Lecture Notes on Classical Gauge Field Theory

Jonas T. Hartwig

Version 0.12 from Apr 29, 2025

Revision	Date	Author(s)	Description
0.1	Jan 21, 2025	JH	Added Lecture 1 (Geometries)
0.2	Jan 28, 2025	JH	Added Lecture 2 (Tensor Fields)
0.3	Feb 6, 2025	JH	Added Lecture 3 (Covariant Derivatives)
0.4	Feb 15, 2025	JH	Added Lecture 4 (Curvature, Metric)
0.5	Feb 20, 2025	JH	Added Lecture 5 (Invariant Integration).
0.7	Mar 7, 2025	JH	Typographical changes. Added Lectures 6,7 (Action Principles and Conserved Quantities)
0.8	Mar 11, 2025	JH	Added Lecture 8 (Energy-Momentum Tensor, Electromagnetism)
0.9	Mar 31, 2025	JH	Added most of Lecture 9 (GR)
0.10	Apr 1, 2025	JH	Added draft of Lecture 10 (Principal Bundles)
0.11	Apr 16, 2025	JH	Added draft of Lectures 11, 12 (Gauge Fields, Spinors)
0.12	Apr 29, 2025	JH	Added Weyl spinors to Lecture 12. Drafts of Lecture 13 (SM and Higgs) and Lecture 14 (GUTs)

Contents

I Geometry and Tensor Calculus **9**

1 Geometries **13**
1.1 Coordinate Transformations . . . 13
1.1.1 The Galilean/Newtonian geometry . . . 14
1.1.2 The Lorentzian geometry (special relativity) . . . 15
1.1.3 The smooth (or general) geometry . . . 18
1.2 Homework #1 . . . 19

2 Tensor Fields **23**
2.1 Vector Fields and Contravariant Tensor Fields . . . 23
2.1.1 Curves, Tangent Vectors, and Vector Fields . . . 23
2.1.2 Addition and Scaling of Vector Fields . . . 25
2.1.3 Contravariant Tensor Fields . . . 25
2.2 Covector Fields and Covariant Tensors Fields . . . 26
2.3 Mixed Tensors . . . 27
2.4 Operations on Tensors: Contraction, Relabeling and Permuting . . . 27
2.5 Connections to Multilinear Algebra . . . 28
2.6 Homework #2 . . . 29

3 Covariant Differentiation **33**
3.1 The Problem of Constant Fields . . . 33
3.2 Affine Connections . . . 34
3.3 Derivative of Contravariant Fields . . . 35
3.4 Derivative of Covariant Fields . . . 36
3.5 Derivative of Mixed Tensor Fields . . . 36
3.6 Locally Inertial Coordinate Systems . . . 37
3.7 Homework #3 . . . 38

4 Curvature, Torsion, Metrics and Vielbeins **39**
4.1 The Curvature and Torsion of an Affine Connection 39
4.2 Metrics and Vielbeins 41
4.3 Raising and Lowering of Indices 42
4.4 The Levi-Civita Connection of a Metric 44
4.5 Homework #4 45

5 Invariant Integration **49**
5.1 Integration and Scalar Densities 49
5.2 Relative Tensor Fields 50
5.3 The Levi-Civita Symbols, and Determinants 52
5.4 The Scalar Density $\sqrt{-g}$ 53
5.5 Useful Formulas Involving the Metric 54
5.6 Homework #5 56

II Lagrangian Field Theories **59**

6 Action Principles **65**
6.1 Natural Units 65
6.2 Lagrangian Densities 66
6.3 Examples 67
6.3.1 Theories in Flat 3+1 Dimensions 67
6.3.2 Theories in Curved Spacetime 68
6.4 Equations of Motion: Euler-Lagrange 68
6.5 Examples 70
6.5.1 Scalar field 70
6.5.2 Scalar Field in One Dimension 70
6.6 Homework #6 71

7 Conserved Quantities **73**
7.1 Global Symmetries and Noether's Theorem 73
7.2 Charge Conservation 74
7.3 Example: A Scalar Doublet 75
7.4 Homework #7 76

8 The Energy-Momentum Tensor **79**
8.1 Coordinate Transformation Symmetries 79
8.2 Spacetime Translation Operators 79

8.3 The Conserved Currents of Translation Symmetry . . . 80
8.4 Alternative Definition of Energy-Momentum Tensor . . . 82
8.5 Homework #8 . . . 84

9 General Relativity 89
9.1 Gravity . . . 89
9.2 Derivation of the Einstein-Hilbert Lagrangian . . . 90
9.2.1 First Order Metric Lagrangians are Trivial . . . 90
9.2.2 Second Order Metric Lagrangians . . . 92
9.3 Field Equations . . . 95
9.4 Newtonian Limit . . . 96
9.5 Linear Approximation and Gravitational Waves . . . 97
9.6 Homework #9 . . . 97

III Local Gauge Field Theories 99

10 Principal Bundles I: Matter Fields 103
10.1 Yang-Mills Theory . . . 103
10.2 Homework #10 . . . 105

11 Principal Bundles II: Gauge Fields 107
11.1 Derivatives of Matter Fields . . . 107
11.2 Geometric Interpretation (optional) . . . 109
11.3 Homework #11 . . . 110

12 Spinors 111
12.1 The Concept of Spin . . . 111
12.2 Clifford Algebra . . . 113
12.3 Dirac Spinors . . . 115
12.4 Weyl Spinors . . . 118
12.5 Homework #12 . . . 118

13 The Standard Model 119
13.1 Toy Models . . . 119
13.1.1 $(1+0)$-Dimensional Theory . . . 119
13.1.2 Charged Scalar Field Case . . . 119
13.2 Standard Model of Particle Physics . . . 122
13.3 Spontaneous Symmetry Breaking . . . 124

13.4 Homework #13 126

14 Grand Unified Theories **127**
14.1 A Cascade of Broken Dreams 127
14.2 $SU(5)$ 127
14.3 $SO(10)$ 129

Acknowledgements

Part I

Geometry and Tensor Calculus

The Origin of Geometry in Physics

Physics is the ongoing attempt to describe "nature", which is to say, the collection of phenomena that seem "external" to us; phenomena which we feel would take place even without the presence of humans. It is therefore clear that physical laws, to the extent possible, should be formulated in a way that is independent of any one person. This includes past, current, and future people, all over the world. More generally, due to the feeling that we have a choice about our future, we include hypothetical people. The umbrella term for these different perspectives is that of an *observer*. Since we are not really concerned with individual details related only to the observer itself, what matters is how the observer interacts with the "external" environment. This is sometimes described as an observer carrying a clock and a marked rod. This is referring to the observer's notion of the passage of time and local measure of distance. More generally, we can imagine an observer carrying other collections of measuring apparatuses, or gauges, through which they can probe the world. In fact, it is the only thing we care about when it comes to an observer. Thus, an **observer** can be abstracted to *a choice of coordinate system for spacetime, along with a choice of gauges*. Collectively, we refer to coordinate systems and choice of gauge as a **reference frame**. These choices extend only locally, as the observer cannot perform measurements far away.

Now we encounter the strange dilemma of how an observer would actually make such a "choice". Consider the situation of a lone observer floating in empty space, carrying a clock and nothing else. Suppose the clock is equipped with a dial, enabling the clock hands to speed up or slow down. Consider two different settings, a slow and a fast mode. What is the difference? You might say that the observer knows roughly how long a second is, and could therefore tell the two settings apart. Surely their heartbeat, if nothing else, can be used to compare the two settings? But this requires internal details about the observer which we are not allowed to refer to. There is no heartbeat, no person, only a clock. You might say, the faster setting would wear out the gears (or battery) of the clock faster and would allow us to distinguish between the two settings. But that again refers to internal structure of the observer (now just

a clock). Therefore the observer must be regarded as a disembodied clock without internal structure. There is therefore *no physical difference between the settings of the clock*. The same reasoning applies to choosing measuring rods, or calibrating various gauges. The inescapable conclusion is that there is actually **no content to a particular observer themself, only the relative comparison between two reference frames has physical meaning**. The manner in which we relate reference frames to each other therefore takes center stage. Typically this is done using the language of coordinates and coordinate transformations. It is in this way that geometry becomes a theoretical foundation for physics.

Lecture 1

Geometries

1.1 Coordinate Transformations

Definition 1.1.1. Let n be a positive integer and M be a set.

i) A (global, n-dimensional) *coordinate system* on M is a mapping from M to $\mathbb{R}^n$, denoted

$$x \mapsto x^i \stackrel{\text{def}}{=} (x^1, x^2, \ldots, x^n) \tag{1.1.1}$$

which is injective (i.e. one-to-one), and whose image is open and connected.[1]

ii) If $x \mapsto x^{i'}$ is another coordinate system, the map

$$x^i \mapsto x^{i'} \tag{1.1.2}$$

is called a *coordinate transformation* or a *change of coordinates*. It is a one-to-one and onto map from a subset of $\mathbb{R}^n$ (the image of $x \mapsto x^i$) to a subset of $\mathbb{R}^n$ (the image of $x \mapsto x^{i'}$).

iii) An (n-dimensional) *geometry* $\mathscr{G}$ is a collection of one-to-one and onto functions between open and connected subsets of $\mathbb{R}^n$, closed under taking inverses, and function compositions (when defined).

iv) A (global, n-dimensional) $\mathscr{G}$*-space* is a set M equipped with a set $\mathscr{C}$ of coordinate systems such that

[1] Requiring open is convenient when we need to differentiate. It also justifies calling it "n-dimensional". If the image is not connected, M would consist of different independent pieces, so we might as well restrict attention to each piece.

1. The coordinate transformations between any coordinate systems from $\mathscr{C}$ belong to $\mathscr{G}$.
2. Composition defines a map $\mathscr{G} \times \mathscr{C} \to \mathscr{C}$.

Broadly speaking, there are three geometries that play important roles in physics.

1.1.1 The Galilean/Newtonian geometry

Here $n = 3+1$ and we define $\mathscr{G}_{\text{Gal}}$ to consist of the following three kinds of transformations and their compositions:

i) Rotations/reflections:

$$\begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & A_i^j & \\ 0 & & & \end{bmatrix} \cdot \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \tag{1.1.3}$$

where $(A_i^j)_{i,j=1}^3$ is an orthogonal 3×3-matrix (see Homework #1, Problem 1):

$$\sum_{j,l=1}^{3} A_i^j A_k^l \delta_{jl} = \delta_{jk}. \tag{1.1.4}$$

ii) *Galilean boosts* parametrized by $v^i = (v^1, v^2, v^3) \in \mathbb{R}^3$ (called *relative velocity*):

$$\begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ v^1 & 1 & 0 & 0 \\ v^2 & 0 & 1 & 0 \\ v^3 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \tag{1.1.5}$$

which can also be written $\begin{cases} t' = t \\ x^{i'} = x^i + v^i \cdot t \end{cases}$

iii) Translations:

$$\begin{cases} t' = t + b \\ x^{i'} = x^i + a^i \end{cases} \tag{1.1.6}$$

where $b, a^1, a^2, a^3 \in \mathbb{R}$.

Remark 1.1.2. We give three examples of notions that are *invariant* under Galilean coordinate transformations. It means that these objects "live on M" (where $(M, \mathscr{C})$ is some fixed $\mathscr{G}$-space, $\mathscr{G}$ being the Galilean geometry defined above).

If p and q are two points on M (since M model a region if spacetime, we also call such p and q *events*), choose one of the coordinate systems on M and write $p^\mu = (t, p^1, p^2, p^3)$ and $q^\mu = (u, q^1, q^2, q^3)$. Then we can define the following two quantities (which depend on p and q):

$$\Delta t = t - u, \qquad \Delta s = \Big(\sum_{i,j=1}^{3} (p^i - q^i)(p^j - q^j)\delta_{ij} \Big)^{1/2} = \Big(\sum_{i=1}^{3} (p^i - q^i)^2 \Big)^{1/2}. \quad (1.1.7)$$

Then, for fixed p and q, the quantity Δt take the same value in all coordinate systems. To see this, recall that by definition of a $\mathscr{G}$-space, all coordinate systems are related to each other by some sequence of coordinate changes of the above three types. For each of those coordinate changes we have $\Delta t' = t' - u' = t - u = \Delta t$ (the only nontrivial case is translations). It means that the *duration between the events p and q is coordinate independent* and therefore is a notion all observers agree on. Consequently we can consider the duration between two events as part of reality (in Galilean physics).

Similarly, but now using that A^j_i are entries from an orthogonal matrix, the quantity Δs can be seen to be independent of coordinate changes and therefore the *(spatial) distance between two events p and q* is something "real" in Galilean physics.

Lastly, the Laplace differential operator

$$\Delta = \sum_{i,j=1}^{3} \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j} \delta_{ij} \quad (1.1.8)$$

is also a coordinate independent quantity. This follows from the chain rule for the coordinate change of type i) above:

$$\frac{\partial}{\partial x^{i'}} = \sum_j A^j_{i'} \frac{\partial}{\partial x^j} \quad (1.1.9)$$

and that $A^j_{i'}$ are entries of an orthogonal matrix. (Details left as homework problem.)

1.1.2 The Lorentzian geometry (special relativity)

Again, $n = 3 + 1$. Here space and time will be considered on equal footing. To facilitate this we need a dimensionfull conversion factor between time and space. We

call it c:

$$x^0 = ct \tag{1.1.10}$$

When we get to Maxwell's equations, c will be identified with the speed of light in vaccuum. The Lorentzian geometry $\mathscr{G}_{\text{Lor}}$ consists of compositions of three types of coordinate transformations:

i) Rotations/reflections in space:

$$\begin{bmatrix} x^{0'} \\ x^{1'} \\ x^{2'} \\ x^{3'} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & A_i^j & \\ 0 & & & \end{bmatrix} \cdot \begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} \tag{1.1.11}$$

where $(A_i^j)_{i,j=1}^3$ is an orthogonal 3×3-matrix as before.

ii) *Lorentz boosts* parametrized by $\lambda \in \mathbb{R}$ (called the *rapidity*):

$$\begin{bmatrix} x^{0'} \\ x^{1'} \\ x^{2'} \\ x^{3'} \end{bmatrix} = \begin{bmatrix} \cosh\lambda & \sinh\lambda & 0 & 0 \\ \sinh\lambda & \cosh\lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma} & \beta\boldsymbol{\gamma} & 0 & 0 \\ \beta\boldsymbol{\gamma} & \boldsymbol{\gamma} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} \tag{1.1.12}$$

(Here for simplicity we only consider Lorentz boost in the x^1-direction. Replacing $(\beta, 0, 0)$ by an arbitrary 3-vector β^i we get abitrary direction. We leave the details to the reader.) By hyperbolic trig identity we have $\boldsymbol{\gamma}^2 - \beta^2\boldsymbol{\gamma}^2 = 1$, therefore

$$\boldsymbol{\gamma} = (1 - \beta^2)^{-1/2}. \tag{1.1.13}$$

Substituting $x^0 = ct$, we may write the Lorentz boost in terms of time and space coordinates as follows

$$\begin{cases} t' = \boldsymbol{\gamma}(t + (\beta/c)x^1) \\ x^{1'} = \boldsymbol{\gamma}(\beta ct + x^1) \\ x^{2'} = x^2 \\ x^{3'} = x^3 \end{cases} \tag{1.1.14}$$

from which we see that

$$v^1 := \frac{dx^{1'}}{dt'} = \frac{dx^{1'}/dt}{dt'/dt} = \beta c \tag{1.1.15}$$

Thus, substituting $\beta = v^1/c$ into (1.1.13),(1.1.14) and taking the limit as $c \to \infty$ we obtain a Galilean boost with relative velocity $v^i = (v^1, 0, 0)$. In this way we see that *Galilean physics is a limiting case of Lorentzian physics.*

iii) Translations, $x^{\mu'} = x^\mu + a^\mu$, $(\mu = 0, 1, 2, 3)$.

The following theorem gives an alternative definition, that is often much simpler to work with (but the connection to Galilean geometry is less clear).

Theorem 1.1.3. *The two types of Lorentz transformations i) and ii) above can be summarized in a single type of transformation of the form*

$$x^{\mu'} = \sum_{\nu=0}^{3} \Lambda^{\mu'}_{\nu} x^\nu \tag{1.1.16}$$

where $\Lambda^{\mu'}_{\nu}$ *are the entries of a* 4×4*-matrix satisfying*

(a)

$$\sum_{\nu',\tau'} \Lambda^{\nu'}_{\mu} \Lambda^{\tau'}_{\sigma} \eta_{\nu'\tau'} = \eta_{\mu\sigma} \tag{1.1.17}$$

where $\eta_{\mu\nu}$ *are the entries of the so called* flat metric*, by definition a diagonal matrix with entries* $(+1, -1, -1, -1)$ *(the "mostly minus" convention. The opposite sign convention yields the same Lorentz transformations.)*

(b) $\Lambda^{0'}_{0} > 0$ *(we say* (Λ^ν_μ) *is* orthochronous*, meaning time-direction-preserving, if this holds)*

(c) $\det(\Lambda^\nu_\mu) = 1$ *(we say* (Λ^ν_μ) *is* proper*, if this holds).*

Proof. Exercise for the reader. □

We introduce some notation:

$$\begin{aligned} \mathrm{O}(1,3) &= \{\Lambda \mid (1.1.17) \text{ holds}\} \\ \mathrm{SO}(1,3) &= \{\Lambda \in \mathrm{O}(1,3) \mid \det \Lambda = 1\} \\ \mathrm{O}^+(1,3) &= \{\Lambda \in \mathrm{O}(1,3) \mid \Lambda^0_0 > 0\} \\ \mathrm{SO}^+(1,3) &= \mathrm{SO}(1,3) \cap \mathrm{O}^+(1,3). \end{aligned}$$

Each of these is a group under matrix multiplication. $\mathrm{SO}^+(1,3)$ is called the *Lorentz group*. It is connected.

Remark 1.1.4. We can easily extend Lorentz geometry to more space dimensions. The same applies to Galilean geometry.

1.1.3 The smooth (or general) geometry

This n-dimensional geometry, denoted $\mathscr{G}_{\text{sm}}$ consists of all one-to-one and onto functions $f : U \to V$ where U and V are open subsets of $\mathbb{R}^n$ such that all partial derivatives of f (and f^{-1}) exist to all orders.

Denoting such a function f by

$$x^i \mapsto x^{i'} = x^{i'}(x^j) = x^{i'}(x^1, x^2, \ldots, x^n) \tag{1.1.18}$$

we let $J^{i'}_j$ denote the entries of the Jacobian matrix:

$$J^{i'}_j = \frac{\partial x^{i'}}{\partial x^j} \tag{1.1.19}$$

and we let J' denote the determinant

$$J' = \det \left(J^{i'}_j \right)^n_{i',j=1} \tag{1.1.20}$$

Entries of the Jacobian matrix of the *inverse* coordinate change $x^{i'} \mapsto x^i(x^{j'})$ are consequently denoted

$$J^i_{j'} = \frac{\partial x^i}{\partial x^{j'}} \tag{1.1.21}$$

with determinant

$$J\prime = (J')^{-1} = \det \left(J^i_{j'} \right)^n_{i,j'=1}. \tag{1.1.22}$$

(Note that $'$ is used a subscript in $J\prime$ although we will not use it much.)

We can get more restricted geometries by imposing conditions on J':

- $J' > 0$ is called *oriented geometry*. It is useful for integration.

Although less important, other examples are:

- $|J| = 1$, leading to *unimodular geometry*, and
- $J = 1$, giving *proper geometry* (in analogy with Lorentz transformations).

1.2 Homework #1

1. Show that a matrix $A = (A_{ij})_{i,j=1}^n$ satisfies $A^T \cdot A = I_n$ (where A^T denotes the transpose) if and only if

$$\sum_{j,l=1}^{n} A_{ij} A_{kl} \delta^{jl} = \delta_{ik}, \qquad \forall i, k \tag{1.2.1}$$

where $\delta_{jl} = \delta^{jl} = \begin{cases} 1, & j = l \\ 0, & j \neq l \end{cases}$ denotes the Kronecker delta.

2. Let $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a symmetric bilinear form on $\mathbb{R}^n$ and define $B_{ij} = B(e_i, e_j)$ where $(e_i)_{i=1}^n$ is the standard ordered basis for $\mathbb{R}^n$. Further, let $(A_i^j)_{i,j=1}^n$ be a real $n \times n$-matrix. Show that B is A-invariant, that is,

$$B(Av, Aw) = B(v, w), \qquad \forall v, w \in \mathbb{R}^n \text{ (regarded as column vectors)} \tag{1.2.2}$$

if and only if the following identity holds:

$$\sum_{j,l} A_i^j A_k^l B_{jl} = B_{ik}, \qquad \forall i, k. \tag{1.2.3}$$

3. Let $(M, \mathscr{C})$ be a $\mathscr{G}$-space where $\mathscr{G}$ is any geometry whose coordinate transformations are smooth (i.e. $\mathcal{C}^\infty$) maps between open subsets of $\mathbb{R}^n$. Let $\phi : M \to \mathbb{R}$ be a function. Show that if ϕ is smooth in some coordinate system, then it is smooth in all coordinate systems. [That is, let $M \ni x \mapsto x^i \in \mathbb{R}^n$ be a coordinate system from the collection $\mathscr{C}$. Abusing notation, denote the corresponding function from (a subset of) $\mathbb{R}^n$ to $\mathbb{R}$ by $\phi(x^i) = \phi(x^1, x^2, \ldots, x^n)$. Show that if $\phi(x^i)$ is smooth then $\phi(x^{i'})$ is smooth for every other coordinate system $x^{i'}$ from $\mathscr{C}$.] (In this case we call ϕ *smooth*.)

4. Show that the Laplace operator

$$\Delta = \sum_{i,j=1}^{3} \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j} \delta_{ij} \tag{1.2.4}$$

is invariant under Galilean coordinate transformations. (Conclude that if $\phi, \psi : M \to \mathbb{R}$ are smooth functions, in the sense of previous problem, then the equation $\Delta\phi = \psi$ makes sense without choosing coordinates. See the Newton-Laplace problem below.)

5. Show that the d'Alembert operator

$$\Box = \sum_{\mu,\nu=0}^{3} \frac{\partial}{\partial x^\mu} \frac{\partial}{\partial x^\nu} \eta_{\mu\nu} \tag{1.2.5}$$

is invariant under Lorentz transformations. [Hint: Use Theorem 1.1.3.]

6. Consider a $\mathscr{G}_{\text{Gal}}$-space $(M, \mathscr{C})$. The *Newton-Laplace* equation for gravity is

$$\Delta\phi = 4\pi G\rho \tag{1.2.6}$$

where

- $\phi : M \to \mathbb{R}$ (in coordinates, $\phi(t, x, y, z)$) is the so-called *gravitational potential*,
- $\rho : M \to \mathbb{R}$ is a mass density distribution,
- G is Newton's constant.

The physical interpretation of ϕ is that $\nabla\phi = -\vec{a}$ where $\vec{a}$ is the acceleration vector field for a test particle.

Show that for $\rho(t, x, y, z) = \delta(x, y, z)M$ (where $\delta(x, y, z)$ denotes the delta distribution on $\mathbb{R}^3$ centered at the origin) we recover Newton's Law of Universal Gravitation.

[Integrate both sides of (1.2.6) over a closed ball of radius r centered at the origin: $\iiint_{B(0,r)} \Delta\phi \, d^3x = \iint_{\partial B(0,r)} \nabla\phi \cdot d\vec{S}$ by the divergence theorem. The latter equals $-\big|(\nabla\phi)(r)\big| 4\pi r^2$ by symmetry. On the other hand, $\iiint_{B(0,r)} 4\pi G\rho \, d^3x = 4\pi GM$. If the test particle has mass m is at distance r from the origin we thus get $F = m|\vec{a}(r)| = G\frac{mM}{r^2}$.]

7. In Lorentz geometry, show that if a curve is given by $x^i = x^i(t)$, $i = 1, 2, 3$, in some coordinate system such that $\sum_{i=1}^{3} (dx^i/dt)^2 = c^2$ then the same is true in any other coordinate system. (This shows that the speed of light in vacuum (the constant c) is the same in all reference frames.)

8. In Lorentz geometry, if two distinct events $x, y \in M$ have the same time coordinates, i.e. $x^0 = y^0$, in some coordinate system, then there is another coordinate system in which $x^{0'} \neq y^{0'}$. (Thus, simultaneity is lost as a "real" concept, when we go from Galilean to Lorentzian physics.)

9. In Lorentz geometry, the *interval* between two events $x, y \in M$ is defined by

$$\Delta\tau = \Big(\sum_{\mu,\nu=0}^{3} (x^\mu - y^\mu)(x^\nu - y^\nu)\eta_{\mu\nu} \Big)^{1/2} \tag{1.2.7}$$

Show that $\Delta\tau$ only depends on the events x and y, and not on the particular choice of coordinate system we used to write down the right hand side.

Lecture 2

Tensor Fields

2.1 Vector Fields and Contravariant Tensor Fields

In this lecture, we assume $\mathscr{G} \subseteq \mathscr{G}_{\mathrm{sm}}$ is a smooth geometry, i.e. consisting of smooth maps between open subsets of $\mathbb{R}^n$ (but not necessarily *all* such maps), and we fix a $\mathscr{G}$-space $M = (M, \mathscr{C})$.

2.1.1 Curves, Tangent Vectors, and Vector Fields

Definition 2.1.1. A *(smooth) curve* in M is a function

$$\gamma : \mathbb{R} \to M \tag{2.1.1}$$

such that, for some coordinate system x^μ (from $\mathscr{C}$) the functions

$$\gamma^\mu : \mathbb{R} \to \mathbb{R}^n, \qquad \gamma^\mu(t) = x^\mu\big(\gamma(t)\big), \tag{2.1.2}$$

are smooth functions of t.

Remark 2.1.2. If $x^{\mu'}$ is another coordinate system (from $\mathscr{C}$), then the corresponding functions $\gamma^{\mu'}(t)$ defined by $\gamma^{\mu'}(t) = x^{\mu'}\big(\gamma(t)\big)$ are also smooth. This follows because any two coordinate systems are related by a smooth coordinate transformation. Explicily, for any $x \in M$, we have $x^{\mu'}(x) = x^{\mu'}(x^\nu(x))$, where in the left hand side $x^{\mu'}$ is the new coordinate system and in the right hand side $x^{\mu'}$ stands for the coordinate transformation from x^μ to $x^{\mu'}$. We thus have

$$\gamma^{\mu'}(t) = x^{\mu'}\big(\gamma^\nu(t)\big). \tag{2.1.3}$$

Since the composition of smooth maps is smooth, $\gamma^{\mu'}$ is indeed smooth.

Definition 2.1.3. The *tangent vector* of a curve γ at a point $\gamma(t)$ in a coordinate system x^μ is given by

$$\dot{\gamma}^\mu(t) = \frac{d}{dt}\gamma^\mu(t). \tag{2.1.4}$$

Remark 2.1.4. If $x^{\mu'}$ is another coordinate system then, by the chain rule,

$$\dot{\gamma}^{\mu'}(t) = \sum_{\nu=1}^{n} \frac{\partial x^{\mu'}}{\partial x^\nu}\dot{\gamma}^\nu(t). \tag{2.1.5}$$

This can be written more succinctly:

$$\dot{\gamma}^{\mu'} = J^{\mu'}_\nu \dot{\gamma}^\nu \tag{2.1.6}$$

using the *Einstein summation convention* that **for repetead indices, one upper and one lower, summation over the index range is implied**.

To summarize, tangent vectors $\dot{\gamma}^\mu$ are given in each coordinate system, and are related by the transformation law (2.1.6). We say that the components $\dot{\gamma}^\mu$ *transform according to* (2.1.6) *under coordinate transformations* $x^\mu \mapsto x^{\mu'}$.

