basic Books, 2014. Print.

Susskind, Leonard and Friedman, Art. *Special relativity and classical field theory: The theoretical minimum*. New York: Basic Books, 2017. Print.

Susskind, Leonard. *The theoretical minimum: Quantum mechanics*. (Especially lectures 6 and 7 on entanglement.) http://theoreticalminimum.com/courses/quantum-mechanics/2012/winter. On-line lectures.