Since any vector in $\mathbb{R}^n$ is the tangent vector of some curve, the following definition makes sense. (From now on we drop the adjective "smooth".)

Definition 2.1.5. A *vector field* on M assigns to each coordinate system x^μ an n-tuple of functions $A^\mu = (A_\mu)_{\mu=1}^n$ (each defined on the image of x^μ) such that, for any coordinate transformation $x^\mu \mapsto x^{\mu'} = x^{\mu'}(x^\nu)$ we have

$$A^{\mu'} = J^{\mu'}_\nu A^\nu \tag{2.1.7}$$

where $A^{\mu'}$ is the n-tuple corresponding to the coordinate system $x^{\mu'}$. The functions A^μ are the *components* of the vector field.

For $n = 2$, if we write out (2.1.7), it reads

$$\begin{cases} A^{1'}(x^{1'}, x^{2'}) = \dfrac{\partial x^{1'}}{\partial x^1}(x^1, x^2) \cdot A^1(x^1, x^2) + \dfrac{\partial x^{1'}}{\partial x^2}(x^1, x^2) A^1(x^1, x^2) \\ A^{2'}(x^{1'}, x^{2'}) = \dfrac{\partial x^{2'}}{\partial x^1}(x^1, x^2) \cdot A^1(x^1, x^2) + \dfrac{\partial x^{2'}}{\partial x^2}(x^1, x^2) A^1(x^1, x^2) \end{cases} \tag{2.1.8}$$

2.1.2 Addition and Scaling of Vector Fields

If A^μ and B^μ are (the components in an arbitrary coordinate system x^μ of) two vector fields, then

$$A^\mu + B^\mu, \qquad cA^\mu, \qquad (c \in \mathbb{R} \text{ a constant}) \tag{2.1.9}$$

define vector fields. To check this, we must see that they obey the transformation law (2.1.7). We have

$$A^{\mu'} + B^{\mu'} = J^{\mu'}_\nu A^\nu + J^{\mu'}_\nu B^\nu = J^{\mu'}_\nu (A^\nu + B^\nu)$$
$$cA^{\mu'} = cJ^{\mu'}_\nu A^\nu = J^{\mu'}_\nu (cA^\nu)$$

It is straightforward to show that the set of all vector fields on M, with these two operations, forms itself a vector space. This vector space is denoted by $\mathrm{Vect}(M)$.

2.1.3 Contravariant Tensor Fields

Perhaps it is natural to wonder whether we can multiply vector fields as well. We immediately run into problems, however. If we try with $A^\mu B^\mu$ (without summation convention) then these objects do not transform according to (2.1.7) (check this!). On the other hand, $A^\mu B^\nu$ have too many indices (two upper). The best solution is to accept that we obtain not a vector field, but a new type of object. We do have the following transformation law for $A^\mu B^\nu$:

$$A^{\mu'} B^{\nu'} = J^{\mu'}_\lambda J^{\nu'}_\rho A^\lambda B^\rho \tag{2.1.10}$$

(summation over both λ and ρ implied, per convention). Similarly, if we multiply three vector fields we get three factors of the Jacobian matrix. We take this as basis for a new definition.

Definition 2.1.6. A *contravariant r-tensor (field)* assigns to each coordinate system x^μ an n^r-tuple of functions $T^{\mu_1\mu_2\cdots\mu_r}$ (defined on the image of x^μ) such that under $x^\mu \mapsto x^{\mu'}$,

$$T^{\mu'_1\mu'_2\cdots\mu'_r} = J^{\mu'_1}_{\nu_1} J^{\mu'_2}_{\nu_2} \cdots J^{\mu'_r}_{\nu_r} T^{\nu_1\nu_2\cdots\nu_r}. \tag{2.1.11}$$

Thus, for example, if A^μ and B^μ are vector fields, then $A^\mu B^\nu$ is a contravariant 2-tensor. However, not every 2-tensor has this form. Just like for vector fields, r-tensors for fixed r can be added and scaled, and form a vector space. For example,

$$A^\mu B^\nu + A^\nu B^\mu$$

is a contravariant 2-tensor.

2.2 Covector Fields and Covariant Tensors Fields

There is a dual notion to the previous subsection. Instead of starting with a curve, which is a function from $\mathbb{R}$ to M, we start with a *scalar field* ϕ which is a function from M to $\mathbb{R}$. We assume ϕ is smooth, in the sense of Problem 3 of Homework #1, and denote by $\phi(x^\mu)$ the function $\mathbb{R}^n \to \mathbb{R}$ obtained by composing the inverse of the coordinate system by ϕ.

Definition 2.2.1. The *gradient* of a scalar field ϕ is

$$\partial_\mu \phi = \frac{\partial}{\partial x^\mu} \phi(x^\mu). \tag{2.2.1}$$

Just as with the tangent vector, the gradient depends on the choice of coordinate system. If $x^{\mu'}$ is another coordinate system, the gradient transforms as follows:

$$\partial_{\mu'} \phi = \frac{\partial}{\partial x^{\mu'}} \phi(x^\nu) = \frac{\partial x^\nu}{\partial x^{\mu'}} \frac{\partial}{\partial x^\nu} \phi(x^\nu) = J^\nu_{\mu'} \partial_\nu \phi \tag{2.2.2}$$

where $J^\nu_{\mu'}$ is now the inverse of the Jacobian matrix $J^{\mu'}_\nu$. That is:

$$J^\nu_{\mu'} J^{\mu'}_\lambda = \delta^\nu_\lambda, \qquad J^\nu_{\mu'} J^{\lambda'}_\nu = \delta^{\lambda'}_{\mu'} \tag{2.2.3}$$

where we used the Kronecker delta (see example below). We take the transformation law (2.2.2), which is opposite to the vector field (2.1.7), as the basis for a new definition.

Definition 2.2.2. A *covector field* on M assigns to each coordinate system x^μ an n-tuple of functions $B_\mu = (B_\mu)_{\mu=1}^n$ (each defined on the image of $x \mapsto x^\mu$) which transform as

$$B_{\mu'} = J^\nu_{\mu'} B_\nu \tag{2.2.4}$$

under a coordinate change $x^\mu \mapsto x^{\mu'}$.

Following the same procedure as for vector fields leads us to the dual notion to contravariant tensors.

Definition 2.2.3. A *covariant s-tensor (field)* assigns to each coordinate system x^μ an n^s-tuple $B_{\mu_1 \mu_2 \cdots \mu_s}$ of functions (defined on the image of $x \mapsto x^\mu$) which transform as

$$B_{\mu'_1 \mu'_2 \cdots \mu'_s} = J^{\nu_1}_{\mu'_1} J^{\nu_2}_{\mu'_2} \cdots J^{\nu_s}_{\mu'_s} B_{\nu_1 \nu_2 \cdots \nu_s} \tag{2.2.5}$$

under a coordinate change $x^\mu \mapsto x^{\mu'}$.

2.3 Mixed Tensors

Lastly, there is nothing preventing us from multiplying contravariant and covariant tensor fields. Since upper and lower indices obey opposite transformation laws, we are led to the following general definition. For non-negative integers r, s we write $\binom{r}{s}$ for the column vector (not a binomial coefficient).

Definition 2.3.1. An $\binom{r}{s}$*-tensor*, also called a *tensor of type* $\binom{r}{s}$, assigns to each coordinate system x^μ on M an n^{r+s}-tuple

$$T^{\mu\cdots}_{\nu\cdots}$$

with r upper and s lower subscripts, which transform as

$$T^{\mu'\cdots}_{\nu'\cdots} = J^{\mu'}_{\lambda}\cdots J^{\rho}_{\nu'}\cdots T^{\lambda\cdots}_{\rho\cdots} \tag{2.3.1}$$

under coordinate transformations $x^\mu \mapsto x^{\mu'}$. The functions $T^{\mu\cdots}_{\nu\cdots}$ are the *components* of the tensor. The *rank* of the tensor is the number $r + s$. The upper indices are called *contravariant* and the lower indices are called *covariant*.

The set of all $\binom{r}{s}$-tensors forms a vector space, that we could denote by $\mathscr{T}\binom{r}{s}$. Furthermore, the product of an $\binom{r}{s}$-tensor by a $\binom{t}{u}$-tensor is an $\binom{r+t}{s+u}$-tensor.

Example 2.3.2 (Kronecker Delta)**.** In all coordinate systems, define

$$\delta^\nu_\mu = \begin{cases} 1, & \mu = \nu \\ 0, & \mu \neq \nu \end{cases}. \tag{2.3.2}$$

Then δ^ν_μ is a $\binom{1}{1}$-tensor field on M. (This is one of very few "universal" tensors.) To check this, we start with the right hand side of (2.3.1):

$$J^{\mu'}_{\lambda} J^{\rho}_{\nu'} \delta^\lambda_\rho = J^{\mu'}_{\lambda} J^{\lambda}_{\nu'} = \delta^{\mu'}_{\nu'} \tag{2.3.3}$$

by one of the two Jacobian identities (2.2.3). This shows δ^μ_ν satisfies the transformation law for a $\binom{1}{1}$-tensor.

2.4 Operations on Tensors: Contraction, Relabeling and Permuting

The process of *contraction* involves setting one upper index equal to a lower index and then summing over that index. The result turns an $\binom{r+1}{s+1}$-tensor into an $\binom{r}{s}$-tensor.

Example 2.4.1. If A^μ_ν is a $\binom{1}{1}$-tensor, then A^μ_μ is a $\binom{0}{0}$-tensor, also known as a scalar field. To check this, consider:

$$A^{\mu'}_{\mu'} = J^{\mu'}_\lambda J^\rho_{\mu'} A^\lambda_\rho = \delta^\rho_\lambda A^\lambda_\rho = A^\lambda_\lambda \tag{2.4.1}$$

As another application, for a vector field ξ^μ and a scalar field ϕ, we can define the *directional derivative of ϕ along ξ^μ* to be

$$\xi^\mu \partial_\mu \phi. \tag{2.4.2}$$

Since this is a contraction of the $\binom{1}{1}$-tensor $\xi^\mu \partial_\nu \phi$, it is a scalar function describing the rate of change of ϕ relative to the vector field x^μ. This quantity is also known as the *Lie derivative of ϕ along ξ^μ* and is denoted $\pounds_\xi \phi$. (One can define the Lie derivative of tensors of rank > 0, but they are more complicated.)

A trivial operation is to simply permute the indices of a tensor. For example, if $A^{\mu\nu\lambda}$ is a tensor (of type $\binom{3}{0}$, needless to say) then, so is $A^{\nu\lambda\mu}$. Combining this with addition we can construct more interesting tensors, for example

$$A^{\mu\nu\lambda} + A^{\nu\lambda\mu} + A^{\lambda\mu\nu}.$$

Similarly, if B^j_i and C^l_k are tensors then a somewhat nontrivial combination of them is the tensor

$$B^j_i C^l_k + B^l_i C^j_k.$$

2.5 Connections to Multilinear Algebra

In this section we assume the reader is familiar with some elements of multilinear algebra. It can be safely skipped; no future sections or problems will rely on it.

If V is a real n-dimensional vector space and we pick a basis $\boldsymbol{e}_i$ [1] for V, and let $\boldsymbol{e}^i$ be the dual basis in V^*, then $V^{\otimes r} \otimes (V^*)^{\otimes s}$ has a basis $(\boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_r}) \otimes (\boldsymbol{e}^{j_1} \otimes \cdots \otimes \boldsymbol{e}^{j_s})$. Consequently, any element $T \in V^{\otimes r} \otimes (V^*)^{\otimes s}$ can be written

$$T = T^{i_1 \cdots i_r}_{j_1 \cdots j_s} \cdot (\boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_r}) \otimes (\boldsymbol{e}^{j_1} \otimes \cdots \otimes \boldsymbol{e}^{j_s}) \tag{2.5.1}$$

for some real numbers $T^{i_1 \cdots i_r}_{j_1 \cdots j_s}$. It is starting to look like a tensor field! (At least, evaluated at a point.) If $\boldsymbol{e}_{i'}$ is any other basis, we have

$$\boldsymbol{e}_j = J^{i'}_j \boldsymbol{e}_{i'} \tag{2.5.2}$$

[1] We use tuple and summation conventions meaning that the $(\cdot)_{i=1}^n$ are dropped around $\boldsymbol{e}_i$. Similarly, A^j_i is a matrix, and $A^j_i e_j$ means $(\sum_{j=1}^n A^j_i e_j)_{i=1}^n$ etc.

for some invertible matrix $J^{i'}_{j}$. Let $\boldsymbol{e}^{i'}$ be the basis for V^* dual to $\boldsymbol{e}_{i'}$. It is straightforward to check that

$$\boldsymbol{e}^{j} = J^{j}_{i'}\boldsymbol{e}^{i'}, \tag{2.5.3}$$

where $J^{j}_{i'}$ is the inverse of the matrix $J^{i'}_{j}$. The same element T can now be expanded in the new basis $(\boldsymbol{e}_{i'_1} \otimes \cdots \otimes \boldsymbol{e}_{i'_r}) \otimes (\boldsymbol{e}^{j'_1} \otimes \cdots \otimes \boldsymbol{e}^{j'_s})$ for $V^{\otimes r} \otimes (V^*)^{\otimes s}$:

$$T = T^{i'_1 \cdots i'_r}_{j'_1 \cdots j'_s}(\boldsymbol{e}_{i'_1} \otimes \cdots \otimes \boldsymbol{e}_{i'_r}) \otimes (\boldsymbol{e}^{j'_1} \otimes \cdots \otimes \boldsymbol{e}^{j'_s}) \tag{2.5.4}$$

for some $T^{i'_1 \cdots i'_r}_{j'_1 \cdots j'_s} \in \mathbb{R}$. Substituting (2.5.2) and (2.5.3) into (2.5.1), using bilinearity, and then equating coefficients with those in (2.5.4), we find that

$$T^{i'_1 \cdots i'_r}_{j'_1 \cdots j'_s} = \left(J^{i'_1}_{a_1} \cdots J^{i'_r}_{a_r}\right)\left(J^{b_1}_{j'_1} \cdots J^{b_s}_{j'_s}\right) T^{a_1 \cdots a_r}_{b_1 \cdots b_s}. \tag{2.5.5}$$

This is exactly the same transformation law as for a $\binom{r}{s}$-tensor field, (2.3.1), except that here there is no dependence on a point $x \in M$. The coefficients of T that we discussed in this section (which are assigned to each basis for V, and transform according to (2.5.5) under a change of basis) are known as *numerical tensors*.

2.6 Homework #2

1. Convert to "tensor index notation" using tuple and summation conventions.

   (i) $\left(\sum_{i=1}^{n} A^{jk}_{il} B^i - 2A^{ij}_{li} B^k\right)^{n}_{j,k=1}$

   (ii) $(X_{11}Y^1 + X_{12}Y^2,\ X_{21}Y^1 + X_{22}Y^2)$ assuming $n = 2$

   (iii) $\mathrm{Tr}(ABC)$ for square matrices A, B, C

2. Let us define the "trace" of a vector field T^μ to be $T^1 + T^2 + \cdots + T^n$. What is wrong with this definition?

3. Assuming all objects involved are tensors, which of the following expressions define tensors?

   (i) $F_{\mu\nu}F_{\lambda\rho}g^{\mu\lambda}g^{\nu\rho}$

   (ii) $A_{ij}A_{ji}$

   (iii) $X^{ij} + Y^{jk} + Z^{ki}$

(iv) $W^a_\mu T^i_{aj} \psi^j$

4. Show that the *covariant Kronecker delta* $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ (remaining the same in all coordinate system) is a $\binom{0}{2}$-tensor if and only if the geometry consists of orthogonal transformations along with translations. (That is, show that all coordinate transformations must have the form $x^{i'} = A^{i'}_j x^j + b^{i'}$, where $A^{i'}_j$ are the entries of an orthogonal matrix.) Consider the analogous problem for the flat metric $\eta_{\mu\nu}$.

5. (a) A $\binom{0}{2}$-tensor $F_{\mu\nu}$ is *symmetric* if $F_{\mu\nu} = F_{\nu\mu}$. Similarly, we say $F_{\mu\nu}$ is *anti-symmetric* if $F_{\mu\nu} = -F_{\mu\nu}$. Analogous definitions apply to $\binom{2}{0}$-tensors. Show that these notions are coordinate-independent.
   (b) If $F_{\mu\nu}$ is anti-symmetric and $G^{\mu\nu}$ is symmetric, show that $F_{\mu\nu} G^{\mu\nu} = 0$.
   (c) Show that any $\binom{0}{2}$-tensor is the sum of one symmetric and one anti-symmetric $\binom{0}{2}$-tensor.

6. If A_μ is any covector field and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\nu$, show that $F_{\mu\nu}$ transforms as a $\binom{0}{2}$-tensor field under general (smooth) coordinate transformations. (Hint: You will have to use the product rule and the definition of $J^\mu_{\nu'}$.)

7. Show that if two $\binom{r}{s}$-tensors are equal in one coordinate system, then they are equal in all coordinate systems. In particular, conclude that whether "a tensor field vanishes on M (or some subset thereof)" is a coordinate-independent property.

8. Let X^μ be a vector field. Suppose that, in some coordinate system, $\partial_\mu X^\nu = 0$ on M, that is to say, all partial derivaties of all components of the vector field vanishes everywhere. Must the same be true in all other coordinate systems? If not, find some sufficient conditions which ensures $\partial_\mu X^\nu$ vanishes in all coordinate systems.

9. A $\binom{1}{1}$-tensor A^ν_μ is *invertible* if there is a $\binom{1}{1}$-tensor B^ν_μ such that $A^\nu_\mu B^\lambda_\nu = \delta^\lambda_\mu$ and $B^\nu_\mu A^\lambda_\nu = \delta^\lambda_\mu$ (by Problem 7, this is a coordinate-independent property). Show that in this case,
$$T^\mu_\lambda A^\nu_\mu = S^\nu_\lambda \iff T^\mu_\lambda = S^\nu_\lambda B^\mu_\nu$$
for any tensor fields S^ν_μ and T^ν_μ. Similarly, show that for any coordinate change $x^\mu \mapsto x^{\mu'}$ we have $S_{\mu'} J^{\mu'}_\nu = T_\nu$ if and only if $S_{\mu'} = T_\nu J^\nu_{\mu'}$.

10. A symmetric tensor $g_{\mu\nu}$ is *non-degenerate* if there is a symmetric tensor $h^{\mu\nu}$ such that $g_{\mu\nu}h^{\nu\lambda} = \delta^{\lambda}_{\mu}$. A *metric* is a symmetric non-degenerate $\binom{0}{2}$-tensor. Show that any metric $g_{\mu\nu}$ gives a well-defined bijection between vector fields and covector fields, sending A^{μ} to $g_{\mu\nu}A^{\nu}$.

11. With $n = 2$, consider the (polar) coordinate change

$$\begin{cases} x^{1'} = x^1 \cos x^2, \\ x^{2'} = x^1 \sin x^2, \end{cases}$$

where $x^1 > 0$ and $x^2 \in (0, \pi/2)$. (x^2 is the second coordinate, not the square of x)

(a) Find the matrices $J^{a}_{b'}$, $J^{b'}_{a}$, and their respective determinants $J_{\prime}$ and J'.

(b) If a vector field A^{μ} is given by $A^1(x^1, x^2) = x^1$ and $A^2(x^1, x^2) = 1$, find the components $A^{\mu'}(x^{1'}, x^{2'})$ of the vector field in the coordinates $x^{\mu'}$.

(c) In each coordinate system, sketch the vector field. Sketch the coordinate curves $t \mapsto (t, 0)$ and $t \mapsto (0, t)$ of one system in the other.

Lecture 3

Covariant Differentiation

Fix a smooth geometry $\mathscr{G} \subseteq \mathscr{G}_{\text{sm}}$ and a $\mathscr{G}$-space $(M, \mathscr{C})$.

3.1 The Problem of Constant Fields

If $\phi : M \to \mathbb{R}$ is a scalar field, we say ϕ is *constant* if

$$\partial_\mu \phi = 0 \tag{3.1.1}$$

Remark 3.1.1. It makes sense to say that $\phi : M \to \mathbb{R}$ is constant, as a function. This does not require a coordinate system on M. Every such constant function is constant in the above sense, and conversely.[1]

Remark 3.1.2. The condition (3.1.1) is coordinate independent, because $\partial_\mu \phi$ is a tensor field (see Problem 7 of Section 2.6). On the other hand, if A^μ is a vector field on M, the condition

$$\partial_\mu A^\nu = 0 \tag{3.1.2}$$

is **not coordinate independent** (see Problem 8 of Section 2.6). In fact, let us see precisely what goes wrong:

$$\begin{aligned} \partial_{\mu'} A^{\nu'} &= J^\alpha_{\mu'} \partial_\alpha (J^{\nu'}_\beta A^\beta) \\ &= J^\alpha_{\mu'} J^{\nu'}_\beta \partial_\alpha A^\beta + J^\alpha_{\mu'} J^{\nu'}_{\alpha\beta} A^\beta \end{aligned} \tag{3.1.3}$$

where we introduce the notation

$$J^{\nu'}_{\alpha\beta} = \partial_\alpha J^{\nu'}_\beta = \frac{\partial^2 x^{\nu'}}{\partial x^\alpha \partial x^\beta}. \tag{3.1.4}$$

[1] We assume the image of the coordinate systems $x \mapsto x^\mu$ are connected subsets of $\mathbb{R}^n$.

The first term in (3.1.3) would say that $\partial_\mu A^\nu$ is a tensor of type $\binom{1}{1}$, but the second term,

$$J^\alpha_{\mu'} J^{\nu'}_{\alpha\beta} A^\beta, \tag{3.1.5}$$

spoils this. We therefore think of this as an "error term" that needs to be corrected for.

Remark 3.1.3. The same problem appears for covector fields. However, the expression $F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ does in fact transform as a $\binom{0}{2}$-tensor (see Problem 6 in Section 2.6), even though neither term does. The lesson we draw here is that *sometimes the sum of two non-tensorial terms is tensorial.*

3.2 Affine Connections

The remarks in the previous section lead us to the idea that we might be able to add a *counteracting term* to $\partial_\mu A^\nu$, which is also non-tensorial but transform in such a way as to counteract the error term. The factor in front of A^β in the error term (3.1.5) has index structure $\binom{1}{2}$ and contracts against the vector field A^β in (3.1.3). Thus, we make the following Ansatz:

$$\nabla_\mu A^\nu := \partial_\mu A^\nu + \Gamma^\nu_{\mu\beta} A^\beta \tag{3.2.1}$$

and the question we face is: *How should $\Gamma^\nu_{\mu\beta}$ transform under coordinate changes, so as to counteract the error term in (3.1.3), making the object $\nabla_\mu A^\nu$ into a tensor of type $\binom{1}{1}$?*

Under a change of coordinates $x^\mu \mapsto x^{\mu'}$, we have by (3.2.1) and (3.1.3),

$$\begin{aligned} \nabla_{\mu'} A^{\nu'} &= \partial_{\mu'} A^{\nu'} + \Gamma^{\nu'}_{\mu'\beta'} A^{\beta'} \\ &= J^\alpha_{\mu'} J^{\nu'}_\beta \partial_\alpha A^\beta + \left(J^\alpha_{\mu'} J^{\nu'}_{\alpha\beta} + \Gamma^{\nu'}_{\mu'\alpha'} J^{\alpha'}_\beta \right) A^\beta \end{aligned} \tag{3.2.2}$$

where we also used that A^μ are the components of a vector field. For $\nabla_\mu A^\nu$ to be a tensor field type $\binom{1}{1}$, we want (3.2.2) to equal

$$J^\alpha_{\mu'} J^{\nu'}_\beta \nabla_\alpha A^\beta = J^\alpha_{\mu'} J^{\nu'}_\beta \partial_\alpha A^\beta + J^\alpha_{\mu'} J^{\nu'}_\lambda \Gamma^\lambda_{\alpha\beta} A^\beta. \tag{3.2.3}$$

Equating the coefficients of A^β in (3.2.2) and (3.2.3) gives

$$J^{\alpha'}_\beta \Gamma^{\nu'}_{\mu'\alpha'} = J^\alpha_{\mu'} J^{\nu'}_\lambda \Gamma^\lambda_{\alpha\beta} - J^\alpha_{\mu'} J^{\nu'}_{\alpha\beta} \tag{3.2.4}$$

or, after multiplying both sides by $J^{\tau}_{\sigma'}$ and contracting along $\tau = \beta$ (see Homework 9 in Section 2.6):

$$\boxed{\Gamma^{\nu'}_{\mu'\rho'} = J^{\alpha}_{\mu'} J^{\beta}_{\rho'} J^{\nu'}_{\gamma} \Gamma^{\gamma}_{\alpha\beta} - J^{\alpha}_{\mu'} J^{\beta}_{\rho'} J^{\nu'}_{\alpha\beta}} \tag{3.2.5}$$

We have discovered a new object, of fundamental importance in geometry and physics.

Definition 3.2.1. An *affine connection* on M assigns to each coordinate system x^{μ} an n^3-tuple of functions $\Gamma^{\nu}_{\mu\rho}$ which transforms according to (3.2.5) under changes of coordinates $x^{\mu} \mapsto x^{\mu'}$.

Definition 3.2.2. An *affine connection space* is a $\mathscr{G}$-space equipped with an affine connection. We denote it by $(M, \Gamma^{\nu}_{\mu\rho})$ or just (M, Γ) for short.

3.3 Derivative of Contravariant Fields

Definition 3.3.1. Let (M, Γ) be an affine connection space. The *covariant derivative* of a vector field A^{μ} is defined to be the $\binom{1}{1}$-tensor field $\nabla_{\mu} A^{\nu}$ defined in (3.2.1).

Because $\nabla_{\mu} A^{\nu}$ is a tensor field, we can now define:

Definition 3.3.2. A vector field A^{μ} on an affine connection space (M, Γ) is said to be *(locally) constant* if $\nabla_{\mu} A^{\nu} = 0$.

(There are of course many more use cases of the covariant derivative.)

For a scalar field ϕ, for the sake of consistent notation, we put

$$\nabla_{\mu} \phi = \partial_{\mu} \phi \tag{3.3.1}$$

No counteracting term is needed here, because $\partial_{\mu}\phi$ is already a tensor field.

To summarize, we have:

$$\begin{array}{rcl} \text{a } \binom{0}{0}\text{-tensor } \phi & \rightsquigarrow & \text{a } \binom{0}{1}\text{-tensor } \nabla_{\mu}\phi \\ \text{a } \binom{1}{0}\text{-tensor } A^{\mu} & \rightsquigarrow & \text{a } \binom{1}{1}\text{-tensor } \nabla_{\mu} A^{\nu} \end{array}$$

What about $\nabla_{\mu} T^{\sigma\tau}$? We want the product rule to hold. So if $T^{\sigma\tau}$ happens to be a simple tensor, i.e., a product of two vector fields $T^{\sigma\tau} = A^{\sigma} B^{\tau}$ then we would want

$$\begin{aligned} \nabla_{\mu}(A^{\lambda} B^{\rho}) &= (\nabla_{\mu} A^{\lambda}) B^{\rho} + A^{\lambda} (\nabla_{\mu} B^{\rho}) \\ &= (\partial_{\mu} A^{\lambda} + \Gamma^{\lambda}_{\mu\alpha} A^{\alpha}) B^{\rho} + A^{\lambda} (\partial_{\mu} B^{\rho} + \Gamma^{\rho}_{\mu\alpha} B^{\alpha}) \\ &= \partial_{\mu}(A^{\lambda} B^{\rho}) + \Gamma^{\lambda}_{\mu\alpha} A^{\alpha} B^{\rho} + \Gamma^{\rho}_{\mu\alpha} A^{\lambda} B^{\alpha}. \end{aligned}$$

This leads to the guess

$$\boxed{\nabla_\mu T^{\lambda\rho} = \partial_\mu T^{\lambda\rho} + \Gamma^\lambda_{\mu\alpha} T^{\alpha\rho} + \Gamma^\rho_{\mu\alpha} T^{\lambda\alpha}} \tag{3.3.2}$$

This does indeed work! That is, if $T^{\lambda\rho}$ are the components of a $\binom{2}{0}$-tensor, then $\nabla_\mu T^{\lambda\rho}$ are the components of a $\binom{2}{1}$-tensor. The same logic leads to covariant derivatives of higher rank contravariant tensor fields. Before stating the general formula, let's address the same question for covariant tensor fields.

3.4 Derivative of Covariant Fields

Similarly to the case of $\binom{2}{0}$-tensors, there is a natural way to arrive at the definition of the covariant derivative $\nabla_\mu A_\nu$ of a covector field A_μ. If we assume B^μ is some vector field, then $A_\mu B^\mu$ is a scalar field, and therefore

$$\nabla_\mu(A_\nu B^\nu) = \partial_\mu(A_\nu B^\nu) = (\partial_\mu A_\nu)B^\nu + A_\nu(\partial_\mu B^\nu). \tag{3.4.1}$$

On the other hand, we would like the covariant derivative to also satisfy the product rule. Therefore would like

$$\begin{aligned} \nabla_\mu(A_\nu B^\nu) &= (\nabla_\mu A_\nu)B^\nu + A_\nu(\nabla_\mu B^\nu) \\ &= (\nabla \mu A_\nu)B^\nu + A_\nu(\partial_\mu B^\nu + \Gamma^\nu_{\mu\alpha} B^\alpha) \end{aligned} \tag{3.4.2}$$

Equating (3.4.1) and (3.4.2), we see that a sufficient condition is that

$$\boxed{\nabla_\mu A_\nu = \partial_\mu A_\nu - \Gamma^\alpha_{\mu\nu} A_\alpha} \tag{3.4.3}$$

Again, this does indeed give a $\binom{0}{2}$-tensor (check!).

3.5 Derivative of Mixed Tensor Fields

Given the previous sections, the following is not surprising.

Definition 3.5.1. Let (M, Γ) be an affine connection space. The *covariant derivative* of an $\binom{r}{s}$-tensor field on $T^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s}$ is

$$\nabla_\lambda T^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s} = \partial_\lambda T^{\mu_1\cdots\mu_r}_{\nu_1\cdots\nu_s} + \sum_{i=1}^{r} \Gamma^{\mu_i}_{\lambda\alpha} T^{\mu_1\cdots\overset{i}{\alpha}\cdots\mu_r}_{\nu_1\cdots\nu_s} - \sum_{j=1}^{s} \Gamma^{\beta}_{\lambda\nu_j} T^{\mu_1\cdots\mu_r}_{\nu_1\cdots\overset{j}{\beta}\cdots\nu_s} \tag{3.5.1}$$

The result is a tensor field of type $\binom{r}{s+1}$ (verify this!).

3.6 Locally Inertial Coordinate Systems

In this section, for simplicity we assume $\mathscr{G} = \mathscr{G}_{\text{sm}}$ so that we have all smooth coordinate transformations available. (As the reader will see, we actually only need polynomial coordinate changes of degree two or less.)

Definition 3.6.1. An affine connection $\Gamma^{\lambda}_{\mu\nu}$ is *symmetric* if $\Gamma^{\lambda}_{\mu\nu} = \Gamma^{\lambda}_{\nu\mu}$.

Symmetric connections are particularly nice, in that, for any point $x \in M$ there is a coortinate system in which the connection vanishes. This property plays a crucial role in the phenomenology of general relativity, and can also greatly simplify mathematical calculations.

Theorem 3.6.2. *Let $\Gamma^{\lambda}_{\mu\nu}$ be a symmetric affine connection on M, and let $y \in M$ be any point. Then there exists a coordinate system x^{μ} on M in which $\Gamma^{\lambda}_{\mu\nu}(y) = 0$.*[2]

Proof. After changing coordinates using a translation, we may without loss of generality assume that $y^{\mu} = 0$. Consider now the coordinate transformation

$$x^{\mu'} = A^{\mu'}_{\alpha} x^{\alpha} + \frac{1}{2} B^{\mu'}_{\alpha\beta} x^{\alpha} x^{\beta}$$

where $A^{\mu'}_{\alpha}$ and $B^{\mu'}_{\alpha\beta}$ are constant arrays of real numbers, to be determined. We may without loss of generality assume $B^{\mu'}_{\alpha\beta} = B^{\mu'}_{\beta\alpha}$ (because its anti-symmetric piece must vanish due to $x^{\alpha}x^{\beta} = x^{\beta}x^{\alpha}$). Computing the first and second Jacobian of this coordinate change we get

$$J^{\mu'}_{\alpha} = A^{\mu'}_{\alpha} + B^{\mu'}_{\alpha\beta} x^{\beta}, \qquad J^{\mu'}_{\alpha\beta} = B^{\mu'}_{\alpha\beta}$$

(recall the second Jacobian from (3.1.4)). Thus, at the point y we have $J^{\mu'}_{\alpha}(y) = A^{\mu}_{\alpha}$. Therefore, as long as $A^{\mu'}_{\alpha}$ is an invertible matrix, the coordinate change is a bijection in a neighborhood of y. The transformation law for an affine connection can be written

$$J^{\mu'}_{\alpha} J^{\nu'}_{\beta} \Gamma^{\lambda'}_{\mu'\nu'} = J^{\lambda'}_{\gamma} \Gamma^{\gamma}_{\alpha\beta} - J^{\lambda'}_{\alpha\beta}.$$

(See Problem 1) At the point y we have $y^{\mu} = 0$, so the right hand side vanishes at y if we choose

$$B^{\lambda'}_{\alpha\beta} = A^{\lambda'}_{\gamma} \Gamma^{\gamma}_{\alpha\beta}(y)$$

Note that this requires $\Gamma^{\gamma}_{\alpha\beta}$ to be symmetric. □

[2] Note that we here regarded $\Gamma^{\lambda}_{\mu\nu}$ as a function on M rather than on the image of x^{μ}. We can of course switch back and forth between viewpoints because x^{μ} is bijective onto its image.

Definition 3.6.3. A coordinate system x^μ on an affine connection space (M, Γ) is *locally inertial at* $y \in M$ if the components of the connection vanish at y:

$$\Gamma^\lambda_{\mu\nu}(y) = 0.$$

Such coordinate systems are also called *normal* and the point y is a *pole* of the coordinate system.

3.7 Homework #3

1. Show that the transformation law for an affine connection can be written
$$J^{\mu'}_{\alpha} J^{\nu'}_{\beta} \Gamma^{\lambda'}_{\mu'\nu'} = J^{\lambda'}_{\gamma} \Gamma^{\gamma}_{\alpha\beta} - J^{\lambda'}_{\alpha\beta}.$$

2. Show that if $\Gamma_\mu{}^\nu{}_\lambda$ and $\gamma_\mu{}^\nu{}_\lambda$ are two affine connections then their difference $\Gamma_\mu{}^\nu{}_\lambda - \gamma_\mu{}^\nu{}_\lambda$ is a $\binom{1}{2}$-tensor. (This suggests that an affine connection is a kind of potential.)

3. The *torsion* of an affine connection $\Gamma_\mu{}^\nu{}_\lambda$ is defined by $T_\mu{}^\nu{}_\lambda = \Gamma_\mu{}^\nu{}_\lambda - \Gamma_\lambda{}^\nu{}_\mu$. Show that the torsion is a $(1,2)$-tensor. In particular, conclude that the property of an affine connection to be symmetric is coordinate independent.

4. Differentiate $J^\mu_{\alpha'} J^{\alpha'}_\nu = \delta^\mu_\nu$ with respect to $x^{\lambda'}$ to show the following sometimes useful identity involving the second Jacobian:
$$J^{\mu}_{\alpha'\lambda'} J^{\alpha'}_{\nu} = -J^{\mu}_{\alpha'} J^{\beta}_{\lambda'} J^{\alpha'}_{\beta\nu}.$$
Use this to rewrite the transformation law for an affine connection so that the second term is a (slightly different looking) term with a plus sign.

5. Show that, in general, the set of affine connections on a space is not closed under multiplication by a scalar.

6. Given a covector field A_μ on an affine connection space, compare the two $\binom{0}{2}$-tensors $G_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\mu A_\nu$ and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. When are they the same?

7. Compute the second order covariant derivative $\nabla_\mu \nabla_\nu A^\lambda$ of a vector field A^λ. (Hint: First apply ∇_μ, using that $\nabla_\nu A^\lambda$ is a tensor field.)

8. Verify that the formula (3.4.3) does define a tensor field of type $\binom{0}{2}$.

Lecture 4

Curvature, Torsion, Metrics and Vielbeins

4.1 The Curvature and Torsion of an Affine Connection

Let $\mathscr{G} \subset \mathscr{G}_{\mathrm{sm}}$ be a smooth geometry and $(M, \mathscr{C})$ a $\mathscr{G}$-space equipped with an affine connection $\Gamma^{\lambda}_{\mu\nu}$. (In other words, let M be an affine connection space in a smooth setting.)

As we saw in the previous lecture, we then have available the covariant derivative ∇_μ, which operates on the space of all tensor fields. It is natural to ask to what extent the different components of ∇_μ commute.

To check this, we let A^λ be a vector field and compute $\nabla_\mu \nabla_\nu A^\lambda$ in order to compare it with $\nabla_\nu \nabla_\mu A^\lambda$. Since $\nabla_\nu A^\lambda$ is a $\binom{1}{1}$-tensor, the general formula of covariant derivative (3.5.1) tells us we get two correction terms, one with a minus sign corresponding to the covariant index ν, and one with a plus sign corresponding to the contravariant index λ:

$$\nabla_\mu \nabla_\nu A^\lambda = \partial_\mu (\nabla_\nu A^\lambda) - \Gamma^{\alpha}_{\mu\nu} \nabla_\alpha A^\lambda + \Gamma^{\lambda}_{\mu\alpha} \nabla_\nu A^\alpha. \tag{4.1.1}$$

Since the middle term already has a clear dependence on switching μ and ν, we keep it as is. For the other two terms we have, by the formula (3.2.1) for the covariant derivative of a vector field,

$$\begin{aligned} \partial_\mu (\nabla_v A^\lambda) &= \partial_\mu (\partial_\nu A^\lambda + \Gamma^{\lambda}_{\nu\alpha} A^\alpha) \\ &= \partial_\mu \partial_\nu A^\lambda + \partial_\mu \Gamma^{\lambda}_{\nu\alpha} A^\alpha + \Gamma^{\lambda}_{\nu\alpha} \partial_\mu A^\alpha \end{aligned} \tag{4.1.2}$$

where we used the product rule for the ordinary partial derivative, and

$$\Gamma^{\lambda}_{\mu\alpha}\partial_\nu A^\alpha + \Gamma^{\lambda}_{\mu\alpha}\Gamma^{\alpha}_{\nu\beta}A^\beta. \tag{4.1.3}$$

Substituting these into (4.1.1) we obtain

$$\begin{aligned}\nabla_\mu\nabla_\nu A^\lambda = {}& \partial_\mu\partial_\nu A^\lambda + \partial_\mu\Gamma^{\lambda}_{\nu\alpha}A^\alpha + \Gamma^{\lambda}_{\nu\alpha}\partial_\mu A^\alpha \\ & - \Gamma^{\alpha}_{\mu\nu}\nabla_\alpha A^\lambda \\ & + \Gamma^{\lambda}_{\mu\alpha}\partial_\nu A^\alpha + \Gamma^{\lambda}_{\mu\alpha}\Gamma^{\alpha}_{\nu\beta}A^\beta.\end{aligned} \tag{4.1.4}$$

When we switch μ and ν and substract, the red terms (the first, third, and fifth) cancel, and the remaining terms get anti-symmetrized in μ and ν:

$$\begin{aligned}&[\nabla_\mu,\nabla_\nu]A^\lambda \\ &= \big(\underbrace{\partial_\mu\Gamma^{\lambda}_{\nu\alpha} - \partial_\nu\Gamma^{\lambda}_{\mu\alpha} + \Gamma^{\lambda}_{\mu\beta}\Gamma^{\beta}_{\nu\alpha} - \Gamma^{\lambda}_{\nu\beta}\Gamma^{\beta}_{\mu\alpha}}_{\stackrel{\text{def}}{=} R^{\lambda}_{\mu\nu\alpha}}\big)A^\alpha - \big(\underbrace{\Gamma^{\alpha}_{\mu\nu} - \Gamma^{\alpha}_{\nu\mu}}_{\stackrel{\text{def}}{=} T^{\alpha}_{\mu\nu}}\big)\nabla_\alpha A^\lambda.\end{aligned} \tag{4.1.5}$$

Definition 4.1.1. $R^{\lambda}_{\mu\nu\alpha}$ is the *Riemann Curvature tensor* and $T^{\alpha}_{\mu\nu}$ is the *torsion tensor* of the affine connection.

In a previous homework problem, it was shown that $T^{\alpha}_{\mu\nu}$ is a tensor. Therefore $T^{\alpha}_{\mu\nu}\nabla_\alpha A^\lambda$ is a tensor. Since the left hand side of (4.1.5) is also a tensor, it follows that $R^{\lambda}_{\mu\nu\alpha}A^\alpha$ is a tensor for every vector field A^α. One can from this draw the conclusion that $R^{\lambda}_{\mu\nu\tau}$ is a tensor of type $\binom{1}{3}$, see Problem 3 of Section 4.5. Alternatively, one can directly verify that $R^{\lambda}_{\mu\nu\tau}$ is a tensor (see Problem 4 of Section 4.5).

Remark 4.1.2.

1. The torsion tensor vanishes if and only if the affine connection is symmetric in its two lower indices.

2. The curvature tensor $R^{\lambda}_{\mu\nu\tau}$ is anti-symmetric in its first two covariant indices μ and ν.

3. The formula for $[\nabla_\mu,\nabla_\nu]$ acting on other tensors is similar in form to the covariant derivative, in that we get a $+$ curvature term for each contravariant index, and a $-$ curvature term for each covariant index, though only a single $-$ torsion term. For example:

$$\begin{aligned}[\nabla_\mu,\nabla_\nu]B^{\sigma\tau}_{\pi} = {}& R^{\sigma}_{\mu\nu\alpha}B^{\alpha\tau}_{\pi} + R^{\tau}_{\mu\nu\alpha}B^{\sigma\alpha} \\ & - R^{\alpha}_{\mu\nu\pi}B^{\sigma\tau}_{\alpha} - T^{\alpha}_{\mu\nu}\nabla_\alpha B^{\sigma\tau}_{\pi}\end{aligned} \tag{4.1.6}$$

(We leave it to the reader to verify this.)

4. The *Ricci tensor* is obtained by contracting the curvature tensor:

$$R_{\mu\nu} = R^{\alpha}_{\alpha\mu\nu}. \tag{4.1.7}$$

   Warning: There is a sign convention here; some authors define $R_{\mu\nu} = R^{\alpha}_{\mu\alpha\nu}$.

5. The *scalar curvature* is defined by

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{4.1.8}$$

4.2 Metrics and Vielbeins

Definition 4.2.1. A *metric* $g_{\mu\nu}$ on M is a symmetric non-degenerate $\binom{0}{2}$-tensor. Symmetric means $g_{\mu\nu} = g_{\nu\mu}$, while non-degenerate means there is a contravariant tensor $\widetilde{g}^{\mu\nu}$ (sometimes called the *inverse metric*) such that

$$g_{\mu\nu}\widetilde{g}^{\nu\lambda} = \delta^{\lambda}_{\mu}. \tag{4.2.1}$$

Example 4.2.2. The flat metric η_{ab} is a metric in the Lorentzian geometry.

If X^{μ} and Y^{μ} are two vector fields, then

$$g_{\mu\nu} X^{\mu} Y^{\nu} \tag{4.2.2}$$

is a scalar field. Thus a metric is a kind of "dot product". More precisely, at each point on M, a metric $g_{\mu\nu}$ defines a non-degenerate bilinear form on the tangent space (defined as the space of all vector fields evaluated at that point). By Gram-Schmidt's orthonormalization process from linear algebra, we know that such a bilinear form can be diagonalized. In our notation, it means there exists a matrix $e^{a}_{\mu}(x)$ defined at each point x on M, such that

$$g_{\mu\nu}(x) = e^{a}_{\mu}(x) e^{b}_{\nu}(x) \eta_{ab} \tag{4.2.3}$$

where η_{ab} is diagonal with ± 1 on the diagonal (since square roots are available in $\mathbb{R}$). The pair (t, s) where t is the number of $+1$'s and s the number of -1's, is the *signature* of $g_{\mu\nu}$. That e^{a}_{μ} are smooth functions on M follows from the fact that the operations in the orthonormalization process are algebraic.

The field e^{a}_{μ} transform as a covector in μ. On the other hand, $e^{a}_{\mu}(x)$ is uniquely determined (by $g_{\mu\nu}$ and the relation (4.2.3)) up to a (not necessarily proper) Lorentz tranformation acting on the top index: if Λ^{b}_{a} is any matrix such that $\Lambda^{a}_{i}\Lambda^{b}_{j}\eta_{ab} = \eta_{ij}$

then $e^a_\mu(x)$ may be replaced by $\Lambda^a_b e^b_\mu(x)$. Conversely, any two solutions to (4.2.3) are related this way.

Thus, the Vielbein is subject to two distinct transformation laws:

$$e^a_{\mu'} = J^\nu_{\mu'} e^a_\nu, \qquad e^{a'}_\mu = \Lambda^{a'}_b e^b_\mu. \tag{4.2.4}$$

The fact that the matrix is invertible says there is a field e^μ_a transforming as a general vector and as a Lorentz covector such that

$$e^a_\mu e^\mu_b = \delta^a_b, \qquad e^a_\mu e^\nu_a = \delta^\nu_\mu. \tag{4.2.5}$$

In e^a_μ we call a a *flat index* and μ is a *curved index*. Conversely, given any such field e^a_μ, we may use (4.2.3) to define a metric.

The field e^a_μ is called the *Vielbein* (German for many-legged) of the metric. Other common names for e^a_μ is *soldering form*, *orthogonal frame section*, *tetrad* or *Vierbein* ($n = 4$), *Dreibein* ($n = 3$), *Tweibein* ($n = 2$).

The Vielbein plays an important role in physics and geometry. For example, it can serve as an (equivalent) starting point for gravity, wherein the Vielbein takes over the role of the graviton field. It also provides a way to define spinor fields in curved spacetime.

4.3 Raising and Lowering of Indices

Let $g_{\mu\nu}$ be a metric. If Y^μ is a vector field, then we define the covector field

$$Y_\mu := g_{\mu\alpha} Y^\alpha. \tag{4.3.1}$$

We say that we have *lowered* the index μ (with respect to the metric $g_{\mu\nu}$). Similarly, we may turn a covector field A_μ into a vector field using the contravariant (what some call the "inverse") metric:

$$A^\mu = \widetilde{g}^{\mu\alpha} A_\alpha, \tag{4.3.2}$$

and we say we have *raised* the index μ in A_μ. These processes are mutually inverse:

$$Y^\mu \rightsquigarrow g_{\mu\alpha} Y^\alpha \rightsquigarrow \widetilde{g}^{\mu\beta} g_{\beta\alpha} Y^\alpha = \delta^\mu_\alpha Y^\alpha = Y^\mu.$$

For a tensor $A^{\mu\nu}$ we wish to be able to distinguish between the result of lowering μ or ν. It is natural to define

$$A_\mu{}^\nu = g_{\mu\alpha} A^{\alpha\nu}$$

$$A^{\mu}{}_{\nu} = g_{\nu\alpha}A^{\mu\alpha}$$

Unless $A^{\mu\nu}$ is symmetric, these are two different $\binom{1}{1}$-tensors. This makes **horizontal placement of indices** important, when we have a metric present and plan to raise and lower indices. Similarly, when a metric is present, we should avoid writing B^{ν}_{μ} because it is completely unclear what $B_{\mu\nu}$ means: $g_{\mu\alpha}B^{\alpha}_{\nu}$ or $g_{\nu\alpha}B^{\alpha}_{\mu}$? The rule is to simply **avoid writing one index directly above another**, and either denote a $\binom{1}{1}$-tensor by $B_{\mu}{}^{\nu}$ or $B^{\nu}{}_{\mu}$. This is purely a notational distinction. There are two conventions that have some logic to them, but many texts mix and match:

In the **prefix** notation, the "default" form of writing a tensor (or connection, etc) is $T_{\nu\ldots}{}^{\mu\cdots}$. This form is preserved by application of a differential operator from the left:

$$\nabla_{\mu}A_{\nu}{}^{\lambda} = B_{\mu\nu}{}^{\lambda}$$
$$2\Gamma_{\mu\nu}{}^{\lambda} = (\partial_{\mu}g_{\nu\alpha} + \cdots)g^{\alpha\lambda}$$
$$R_{\mu\nu\lambda}{}^{\rho} = \partial_{\mu}\Gamma_{\nu\lambda}{}^{\rho} + \cdots$$

We will mostly follow this convention.

In the **postfix** notation, one defaults to $T^{\mu\cdots}{}_{\nu\ldots}$. In this convention, it is logical to write the result of derivatives with an index on the right (to preserve the default index form). For partial derivatives, one defines

$$A^{\mu}{}_{\nu,\lambda} \stackrel{\text{def}}{=} \partial_{\lambda}A^{\mu}{}_{\nu}$$

and use ; or | for covariant derivatives:

$$A^{\mu}{}_{\nu;\lambda} = A^{\mu}{}_{\nu|\lambda} \stackrel{\text{def}}{=} \nabla_{\lambda}A^{\mu}{}_{\nu}$$

The second covariant derivative of a vector field X^{μ} is written $X^{\mu}{}_{|\nu\lambda}$ or $X^{\mu}{}_{|\nu|\lambda}$. Connections and curvature tensors are then typically written

$$2\Gamma^{\lambda}{}_{\mu\nu} = g^{\lambda\alpha}(g_{\alpha\mu,\nu} + \cdots)$$
$$Y^{\lambda}{}_{|\mu|\nu} - Y^{\lambda}{}_{|\nu|\mu} = Y^{\alpha}R^{\lambda}{}_{\alpha\mu\nu} - Y^{\lambda}{}_{|\alpha}T^{\alpha}{}_{\mu\nu}$$
$$R^{\lambda}{}_{\mu\nu\rho} = \Gamma^{\lambda}{}_{\mu\nu,\rho} + \cdots$$

Here there one needs to use slight caution: $\nabla_{\mu}\nabla_{\nu}Y^{\lambda} = Y^{\lambda}{}_{|\nu|\mu}$. Thus, one can encounter a different sign convention for the curvature tensor, justified by the notation. (In addition, there is a different sign convention in the definition of the Ricci tensor). Ultimately these signs only matter once we compute the energy component of the energy-momentum tensor of our lagrangian density: we want the energy to be non-negative.

Remark 4.3.1. A curious fact is revealed when we raise both indices in the metric itself:

$$g^{\mu\nu} = \widetilde{g}^{\mu\alpha}\widetilde{g}^{\nu\beta}g_{\alpha\beta} = \widetilde{g}^{\mu\alpha}\delta^{\nu}_{\alpha} = \widetilde{g}^{\mu\nu}.$$

In other words, raising the metric gives the contravariant metric. For this reason, the contravariant metric is denoted by the simpler form $g^{\mu\nu}$. Similarly, the raising μ (using $g_{\mu\nu}$) and lowering a (using the flat metric) in the Vielbein e^{a}_{μ} gives the inverse Vielbein e^{μ}_{a}.

4.4 The Levi-Civita Connection of a Metric

Definition 4.4.1. (i) A $\mathscr{G}$-space M equipped with a metric $g_{\mu\nu}$ is a *pseudo-Riemannian* $\mathscr{G}$-space. If the metric is positive definite (i.e. its diagonal form η_{ab} has all $+1$'s on the diagonal) then M is a Riemannian $\mathscr{G}$-space.

(ii) A connection $\Gamma_{\mu\nu}{}^{\lambda}$ on a pseudo-Riemannina $\mathscr{G}$-space $(M, g_{\mu\nu})$ is called *metric* (or *metric-compatible*) if

$$\nabla_{\lambda} g_{\mu\nu} = 0 \tag{4.4.1}$$

where the covariant derivative is computed using $\Gamma_{\mu\nu}{}^{\lambda}$.

Theorem 4.4.2 (The Fundamental Theorem of (pseudo-)Riemannian Geometry)**.** *Let $(M, g_{\mu\nu})$ be a pseudo-Riemannian $\mathscr{G}$-space (where $\mathscr{G} \subset \mathscr{G}_{\mathrm{sm}}$ is any smooth geometry). Then there exists a unique affine connection $\Gamma_{\mu\nu}{}^{\lambda}$ on M satisfying the following two conditions:*

1) *The torsion tensor $T_{\mu\nu}{}^{\lambda} = \Gamma_{\mu\nu}{}^{\lambda} - \Gamma_{\nu\mu}{}^{\lambda}$ vanishes identically.*

2) *$\Gamma_{\mu\nu}{}^{\lambda}$ is metric-compatible.*

This affine connection is explicitly given by

$$\Gamma_{\mu\nu}{}^{\lambda} = \frac{1}{2}\big(\partial_{\mu} g_{\nu\alpha} + \partial_{\nu} g_{\mu\alpha} - \partial_{\alpha} g_{\mu\nu}\big) g^{\lambda\alpha}. \tag{4.4.2}$$

Proof. Outlined in the Homework #4 below. □

Definition 4.4.3. The affine connection is called the *Levi-Civita connection* on $(M, g_{\mu\nu})$. It is sometimes denoted by

$$\overset{g}{\Gamma}_{\mu\nu}{}^{\lambda} \quad \text{or} \quad \gamma_{\mu\nu}{}^{\lambda} \quad \text{or} \quad \left\{ \begin{matrix} \lambda \\ \mu \ \ \nu \end{matrix} \right\}$$

and the components called *Christoffel symbols*.

4.5 Homework #4

1. Let S^μ and T^μ be two vector fields on a pseudo-Riemannian $\mathscr{G}$-space $(M, g_{\mu\nu})$. Using notation of lowering and raising, show that $S^\mu T_\mu = S_\mu T^\mu$.

2. Show that the Jacobian $J_\mu^{\nu'}$ (for fixed μ) transforms as a vector in ν', and (for fixed ν') as a covector in μ:

$$J_\mu^{\nu''} = J_{\alpha'}^{\nu''} J_\mu^{\alpha'}, \qquad J_{\mu''}^{\nu'} = J_{\mu''}^{\alpha'} J_{\alpha'}^{\nu'} \tag{4.5.1}$$

   [Thus, on the one hand, for a fixed coordinate system $x \mapsto x^\mu$, the Jacobian defines n vector fields, denoted $\boldsymbol{\partial}_\mu$. The components of $\boldsymbol{\partial}_\mu$ in an arbitrary coordinate system $x \mapsto x^{\nu'}$ are $(\boldsymbol{\partial}_\mu)^{\nu'} = J_\mu^{\nu'}$. These are the *basis vector fields* relative to the coordinate system x^μ. On the other hand, for a fixed coordinate system $x \mapsto x^\mu$, the Jacobian defines n covector fields, denoted dx^μ with components in an arbitrary coordinate system $x \mapsto x^{\mu'}$ given by $(dx^\mu)_{\nu'} = J_{\nu'}^\mu$. These are the *basis covector fields* (or *basis* 1*-forms*) relative to the coordinate system x^μ. Thus,

$$(\boldsymbol{\partial}_{\nu'})^\mu = (dx^\mu)_{\nu'} = J_{\nu'}^\mu. \tag{4.5.2}$$

3. Suppose that a tuple of functions $R^\lambda_{\mu\nu\tau}$ are given in each coordinate system, but no transformation law is assumed between different coordinate systems. But, suppose that we do know that the quantities $R^\lambda_{\mu\nu\alpha}A^\alpha$ define a $\binom{1}{3}$-tensor for each vector field A^α. Prove that in that case the quantities $R^\lambda_{\mu\nu\tau}$ also transform as a $\binom{1}{3}$-tensor. (The same is true for any other type, but this case is of particular relevance to the curvature tensor.) (Hint: Fix a coordinate system $x \mapsto x^{\mu'}$ and take the vector field A^μ to be one of the corresponding coordinate vector fields: $A^\mu = (\boldsymbol{\partial}_{\nu'})^\mu = J_{\nu'}^\mu$.)

4. Show directly, using the transformation laws for an affine connection, that the Riemann curvature tensor $R^\lambda_{\mu\nu\tau}$ is in fact a $\binom{1}{3}$-tensor. (Hint: Calculations can be cut in half by writing "$R^\lambda_{\mu\nu\tau} = \partial_\mu \Gamma^\lambda_{\nu\tau} + \Gamma^\lambda_{\mu\alpha}\Gamma^\alpha_{\nu\tau} - \{\mu \leftrightarrow \nu\}$".)

5. Prove the Fundamental Theorem of (pseudo-)Riemannian Geometry as follows:

   (a) Suppose that $g_{\mu\nu}$ is a metric, and that $\Gamma_{\mu\nu}{}^\lambda$ is any symmetric affine connection such that $\nabla_\lambda g_{\mu\nu} = 0$. Using the definition of the covariant derivative (see e.g. (3.5.1)), find an equation involving the metric and the affine connection. It should have the symbolic form $\partial g - \Gamma g - \Gamma g = 0$.

(b) Let $\Gamma_{\mu\nu\lambda} = \Gamma_{\mu\nu}{}^{\alpha} g_{\alpha\lambda}$ be the so-called *covariant connection.*[1] Use the part (a) to show that

$$\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} = 2\Gamma_{\mu\nu\lambda}.$$

(c) Conclude that the connection is therefore determined by the metric:

$$\Gamma_{\mu\nu}{}^{\lambda} = \frac{1}{2} g^{\lambda\alpha}\big(\partial_\mu g_{\nu\alpha} + \partial_\nu g_{\mu\alpha} - \partial_\alpha g_{\mu\nu}\big) \tag{4.5.3}$$

(d) Conversely, prove that for any metric $g_{\mu\nu}$, the formula (4.5.3) actually defines an affine connection. This is the Levi-Civita connection on a pseudo-Riemannian space $(M, g_{\mu\nu})$.

6. Let e^a_μ be a Vielbein field. Show that $\Gamma^\lambda_{\mu\nu} = e^\lambda_a \partial_\mu e^a_\nu$ defines an affine connection. [This is actually a much easier formula than the traditional one (4.5.3) in terms of the metric! Note that the torsion does not necessarily vanish here.]

7. Let $\Gamma_{\mu\nu}{}^{\lambda}$ be the Levi-Civita connection of a metric $g_{\mu\nu}$. Fix a coordinate system x^μ and a point $x \in M$. Show that the following two statements are equivalent:

(a) $\Gamma_{\mu\nu}{}^{\lambda}(x) = 0$ (that is, this is a locally inertial coordinate system at x, see Definition 3.6.3)

(b) $\partial_\lambda g_{\mu\nu}(x) = 0$ (that is, in this coordinate system, the metric is constant to first order at x).

8. Let $g_{\mu\nu}$ be a metric and $\Gamma_{\mu\nu}{}^{\lambda}$ the corresponding Levi-Civita connection. Let $R_{\mu\nu\lambda\rho} = g_{\rho\alpha} R_{\mu\nu\lambda}{}^{\alpha}$ be the *covariant curvature tensor.* Prove that in a locally inertial coordinate system at x,

$$R_{\mu\nu\lambda\rho} = \frac{1}{2}\big(\partial_\mu\partial_\lambda g_{\nu\rho} + \partial_\nu\partial_\rho g_{\mu\lambda} - \partial_\nu\partial_\lambda g_{\mu\rho} - \partial_\mu\partial_\rho g_{\nu\lambda}\big) \tag{4.5.4}$$

9. Show that the covariant curvature tensor of the Levi-Civita connection satisfies the C_2-symmetries[2]

$$R_{\mu\nu\lambda\rho} = -R_{\nu\mu\lambda\rho}, \qquad R_{\mu\nu\lambda\rho} = -R_{\mu\nu\rho\lambda}, \qquad R_{\mu\nu\lambda\rho} = R_{\lambda\rho\mu\nu}, \tag{4.5.5}$$

and the C_3-symmetry

$$R_{\mu\nu\lambda\rho} + R_{\mu\lambda\rho\nu} + R_{\mu\rho\nu\lambda} = 0. \tag{4.5.6}$$

[1] It is only "covariant" in the sense that it has all indices downstairs. It is not a $\binom{0}{3}$-tensor but satisfies some other transformation law.

[2] By C_k we mean the cyclic group of order k.

[Hint: The first one in (4.5.5) holds by definition of the curvature tensor. For the second one, by the previous problem $R_{\mu\nu\lambda\rho}(x) = R_{\mu\lambda\rho\nu}(x)$ at the pole x of an inertial coordinate system. Since that identity is tensorial, it must hold (at x) in any coordinate system. But x was arbitary so it holds everywhere. The other identity is proved similarly.]

10. Use (4.5.5) to show that the Ricci tensor of the Levi-Civita connection is symmetric.

11. (a) Show that if x, y, z are three linear operators on a vector space, then

$$[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$$

where $[a,b] = a \circ b - b \circ a$ is the commutator. (This is the *Jacobi identity.*)

(b) Assuming $\Gamma^\lambda_{\mu\nu}$ is a symmetric affine connection, apply part (a) to the covariant derivatives $x = \nabla_\mu, y = \nabla_\nu, z = \nabla_\lambda$, to prove the *Bianchi Identity*

$$\nabla_\mu R^\sigma_{\nu\lambda\tau} + \nabla_\nu R^\sigma_{\lambda\mu\tau} + \nabla_\lambda R^\sigma_{\mu\nu\tau} = 0. \tag{4.5.7}$$

12. Give a second proof of the Bianchi Identity using the hint in Problem 9.

13. Show that under a spatial spherical change of coordinates $(x^0, x^1, x^2, x^3) \to (t, r, \theta, \phi)$, the Lorentzian flat metric is transformed to

$$ds^2 = dt^2 - dr^2 - r^2 d\Omega^2$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. [Here we use the *line element notation* $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$ to specify the metric.]

14. For constant r and t, we have $dt = dr = 0$ in the previous problem, and we get (after negating for simplicity) the metric $ds^2 = r^2 d\Omega^2$ on the static 2-sphere. In other words, $n = 2$, and

$$g_{\theta\theta} = r^2, \qquad g_{\phi\phi} = r^2 \sin^2\theta, \qquad g_{\phi\theta} = g_{\theta\phi} = 0.$$

Compute the Levi-Civita connection, curvature tensor, Ricci tensor. Show that the curvature scalar R here is constant, equal to $\frac{2}{r^2}$.

Lecture 5

Invariant Integration

5.1 Integration and Scalar Densities

Let $K \subset M$ and consider the idea of an integral

$$I = \int_K \phi \, d^n x \tag{5.1.1}$$

where ϕ is some object to be determined. To make sense of this, we choose a coordinate system $x \mapsto x^\mu$ (from the collection $\mathcal{C}$ that M comes equipped with), and we define

$$\int_K \phi \, d^n x \stackrel{\text{def}}{=} \int_{K_1} \phi(x^\mu) \, dx^1 dx^2 \cdots dx^n \tag{5.1.2}$$

where $\phi(x^\mu)$ is a function expressing ϕ in the coordinate system $x \mapsto x^\mu$, and $K_1 = \{x^\mu \mid x \in K\}$. But, we want this definition to be **independent of the choice of coordinate system**. This imposes a relationship between the different coordinate expressions of ϕ. More precisely, suppose $x \mapsto x^{\mu'}$ is another coordinate system (from $\mathcal{C}$) on M. Then we need

$$\int_K \phi \, d^n x = \int_{K_2} \phi(x^{\mu'}) \, dx^{1'} dx^{2'} \cdots dx^{n'}.$$

The two coordinate systems are related by a transformation belonging to the geometry $\mathscr{G}$ under consideration. We will make the assumption on $\mathscr{G}$ that all coordinate transformations have a positive Jacobian determinant $J\prime = \det(\partial x^\mu / \partial x^{\nu'})$. Making the substitution $x^\mu = x^\mu(x^{\mu'})$ in (5.1.2), we obtain

$$\int_K \phi \, d^x = \int_{K_2} \phi(x^\mu(x^{\nu'})) \cdot J\prime \, dx^{1'} dx^{2'} \cdots dx^{n'} \tag{5.1.3}$$

Comparing (5.1.2) with (5.1.3), we see that we want

$$\phi(x^{\mu'}) = \phi(x^{\mu}(x^{\nu'})) \cdot J_{\prime} \tag{5.1.4}$$

We arrive at the following definition.

Definition 5.1.1. A *scalar density* ϕ assigns to each coordinate system x^{μ} a (real-valued, smooth) function $\phi(x^{\mu})$, called the *expression of* ϕ *in the coordinates* x^{μ}. Furthermore, if x^{μ} and $x^{\mu'}$ are any two coordinate systems on M, their respective expressions are related by the transformation law (5.1.4).

Example 5.1.2. Suppose that the expression of $\phi : M \to \mathbb{R}$ in a coordinate system (x, y) is $\phi(x, y) = x^3 + xy^2$. Suppose $K \subset M$ is a region whose image under the coordinate map $p \mapsto (x(p), y(p))$ is $K_1 = \{(x, y) \mid x^2 + y^2 \leq 1\}$ (the unit disk). Then, by definition,

$$\int_K \phi \, d^2x = \iint_{K_1} (x^3 + xy^2) \, dxdy$$

Suppose further that (r, θ) is another coordinate system, related to the first by polar change of coordinates $x = r\cos\theta$, $y = r\sin\theta$. The Jacobian determinant is $\det \left(\begin{smallmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{smallmatrix} \right) = r$ and therefore, with $K_2 = \{(r, \theta) \mid r \leq 1\}$ and the "naive substitution" is $\phi(x(r, \theta), y(r, \theta)) = r^3 \cos\theta$. As we know from calculus, therefore,

$$\int_K \phi \, d^n x = \iint_{K_2} r^3 \cos\theta \cdot r \, drd\theta.$$

This illustrates that the *expression* $\phi(r, \theta)$ *of* ϕ *in* (r, θ) *is not merely the "naive substitution", but we also have to multiply by a Jacobian determinant.* In other words: $\phi(r, \theta) = r^3 \cos\theta \cdot r$.

5.2 Relative Tensor Fields

If ϕ and ψ are two scalar densities, then ϕ/ψ is a scalar field. Therefore we also call them relative scalars. However, the product $\phi\psi$ transforms as $\phi'\psi' = \phi\psi(J_{\prime})^2$. We say that $\phi\psi$ is a relative scalar of weight 2. Multiplying a relative scalar and a tensor gives a "relative tensor". The definition that cover everything we need is as follows.

Definition 5.2.1. A *relative tensor (field) of type* $\binom{r}{s}$ *and weight* w is an n^{r+s}-tuple of functions $S_{\nu\ldots}{}^{\mu\cdots}$ given in each coordinate system x^{μ}, which are related by the transformation law

$$S_{\nu'\ldots}{}^{\mu'\cdots} = (J_{\prime})^w \cdot J^{\alpha}_{\nu'} \cdots J^{\mu'}_{\beta} \cdots S_{\alpha\ldots}{}^{\beta\cdots}. \tag{5.2.1}$$

Relative tensors/vectors/covectors of weight 1 are called *tensor/vector/covector densities*. (We will only need the case when w is an integer. In geometries where all Jacobian determinants are positive, w could in principle be any real number.)

Theorem 5.2.2. *Relative tensors have the following properties.*

(i) *The set of all relative tensors of type* $\binom{r}{s}$ *and weight* w *forms a vector space (i.e. it is closed under addition and multiplication by constant scalars).*

(ii) *Relative tensors can be multiplied; the resulting weight is the sum of the respective weights (just like the types).*

(iii) *The* covariant derivative *of a relative tensor field of type* $\binom{r}{s}$ *and weight* w *is defined by*

$$\begin{aligned}\nabla_\lambda T_{\mu\cdots}{}^{\nu\cdots} &= \partial_\lambda T_{\mu\cdots}{}^{\nu\cdots} \\ &\quad + \Gamma_{\lambda\alpha}{}^{\nu} T_{\mu\cdots}{}^{\alpha\cdots} + \cdots \quad \textit{(one term for each contravariant index)} \\ &\quad - \Gamma_{\lambda\mu}{}^{\alpha} T_{\alpha\cdots}{}^{\nu\cdots} - \cdots \quad \textit{(one term for each covariant index)} \\ &\quad - w\Gamma_{\alpha\lambda}{}^{\alpha} T_{\mu\cdots}{}^{\nu\cdots} \quad \textit{(a single extra term)}\end{aligned} \tag{5.2.2}$$

The result is a tensor of type $\binom{r}{s+1}$ *and weight* w.

(iv) *The product rule works as usual for taking the covariant derivative of any products of relative tensors. Dropping all indices for brevity, the rule reads*

$$\nabla(TS) = (\nabla T)S + S(\nabla T). \tag{5.2.3}$$

Example 5.2.3. If j^μ is a vector density, then its "covariant divergence" is

$$\nabla_\mu j^\mu = \partial_\mu j^\mu + \Gamma_{\mu\alpha}{}^{\mu} j^\alpha - \Gamma_{\beta\mu}{}^{\beta} j^\mu = \partial_\mu j^\mu \tag{5.2.4}$$

In particular, the usual partial derivative $\partial_\mu j^\mu$ is a scalar density! [1]

As a consequence of this remarkable fact is that the integral of the covariant divergence of a vector density has coordinate independent meaning. An important result related to this is the divergence theorem. We only need the case when j^μ vanishes on the boundary of K.

[1] Furthermore, this fact — although very special to the vector density case — can be a useful way of remembering the formula for the covariant derivative of a relative tensor.

Theorem 5.2.4 (Special Case of the Divergence Theorem from Calculus)**.** *For any vector density j^μ vanishing on the boundary ∂K of a region $K \subset M$, we have*

$$\int_K \partial_\mu j^\mu \, d^n x = 0. \tag{5.2.5}$$

Remark 5.2.5. In practice, we will only need to consider regions $K \subset M$ which in some coordinate system is a closed ball.

5.3 The Levi-Civita Symbols, and Determinants

We define, in all coordinate systems,

$$\varepsilon_{\mu_1 \cdots \mu_n} = \varepsilon^{\mu_1 \mu_2 \cdots \mu_n} = \begin{cases} +1, & \text{if } \mu_1 \cdots \mu_n \text{ is an even permutation of } 12 \cdots n, \\ -1, & \text{if } \mu_1 \cdots \mu_n \text{ is an odd permutation of } 12 \cdots n, \\ 0, & \text{if } \mu_1 \cdots \mu_n \text{ is not a permutation of } 12 \cdots n. \end{cases} \tag{5.3.1}$$

These are called the *Levi-Civita symbols*. Note that the number of indices is equal to the dimension n of the space M.

Example 5.3.1. When $n = 2$, we have $\varepsilon_{12} = 1, \varepsilon_{21} = -1, \varepsilon_{11} = \varepsilon_{22} = 0$.

When $n = 3$, $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1$, $\varepsilon_{132} = \varepsilon_{321} = \varepsilon_{213} = -1$, and all other components (having at least two indices equal) are zero.

This notation is convenient for expressing determinants. For example,

$$\det(A_\mu{}^\nu) = \varepsilon^{\nu_1 \cdots \nu_n} A_{\nu_1}{}^1 \cdots A_{\nu_n}{}^n. \tag{5.3.2}$$

(Check this for $n = 2, 3$!) It can also be written in a more tensorial way:

$$\det(A_\mu{}^\nu) \varepsilon^{\mu_1 \cdots \mu_n} = \varepsilon^{\nu_1 \cdots \nu_n} A_{\nu_1}{}^{\mu_1} \cdots A_{\nu_n}{}^{\mu_n} \tag{5.3.3}$$

Something interesting is learned when this formula is applied to the Jacobian matrix $J_\nu^{\mu'}$ associated to a coordinate change:

$$J_{\nu_1}^{\mu'_1} \cdots J_{\nu_n}^{\mu'_n} \varepsilon^{\nu_1 \cdots \nu_n} = J' \cdot \varepsilon^{\mu'_1 \cdots \mu'_n}$$

Multiplying both sides by $J_\prime = (J')^{-1}$ we get

$$J_\prime \cdot J_{\nu_1}^{\mu'_1} \cdots J_{\nu_n}^{\mu'_n} \varepsilon^{\nu_1 \cdots \nu_n} = \varepsilon^{\mu'_1 \cdots \mu'_n}. \tag{5.3.4}$$

This shows the first part of the following proposition.

Proposition 5.3.2.

(a) *The contravariant Levi-Civita symbol*

$$\varepsilon^{\mu_1\cdots\mu_n}$$

is a relative tensor of type $\binom{n}{0}$ *and weight* 1 *(i.e. an* $\binom{n}{0}$*-tensor density).*

(b) *The covariant Levi-Civita symbol*

$$\varepsilon_{\mu_1\cdots\mu_n}$$

is a relative tensor of type $\binom{n}{0}$ *and weight* -1*.*

An importan application of this is that by anti-symmetrizing (i.e. contracting against a Levi-Civita symbol) we can create scalar densities, which can be integrated in a coordinate-independent manner:

Corollary 5.3.3. *If* $C_{\mu_1\mu_2\cdots\mu_n}$ *is any* $\binom{0}{n}$*-tensor, then*

$$\varepsilon^{\mu_1\cdots\mu_n}C_{\mu_1\cdots\mu_n} \tag{5.3.5}$$

is a scalar density.

5.4 The Scalar Density $\sqrt{-g}$

Another important way to build scalar densities comes about when we have a metric available. The determinant of the metric $g_{\mu\nu}$ is denoted by g. It can be written

$$g = \det(g_{\mu\nu}) = \frac{1}{n!}\varepsilon^{\mu_1\cdots\mu_n}\varepsilon^{\nu_1\cdots\nu_n}g_{\mu_1\nu_1}\cdots g_{\mu_n\nu_n}. \tag{5.4.1}$$

Since this is a fully contracted product of two relative tensors of weight 1 and a one of weight 0, it follows that g is a relative scalar of weight 2. That is,

$$g' = (J\prime)^2 g, \tag{5.4.2}$$

where $g' = \det(g_{\mu'\nu'})$. Assuming that $J\prime > 0$ for all coordinate changes, taking absolute values on both sides and then the square roots, we conclude that

$$\sqrt{|g'|} = (J\prime)\sqrt{|g|}. \tag{5.4.3}$$

This proves the following result about the metric.

Proposition 5.4.1. *Let $g_{\mu\nu}$ be a metric on a $\mathscr{G}$-space M, where $\mathscr{G}$ is an oriented geometry (i.e. all coordinate transformations in $\mathscr{G}$ have positive determinant). Then the field*

$$\sqrt{|g|} \tag{5.4.4}$$

is a scalar density. In particular, when $n = 4$ and $g_{\mu\nu}$ has Lorentzian signature (either $+---$ or $-+++$), then $g < 0$ and consequently

$$\sqrt{-g} \tag{5.4.5}$$

is a scalar density.

5.5 Useful Formulas Involving the Metric

In this section we provide a number of useful formulas involving the metric.

Let $g_{\mu\nu}$ be any metric and g its determinant.

Lemma 5.5.1.

$$\boxed{\partial_\mu g = g g^{\alpha\beta} \partial_\mu g_{\alpha\beta}} \tag{5.5.1}$$

Proof. By cofactor expansion of the determinant,

$$g\delta^\rho_\lambda = g_{\lambda\alpha} C^{\alpha\rho} \tag{5.5.2}$$

where

$$C^{\lambda\rho} = g \cdot g^{\lambda\rho} \tag{5.5.3}$$

Regarding the g as a polynomial in n^2 variables $g_{\mu\nu}$, we have by the chain rule [2]

$$\partial_\mu g = \frac{\partial g}{\partial g_{\alpha\beta}} \partial_\mu g_{\alpha\beta} \tag{5.5.4}$$

By (5.5.2), and that each variable $g_{\mu\nu}$ only once in a single monomial of g, we have

$$\partial_\mu g = C^{\alpha\beta} \partial_\mu g_{\alpha\beta}$$

Now, using (5.5.3) we get

$$\partial_\mu g = g g^{\alpha\beta} \partial_\mu g_{\alpha\beta}.$$

□

[2] Note that $\partial g / \partial g_{\alpha\beta}$ should be regarded as a contravariant object, i.e. it has two *upper* indices. This is for three reasons: In this way the contraction against $\partial_\mu g_{\alpha\beta}$ is valid. Secondly, (5.5.4) shows $\partial g / \partial g_{\alpha\beta}$ is a $\binom{2}{0}$-tensor in any geometry with unimodular ($J = 1$) transformations. Third, it is analogous to how $\partial_\mu \phi = \partial \phi / \partial x^\mu$ is a $\binom{0}{1}$-tensor for scalars ϕ.

Let $|g|$ be the absolute value of the determinant g of the metric.

Lemma 5.5.2.

$$\boxed{\partial_\mu \sqrt{|g|} = \frac{1}{2}\sqrt{|g|} \cdot g^{\alpha\beta}\partial_\mu g_{\alpha\beta}} \tag{5.5.5}$$

Proof. We write the proof when $|g| = -g$.

$$\begin{aligned}
\partial_\mu \sqrt{-g} &= \frac{1}{2}(-g)^{-1/2}(-\partial_\mu g) && \text{by the chain rule} \\
&= \frac{1}{2}(-g)^{-1/2} \cdot (-g g^{\alpha\beta}\partial_\mu g_{\alpha\beta}) && \text{by Lemma 5.5.1} \\
&= \frac{1}{2}\sqrt{-g} g^{\alpha\beta}\partial_\mu g_{\alpha\beta}.
\end{aligned}$$

□

Lemma 5.5.3. *Let ${\Gamma_{\mu\nu}}^\lambda$ be the Levi-Civita connection corresponding $g_{\mu\nu}$. Then*

$$\boxed{{\Gamma_{\alpha\nu}}^\alpha = \frac{\partial_\nu \sqrt{|g|}}{\sqrt{|g|}} = \partial_\nu \ln \sqrt{|g|}} \tag{5.5.6}$$

Proof. The covariant Levi-Civita connection is given by

$$2\Gamma_{\mu\nu\lambda} = \partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}.$$

Since the first plus the third term is $\partial_\mu g_{\nu\lambda} - \partial_\lambda g_{\nu\mu}$, those terms vanish upon contraction against the symmetric tensor $g^{\mu\lambda}$:

$${\Gamma_{\alpha\nu}}^\alpha = g^{\mu\lambda}\Gamma_{\mu\nu\lambda} = \frac{1}{2}g^{\mu\lambda}\partial_\nu g_{\mu\lambda} = \frac{\partial_\nu \sqrt{|g|}}{\sqrt{|g|}}, \tag{5.5.7}$$

where we used Lemma 5.5.2 in the last equality. □

Lemma 5.5.4. *Let ∇_μ be the covariant derivative with respect to the Levi-Civita connection corresponding to $g_{\mu\nu}$. Then*

$$\boxed{\nabla_\mu \sqrt{|g|} = 0.} \tag{5.5.8}$$

Proof. Since $\sqrt{|g|}$ is a scalar density,

$$\nabla_\mu \sqrt{|g|} = \partial_\mu \sqrt{|g|} - {\Gamma_{\alpha\mu}}^\alpha \sqrt{|g|}$$

which is zero by Lemma 5.5.3. □

Lemma 5.5.5. *Let ∇_μ be the covariant derivative with respect to the Levi-Civita connection corresponding to $g_{\mu\nu}$. Then, for any scalar field ϕ, we have*

$$\boxed{\nabla_\mu(\phi\sqrt{|g|}) = (\partial_\mu\phi)\sqrt{|g|}.} \tag{5.5.9}$$

Proof. By the product rule for the covariant derivative,

$$\nabla_\mu(\phi\sqrt{|g|}) = (\nabla_\mu\phi)\sqrt{|g|} + \phi\cdot\nabla_\mu\sqrt{|g|}$$

Since ϕ is a scalar field, $\nabla_\mu\phi = \partial_\mu\phi$, and the other term is zero by Lemma 5.5.4. □

5.6 Homework #5

1. Show that when $n = 3$, the curl of a covector field A_j can be written $\varepsilon^{ijk}\nabla_i A_j$. What is the type and weight of this relative tensor? If you had to, how would you define the "curl" of a covector field A_μ in n dimensions?

2. Suppose we are given an affine connection $\Gamma_{\mu\nu}{}^\lambda$. Let A_μ be a relative covector of weight w. Show that the formula (5.2.2) for the covariant derivative of A_μ does indeed give a relative $\binom{0}{2}$-tensor of weight w.

3. Let $\Gamma_{\mu\nu}{}^\lambda$ be an affine connection and ∇_μ the corresponding covariant derivative. Show that if ϕ is a relative scalar of weight w and A_μ is a relative covector field of weight v, then the following product rule holds:

$$\nabla_\mu(\phi A_\nu) = (\nabla_\mu\phi)A_\nu + \phi\nabla_\mu A_\nu.$$

4. Prove the claim that $\varepsilon_{\mu_1\cdots\mu_n}$ is a relative tensor type $\binom{0}{n}$ and weight -1.

5. (Raised and lowered Levi-Civitas) Suppose a metric $g_{\mu\nu}$ has been chosen. Define the *raised Levi-Civita symbol* $\varepsilon_g^{\mu_1\cdots\mu_n}$ as the relative tensor obtained by raising all the indices in $\varepsilon_{\mu_1\cdots\mu_n}$ using the metric. Show that this gives a relative $\binom{n}{0}$-tensor of weight 1 and that

$$\varepsilon_g^{\mu_1\cdots\mu_n} = g^{-1}\varepsilon^{\mu_1\cdots\mu_n}$$

   Similarly, the *lowered Levi-Civita symbol* $\varepsilon^g_{\mu_1\cdots\mu_n}$ is obtained by lowering all indices in the contravariant Levi-Civita symbol and $\varepsilon^g_{\mu_1\cdots\mu_n} = g\varepsilon_{\mu_1\cdots\mu_n}$. [Warning: Some authors use $\varepsilon^{\mu_1\cdots\mu_n}$ to denote the raised Levi-Civita symbol.]

6. Let η_{ab} be the flat metric (or, if you want, any constant metric, regarded as a Lorentz $\binom{0}{2}$-tensor), $g_{\mu\nu}$ a general metric, and e^a_μ be a Vielbein field connecting the two as in (4.2.3). Let $\eta = \det(\eta_{ab})$, $g = \det(g_{\mu\nu})$, and $e = \det(e^a_\mu)$ be the respective determinants. Show that

$$g = \eta e^2 \tag{5.6.1}$$

In particular, for $n = 4$ and Lorentzian signature, conclude that

$$e = \sqrt{-g}. \tag{5.6.2}$$

Part II

Lagrangian Field Theories

Introduction to Action Principles

Suppose we have a falling rock, its height above ground described by $y = y(t)$. What is the position $y(t)$ at time t? To find out, we must solve the *equation of motion*. In this case the equation is $\ddot{y} = -g$, where g is Earth's constant of gravitational acceleration. But *why* is this the equation of motion? Where does it come from? Is it just a law of nature? Or, is there some deeper principle at work here, that can be used to derive equations of motions for a whole range of physical systems?

Let us consider the *action*, S, (=work=energy×time) performed by the falling rock along its trajectory. At time t, it has velocity $v(t) = \dot{y}(t)$, and therefore has a kinetic energy of $v(t)^2/2$. On the other hand, the gravitational field grants the rock a potential energy of $gy(t)$. This contributes negatively to the work done by the rock, because it represents energy *given* to it. Thus the action-per-time done by the rock is

$$L(t) = \frac{1}{2}\dot{y}^2 - gy.$$

Therefore, the action is the time integral of this:

$$S = \int L(t)dt$$

(from $t = 0$ to $t = 1$, say). For each possible trajectory $y(t)$, we can calculate the action S. Thus S is a *functional*[3] of $y(t)$. We denote this by square brackets:

$$S = S[y(t)] \quad \text{or just} \quad S[y].$$

If we vary the trajectory $y(t)$, the value of the functional will change. Let us consider all possible (smooth) trajectories with given boundary conditions: $y(0) = h$ and $y(1) = 0$, for example. We observe:

- If the rock spends a long time at high altitude, then once t gets closer to 1, it has to eventually move quickly to reach the earth before the time is up. In other words, it has to spend a lot of kinetic energy, driving up S.

[3]It is worth noting that S is a nonlinear functional, because of the square.

- On the other hand, if the rock descends too early, it will miss out on utilizing the negative contribution of the potential energy, again making S bigger than it has to be.

So, for some trajectory among these extremes (and other even stranger that would travel upwards for some interval of time) there might be a trajectory $y(t)$ with the *least action*. For such trajectories, any small variation $\delta y(t)$ of the trajectory gives a larger action, just like a local minimum in calculus. And similarly to calculus the action must be stationary at those trajectories. We now define what that means. The *variation* of S is

$$\delta S = S[y + \delta y] - S[y] \tag{5.6.3}$$

where $\delta y = \delta y(t)$ is a smooth function vanishing at the endpoints $t = 0, 1$. We say that S is *stationary* at the trajectory $y(t)$ if $\delta S = 0$ for any δy. Equivalently, if

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} S[y(t) + \varepsilon z(t)] = 0 \tag{5.6.4}$$

for any $z(t)$ vanishing at $t = 0, 1$. Substituting $L = L(y, \dot{y})$ into this, we get

$$0 = \int_0^1 \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} L(y + \varepsilon z, \dot{y} + \varepsilon \dot{z})\, dt = \int_0^1 \Big(\frac{\partial L}{\partial y} z + \frac{\partial L}{\partial \dot{y}} \dot{z}\Big)\, dt$$

where we regard L as a polynomial in two indeterminates y and $\dot{y}$. Integrating by parts in the second term we get

$$0 = \int_0^1 \Big(\frac{\partial L}{\partial y} z - \frac{d}{dt}\frac{\partial L}{\partial \dot{y}} z\Big)\, dt + \Big[\frac{\partial L}{\partial \dot{y}} z\Big]_0^1$$

Since z vanishes at $t = 0, 1$, the last term is zero. Furthermore, since the integral has to vanish for all such z, we conclude that $y(t)$ is a stationary trajectory for S if and only if $y(t)$ satisfies the *Euler-Lagrange equation*

$$-\frac{\partial L}{\partial y} + \frac{d}{dt}\frac{\partial L}{\partial \dot{y}} = 0. \tag{5.6.5}$$

If we substitute $L(y, \dot{y}) = \frac{1}{2}\dot{y}^2 - gy$ into this we *deduce* the equation of motion $gy + \ddot{y} = 0$.

It turns out, the equations of motions for *all* fundamental theories (accepted ones like Einstein gravity and the standard model of particle physics, as well as theoretical models such as supergravity, string theory, etc.) arise this way. The starting point for

all these theories is a *lagrangian density* $\mathscr{L}$, which is a smooth function (typically a polynomial, possibly involving a square root, of a number of variables into which one is to substitute the field components and their derivatives. We say $\mathscr{L}$ is of k:th order if k is the highest order derivative occuring in $\mathscr{L}$. The lagrangian density is **highly constrained**. First, it has to be such that (after substituting the field components) it defines a scalar density, so that the action integral is coordinate-independent. Second, it has to have the right dimensionality: in natural units it has mass dimension n. For any terms with the wrong mass dimension, a dimensionfull (coupling) constant has to be introduced (and we want to have as few of those as possible, as their value must be obtained by fitting data). Third, if $\mathscr{L}$ is a k:th order lagrangian density, then the equations of motion typically have order $2k$. All equations of motion in known physics are second order PDEs, so usually $k = 1$. Fourth, in quantum field theory, the exact same action for a classical theory is used in the corresponding quantum theory. All we have to do is replace the variational problem by a rule in which each field configuration φ receives a probability amplitude of $\exp(iS[\varphi])$. In order to extract numbers from the quantum theory which can be compared with experiments, the theory must be so-called *renormalizable*. This imposes further restrictions on the lagrangian. In particular, no coupling constants can have negative mass dimension. [4] Lastly, some fields have additional transformation properties called (local and global) *gauge transformations*. It is deemed that such transformations leave the physics unchanged, similarly to coordinate transformations. This puts further restrictions on the lagrangian to be gauge invariant.

The fact that the lagrangian is so highly constrained means that rather than already having the equations of motion and an understanding of what kinetic energy is and so on, we can flip things around and start with a set of field components, subject to transformations properties (tensorial, and/or gauge transformations) and write down the most general lagrangian satisfying all the constraints. After we have the lagrangian, we can derive everything from it including the equations of motion (as the Euler-Lagrange equations of the action), conserved quantities (energy, momentum, charge, etc.) and start producing predictions to be compared with observations.

Two final remarks. Model building with lagrangians has a modular character. There are standard kinetic and mass terms for all basic field types. To combine two field theories simply add up the lagrangians. To make the two theories interact, an

[4] However, there are important so-called effective field theories that are not renormalizable, but nevertheless extremely valuable. General relativity is a non-renormalizable field theory, yet is spectacularly successful in its classical predictions. Many believe that general relativity is an effective field theory of some renormalizable quantum field theory.

interaction lagrangian is added which should involve products of field components from each theory. In local gauge theory, as we shall see, such interaction terms arise naturally.

Lastly, different lagrangians can give rise to equivalent theories. Ultimately, it is the equations of motion of the field components that determine the theory. The simplest way in which this can happen is that the two lagrangians differ by a total divergence which integrates to zero in the action (assuming vanishing boundary conditions). But, there are also equivalent lagrangians defined in completely different spacetime dimensions, such as in Kaluza-Klein theories.

Lecture 6

Action Principles

6.1 Natural Units

For a physical quantity q, we let $[q]$ denote the dimensionality (e.g., $\mathrm{kg}\cdot\mathrm{m}\cdot\mathrm{s}^{-2}$). In a natural system of units, $\hbar = c = 1$. This means that the dimensions of mass, energy, length, time are related by

$$M = E = L^{-1} = T^{-1}.$$

(due to $E = mc^2$, $[c] = L/T$, $[\hbar] = ET$). Thus every dimensionfull quantity can be expressed in terms of mass, and we write $[q] = M^d$ or simply

$$[q] = d \tag{6.1.1}$$

where $d \in \mathbb{Z}$ (or, sometimes, $d \in \frac{1}{2}\mathbb{Z}$). We call d the *mass dimension* of q.[1] For example, dx = a small length displacement so $[dx] = L = M^{-1}$ and thus $[d^n x] = M^{-n}$. Therefore, if $S = \int d^n x \mathscr{L}$ and $[S] = M^0$, then we want $[\mathscr{L}] = M^n$. Another useful fact is that $[\partial_\mu \varphi] = 1 + [\varphi]$, because we divide by a length when computing derivatives.

Remark 6.1.1. The introduction of a mass dimension can be implemented algebraically as follows. If $\mathscr{T}$ denotes the commutative algebra of all relative tensor fields (of all types and all weights), we form the Laurent polynomial algebra

$$\mathscr{T}[M^{1/2}, M^{-1/2}].$$

A field with a definite mass dimension $d \in \frac{1}{2}\mathbb{Z}$ is then an element of $\mathscr{T}M^d$.

[1] The mass dimension is also called *operator dimension* or *canonical dimension*.

6.2 Lagrangian Densities

Definition 6.2.1. Let k be a positive integer.

(i) A *k:th order Lagrangian density for a scalar field φ* is a smooth function

$$\mathscr{L} : \mathbb{R}^m \to \mathbb{R} \tag{6.2.1}$$

(for appropriate m) such that, substituting the partial derivatives for φ up to order k into $\mathscr{L}$

$$\mathscr{L}(\varphi,\, \partial_\mu \varphi,\, \partial_\mu \partial_\nu \varphi,\, \ldots) \tag{6.2.2}$$

yields a scalar density of mass dimension n.

(ii) More generally, a k:th order Lagrangian density for a tuple $\varphi^{(a)} = \left(\varphi^{(a)}\right)_{a=1}^{N}$ of N field components (each component a real-valued function, given in each coordinate system, and typically transforming according to some specification, and having some mass dimension d_a) is a smooth function

$$\mathscr{L} : \mathbb{R}^m \to \mathbb{R} \tag{6.2.3}$$

(for appropriate m) such that, subtituting partial derivatives up to order k,

$$\mathscr{L}(\varphi^{(a)},\, \partial_\mu \varphi^{(a)},\, \partial_\mu \partial_\nu \varphi^{(a)},\, \ldots) \tag{6.2.4}$$

yields a scalar density of mass dimension n.

Definition 6.2.2. The *field content* of a theory is the specification of a set of (possibly relative) tensor field types and their mass dimensions, along with designating each field as either *dynamic* or a *background* field.

Definition 6.2.3. A *field theory* $F = (\mathscr{G}, \varphi^{(a)}, \mathscr{L})$ is a triple, where

1. $\mathscr{G}$ is a geometry,
2. $\varphi^{(a)}$ is the field content,
3. $\mathscr{L}$ is a Lagrangian density for that field content.

We say F is a k:th order theory if $\mathscr{L}$ has order k. We say F is *Galilean* (respectively *Lorentzian*, *generally covariant*) if $\mathscr{G} = \mathscr{G}_{\text{Gal}}$ (respectively $\mathscr{G}_{\text{Lor}}$, $\mathscr{G}_{\text{sm}}$).

6.3 Examples

6.3.1 Theories in Flat 3+1 Dimensions

In all these, $n = 4$ and $\mathscr{G} = \mathscr{G}_{\text{Lor}}$.

1. Dynamic scalar field φ of mass dimension 1 with lagrangian

$$\mathscr{L} = \frac{1}{2}\partial^\mu\varphi\partial_\mu\varphi - \frac{1}{2}m^2\varphi^2 \tag{6.3.1}$$

   where m is the *mass*, $[m] = 1$. The $1/2$ are arbitrary normalization constants, but the signs in front of the terms are determined by the requirement that the energy should be non-negative (we will discuss the energy-momentum tensor in a future chapter).

2. Dirac lagrangian

$$\mathscr{L} = \overline{\psi}(i\boldsymbol{\gamma}^\mu\partial_\mu - m)\psi \tag{6.3.2}$$

   Here $\boldsymbol{\gamma}^\mu$ are four complex 4×4-matrices satifying[2] $\{\boldsymbol{\gamma}^\mu, \boldsymbol{\gamma}^\nu\} = 2\eta^{\mu\nu}$, and ψ is a complex column vector with 4 entries, and $\overline{\psi} = \psi^\dagger\boldsymbol{\gamma}^0$. Thus, ψ is really 8 dynamic real field components bundled together. All leptons and quarks in the standard model are described using this lagrangian. We will study it in detail in a future lecture. Note that $[\psi] = 3/2$ in order for $[\mathscr{L}] = 4$ to hold (assuming $[\boldsymbol{\gamma}^\mu] = 0$).

3. Maxwell/Proca lagrangian for a dynamic (co)vector field A_μ of mass dimension 1, and a background vector field J^μ of mass dimension 2:

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + m^2A_\mu A^\mu + qJ^\mu A_\mu \tag{6.3.3}$$

   where m is the *mass* of A_μ, $[m] = 1$, and q is the *coupling constant* to the background field, $[q] = 1$. When $m = 0$ we have the Maxwell lagrangian describing the photon field. For $m > 0$, the lagrangian describes a massive vector boson. There are three such fields in the standard model, the $W^\pm$ and Z^0 bosons, which are the carriers of the weak interaction. If $q = 0$, we say it is a field theory *in vacuo*.

[2] $\{a, b\} = ab + ba$ denotes the *anti-commutator* of a and b.

6.3.2 Theories in Curved Spacetime

Here $\mathscr{G} = \mathscr{G}_{\text{sm}}$, and we usually have a background metric field $g_{\mu\nu}$, and one should replace ∂_μ by the covariant derivative ∇_μ (with respect to the Levi-Civita connection) and multiply by $\sqrt{|g|}$ at the end to ensure the lagrangian is a scalar density.

For example, the standard lagrangian for a scalar field theory in curved spacetime would be

$$\mathscr{L} = \big(\frac{1}{2}\nabla^\mu\varphi\nabla_\mu\varphi - \frac{1}{2}m^2\varphi^2\big)\sqrt{-g} \tag{6.3.4}$$

where, of course, $\nabla^\mu\varphi = g^{\mu\nu}\nabla_\nu\varphi$. Note that for scalar fields, $\partial_\mu\varphi = \nabla_\mu\varphi$.

Similarly, Maxwell's theory of electromagnetism (in vacuo) for (3+1)-dimensional curved spacetime reads

$$\mathscr{L}_{\text{Maxwell}} = -F_{\mu\nu}F^{\mu\nu}\sqrt{-g} \tag{6.3.5}$$

where $F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu$ (which again happens to be $\partial_\mu A_\nu - \partial_\nu A_\mu$ due to vanishing of torsion in the Levi-Civita connection).

In Einstein's gravity, the graviational field is identified with the metric field itself which is now dynamic, and

$$\mathscr{L}_{\text{EH}} = (\alpha_1 R + \alpha_0)\sqrt{-g} \tag{6.3.6}$$

for certain constants α_0, α_1. [3] If $\mathscr{L}_{\text{n.g.}}$ is another lagrangian governing non-gravitational fields (in this context called matter fields), then

$$\mathscr{L} = \mathscr{L}_{\text{EH}} + \mathscr{L}_{\text{n.g.}}$$

turns the curved matter theory into a gravitational theory. Through the raising and lowering of indices necessary to make a scalar density, the coupling to gravity is already implicitly present in the matter lagrangian. For example, with $\mathscr{L} = \mathscr{L}_{\text{EH}} + \mathscr{L}_{\text{Maxwell}}$ one obtains a generally covariant theory of gravitation and electromagnetism, which includes the bending of light near a large mass. This was one of the first predictions of Einstein's theory of general relativity to be observationally confirmed.

6.4 Equations of Motion: Euler-Lagrange

If $(\mathscr{G}, \varphi^{(a)}, \mathscr{L})$ is a field theory, the *action* is defined as

$$S = S[\varphi^{(a)}] = \int \mathscr{L}\, d^n x. \tag{6.4.1}$$

[3] The constant α_1 is determined using the Newtonian approximation, and is related to Newton's constant of gravity, while α_0 is extremely close to zero, and related to the cosmological constant.

The region of integration is any small region in which we wish to study the field theory. In any coordinate system it can be taken to be a closed ball of some radius. S is a functional of the field components. This means that S takes as input a set of fields and outputs a real number.

The *principle of stationary action* is that the laws of nature are such that the fields satisfying those laws minimize the action.

For each field component $\varphi^{(a)}$ that the lagrangian depends on, and is designated dynamic in the theory, there is an associated equation of motion, obtained through the variational principle that the action be stationary with respect to varying that field component.

For a first order lagrangian we have

$$\delta S = S[\varphi^{(a)}+\delta\varphi^{(a)}]-S[\varphi^{(a)}] = \int \Big(\mathscr{L}(\varphi^{(a)}+\delta\varphi^{(a)}, \partial_\mu\varphi^{(a)}+\partial_\mu\delta\varphi^{(a)})-\mathscr{L}(\varphi^{(a)}, \partial_\mu\varphi^{(a)})\Big)\, d^n x$$

Taylor expanding $\mathscr{L}$, discarding[4] terms quadratic in the variations $\delta\varphi^{(a)}$, we get

$$\delta S = \int \Big(\frac{\partial\mathscr{L}}{\partial\varphi^{(a)}}\delta\varphi^{(a)} + \frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^{(a)})}\delta\partial_\mu\varphi^{(a)}\Big)\, d^n x$$

Since ∂_μ is linear, $\delta\partial_\mu\varphi^{(a)} = \partial_\mu\delta\varphi^{(a)}$. Integrating by parts in the second term we get

$$\delta S = \int \Big(\frac{\partial\mathscr{L}}{\partial\varphi^{(a)}} - \partial_\mu\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^{(a)})}\Big)\delta\varphi^{(a)}\, d^n x$$

since $\delta\varphi^{(a)}$ are assumed to vanish on the boundary. Thus, for each dynamic field component $\varphi^{(a)}$, its equation of motion is the Euler-Lagrange equation

$$-\frac{\partial\mathscr{L}}{\partial\varphi^{(a)}} + \partial_\mu\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^{(a)})} = 0. \tag{6.4.2}$$

Sometimes the fields are such that "motion" does not make sense (eg. if no time coordinate is present). A more general term is to simply call (6.4.2) the *field equations* for the field component $\varphi^{(a)}$. Fields designated *dynamic* are those whose components are subject to the field equations. On the other hand, fields considered part of the *background* are not subject the field equations. One reason is that we may not know, or do not care about, their precise nature, and therefore leave out terms from the lagrangian that would describe them fully.

[4] Readers uncomfortable with this may instead define $\frac{\delta S}{\delta\varphi^{(a)}} = \frac{d}{d\varepsilon}\big|_{\varepsilon=0} S[\varphi^{(b)} + \epsilon\delta^b_a\psi^{(a)}]$ where $\psi^{(a)}$ is an arbitrary smooth function vanishing on the integration boundary. Setting this to zero, and moving $\frac{d}{d\varepsilon}\big|_{\varepsilon=0}$ inside the integral, leads to the same result (6.4.2).

6.5 Examples

6.5.1 Scalar field

The Euler-Lagrange equation for φ in the scalar field lagrangian

$$\mathscr{L}_{\text{scalar}} = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - \frac{1}{2}m^2\varphi^2$$

is computed as follows. First, $\partial\mathscr{L}/\partial\varphi = -m^2\varphi$. This is because we should regard φ and the n partial derivatives $\partial_\mu\varphi$ as independent variables. Second,

$$\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi)} = \frac{\partial}{\partial(\partial_\mu\varphi)}\left(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\varphi\partial_\beta\varphi\right) = \frac{1}{2}g^{\alpha\beta}\left(\delta^\mu_\alpha\partial_\beta\varphi + \partial_\alpha\varphi\delta^\mu_\beta\right) = \partial^\mu\varphi.$$

Here we used the product rule, and that $\frac{\partial}{\partial(\partial_\mu\varphi)}(\partial_\nu\varphi) = \delta^\mu_\nu$. Putting it all together, the Euler-Lagrange equation for $\mathscr{L}_{\text{scalar}}$ reads

$$m^2\varphi + \partial_\mu\partial^\mu\varphi = 0. \tag{6.5.1}$$

The differential operator $\partial_\mu\partial^\mu$ is the d'Alembertian $\Box$, so this differential equation is often written

$$(\Box + m^2)\varphi = 0. \tag{6.5.2}$$

This is the *Klein-Gordon equation* for a free scalar field of mass m. [5]

6.5.2 Scalar Field in One Dimension

Taking $n = 1$ with metric $\eta_{00} = 1$ (i.e. one time and zero space dimensions) in the previous example, and putting $y = \varphi$ and $\omega = m$, we get the lagrangian

$$\mathscr{L} = \frac{1}{2}\dot{y}^2 - \frac{1}{2}\omega^2 y^2$$

This is known as the harmonic oscillator. Its Euler-Lagrange equation for y is

$$\ddot{y} = -\omega^2 y$$

and thus has the general solution

$$y = A\cos(\omega t) + B\sin(\omega t)$$

describing frequency ω oscillations from an equilibrium.

[5] In quantum field theory, quanta in this field would be particles of mass m. There are no elementary particles in the standard model of exactly this kind. However, at lower energies, there are effective field theories involving pi mesons (pions) that are quanta in such fields.

6.6 Homework #6

1. Fix a metrix $g_{\mu\nu}$. A curve $t \mapsto x^\mu(t)$ is a *geodesic* if $\nabla_\mu \dot{x}^\nu = 0$.[6] Show that this so-called *geodesic equation* can be obtained as the Euler-Lagrange equation for the lagrangian

$$\mathscr{L} = \mathscr{L}(x^\mu, \dot{x}^\mu) = \left(\dot{x}^\mu \dot{x}^\nu g_{\mu\nu}(x^\lambda)\right)^{1/2}.$$

   (Note: Here, $x^\mu = x^\mu(t)$ play the role of the field components in our theory. Thus there should be n Euler-Lagrange equations.) (Also note: The "action" in this case is nothing but the *length of the curve*!)

2. Show that the Newton-Laplace equation for the gravitational scalar potential ϕ from (1.2.6) can be obtained as the Euler-Lagrange equation for the Galilean lagrangian $\mathscr{L} = \mathscr{L}(\phi, \partial_i\phi) = -\frac{1}{2}\delta^{ij}\partial_i\phi\partial_j\phi - 4\pi G\rho\phi$.

3. Show that the mass dimension of the scalar curvature R is M^2. That is, $[R] = 2$. Conclude that $[\alpha_1] = 2$ and $[\alpha_0] = 4$. (See (6.3.6).)

4. (Alternative system of mass dimension) In general relativity, the coordinates have no physical meaning, and therefore it makes sense to declare them dimensionless. Instead, one assigns mass dimension of -2 (equivalently, length squared) to the metric field components $g_{\mu\nu}$. That is, $[g_{\mu\nu}] = -2$, and thus $[g^{\mu\nu}] = 2$. In this system, partial derivatives do not change mass dimension, and instead raising an index increases mass dimension by 2. Using this alternative system, find the mass dimensions of the various lagrangian terms found in chapter. (Be sure to include $\sqrt{|g|}$, and even in flat spacetime regard $[\eta_{\mu\nu}] = -2$.)

5. (On-shell vs. off-shell) Plug in $\varphi(x^\mu) = \exp(ik_\mu x^\mu)$ (where k_μ is constant) into the Klein-Gordon equation $\partial_\mu\partial^\mu\varphi + m^2\varphi = 0$ to deduce a necessary and sufficient condition on k_μ to have a solution. When $n = 2$ and $\eta_{\mu\nu} = \mathrm{diag}(1, -1)$, sketch all solutions (k_0, k_1) for fixed positive m. Do the same for $n = 3, \eta_{\mu\nu} = \mathrm{diag}(1, -1, -1)$. This set is called the *mass-shell*. Quite generally, we say that something is true *on-shell*, if it holds when the equations of motions are imposed. If a statement is true even without imposing the equations of motion, we say it is true *off-shell*.

[6] The covariant derivative should of course be computed using the Levi-Civita connection of the metric.

6. Assume flat spacetime. Show that if $\mathscr{L}(\varphi, \partial_\mu\varphi, \partial_\mu\partial_\nu\varphi)$ is a second order lagrangian density for a single scalar field φ, then $\delta S = 0$, with both variations $\delta\varphi$ and $\delta\partial_\mu\varphi$ vanishing on the integration boundary, is equivalent to the *second order Euler-Lagrange equation*

$$-\frac{\partial\mathscr{L}}{\partial\varphi} + \partial_\mu\frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi)} - \partial_\mu\partial_\nu\frac{\partial\mathscr{L}}{\partial(\partial_\mu\partial_\nu\varphi)} = 0. \tag{6.6.1}$$

   Generalize to k:th order Lagrangians.

7. Generalize the previous problem to curved spacetime with fixed background metric $g_{\mu\nu}$, requiring vanishing of $\delta\varphi, \delta\nabla_\mu\varphi, \ldots$ up to order $k-1$ on the integration boundary.

Lecture 7

Conserved Quantities

7.1 Global Symmetries and Noether's Theorem

It turns out that the lagrangian contains more information than just the equations of motion. Conserved quantities of physical significance such as energy, momentum, charge, etc., all arise as a consequence of certain symmetries of the lagrangian. This is a profound realization, expressed in Noether's First Theorem.

Definition 7.1.1. A *current* is a relative vector field of weight 1.

We have noted earlier that currents satisfy $\nabla_\mu j^\mu = \partial_\mu j^\mu$ for any covariant derivative, by definition of covariant derivative of relative tensor fields. In particular, $\partial_\mu j^\mu = 0$ is a coordinate-independent property.

Definition 7.1.2. A current j^μ is *conserved* if $\partial_\mu j^\mu = 0$.

Recall from the previous homework set that *on-shell* means "assuming the equations of motions for the dynamic field components".

Theorem 7.1.3 (Noether's First Theorem (fields-to-fields case))**.** *To each one-parameter group of symmetries of a lagrangian density $\mathscr{L}$ there corresponds a conserved current. More precisely, if*

$$\varphi^{(a)} \mapsto \overline{\varphi}^{(a)} = \varphi^{(a)} + \zeta^{(a)}(\varphi^{(b)})\varepsilon + \mathcal{O}(\varepsilon^2) \tag{7.1.1}$$

is a transformation among dynamic field components that leaves a first order lagrangian density $\mathscr{L}(\varphi^{(a)}, \partial_\mu \varphi^{(a)})$ unchanged, then the quantity

$$j^\mu = \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi^{(a)})}\zeta^{(a)} \tag{7.1.2}$$

is a current which is conserved on-shell. That is, if the field components $\varphi^{(a)}$ each satisfy the Euler-Lagrange equation, then $\partial_\mu j^\mu = 0$.

Proof. Put $\overline{\mathscr{L}} = \mathscr{L}(\overline{\varphi}^{(a)}, \partial_\mu \overline{\varphi^{(a)}})$. By assumption, $\overline{\mathscr{L}} = \mathscr{L}$ and thus

$$0 = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \overline{\mathscr{L}} = \frac{\partial \mathscr{L}}{\partial \varphi^{(a)}} \zeta^{(a)} + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi^{(a)})} \partial_\mu \zeta^{(a)}.$$

Using the Euler-Lagrange equation for $\varphi^{(a)}$ in the first term we get

$$\partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi^{(a)})} \zeta^{(a)} + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi^{(a)})} \partial_\mu \zeta^{(a)} = 0.$$

Using the product rule this can be written

$$\partial_\mu j^\mu = 0$$

where j^μ is given by (7.1.2). That j^μ is a current follows from the fact that $\mathscr{L}$ is a scalar density. □

Definition 7.1.4. The conserved current j^μ is called the *Noether current* corresponding to the symmetry.

7.2 Charge Conservation

In this section we explain how vanishing of divergence is a conservation law.

Suppose that we are in flat spacetime with metric $\eta_{\mu\nu} = \mathrm{diag}(1, -1, -1, \ldots, -1)$ of signature $(1, n-1)$, and let j^μ be a current. The time-component, j^0, is called the *charge density* of the current, and its space integral

$$Q = Q(x^0) = \int_\Omega j^0 \, dx^1 dx^2 \cdots d^{n-1} \tag{7.2.1}$$

is called the *charge*. This is a function of time (and it depends on the space region Ω).

Now suppose that j^μ is conserved, that is, its divergence vanishes. Due to the form of the metric, this can be written

$$\partial_0 j^0 = \partial_a j^a$$

where a runs over the space indices $1, 2, \ldots, n-1$. Suppose furthermore that the space components j^a vanish far away, i.e. on the boundary of Ω. Taking the time-derivative on both sides of (7.2.1) we get

$$\frac{d}{dt}Q = \int_\Omega \partial_0 j^0 \, d^{n-1}x = \int_\Omega \partial_a j^a \, d^{n-1}x = 0$$

by the divergence theorem in $(n-1)$-dimensional space. In other words, Q is a quantity that is constant in time.

Definition 7.2.1. Q is called the *Noether charge* corresponding to the symmetry.

Note that the charge is coordinate-*dependent*. The Lorentzian quantity is the Noether current j^μ, and extracting the zero:th component j^0 from it violates Lorentz invariance. However, for each observer, the charge is a real quantity that is measured to be constant over time. It is furthermore invariant under rotations and translations in space (but not under Lorentz boosts). Thus, observers moving relative to each other at constant velocity will measure different values of Q.

Another point here is that Q depends on the region Ω. Suppose that Ω is partitioned up into two regions Ω_1 and Ω_2, and that we form $Q_i = \int_{\Omega_i} j^0 \, d^{n-1}x$. Then $Q_1 + Q_2$ is conserved (i.e. constant over time) but in general Q_1 and Q_2 are not (unless j^a vanishes on $\partial\Omega_i$). This gives a much more dynamic interpretation of the charge Q; it can flow from one region to another, but the overall sum of all charge remains constant in Ω.

7.3 Example: A Scalar Doublet

In flat spacetime, consider two free scalar fields ϕ^1 and ϕ^2 with the same mass m. The lagrangian density is

$$\mathscr{L} = \frac{1}{2}\partial_\mu\varphi^1\partial^\mu\varphi^1 + \frac{1}{2}\partial_\mu\varphi^2\partial^\mu\varphi^2 - \frac{1}{2}m^2(\varphi^1\varphi^1 + \varphi^2\varphi^2). \tag{7.3.1}$$

By introducing the scalar doublet

$$\Phi = \begin{pmatrix}\varphi^1 \\ \varphi^2\end{pmatrix} \tag{7.3.2}$$

the lagrangian can be more succinctly expressed as

$$\mathscr{L} = \frac{1}{2}(\partial_\mu\Phi)^{\mathrm{T}}\partial^\mu\Phi - \frac{1}{2}m^2\Phi^{\mathrm{T}}\Phi, \tag{7.3.3}$$

where it is understood that ∂_μ acts on both components of Φ. Let

$$R = R(\varepsilon) = \begin{pmatrix} \cos\varepsilon & -\sin\varepsilon \\ \sin\varepsilon & \cos\varepsilon \end{pmatrix} = I_2 + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \varepsilon + \mathcal{O}(\varepsilon^2) \tag{7.3.4}$$

be counter-clockwise rotation by ε radians. This is an orthogonal 2×2-matrix of determinant 1. Those form a group denoted $\mathrm{SO}(2)$ called the *special orthogonal group*. Topologically this group just the unit circle. In particular, $R(\varepsilon + 2\pi) = R(\varepsilon)$. That R is orthogonal means $R^{\mathrm{T}}R = I_2$, where I_2 is the identity matrix. Consequently, substituting $\Phi \mapsto \bar{\Phi} = R\Phi$ into $\mathscr{L}$ and using that ε does not depend on x^μ, we get[1]

$$\frac{1}{2}(R\partial_\mu\Phi)^{\mathrm{T}}R\partial^\mu\Phi - \frac{1}{2}m^2(R\Phi)^{\mathrm{T}}(R\Phi) = \frac{1}{2}(\partial_\mu\Phi)^{\mathrm{T}}R^{\mathrm{T}}R\partial^\mu\Phi - \frac{1}{2}m^2\Phi^{\mathrm{T}}(R^{\mathrm{T}}R)\Phi = \mathscr{L}$$

We say that $\mathscr{L}$ has a *global* $\mathrm{SO}(2)$*-symmetry*. Global, because R does not depend on the spacetime coordinate. The same transformation is simultaneously applied all throughout space and time.

In the notation of Noether's First Theorem (though dropping parenthesis around superscripts), we have by (7.3.4),

$$\zeta^1(\varphi^1, \varphi^2) = -\varphi^2, \qquad \zeta^2(\varphi^1, \varphi^2) = \varphi_1. \tag{7.3.5}$$

Consequently, the Noether current corresponding to this $\mathrm{SO}(2)$-symmetry is[2]

$$j^\mu = \frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^1)}(-\varphi^2) + \frac{\partial\mathscr{L}}{\partial(\partial_\mu\varphi^2)}\varphi^1. \tag{7.3.6}$$

Substituting our particular Lagrangian (7.3.1), we obtain

$$j^\mu = \varphi^1\partial^\mu\varphi^2 - \varphi^2\partial^\mu\varphi^1. \tag{7.3.7}$$

In this form, one can verify directly that $\partial_\mu j^\mu = 0$ provided the fields satisfy their respective (Klein-Gordon) equation of motion.

7.4 Homework #7

1. In dimension $n = 1$ (one time, no space), consider the following lagrangian:

$$L = L(x, \dot{x}, y, \dot{y}) = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\dot{y}^2 - \frac{1}{2}\omega^2(x^2 + y^2), \tag{7.4.1}$$

where $\omega > 0$ is a constant.

[1] using $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$

[2] Note that this expression is a general formula for the Noether current of *any* $\mathrm{SO}(2)$-invariant lagrangian mixing two scalar fields φ^1 and φ^2.

(a) Find the Euler-Lagrange equations obtained by varying x and y and give the general solution of each equation as a linear combination of sine and cosine functions.

(b) Show that L remains unchanged upon the substitution $x \mapsto x\cos\varepsilon - y\sin\varepsilon$, $y \mapsto x\cos\varepsilon + y\sin\varepsilon$.

(c) Compute the corresponding Noether current (which is equal to the charge in this case, since $n = 1$). Substitute an arbitrary solution $(x(t), y(t))$ to the equations of motion and verify that the Noether current/charge is constant.

2. $U(1)$-symmetry: Let $\phi = \frac{1}{\sqrt{2}}(\varphi_1 + i\varphi_2)$ be a complex field with lagrangian density

$$\mathscr{L} = (\partial^\mu \phi)^* \cdot \partial_\mu \phi - m^2 \varphi^* \cdot \varphi$$

where $*$ means complex conjugate and $m > 0$. Show that this lagrangian is invariant under the global symmetry $\phi \mapsto \exp(i\varepsilon)\phi$ for any real ε. What is the corresponding Noether current and charge?

3. SU(2) scalar doublet: Consider a theory with two complex fields $\phi^1 = \frac{1}{\sqrt{2}}(\varphi^1 + i\varphi^2)$ and $\phi^2 = \frac{1}{\sqrt{2}}(\varphi^3 + i\varphi^4)$ with Lagrangian density (assume flat spacetime)

$$\mathscr{L} = (\partial^\mu \Phi)^\dagger \cdot (\partial_\mu \Phi) - m^2 \Phi^\dagger \cdot \Phi$$

where $\Phi = \begin{pmatrix} \phi^1 \\ \phi^2 \end{pmatrix}$ and $\dagger$ means transpose and complex conjugate.

(a) Let SU(2) be the *special unitary group* consisting of matrices U that are unitary ($U^\dagger \cdot U = I_2$) and have determinant one. Show that $\mathscr{L}$ is invariant under global SU(2)-transformations.

(b) Three one-parameter groups of such matrices are $U_j(\varepsilon) = \exp(i\varepsilon\sigma_j)$ for $j \in \{x, y, z\}$, where σ_j are the *Pauli matrices*

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

Find the Noether currents j_x^μ, j_y^μ, j_z^μ expressed in terms of the field components and their derivatives.

4. Derive a formula for the Noether current associated with a continuous symmetry of a second order lagrangian in flat spacetime.

Lecture 8

The Energy-Momentum Tensor

8.1 Coordinate Transformation Symmetries

Last lecture we looked at global symmetries of the Lagrangian which rotates the various (dynamic) field components into each other, and how that gives rise to conserved currents and charges. Similarly, we can look at symmetries that arise from coordinate transformations. [1] and there is a corresponding theorem about conserved currents associated to such symmetries. Important cases include Lorentz transformations $\mathrm{SO}(1,3)$, and spacetime translations $\mathbb{R}^4$, which together make up the Poincaré group $\mathrm{SO}(1,3) \ltimes \mathbb{R}^4$.

From $\mathrm{SO}(1,3)$ we get conserved currents $\Sigma_{\mu\nu}{}^{\lambda} = -\Sigma_{\nu\mu}{}^{\lambda}$ (6 total, corresponding to the dimension of $\mathrm{SO}(1,3)$). We will not discuss these here. From the group $\mathbb{R}^4$ of spacetime translations we get 4 currents $T_{\mu}{}^{\nu}$ making up the *energy-momentum tensor*.

8.2 Spacetime Translation Operators

We work in n spacetime dimensions. Fix a constant vector a^{μ} in $\mathbb{R}^n$, which is to be the direction of translation, and let $\varepsilon > 0$ be small real number. On *coordinates*, we define

$$T : \mathbb{R}^n \to \mathbb{R}^n, \qquad T(x^{\mu}) = x^{\mu} + \delta x^{\mu}$$

where we put

$$\delta x^{\mu} = \varepsilon a^{\mu}.$$

[1] One can also look at more general transformations which include both fields and coordinates, but we won't need that generality. We refer to [2, Chapter 6] for details.

On *fields*, we define

$$(T\varphi)(x^\mu) = \varphi(x^\mu - \delta x^\mu). \tag{8.2.1}$$

The reader might wonder about the minus sign here, and it is actually quite important, so we address it. One answer is that if a group G acts on a set M, then it acts on (real-valued, say) functions on M by $(g \cdot \varphi)(x) = \varphi(g^{-1} \cdot x)$. The inverse is needed for non-abelian groups in order to make it a left action. In our case, of course, G is the additive group of $\mathbb{R}^4$ which is abelian so it wouldn't matter. But, the inverse (here negative) is still the natural thing to do. Another answer is that, when $n = 1$ and $\delta t > 0$, replacing $\varphi(t)$ by $\varphi(t - \delta t)$ moves the field *forwards* (i.e. to the *right*). Compare the graphs of $\exp(-t^2/2)$ and $\exp(-(t-1)^2/2)$; the latter is centered around $t = 1$.

We will need infinitesimal translations. Expanding $T\varphi$ to first order in ε gives

$$T\varphi = \varphi - \partial_\mu\varphi \cdot \delta x^\mu + \mathcal{O}(\varepsilon^2). \tag{8.2.2}$$

Putting $\delta\varphi = T\varphi - \varphi$ and dropping the big $\mathcal{O}$ notation we can write this as

$$\delta\varphi = -\partial_\mu\varphi \cdot \delta x^\mu. \tag{8.2.3}$$

Alternatively, directly differentiating (8.2.1) using the chain rule,

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} T\varphi = -\partial_\mu\varphi \cdot a^\mu. \tag{8.2.4}$$

8.3 The Conserved Currents of Translation Symmetry

Suppose $\mathscr{L} = \mathscr{L}(\varphi^{(a)}, \partial_\mu\varphi^{(a)})$ is a first order Lagrangian density. Coordinate symmetries work differently from the field-component symmetries of previous lecture. Requiring $\mathscr{L}$ to remain unaffected under translation of the fields is too strong, for example a mass term $m^2\varphi^2$ would get mapped to $m^2(T\varphi)^2$ which of course is different. Instead, the correct notion of symmetry is that *the Lagrangian density commutes with the translation operator*:

$$[T, \mathscr{L}](\varphi^{(a)}, \partial_\mu\varphi^{(a)}) = 0. \tag{8.3.1}$$

Written out, this means that (and from now on we drop the superscript (a) from the notation for simplicity)

$$T\big(\mathscr{L}(\varphi, \partial_\mu)\big) - \mathscr{L}(T\varphi, \partial_\mu T\varphi) = 0, \tag{8.3.2}$$

where we used that T and ∂_μ commute. In the left term, we regard the output of the lagrangian as a scalar field in its own right, that can be translated. Differentiating both sides with respect to ε, and setting $\varepsilon = 0$, we get

$$-\partial_\mu \mathscr{L} \cdot a^\mu - \Big(\frac{\partial \mathscr{L}}{\partial \varphi}(-\partial_\mu \varphi \cdot a^\mu) + \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi}(-\partial_\mu \partial_\nu \varphi \cdot a^\mu) \Big) = 0. \tag{8.3.3}$$

Assuming φ satisfy the Euler-Lagrange equation $\frac{\partial \mathscr{L}}{\partial \varphi} = \partial_\nu \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi}$, this is equivalent to

$$-\partial_\mu \mathscr{L} \cdot a^\mu + \partial_\nu \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi} \partial_\mu \varphi \cdot a^\mu + \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi} \partial_\mu \partial_\nu \varphi \cdot a^\mu = 0. \tag{8.3.4}$$

Using the product rule, this can be written (where we reintroduce the (a))

$$\partial_\nu \big(-\delta^\nu_\mu \mathscr{L} + \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi^{(a)}} \partial_\mu \varphi^{(a)} \big) \cdot a^\mu. \tag{8.3.5}$$

We have proved the following result.

Theorem 8.3.1 (Noether's First Theorem (coordinate translation symmetry case))**.** *If $\mathscr{L} = \mathscr{L}(\varphi^{(a)}, \partial_\mu \varphi^{(a)})$ is a first order Lagrangian density which commutes with any translation operator, then*

$$T_\mu{}^\nu \overset{def}{=} -\delta^\nu_\mu \mathscr{L} + \frac{\partial \mathscr{L}}{\partial \partial_\nu \varphi^{(a)}} \partial_\mu \varphi^{(a)} \tag{8.3.6}$$

is a conserved current in ν:

$$\partial_\nu T_\mu{}^\nu = 0 \tag{8.3.7}$$

provided the field components $\varphi^{(a)}$ satisfy their Euler-Lagrange equations.

Definition 8.3.2. The relative tensor $T_\mu{}^\nu$ of weight 1 is called the *energy-momentum tensor*.

The reason for this terminology is that in Lorentzian signature $+ - - \cdots -$ i.e. $(1, n-1)$, is that the Noether current $T_0{}^\nu$ is related to energy conservation, while the other currents $T_i{}^\nu$ are related to conservation of momentum ($i = 1, 2, \ldots, n-1$). The *energy density* of the system is the component $T_0{}^0$ and consequently the *energy* is the space integral of this density:

$$E = \int T_0{}^0 \, d^{n-1}x. \tag{8.3.8}$$

Like other conserved charges, it is a quantity calculated from the fields that is constant in time, provided the fields satisfy their field equations. But, of course, energy plays a particularly important role in physics.

Sometimes people raise μ or lower ν to make the energy-momentum tensor co(ntra)variant. If it's the flat diagonal metric then $T_{00} = \eta_{0\mu} T_0{}^{\mu} = T_0{}^0$ and similarly equal to T^{00}. (With signature $-+++$ you get sign changes!)

Example 8.3.3. Let's compute the expression for the energy density of a free scalar theory in curved spacetime. The Lagrangian density is

$$\mathscr{L} = \Big(\frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - \frac{1}{2}m^2\varphi^2\Big)\sqrt{|g|}. \tag{8.3.9}$$

Therefore,

$$\frac{\partial\mathscr{L}}{\partial\partial_\nu\varphi} = \frac{\partial}{\partial\partial_\nu\phi}\big(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\varphi\partial_\beta\varphi - \frac{1}{2}m^2\varphi^2\big)\sqrt{|g|} = \frac{1}{2}g^{\alpha\beta}(\delta^\nu_\alpha\partial_\beta\varphi + \partial_\alpha\varphi\delta^\nu_\beta)\sqrt{|g|} = \partial^\nu\varphi\sqrt{|g|}. \tag{8.3.10}$$

Substituting (8.3.9)-(8.3.10) into (8.3.6), we find that the energy-momentum tensor is

$$\begin{aligned} T_\mu{}^\nu &= \Big(-\delta^\nu_\mu\frac{1}{2}\partial_\alpha\varphi\partial^\alpha\varphi + \delta^\nu_\mu\frac{1}{2}m^2\varphi^2 + \partial^\nu\varphi\partial_\mu\varphi\Big)\sqrt{|g|} \\ &= \Big(\big(\delta^\alpha_\mu g^{\beta\nu} - \frac{1}{2}\delta^\nu_\mu g^{\alpha\beta}\big)\partial_\alpha\varphi\partial_\beta\varphi + \frac{1}{2}\delta^\nu_\mu m^2\varphi^2\Big)\sqrt{|g|} \end{aligned} \tag{8.3.11}$$

Therefore, now assuming $\eta_{\mu\nu} = \mathrm{diag}(1,-1,-1,\ldots,-1)$, we obtain

$$T_0{}^0 = \frac{1}{2}m^2\varphi^2 + \frac{1}{2}\delta^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi \geq 0. \tag{8.3.12}$$

This is non-negative, the second term is a sum of squares. This is the reason for the choices of signs in front of the terms in the scalar lagrangian (8.3.9).

8.4 Alternative Definition of Energy-Momentum Tensor

There is a different, inequivalent, definition of energy-momentum tensor which is closely related to the one above for many lagrangians, and plays an important role

in gravitational theories. The alternative definition is the functional derivative (variation) of the lagrangian density with respect to the (contravariant) metric:

$$\overset{g}{T}_{\mu\nu} = \frac{\delta \mathscr{L}}{\delta g^{\mu\nu}}. \tag{8.4.1}$$

In the case when $\mathscr{L}$ does not involve derivatives of the metric, this is the same thing as

$$\overset{g}{T}_{\mu\nu} = \frac{\partial \mathscr{L}}{\partial g^{\mu\nu}}. \tag{8.4.2}$$

Example 8.4.1. Let us compute this alternative tensor for the free scalar theory.

$$\begin{aligned}\overset{g}{T}_{\mu\nu} &= \frac{\partial}{\partial g^{\mu\nu}}\left(\left(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\varphi\partial_\beta\varphi - \frac{1}{2}m^2\varphi^2\right)\sqrt{|g|}\right)\\ &= \frac{1}{2}\partial_\mu\varphi\partial_\nu\varphi\sqrt{|g|} + \frac{\mathscr{L}}{\sqrt{|g|}}\frac{\partial\sqrt{|g|}}{\partial g^{\mu\nu}}.\end{aligned}$$

By a Problem 2 of Section 8.5, we have

$$\frac{\partial\sqrt{|g|}}{\partial g^{\mu\nu}} = -\frac{1}{2}g_{\mu\nu}\sqrt{|g|}. \tag{8.4.3}$$

Substituting this, we get

$$\overset{g}{T}_{\mu\nu} = \frac{1}{2}\partial_\mu\varphi\partial_\nu\varphi\sqrt{|g|} - \frac{1}{2}g_{\mu\nu}\mathscr{L}.$$

Raising the second index, we obtain

$$\overset{g}{T}_{\mu}{}^{\nu} = -\frac{1}{2}\delta^\nu_\mu\mathscr{L} + \frac{1}{2}\partial_\mu\varphi\partial^\nu\varphi\sqrt{|g|}. \tag{8.4.4}$$

Comparing this with (8.3.6), and using (8.3.10), we see that

$$\overset{g}{T}_{\mu}{}^{\nu} = \frac{1}{2}T_\mu{}^\nu. \tag{8.4.5}$$

So, up to a factor of $1/2$, the two notions of energy-momentum tensor coincide for the free scalar theory in curved spacetime. This is not always the case, though. In particular, the tensor $\overset{g}{T}_{\mu\nu}$ is always symmetric, while there is no a priori reason the covariant energy-momentum tensor $T_{\mu\nu} = T_\mu{}^\alpha g_{\alpha\nu}$ would be symmetric.

8.5 Homework #8

1. In $1+0$ dimension, consider the toy model $L = \frac{1}{2}\dot{x}^2 - \frac{1}{2}\omega^2 x^2$ where $x = x(t)$ is the dynamic field content. Show that the energy of the solution $x(t) = A\cos(\omega t) + B\sin(\omega t)$ is $\frac{\omega^2}{2}(A^2 + B^2)$.

2. Prove the identity (8.4.3). (Hint: You may find some tricks in Section 5.5. First assume $g > 0$ for simplicity, and show that $\frac{\partial(g^{-1})}{\partial g^{\mu\nu}} = g^{-1}g_{\mu\nu}$, then write $g = (g^{-1})^{-1}$ and use the chain rule.)

3. Consider electromagnetism in curved spacetime. The *Maxwell lagrangian* is

$$\begin{aligned}\mathscr{L} = \mathscr{L}(A_\mu,\, \partial_\nu A_\mu,\, g_{\mu\nu}) &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}\sqrt{|g|} \\ &= -\frac{1}{4}g^{\mu\alpha}g^{\nu\beta}F_{\mu\nu}F_{\alpha\beta}\sqrt{|g|} \qquad (8.5.1)\end{aligned}$$

   where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and $g_{\mu\nu}$ is a general background metric. Show that the energy-momentum tensor defined in the alternative way by (8.4.1) is

$$\overset{g}{T}_{\mu\nu} = \left(\frac{1}{8}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta} - \frac{1}{2}g^{\alpha\beta}F_{\mu\alpha}F_{\nu\beta}\right)\sqrt{|g|}. \qquad (8.5.2)$$

   (Hint: If you get confused by the symmetry of the metric tensor, regard its component as n^2 independent variableas and read $g^{\mu\nu}$ in the lagrangian as $\frac{1}{2}(g^{\mu\nu} + g^{\nu\mu})$.

4. Given the result in the previous problem, show that when $n = 4$ and $g_{\mu\nu} = \eta_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$, the energy density $\overset{g}{T}_{00}$ for electromagnetism is given by

$$\overset{g}{T}_{00} = \frac{1}{4}\left(\mathbb{E}^2 + \mathbb{B}^2\right) \qquad (8.5.3)$$

   where

$$\mathbb{E}^2 = (F_{01})^2 + (F_{02})^2 + (F_{03})^2, \qquad \mathbb{B}^2 = (F_{12})^2 + (F_{23})^2 + (F_{31})^2.$$

   [In particular, it is non-negative. This justifies the minus sign in the lagrangian density (8.5.1).]

5. Continuing with the Maxwell Lagrangian (8.5.1), compute the covariant energy-momentum tensor $T_{\mu\nu} = T_\mu{}^\alpha g_{\alpha\nu}$ where $T_\mu{}^\nu$ is given by (8.3.6). Compare $T_{\mu\nu}$ and the symmetrization $T_{(\mu\nu)} \overset{\text{def}}{=} \frac{1}{2}(T_{\mu\nu} + T_{\nu\mu})$ with (8.5.2).

6. Show that the lagrangian density (8.5.1) is invariant under the so-called *gauge transformations* $A_\mu \mapsto A_\mu + \partial_\mu \Lambda$, where Λ is an arbitrary (smooth) real-valued function $\Lambda = \Lambda(x^\mu)$. [The function Λ is called the *parameter* of the transformation.]

7. Prove the following uniqueness theorem of the Maxwell Lagrangian: Suppose that $\mathscr{L} = \mathscr{L}(A_\mu, \partial_\nu A_\mu, g_{\mu\nu})$ is a first order Lagrangian density on curved spacetime (with $g_{\mu\nu}$ a background metric), such that

   (i) $\mathscr{L}$ is invariant with respect to gauge transformations $A_\mu \mapsto A_\mu + \partial_\mu \Lambda$,

   (ii) The Euler-Lagrange equations for the field components A_μ are linear.

   Prove that these assumptions imply that

$$\mathscr{L} = \big(\alpha_1 F_{\mu\nu}F^{\mu\nu} + \alpha_0\big)\sqrt{|g|} \tag{8.5.4}$$

   where α_0, α_1 are constant real numbers. (Hint: Work with $\mathscr{L}' = \mathscr{L}/\sqrt{|g|}$ which has to be a scalar field. Argue $\partial_\mu A_\nu$ appears at most quadratically; thus $\mathscr{L}' = B^{\mu\nu\sigma\tau}\partial_\mu A_\nu \partial_\sigma A_\tau +$ other terms. Since no other tensor fields are available, the coefficients have to be made up of the metric. Coefficients of odd rank like $C^{\mu\nu\lambda}$ or E^λ must be zero. Show that gauge invariance implies $B^{\mu\nu\sigma\tau} = -B^{\sigma\tau\mu\nu} = -B^{\nu\mu\sigma\tau}$ and that consequently $B^{\mu\nu\sigma\tau}$ is proportional to $g^{\mu\sigma}g^{\nu\tau} - g^{\nu\sigma}g^{\mu\tau}$. Show similarly that remaining non-constant terms like $D^{\mu\nu}A_\mu A_\nu$ or $G^{\mu\nu}\partial_\mu A_\nu$ must be zero.)

8. The Proca Lagrangian, is essentially Maxwell plus a nonzero mass term:

$$\mathscr{L}_{\text{Proca}} = \Big(-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu\Big)\sqrt{|g|} \tag{8.5.5}$$

   where $F_{\mu\nu}$ is as before. Find the equations of motion for the field components A_μ and show that all solutions satisfy the condition $\partial_\mu A^\mu = 0$. [When $m = 0$, this is a gauge fixing condition, called the Lorentz gauge. For Proca, we lose gauge invariance and the field appears in a fixed gauge.]

9. The *Schwinger Lagrangian* is a first order Lagrangian for *two* dynamical fields A_μ and $F_{\mu\nu}$ given (let's do this in flat spacetime for simplicity) by

$$\mathscr{L}_{\text{S}} = \frac{1}{2}\big(A_\nu \partial_\mu F^{\mu\nu} - F^{\mu\nu}\partial_\mu A_\nu\big) + \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{8.5.6}$$

   Show the following interesting facts:

(a) The field equations for A_μ and $F_{\mu\nu}$ are, respectively,

$$\partial_\mu F^{\mu\nu} = 0$$
$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.$$

Thus, what is a *definition* in the Maxwell lagrangian, is a *field equation* in the Schwinger lagrangian.

(b) Subtracting from the Schwinger lagrangian the total divergence $\frac{1}{2}\partial_\mu\big(A_\nu F^{\mu\nu}\big)$ we obtain $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, recovering the Maxwell theory.

This illustrates that two lagrangians with completely different field contents can end up describing equivalent field theories.

10. Consider the following *second order* Lagrangian (in flat spacetime)

$$\mathscr{L} = \mathscr{L}(A_\mu,\, \partial_\nu A_\mu,\, \partial_\lambda\partial_\nu A_\mu) = \frac{1}{2}A_\nu\partial_\mu F^{\mu\nu} \tag{8.5.7}$$

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

(a) Show that the field equations for A_μ end up being equivalent to the same as for the Maxwell Lagrangian: $\partial_\mu F^{\mu\nu} = 0$. (Hint: You must use the second order Euler-Lagrange equations here, see (6.6.1).)

(b) Show that we can again subtract a total divergence of $\frac{1}{2}\partial_\mu(A_\nu F^{\mu\nu})$ to recover the Maxwell Lagrangian.

[This shows that first and second order lagrangians can give equivalent field theories.]

11. In flat spacetime, the Maxwell lagrangian with a source field J^μ is

$$\mathscr{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + J^\mu A_\mu.$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ as usual.

(a) Show that the field equations are

$$\partial_\mu F^{\mu\nu} = J^\nu.$$

(b) Use the definition of $F_{\mu\nu}$ to show that

$$\partial_\mu {}^*F^{\mu\nu} = 0$$

where ${}^*F_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}F^{\alpha\beta}$ is the (Hodge) dual field.

(c) Show that you can correct the signs and constants in the following guess: By setting

$$E_i = F_{0i}, \qquad B_i = \varepsilon_{ijk} F^{jk}, \qquad \text{for } i = 1, 2, 3$$

or equivalently

$$F_{\mu\nu} \sim \begin{bmatrix} 0 & E_1 & E_2 & E_3 \\ & 0 & B_3 & -B_2 \\ & & 0 & B_1 \\ & & & 0 \end{bmatrix} \text{(skew-symmetric matrix)}$$

the two equations from parts (a) and (b) together produce the four equations for electric and magnetic fields E_i and B_i commonly known as Maxwell equations. [This is why $F_{\mu\nu}$ is called the *field strength tensor*. It is simply the electric and magnetic fields packaged together in one tensor. From this perspective, A_μ is called the *electromagnetic 4-potential*, because its derivatives give the field strengths. Lastly, the effect of $F_{\mu\nu}$ being a Lorentz tensor is that electric and magnetic fields get mixed up for comoving observers.]

Lecture 9

General Relativity

9.1 Gravity

We have seen examples of Lagrangian densities in curved spacetime $(M, g_{\mu\nu})$. In those theories, $g_{\mu\nu}$ is a fixed background metric (that is, $g_{\mu\nu}$ is not subject to field equations) and is only used to raise and lower indices, along with $\sqrt{|g|}$, in order to construct the Lagrangian as a scalar density.

Einstein's theory of general relativity is both a theory of pure gravity, as well as a framework for incorporating gravitational effects into any classical field theory.

In pure gravity, Einstein's field equations, which are the equations of motion for $g_{\mu\nu}$, can be derived from a Lagrangian density $\mathscr{L}_G$ depending only on the metric $g_{\mu\nu}$ and its derivatives. The gravitational field is in other words identified with the metric tensor itself, which is now a dynamic field.

To incorporate gravity into a (Lorentzian) theory of "matter" (i.e., non-gravitational fields) with Lagrangian $\mathscr{L}_M$ the following steps should be performed on $\mathscr{L}_M$:

1. Replace all $\eta_{\mu\nu}$ by $g_{\mu\nu}$,
2. replace all ∂_μ by ∇_μ (covariant derivative with respect to Levi-Civita connection for $g_{\mu\nu}$),
3. multiply the whole thing by $\sqrt{|g|}$,
4. add on $\mathscr{L}_G$.

For example, if the matter lagrangian is the Maxwell Lagrangian

$$\mathscr{L}_M = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} = -\frac{1}{4} F_{\mu\nu} F_{\lambda\rho} \eta^{\mu\lambda} \eta^{\nu\rho}$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, the corresponding Einstein-Maxwell Lagrangian reads

$$\mathscr{L} = \mathscr{L}_G - \frac{1}{4}\widetilde{F}_{\mu\nu}\widetilde{F}_{\lambda\rho}g^{\mu\lambda}g^{\nu\rho}\sqrt{|g|}$$

where $\widetilde{F}_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu$. (It is, in this context, a lucky "coincidence" that, in fact, $\widetilde{F}_{\mu\nu} = F_{\mu\nu}$ because the Levi-Civita connection is symmetric.) The beauty of this is that there is no need to add an interaction term here. The coupling to gravity is already present in the updated matter Lagrangian.

Remark 9.1.1. In a second order matter Lagrangian this process is not well-defined, as it is unclear whether $\partial_\mu\partial_\nu$ should be replaced by $\nabla_\mu\nabla_\nu$ or $\nabla_\nu\nabla_\mu$ (which are different, since covariant derivatives do not commute). In some cases the right thing to do then is to replace $\partial_\mu\partial_\nu$ by the symmetrized second order covariant differential operator $\nabla_{(\mu\nu)} = \frac{1}{2}(\nabla_\mu\nabla_\nu + \nabla_\nu\nabla_\mu)$.

9.2 Derivation of the Einstein-Hilbert Lagrangian

In this section we provide a mathematical derivation of the Einstein-Hilbert Lagrangian. Such a derivation can never prove that it is the "correct" theory, as ultimately the predictions need to be checked against experiments. However, given that this has been done, the derivation serves to illustrate that the theory is highly constrained.

A fundamental principle of general relativity is that the laws of physics should be invariant under general coordinate transformations. In terms of lagrangian field theory, this means we need a Lagrangian which is a scalar density with respect to the general smooth geometry $\mathscr{G}_{\text{sm}}$.

We add to this one requirement, that the field equations (that is, the Euler-Lagrange equations) should be second order partial differential equations. Ultimately, this is an arbitrary assumption,[1] but can be motivated by the desire to, in some way, recover the Newton-Laplace equation $\Delta\phi = 4\pi G\rho$ for the gravitational (scalar) potential ϕ, which is a second order PDE (see (1.2.6) for details).

9.2.1 First Order Metric Lagrangians are Trivial

At first sight, it seems very natural to look for a first order Lagrangian for $g_{\mu\nu}$:

$$\mathscr{L} = \mathscr{L}(g_{\mu\nu},\, g_{\mu\nu,\lambda}).$$

[1] there are indeed so-called higher-derivative theories of gravity, but none of those so far produce measurably better predictions than standard GR

The field equations would then be $-\frac{\mathscr{L}}{\partial g_{\mu\nu}} + \partial_\lambda \frac{\partial \mathscr{L}}{\partial g_{\mu\nu,\lambda}} = 0$ which is a system of second order PDEs. However, the only such Lagrangian density turns out to be $\mathscr{L} = c\sqrt{|g|}$ where c is a constant. [2] This is a consequence of the following precise statement, applied to $\mathscr{L}/\sqrt{|g|}$:

Proposition 9.2.1. *Suppose that f is any smooth real-valued function of $n^2 + n^3$ variables X_{ij}, Y_{ijk}, with the following property:*

- *Whenever we substitute a metric and its derivatives into f, that is, whenever we substitute $X_{\mu\nu} = g_{\mu\nu}$ and $Y_{\mu\nu\lambda} = g_{\mu\nu,\lambda}$ into f, where $g_{\mu\nu}$ is a metric on M, the resulting collection $\big(f(g_{\mu\nu}, g_{\mu\nu,\lambda})\big)_{x\mapsto x^\mu}$ of real-valued functions[3] (one for each coordinate system on M) constitute a scalar field on M. That is to say, $f(g_{\mu\nu}, g_{\mu\nu,\lambda}) = f(g_{\mu'\nu'}, g_{\mu'\nu',\lambda'})$ for any two coordinate systems $x \mapsto x^\mu$ and $x \mapsto x^{\mu'}$ on M.*

Then the scalar field $f(g_{\mu\nu}, g_{\mu\nu,\lambda})$ is constant on M. Furthermore, this constant only depends on the signature of $g_{\mu\nu}$.[4]

Proof. Let y be a point on M. Let x^μ be a locally inertial coordinate system at y (see Theorem 3.6.2). By Homework 7 of Section 4.5, we have $g_{\mu\nu,\lambda}(y) = 0$. By the proof of Theorem 3.6.2, we can also assume that $g_{\mu\nu}(y) = \eta_{\mu\nu}$ which is diagonal with a number of $+1$'s followed by a number of -1's on the diagonal, according to the signature of $g_{\mu\nu}$. Therefore, we have $f(g_{\mu\nu}(y), g_{\mu\nu,\lambda}(y)) = c$, where $c = f(\eta_{\mu\nu}, 0)$ (where 0 means the n^3-tuple of zeroes). Note that c does not depend on y: It is the real constant number we would get by substituting the flat metric $\eta_{\mu\nu}$ for $g_{\mu\nu}$ and 0 for $g_{\mu\nu,\lambda}$ into f. But if two scalar fields are equal at some point y in one coordinate system, they are equal at that point in all coordinate systems. Therefore, we have $f(g_{\mu'\nu'}(y), g_{\mu'\nu',\lambda'}(y)) = c$ in all coordinate systems $x \mapsto x^{\mu'}$. But, y was an arbitrary point on M, and therefore the scalar field $f(g_{\mu\nu}, g_{\mu\nu,\lambda})$ is constant equal to c. By construction, c only depends on the signature of $g_{\mu\nu}$. □

[2] Therefore, by (8.4.3), the field equations are equivalent to $cg_{\mu\nu}\sqrt{|g|} = 0$ forcing either $g = 0$ (not a metric) or $c = 0$ ($g_{\mu\nu}$ is a background field).

[3] As we mentioned before, we can either think of them all as defined on M, or, that each function in the collection is defined on the image of the corresponding coordinate system.

[4] If we allowed coordinate charts (hence M) to be disconnected, a metric could have different signatures on each connected component of M. In that case the scalar field $f(g_{\mu\nu}, g_{\mu\nu,\lambda})$ could take different values on each connected component.

9.2.2 Second Order Metric Lagrangians

The following section is based on [3, Section 11.2].

Theorem 9.2.2. *Let n be a positive integer, let $\mathscr{G}$ be the n-dimensional smooth geometry (from Section 1.1.3), and M be a $\mathscr{G}$-space.*

Let $\mathscr{L} = \mathscr{L}(g_{\mu\nu},\, g_{\mu\nu,\lambda},\, g_{\mu\nu,\lambda\rho})$ be a second order Lagrangian density for the metric $g_{\mu\nu}$ such that the field equations are second order PDEs. Then there exist constants a and c such that

$$\mathscr{L} = (a \cdot R + c)\sqrt{|g|} \tag{9.2.1}$$

where $R = g^{\mu\nu}R_{\lambda\mu\nu}{}^{\lambda}$ is the scalar curvature of the (Levi-Civita connection of the) metric $g_{\mu\nu}$.

Proof. By definition of what a Lagrangian density is, $\mathscr{L}/\sqrt{|g|}$ is scalar field, say $L = L(x)$, on M of the form $x \mapsto L(x) = f(g_{\mu\nu}(x),\, g_{\mu\nu,\lambda}(x),\, g_{\mu\nu,\lambda\rho}(x))$ where f is a smooth real-valued function depending on $n^2 + n^3 + n^4$ variables. As in the proof of Proposition 9.2.1, that $\mathscr{L}/\sqrt{|g|}$ is a scalar field means that

$$f(g_{\mu\nu},\, g_{\mu\nu,\lambda},\, g_{\mu\nu,\lambda\rho}) = f(g_{\mu'\nu'},\, g_{\mu'\nu',\lambda'},\, g_{\mu'\nu',\lambda'\rho'}) \tag{9.2.2}$$

for any two coordinate systems $x \mapsto x^{\mu}$ and $x \mapsto x^{\mu'}$ on M.

Fix a point $y \in M$. As in the proof of Proposition 9.2.1, there exists a coordinate system on M such that $g_{\mu\nu}(y) = \eta_{\mu\nu}$ and $g_{\mu\nu,\lambda}(y) = 0$. If necessary, after the further change of coordinates (a simple translation) $x^{\mu} \mapsto x^{\mu} - y^{\mu}$, we may also without loss of generality assume $y^{\mu} = 0$. In this coordinate system, the scalar field L, when evaluated at y, is

$$L(y) = f\big(\eta_{\mu\nu},\, 0,\, g_{\mu\nu,\lambda\rho}(y)\big) \tag{9.2.3}$$

The field equations $\mathscr{L}$ read (cf., (6.6.1))

$$-\frac{\partial \mathscr{L}}{\partial g_{\mu\nu}} + \partial_{\lambda}\frac{\partial \mathscr{L}}{\partial g_{\mu\nu,\lambda}} - \partial_{\rho}\partial_{\lambda}\frac{\partial \mathscr{L}}{\partial g_{\mu\nu,\lambda\rho}} = 0. \tag{9.2.4}$$

The only way these can be second order PDEs is if $\frac{\partial \mathscr{L}}{\partial g_{\mu\nu,\lambda\rho}}$ is independent of the second order partial derivatives $g_{\mu\nu,\lambda\rho}$. Otherwise (9.2.4) are guaranteed to be fourth order PDEs. This implies that[5]

$$f(g_{\mu\nu},\, g_{\mu\nu,\lambda},\, g_{\mu\nu,\lambda\rho}) = g_{\mu\nu,\lambda\rho}B^{\mu\nu\lambda\rho}(g_{\alpha\beta},\, g_{\alpha\beta,\gamma}) + C(g_{\alpha\beta},\, g_{\alpha\beta,\gamma}). \tag{9.2.5}$$

[5]If you doubt this, consider the fact that if a smooth function $f(s_1, s_2)$ is such that $\partial f/\partial s_2 = g(s_1)$ then, by integrating, $f(s_1, s_2) = s_2 g(s_1) + h(s_1)$ for some smooth functions g and h.

for some set of real-valued smooth functions $B^{\mu\nu\lambda\rho}$ and C (all depending on n^2+n^3 variables). We remark that $B^{\mu\nu\lambda\rho}$ typically will not transform as a $\binom{4}{0}$-tensor under general coordinate transformations.

Thus, at the point y, still working in our special coordinate system,

$$L(y) = g_{\mu\nu,\lambda\rho}(y)B^{\mu\nu\lambda\rho}(\eta_{\alpha\beta},0) + c, \tag{9.2.6}$$

where $c = C(\eta_{\alpha\beta},0)$ is some constant (that does not depend on y). Because Lorentz tranformations are linear, $g_{\mu\nu,\lambda\rho}$ transform as a Lorentz $\binom{0}{4}$-tensor (but not as a general tensor). In order for $L(y)$ to be coordinate independent, this forces $B^{\mu\nu\lambda\rho}(\eta_{\alpha\beta},0)$ to transform as a Lorentz $\binom{4}{0}$-tensor (unless $L(y) = c$ which, if true for all y, leads to the case of Proposition 9.2.1). Thus we have

$$B^{\mu'\nu'\lambda'\rho'}(\eta_{\alpha'\beta'}) = \Lambda^{\mu'}_{\sigma}\Lambda^{\nu'}_{\tau}\Lambda^{\lambda'}_{\pi}\Lambda^{\rho'}_{\xi}B^{\sigma\tau\pi\xi}(\Lambda^{\gamma}_{\alpha'}\Lambda^{\delta}_{\beta'}\eta_{\gamma\delta}). \tag{9.2.7}$$

This implies (see Problem 3 of Section 9.6) that $B^{\mu\nu\lambda\rho}$ is a quadratic form in the contravariant flat metric $\eta^{\alpha\beta}$ and consequently

$$B^{\mu\nu\lambda\rho} = \sum(\text{constants})\cdot\eta^{\alpha\beta}\eta^{\gamma\delta} \tag{9.2.8a}$$

where we sum over all permutations $(\alpha,\beta,\gamma,\delta)$ of (μ,ν,λ,ρ). Equivalently, due to the symmetries $\eta^{\mu\nu} = \eta^{\nu\mu}$ and $\eta^{\mu\nu}\eta^{\lambda\rho} = \eta^{\lambda\rho}\eta^{\mu\nu}$ there are only three terms corresponding to which index μ is paired with:

$$B^{\mu\nu\lambda\rho} = a_1\eta^{\mu\nu}\eta^{\lambda\rho} + a_2\eta^{\mu\nu}\eta^{\nu\rho} + a_3\eta^{\mu\rho}\eta^{\nu\lambda} \tag{9.2.8b}$$

for some constants a_i (that can't depend on y). Furthermore, after pairing this with $g_{\mu\nu,\lambda\rho}$ and using symmetries in the indices, we obtain

$$L(y) = g_{\mu\nu,\lambda\rho}(y)\big(a\cdot\eta^{\mu\lambda}\eta^{\nu\rho} + b\cdot\eta^{\mu\nu}\eta^{\lambda\rho}\big), \tag{9.2.9}$$

where $a = a_2$, $b = \frac{1}{2}(a_1+a_3)$. Now change cordinates as follows:

$$x^{\mu} = \bar{x}^{\mu} + \frac{1}{3!}C^{\mu}{}_{\alpha\beta\gamma}\bar{x}^{\alpha}\bar{x}^{\beta}\bar{x}^{\gamma} \tag{9.2.10}$$

where C is completely symmetric in the lower indices. Then $\bar{x}^{\mu}$ is also a locally inertial coordinate system at y, and furthermore, using the chain rule and product

rule,

$$\begin{aligned}\bar{g}_{\mu\nu,\lambda\rho}(y) &= \bar{\partial}_\rho\bar{\partial}_\lambda\big((\bar{\partial}_\mu x^\alpha)(\bar{\partial}_\nu x^\beta)g_{\alpha\beta}\big)\\ &= (\bar{\partial}_\rho\bar{\partial}_\lambda\bar{\partial}_\mu x^\alpha)(\bar{\partial}_\nu x^\beta)g_{\alpha\beta}\\ &\quad + (\bar{\partial}_\mu x^\alpha)(\bar{\partial}_\rho\bar{\partial}_\lambda\bar{\partial}_\nu x^\beta)g_{\alpha\beta}\\ &\quad + (\bar{\partial}_\mu x^\alpha)(\bar{\partial}_\nu x^\beta)\bar{\partial}_\rho\bar{\partial}_\lambda x_{\alpha\beta} + (\text{terms vanishing at } y)\\ &= \big(C^\alpha{}_{\rho\lambda\mu}\delta^\beta_\nu + \delta^\alpha_\mu C^\beta{}_{\rho\lambda\nu}\big)\eta_{\alpha\beta} + \delta^\alpha_\mu\delta^\beta_\nu g_{\alpha\beta,\lambda\rho}(y)\\ &= C_{\mu\nu\lambda\rho} + C_{\nu\mu\lambda\rho} + g_{\mu\nu,\lambda\rho}(y)\end{aligned} \tag{9.2.11}$$

where we used the flat metric to lower the first index in C. Since $L(y)$ is independent of coordinate system,

$$\bar{g}_{\mu\nu,\lambda\rho}(y)\big(a\cdot\eta^{\mu\lambda}\eta^{\nu\rho} + b\cdot\eta^{\mu\nu}\eta^{\lambda\rho}\big) - g_{\mu\nu,\lambda\rho}(y)\big(a\cdot\eta^{\mu\lambda}\eta^{\nu\rho} + b\cdot\eta^{\mu\nu}\eta^{\lambda\rho}\big) = 0.$$

Substituting (9.2.11) into this, and using that $C_{\mu\alpha\beta\gamma}$ is completely symmetric in $\alpha\beta\gamma$, we obtain

$$2C_{\mu\nu\lambda\rho}\eta^{\mu\nu}\eta^{\nu\rho}\cdot(a+b) = 0$$

and consequently, $b = -a$. We conclude that in a coordinate system which is locally inertial at $y \in M$, the Lagrangian must take the form

$$\frac{\mathscr{L}}{\sqrt{|g|}} = a\cdot g_{\mu\nu,\lambda\rho}\big(\eta^{\mu\lambda}\eta^{\nu\rho} - \eta^{\mu\nu}\eta^{\lambda\rho}\big) + c \tag{9.2.12}$$

for some constants a and c.

By Problem 8 of Section 4.5, the covariant curvature tensor $R_{\mu\nu\lambda\rho}$ when expressed in a locally inertial coordinate system at y, has the form

$$R_{\mu\nu\lambda\rho}(y) = \frac{1}{2}\big(g_{\nu\rho,\lambda\mu} + g_{\mu\lambda,\rho,\nu} - g_{\mu\rho,\lambda\nu} - g_{\nu\lambda,\rho\mu}\big).$$

Therefore, the scalar curvature $R = g^{\alpha\beta}g^{\mu\nu}R_{\mu\alpha\beta\nu}$ at y is

$$\begin{aligned}R(y) &= \frac{1}{2}\eta^{\alpha\beta}\eta^{\mu\nu}\big(g_{\nu\rho,\lambda\mu} + g_{\mu\lambda,\rho,\nu} - g_{\mu\rho,\lambda\nu} - g_{\nu\lambda,\rho\mu}\big)\\ &= \frac{1}{2}\eta^{\alpha\beta}\eta^{\mu\nu}\big(g_{\alpha\nu,\beta\mu} - g_{\alpha\beta,\mu\nu}\big)\\ &= g_{\mu\nu,\lambda\rho}\big(\eta^{\mu\lambda}\eta^{\nu\rho} - \eta^{\mu\nu}\eta^{\lambda\rho}\big).\end{aligned} \tag{9.2.13}$$

Comparing (9.2.12) and (9.2.13), we find that, for any point $y \in M$, in a locally inertial coordinate system at y we have

$$\frac{\mathscr{L}}{\sqrt{|g|}} = a \cdot R + c \qquad \text{(at the point } y\text{)} \tag{9.2.14}$$

where a and c are constants not depending on y. However, both sides of (9.2.14) are scalar fields. Therefore, equality must hold (at y) in any other coordinate system, locally inertial or not. But y was an arbitrary point of M, and a and c are constants which are independent of y. Therefore equality holds everywhere on M. This finishes the proof of the theorem. □

9.3 Field Equations

The variational problem for the second order lagrangian $\mathscr{L}_G = (a \cdot R + c)\sqrt{|g|}$ is problematic for the following reason. If we consider the action $S_G[g_{\mu\nu}] = \int \mathscr{L}_G \, d^n x$ as a functional of the metric alone, the variation is

$$\delta S_G = \int \Big(\frac{\partial \mathscr{L}_G}{\partial g_{\mu\nu}} \delta g_{\mu\nu} + \frac{\partial \mathscr{L}_G}{\partial g_{\mu\nu,\lambda}} \partial_\lambda \delta g_{\mu\nu} + \frac{\partial \mathscr{L}_G}{\partial g_{\mu\nu,\lambda\rho}} \partial_\lambda \partial_\rho \delta_{\mu\nu} \Big) \, d^n x$$

where we used that $\delta g_{\mu\nu,\lambda} = \partial_\lambda \delta g_{\mu\nu}$. Integrating by parts once in the second term and twice in the third term, and assuming $\partial \partial K = 0$ we get

$$\begin{aligned} \delta S_G = &\int_K \Big(\frac{\partial \mathscr{L}_G}{\partial g_{\mu\nu}} - \partial_\lambda \frac{\partial \mathscr{L}_G}{\partial g_{\mu\nu,\lambda}} + \partial_\lambda \partial_\rho \frac{\partial \mathscr{L}_G}{\partial g_{\mu\nu,\lambda\rho}} \Big) \delta g_{\mu\nu} \, d^n x \\ &+ \int_{\partial K} \partial_\lambda \Big(\frac{\partial \mathscr{L}_G}{\partial g_{\mu\nu,\lambda}} \delta g_{\mu\nu} + \frac{\partial \mathscr{L}_G}{\partial g_{\mu\nu,\lambda\rho}} \partial_\rho \delta g_{\mu\nu} \Big) \, d^{n-1} y. \end{aligned}$$

In order to throw away the boundary term, we need the inside of ∂_λ to be a current vanishing on the boundary ∂K. By restricting to variations $\delta g_{\mu\nu}$ vanishing on ∂K, the first term can be removed. But we also need to require that $\partial_\rho \delta g_{\mu\nu}$ vanish on ∂K. Herein lies the problem: $\partial_\rho \delta g_{\mu\nu}$ is not a tensor. The variation $\delta g_{\mu\nu}$, being a difference of two metrics, is a tensor and therefore its vanishing at a point or subset of M is a coordinate-independent quality.

There are several approaches to this issue. Einstein added an integral over the boundary to the action involving sectional curvature.

We will follow the *Palatini formalism* which regards the action as a functional of two independent fields: a metric[6] $g^{\mu\nu}$ and a (symmetric) affine connection $\Gamma_{\mu\nu}{}^{\lambda}$.

$$S_G = S_G[g^{\mu\nu},\, \Gamma_{\mu\nu}{}^{\lambda}] = \int (a \cdot R + c)\sqrt{|g|}\, d^n x$$

wherein $R = g^{\mu\nu} R_{\lambda\mu\nu}{}^{\lambda}$ and the curvature tensor is regarded as a function of the affine connection and its derivatives explicitly given in (4.1.5) (this was before we introduced raising/lowering; the upper indices should be moved to the right in that formula). The upshot here is that the Lagrangian is now a first order lagrangian in the contravariant metric and affine connection. The variational problem is well-defined, because even though the connection is not a tensor, the *variation* $\delta\Gamma_{\mu\nu}{}^{\lambda}$ is, being a difference of affine connections.

Assuming $a = 1$ and $c = 0$, the field equations are as follows (see Problems below):

$$\frac{\delta S_G}{\delta g^{\mu\nu}} = 0 \quad \Leftrightarrow \quad R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 0 \tag{9.3.1}$$

and

$$\frac{\delta S_G}{\delta \Gamma_{\mu\nu}{}^{\lambda}} = 0 \quad \Leftrightarrow \quad \nabla_\lambda g^{\mu\nu} = 0 \tag{9.3.2}$$

where ∇_λ is the covariant derivative with respect to the affine connection. Since the affine connection is symmetric, the vanishing of $\nabla_\lambda g^{\mu\nu}$ is equivalent to the connection being the Levi-Civita connection.

9.4 Newtonian Limit

We follow [**?**]. We assume $n = 4$ and that $g_{\mu\nu}$ has Lorentzian signature.

Let us consider a static, weak, gravitational field, such as produced by non-relativistic matter moving with low velocities. For weak fields, we can choose a coordinate system which is almost Euclidean. That is,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{9.4.1}$$

where $|h_{\mu\nu}|$ is much smaller than 1, and the non-relativistic limit ensures that $|T_{00}| \gg |T_{ij}|$. In fact, we will assume $T_{ij} = 0$ for $i = j = 1, 2, 3$.

.....

[6] It is somewhat more convenient to use the contravariant metric. Each can of course be expressed in terms of the other.

9.5 Linear Approximation and Gravitational Waves

.....

9.6 Homework #9

1. Prove that the field equation for the metric in the Palatini formulation is given by (9.3.1).
2. Prove that the field equation for the affine connection in the Palatini formulation is given by (9.3.2).
3. Prove that (9.2.7) implies (9.2.8).

Part III

Local Gauge Field Theories

Local Symmetry: Electromagnetism Revisited

Historically, the classical (Maxwell) equations governing electromagnetism were understood before the discovery of the electrically charged particles. From a modern perspective, however, it is more natural to begin with a lagrangian describing, say, an electron:

$$\mathscr{L} = \overline{\psi}(i\boldsymbol{\gamma}^\mu \partial_\mu - m)\psi.$$

Here ψ is a Dirac spinor, $\boldsymbol{\gamma}^\mu$ are Dirac's gamma matrices, and m is the mass of the electron. We will cover spinors in Lecture 12 in detail. For this discussion, the reader unfamiliar with spinors may safely assume that ψ is a complex scalar field, $\bar{\psi} = \psi^*$ is the complex conjugate, and $\boldsymbol{\gamma}^\mu$ are certain imaginary numbers. The following observations are made:

(1) The above lagrangian remains unchanged after the substitution $\psi \mapsto e^{i\theta}\psi$, for any constant real number θ.

This is because $\overline{e^{i\theta}\psi} = e^{-i\theta}\overline{\psi}$ so the two exponential factors cancel out. In other words, there is a *global* $U(1)$*-symmetry*, where $U(1) = \{e^{i\theta} \mid \theta \in \mathbb{R}\}$ is the group of complex numbers of modulus 1.

(2) There is therefore a corresponding conserved *Noether current* j^μ. If we fix a parameter (coupling constant) q, and write the one-parameter group as $\theta \mapsto \exp(-iq\theta)$ then by (7.1.2), we have

$$j^\mu = \frac{\partial \mathscr{L}}{\partial \partial_\mu \psi}(-iq\psi) = q\overline{\psi}\boldsymbol{\gamma}^\mu\psi.$$

Now, we wish to upgrade the global $U(1)$-symmetry to a *local symmetry*. That is:

(3) We want to modify the lagrangian in such a way that it remains unchanged after the substitution $\psi \mapsto e^{i\theta(x)}\psi$, for any smooth function θ defined on spacetime.

A physical justification for this (at first sight rather strong) requirement is that of locality: if ψ and $\exp(i\theta)\psi$ represent the same physical field, then we should also be allowed to replace ψ by $\exp(i\theta(x))\psi$ where $\theta(x)$ is constant in a region of spacetime encompassing the experimental setup, and decays smoothly to zero outside this compact region. Since the experiment occupies an arbitrary region, it seems $\theta(x)$ might as well be any smooth function.

Let us see how far away we are from local symmetry:

(4) Replacing ψ by $E\psi$ where $E = \exp(-iq\theta(x))$ in the lagrangian $\mathscr{L}$, we get

$$\mathscr{L}' = E^*\overline{\psi}(i\boldsymbol{\gamma}^\mu\partial_\mu - m)(E\psi) = \mathscr{L} + E^* \cdot \overline{\psi}i\boldsymbol{\gamma}^\mu\psi \cdot (\partial_\mu E) = \mathscr{L} + j^\mu\partial_\mu\theta$$

The need arises for adding a term $j^\mu A_\mu$ to the lagrangian involving a new field A_μ which also transforms under gauge transformation but in a way that cancel out the extra term. This transformation law is $A'_\mu = A_\mu \pm \partial_\mu\theta(x)$ (up to constants to be checked).

(5) We recognize that the transformation law for A_μ is exactly the same as the one for the electromagnetic 4-potential previously discuss. We also recognize that the correction term $j^\mu A_\mu$ we had to add in, is an interaction term between the spinor field ψ and the vector field A_μ. *This is how local gauge symmetry gives rise to force carriers and naturally provide interaction terms.*

(6) As a last step, we add in a kinetic term for A_μ, giving the final form of the lagrangian that go into quantum electrodynamics of a Dirac spinor ψ (the electron-positron field) and a $U(1)$-gauge field A_μ (the photon field):

$$\mathscr{L}_{QED} = \overline{\psi}(i\boldsymbol{\gamma}^\mu D_\mu + m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

where D_μ is the $U(1)$-covariant derivative $\partial_\mu + A_\mu$ (up to constants to be checked).

Lecture 10

Principal Bundles I: Matter Fields

10.1 Yang-Mills Theory

Yang-Mills theory is a form of generalized electromagnetism, where the gauge group $U(1)$ is replaced by a general Lie group G.

Let M be a smooth manifold, called the *spacetime manifold*. Points $x \in M$ are called *events*.

Let G be a compact Lie group with Lie algebra $\mathfrak{g}$. [1]

Let $\pi : P \to M$ be a principal G-bundle.[2]

Let V be a representation of G, called the *matter representation*.

To this data we have the *associated vector bundle* $\pi_V : P \times_G V \to M$, $[(p, v)] \mapsto \pi(p)$. Here $[(p, v)]$ denotes the equivalence class in $P \times_G V$ containing the point $(p, v) \in P \times G$ with respect to the equivalence relation saying $(p.g, v) \sim (p, g.v)$ for all $g \in G$.

A *matter field* is a (possibly local[3]) section $\Psi : M \dashrightarrow P \times_G V$ of the associated vector bundle.

Excitations in the matter field Ψ are percieved as particles. Each irreducible summand of V corresponds to a distinct particle type. For example, if $G = U(1)_Y \times SU(2)_L \times SU(3)_C$ (subscripts are labels and can be ignored; Y refers to *hypercharge*, a primordial analog of electromagnetism; L to *isospin* and C to *color charge*) we can

[1] It will turn out that irreducible summands in the adjoint representation $\mathfrak{g}$ correspond to *force carriers* (gauge bosons) in the theory. For example, $G = U(1) \times SU(2) \times SU(3)$ would yield three forces. If $G = SU(5)$ or $SO(1)$, there would be a single force.

[2] Whether or not this bundle is trivial (i.e. $P = M \times G \ni (x, g) \mapsto x$) is of no consequence in what follows, other than psychologically. We don't want there to be any "special" map $M \to P$.

[3] Mathematically we should think of Ψ as a global object, as it doesn't depend on choice of coordinates or gauge (to be defined momentarily).

consider the irrep $V = \mathbb{C}_1 \otimes \mathbb{C}^2 \otimes \mathbb{C}$ (the subscript k of $\mathbb{C}_k$ label the integral weight, eg. $\mathbb{C}_k$ means the $U(1)$-module with action $z.1 = z^k$) with appropriate lagrangian would provide a classical model for the Higgs doublet field $\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}$ (as it appears before electroweak symmetry breaking).

To write down a lagrangian producing equations of motions, we need the ability to express changes in the matter fields across the spacetime manifold M. In other words, we need know how to *differentiate sections of the associated vector bundle*.

To this end, we choose a local section $\sigma : M \dashrightarrow P$ of the principal bundle map π. Such a section is called *a choice of gauge*. Once we have chosen σ, there is for each x a unique vector ${}^{\sigma}\psi(x) \in V$ such that

$$\Psi(x) = [\big(\sigma(x), {}^{\sigma}\psi(x)\big)]. \tag{10.1.1}$$

Indeed, if $\Psi(x) = [(p, v)]$, then $p = \sigma(x).g$ for some unique $g \in G$ (due to free and transitive action of G on the fiber $\pi^{-1}(\{x\})$ above x) and hence $[(p, v)] = [(\sigma(x), g.v)]$ and we set ${}^{\sigma}\psi(x) = g.v$.

The upshot of this is that ${}^{\sigma}\psi$ is a V-valued smooth function defined on a region[4] of the manifold M, and we know how to differentiate those. The downside is that we made a choice of gauge. Thus, we don't have one function but a whole family $\psi = \{{}^{\sigma}\psi\}_{\sigma}$. Let us first see how these are related to each other.

Suppose, therefore, that $\sigma' : M \dashrightarrow P$ is different choice of gauge. Then, for each x (in the overlap of domains of σ, σ') there is a unique $g(x) \in G$ such that

$$\boxed{\sigma(x) = \sigma'(x).g(x).} \tag{10.1.2}$$

The smooth G-valued function $g(x)$ is called a *gauge transformation*. Note now that we have (and here we write $[p, v]$ for $[(p, v)]$)

$$[\sigma'(x), {}^{\sigma'}\psi(x)] = \Psi(x) = [\sigma(x), {}^{\sigma}\psi(x)] \overset{(10.1.2)}{=} [\sigma'(x).g(x), {}^{\sigma}\psi(x)] = [\sigma'(x), g(x).{}^{\sigma}\psi(x)] \tag{10.1.3}$$

which, by $[p, v] = [p, w] \Rightarrow v = w$, implies the important *transformation law for the family $\psi = ({}^{\sigma}\psi)_{\sigma}$ under change of gauge*

$$\boxed{{}^{\sigma'}\psi(x) = g(x).{}^{\sigma}\psi(x).} \tag{10.1.4}$$

Furthermore, this calculation shows that given any family $\psi = \{{}^{\sigma}\psi\}_{\sigma}$ of smooth V-valued functions ${}^{\sigma}\phi$ (defined on the domain of σ) for each choice of gauge (local section) $\sigma : M \dashrightarrow P$, they piece together consistently to form a matter field $\Psi : M \dashrightarrow P \times_G V$ defined on the union of domains of the σ's, by the formula (10.1.1).

[4] The domain of definition of ${}^{\psi}$ is the intersection of the domains of Ψ and σ.

10.2 Homework #10

...

Lecture 11

Principal Bundles II: Gauge Fields

11.1 Derivatives of Matter Fields

As a first naïve attempt for the partial derivative of Ψ, let us consider the family

$$\partial_\mu \psi = \{\partial_\mu({}^\sigma\psi)\}_\sigma.$$

To see whether this family piece together to form a new vector bundle section "$\partial_\mu\Psi$" we must check if the components transform correctly under gauge transformations. Let σ, σ' be two choices of gauge related by a gauge transformation $g(x)$ as in (10.1.4). We have [1]

$$\begin{aligned}\partial_\mu({}^{\sigma'}\psi) &= \partial_\mu\big(g(x).{}^\sigma\psi(x)\big) \\ &= \big[g_*(x)\cdot\partial_\mu\big].{}^\sigma\psi(x) + g(x).\big[\partial_\mu{}^\sigma\psi(x)\big]\end{aligned}$$

The first term is an unwanted "error term". Had it been zero, we would have been done. To remedy the situation, we attempt to add a *correction term*, and see what transformation law the correction term needs to have in order for unwanted terms to cancel upon a gauge transformation. Since the error term involves the action of the Lie algebra basis vectors T_a it is natural to make the following Ansatz. We define a new operator D_μ on families ψ by

$$D_\mu\psi = \{{}^\sigma(D_\mu\psi)\}_\sigma \tag{11.1.1}$$

where the components are defined by

$${}^\sigma(D_\mu\psi) = \partial_\mu({}^\sigma\psi)(x) + {}^\sigma A_\mu{}^a(x)T_a.{}^\sigma\psi(x). \tag{11.1.2}$$

[1] We use the notational convention that differentiation has priority over multiplication when semantically interpreting an expression: $\partial_\mu\varepsilon\phi$ means $(\partial_\mu\varepsilon)\cdot\phi$.

We have, dropping x from the notation, and abbreviating $\psi = {}^{\sigma}\psi$, $\psi' = {}^{\sigma'}\psi$; similarly put $A_\mu{}^a = {}^{\sigma}A_\mu{}^a$, $A'_\mu{}^a = {}^{\sigma'}A_\mu{}^a$:

$$\begin{aligned} {}^{\sigma'}(D_\mu\psi) &= \partial_\mu\psi' + A'_\mu{}^a T_a.\psi' \\ &= (g_* \cdot \partial_\mu).\psi + g.\partial_\mu\psi + A'_\mu{}^a T_a.(g.\psi). \end{aligned} \tag{11.1.3}$$

We want this to equal

$$g.{}^{\sigma}(D_\mu\psi) = g.\partial_\mu\psi + A_\mu{}^a g.(T_a.\psi). \tag{11.1.4}$$

Equating (11.1.3) and (11.1.4), a sufficient condition is that

$$(g_* \cdot \partial_\mu).\psi + A'_\mu{}^a T_a.(g.\psi) = A_\mu{}^a g.(T_a.\psi) \tag{11.1.5}$$

or, equivalently, replacing ${}^{\sigma}\psi$ by $g^{-1}.{}^{\sigma}\psi$ and using that $g.(X.(g^{-1}.v)) = \mathfrak{Ad}_g(X).v$ for $v \in V$, $h \in G, X \in \mathfrak{g}$,and restoring x,

$$\boxed{{}^{\sigma'}A_\mu{}^a(x)T_a = {}^{\sigma}A_\mu{}^a(x)\,\mathrm{Ad}_{g(x)*}(T_a) - r_{g(x)^{-1}*}g_*(x)\cdot\partial_\mu} \tag{11.1.6}$$

Putting ${}^{\sigma}\boldsymbol{A}_\mu = {}^{\sigma}A_\mu{}^a T_a$ and $\boldsymbol{\varepsilon} = \varepsilon^a T_a$, the identity can be written

$$\boxed{{}^{\sigma'}\boldsymbol{A}_\mu(x) = \mathrm{Ad}_{g(x)*}\,{}^{\sigma}\boldsymbol{A}_\mu(x) - r_{g(x)^{-1}*}g_*(x)\partial_\mu} \tag{11.1.7}$$

Another form of this is obtained if we put $\boldsymbol{A} = \boldsymbol{A}_\mu dx^\mu$, and observe that $-r_{g^{-1}*}g_* = \ell_g g_*^{-1}$:

$$\boxed{\boldsymbol{A}' = (\mathrm{Ad}_g)^*\boldsymbol{A} + \ell_{g*}dg^{-1}.} \tag{11.1.8}$$

If we further assume G is a matrix group:

$$\boxed{\boldsymbol{A}' = g\boldsymbol{A}g^{-1} + gdg^{-1}.} \tag{11.1.9}$$

Definition 11.1.1. Let $\pi : P \to M$ be a principal G-bundle on M. A *Yang-Mills (YM) field* $\boldsymbol{A}_\mu$ on M is a collection $\left({}^{\sigma}\boldsymbol{A}_\mu\right)_\sigma$ of $\mathfrak{g}$-valued 1-forms, one for each section $\sigma : M \dashrightarrow P$, such that if two sections are related by a gauge transformation via $\sigma(x) = \sigma'(x).g(x)$ for x in the overlap of domains of σ, σ', then (11.1.7) holds.

Definition 11.1.2. Let $\pi : P \to M$ be a principal G-bundle on M and fix a YM field $\boldsymbol{A}_\mu$. Let V be a representation of G and $\pi_V : P \times_G V \to M$ the associated vector bundle. The *covariant derivative* of a section $\Psi : M \dashrightarrow P \times_G V$ is defined by (11.1.2).

11.2 Geometric Interpretation (optional)

An *Ehresmann connection* on a principal bundle P is a smooth distribution $(H_p)_p$, $H_p \subseteq T_p(P)$ such that $T_p(P) = H_p \oplus V_p$, which is G-equivariant in the sense that $r_{g*}(p)H_p \subseteq H_{p.g}$ for all $g \in G$, where $r_g : P \to P$ denotes the right action of g on P and r_{g*} is its differential. Here $V_p \subseteq T_p(P)$ denotes the vertical subspace, defined as the image of $\ell_{p*} : \mathfrak{g} = T_e(G) \to T_p(P)$ where $\ell_p : G \to P$ is "left multiplication by p" i.e. $g \mapsto p.g$.

Let $p \in P$ and $x = \pi(p)$. The restriction of π_* to H_p is then an isomorphism

$$\pi_*\big|_{H_p} : H_p \xrightarrow{\cong} T_x(M). \tag{11.2.1}$$

We let

$$e_\mu(p) \in H_p \tag{11.2.2}$$

be the unique tangent vector that maps to $\partial_\mu(x)$ under π_*.

Together, we have a basis for $T_p(P)$:

$$e_M = (e_\mu, e_a). \tag{11.2.3}$$

Given a section $\sigma : M \dashrightarrow P$ we may push tangent vectors ∂_μ on M to P. We define

$${}^\sigma\partial_\mu = \sigma_*(\partial_\mu) = d\sigma_x(\partial_\mu(x)) \in T_pP, \quad p = \sigma(x). \tag{11.2.4}$$

Another way to think about this is that ${}^\sigma\partial_\mu$ is nothing but $\partial_\mu\sigma$, the derivative of the map σ. Indeed, if $\gamma : \mathbb{R} \dashrightarrow M$ with $\gamma(0) = x$, $\dot\gamma(0) = \partial_\mu$ then

$$(\partial_\mu\sigma)(x) = \frac{d}{dt}\Big|_{t=0}\sigma(\gamma(t)) = \sigma_* \cdot \partial_\mu = {}^\sigma\partial_\mu. \tag{11.2.5}$$

We may decompose these pushed tangent vectors in horizontal and vertical components. Since ${}^\sigma\partial_\mu$ and e_μ both map to ∂_μ under π_*, their difference is in the kernel of π_* which equals V_p. Thus there are coefficients ${}^\sigma A_\mu{}^a(x)$ such that

$${}^\sigma\partial_\mu(p) = e_\mu(p) + {}^\sigma A_\mu{}^a(x)\, e_a(p), \quad p = \sigma(x), x \in \operatorname{dom}\sigma. \tag{11.2.6}$$

We wish to see how the coefficients transform under change of gauge. If $\sigma' : M \dashrightarrow P$ is another choice of gauge, there is a unique $g(x) \in G$, defined for $x \in \operatorname{dom}\sigma \cap \operatorname{dom}\sigma'$, such that

$$\sigma(x) = \sigma'(x).g(x) \tag{11.2.7}$$

Fix $x \in \operatorname{dom}\sigma \cap \operatorname{dom}\sigma'$ and put $p = \sigma(x), p' = \sigma'(x), g = g(x)$. Let $\gamma : \mathbb{R} \dashrightarrow M$ be a curve through x in the direction of ∂_μ. Then we have

$$\begin{aligned} {}^{\sigma}\partial_\mu(p) &= \frac{d}{dt}\Big|_{t=0}\sigma\big(\gamma(t)\big) \\ &= \frac{d}{dt}\Big|_{t=0}\sigma'\big(\gamma(t)\big).g\big(\gamma(t)\big) \end{aligned}$$

Introduce $u = t$, $v = t$ and use the chain rule to get

$$\begin{aligned} \frac{d}{dt}\Big|_{t=0}\sigma'\big(\gamma(t)\big).g + \frac{d}{dt}\Big|_{t=0}p'.g\big(\gamma(t)\big) &= r_{g*}\circ\sigma'_*\cdot\partial_\mu + \ell_{p'*}\circ g_*\cdot\partial_\mu \\ &= r_{g*}\big({}^{\sigma'}\partial_\mu\big) + \ell_{p'*}\partial_\mu g \end{aligned}$$

Furthermore, $\ell_{p'} = \ell_{p.g^{-1}} = \ell_p \circ \ell_{g^{-1}}$ so, abusing notation slightly,

$${}^{\sigma}\partial_\mu(p) = r_{g*}\big({}^{\sigma'}\partial_\mu\big) + \ell_{p*}(g^{-1}\partial_\mu g) \tag{11.2.8}$$

We thus have

$$\begin{aligned} e_\mu(p) + {}^{\sigma}A_\mu{}^a(x)\, e_a(p) &\overset{(11.2.6)}{=} {}^{\sigma}\partial_\mu(p) \\ &\overset{(11.2.8)}{=} r_{g*}\big({}^{\sigma'}\partial_\mu\big) + \ell_{p*}(g^{-1}\partial_\mu g) \\ &\overset{(11.2.6)}{=} r_{g*}\big(e_\mu(p') + {}^{\sigma'}A_\mu{}^a(x)\, e_a(p')\big) + \ell_{p*}(g^{-1}\partial_\mu g) \\ &= e_\mu(p) + {}^{\sigma'}A_\mu{}^a r_{g*}e_a(p') + \ell_{p*}g^{-1}(\partial_\mu g) \end{aligned}$$

Canceling $e_\mu(p)$, and applying $(\ell_{p*})^{-1}$ to both sides, we get an identity in $\mathfrak{g}$ (using that $\ell_{p*}(T_a) = e_a(p)$):

$${}^{\sigma}A_\mu{}^a T_a = {}^{\sigma'}A_\mu{}^a\, \mathrm{Ad}_{g^{-1}{}*}(T_a) + g^{-1}\partial_\mu g \tag{11.2.9}$$

Applying Ad_{g*} on both sides (or replacing g by g^{-1} and σ by σ') we get:

$$\boxed{{}^{\sigma'}A_\mu{}^a T_a = {}^{\sigma}A_\mu{}^a\, \mathrm{Ad}_{g*}(T_a) + g\partial_\mu g^{-1}.} \tag{11.2.10}$$

This is exactly the same transformation law as for the Yang-Mills fields, as derived in the previous section.

11.3 Homework #11

Lecture 12

Spinors

12.1 The Concept of Spin

Recall that in Lorentzian geometry, our coordinate transformations are generated by Lorentz transformations (elements of the group $\mathrm{SO}^+(1,3)$, S means determinant one, $+$ means ${\Lambda_0}^0 > 0$) along with spacetime translations (making up the additive group $\mathbb{R}^4$ of 4-vectors). An element of the Lorentz group can be expanded as

$$\Lambda_\mu{}^\nu = \delta_\mu^\nu + \lambda_\mu{}^\nu + \cdots$$

Substituting this into the Lorentz group condition $\Lambda_\mu{}^\nu \Lambda_\lambda{}^\rho \eta_{\nu\rho} = \eta_{\mu\lambda}$ and keeping only terms linear in $\lambda_\mu{}^\nu$ we obtain

$$\lambda_{\mu\nu} = -\lambda_{\nu\mu} \tag{12.1.1}$$

where used lowering notation $\lambda_{\mu\nu} = \lambda_\mu{}^\tau \eta_{\tau\nu}$. Thus we have six independent parameters $\lambda_{\mu\nu}$ ($\mu < \nu$). This illustrates that $\mathrm{SO}^+(1,3)$ is 6-dimensional.

The group $\mathbb{R}^4$ is of course 4-dimensional.

Therefore, if we have a Lagrangian which is invariant under the above symmetries (that is, it is a scalar field with respect to this Lorentzian geometry), then we have a corresponding set of 10 Noether currents $\Sigma_{\mu\nu}{}^\lambda = -\Sigma_{\nu\mu}{}^\lambda$ and $T_\mu{}^\nu$.

We encountered the energy-momentum tensor $T_\mu{}^\nu$ previously.

Let $M_{\mu\nu}$ and P_μ be the associated conserved charges:

$$M_{\mu\nu} = \int \Sigma_{\mu\nu}{}^0 \, d^3x, \qquad P_\mu = \int T_\mu{}^0 \, d^3x$$

P_μ represents 4-momentum (so P_0 is energy and P_i is usual 3-momentum $i = 1, 2, 3$), the quantities M_{ij} are called angular momenta and M_{0i} are boost charges.

As with all charges, a priori they still depend on the time (x^0) coordinate, because we only integrated over space.

However, by Noether's theorem (coordinate transformation version), along each solution to the (Euler-Lagrange) field equations, these quantities are constant. When we say "along" we mean take a set of fields $\varphi^{(a)}$ which solve the field equations. Substitute them into the expression for the conserved currents (such expressions involve the lagrangian and hence the field components), integrate the result over space to get the charge. Then, this resulting charge is constant in time. It is a number associated to each solution to the field equations.

It is worth pointing out again that different regions can exchange charge, but the overall charge remains constant.

For example, the charge M_{12} represents angular momentum around the z-axis.

Note that the charges $M_{\mu\nu}$ and P_μ transform as covariant tensors under Lorentz transformations. Therefore M_{12} does not have independent meaning. However, we can combine them to form Lorentz scalars which do have coordinate independent meaning.

Definition 12.1.1.

(1) The *invariant mass squared* is the Lorentz-invariant quantity

$$m^2 = P_\mu P^\mu \tag{12.1.2}$$

(2) The *spin 4-(co)vector* S_μ is defined by

$$S_\mu = \varepsilon_{\mu\nu\lambda\rho}\Sigma^{\nu\lambda}P^\rho. \tag{12.1.3}$$

(3) If $m^2 \neq 0$, the *helicity* h is defined by .....

$$..... \tag{12.1.4}$$

Lemma 12.1.2. *Suppose we have a Lorentzian field theory.*

(i) *If $m^2 \neq 0$ then there is a non-negative integer or half integer $s \in \frac{1}{2}\mathbb{Z}_{\geq 0}$ such that*

$$-\frac{1}{2}m^2 s(s+1) = S_\mu S^\mu \tag{12.1.5}$$

(ii) *If $m^2 = 0$ then there is a similar identity relating s to .....*

Definition 12.1.3. The quantity s is known as *spin*.

There is now a general theorem that says that the irreducible representations of the Poincaré group fall into two types: The *massive* ones are those where $m^2 \neq 0$. They are characterized by (m^2, s). The *massless* irreps need one more piece of data, the *helicity* $h \in \pm 1$. Thus (s, h) characterize the massless irreps of the Poincaré group.

In a scalar field theory described by the Klein-Gordon Lagrangian $\mathscr{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2$, one can verify that the invariant mass squared is in fact equals to the parameter appearing in the equation. Furthermore, for this theory the spin is zero.

In a vector field theory (like Maxwell), the spin will be $s = 1$.

It turns out that to properly describe matter fields, we need to come up with a spin 1/2 theory.

In supergravity, the fundamental field is the *gravitini*, the super-partner of the graviton. The gravitini is a field of spin 3/2.

(There is a theorem that says that 3/2 is the highest possible spin occuring in a finitely generated quantum field theory.)

12.2 Clifford Algebra

To motivate the Clifford algebra, we follow an often used line of reasoning. We work in n spacetime dimensions with a flat metric $\eta_{\mu\nu}$ of any signature.

We know that the wave equation $\Box\phi = 0$ plays a fundamental role in field theory. Scalar fields obey it, as do vector fields (in the Lorentz gauge).

Rougly speaking, the idea is to take the square root of the d'Alembertian $\Box$. Since $\Box$ is a constant coefficient second order PDE, we could try to find numbers $\boldsymbol{\gamma}^\mu$ such that

$$(\boldsymbol{\gamma}^\mu\partial_\mu)^2 = \Box$$

Since $\Box = \partial_\mu\partial^\mu = \eta^{\mu\nu}\partial_\mu\partial_\nu$, this equation can be rewritten

$$\boldsymbol{\gamma}^\mu\boldsymbol{\gamma}^\nu\partial_\mu\partial_\nu = \eta^{\mu\nu}\partial_\mu\partial_\nu$$

It seems that we want

$$\boldsymbol{\gamma}^\mu\boldsymbol{\gamma}^\nu = \eta^{\mu\nu} \tag{12.2.1}$$

Unfortunately, this doesn't work: if we take $\eta^{\mu\nu}$ to be diagonal and $n > 1$, then we get for example $\boldsymbol{\gamma}^1\boldsymbol{\gamma}^2 = 0$ so either $\boldsymbol{\gamma}^1 = 0$ or $\boldsymbol{\gamma}^2 = 0$. But this would contradict that $\boldsymbol{\gamma}^\mu\boldsymbol{\gamma}^\mu = \eta^{\mu\mu} = \pm 1$ (no summations).

The problem here is that a product of complex numbers can only be zero if one of the factors is zero. The idea arises to enlarge our search domain. What if $\boldsymbol{\gamma}^\mu$

were matrices (or more generally, some elements of some possibly non-commutative algebra)? Going back to (12.2.1), we observe that the right hand side is symmetric in μ and ν. Thus, we may subtract away the anti-symmetric part from the left hand side, and what remains is the *Clifford algebra relations*

$$\frac{1}{2}(\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu) = \eta^{\mu\nu}\mathbf{1}. \tag{12.2.2}$$

We have interpreted the right hand side to be a multiple of the identity element of the algebra.

Definition 12.2.1. The abstract (real) *Clifford algebra* $\mathrm{Cl}(\eta)$ associated to the metric $\eta_{\mu\nu}$ is defined to be the (real) algebra generated by n symbols $\gamma^1, \gamma^2, \ldots, \gamma^n$ subject to the defining relations (12.2.2). A set of n square matrices satisfying those relation is said to be a (linear) *representation* of the Clifford algebra.

Thus, in fact, it is possible to take the square root of the d'Alembertian $\Box$, provided we can find a representation of the Clifford algebra.

Example 12.2.2. Let us try this when $n = 2$. That is, $1 + 1$ dimensions. Choose $\eta_{\mu\nu} = \mathrm{diag}(+1, -1)$. We must find two square matrices γ^0 and γ^1 such that

$$\gamma^0\gamma^0 = \mathbf{1}, \qquad \gamma^0\gamma^1 = -\gamma^1\gamma^0, \qquad \gamma^1\gamma^1 = -\mathbf{1}. \tag{12.2.3}$$

Because they can't be 1×1-matrices, it makes sense to look for 2×2-matrices. But, even without this assumption, note that the matrix $\widetilde{\gamma} = \gamma^0\gamma^1$ squares to the identity matrix. To make progress, assume that e_1 is an eigenvector to $\widetilde{\gamma}$ with eigenvalue $+1$. Furthermore, $\widetilde{\gamma}$ anti-commutes with both γ^0 and γ^1. Therefore $\gamma_0 e_1$ is an eigenvector of $\widetilde{\gamma}$ with eigenvalue -1. Put $e_2 = \gamma^0 e_1$. Now it turns our that the products $\gamma^\mu e_\alpha$ are completely determined by the relations and the $\widetilde{\gamma}$-eigenvalue properties we have. The result is always a linear combination of e_1 and e_2. (Check this!) Thus, restricting to thie subspace and writing γ^μ in this basis we find

$$\gamma^0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \gamma^1 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{12.2.4}$$

and $\mathbf{1}$ of course the 2×2 identity matrix. Note that $\widetilde{\gamma} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

In n spacetime dimensions, one can always find a representation of the Clifford algebra by some $N \times N$-matrices. For given n, the smallest possible N one can have

is $N = 2^{\lfloor n/2 \rfloor}$. (Here $\lfloor x \rfloor$ is the floor function.) For example, in $n = 11$ dimensions we must deal with eleven 32×32-matrices. This might sound strange but it is not so strange when you notice that the elements

$$\{\boldsymbol{\gamma}^{i_1}\boldsymbol{\gamma}^{i_2}\cdots\boldsymbol{\gamma}^{i_m} \mid 0 \leq m \leq n,\, 1 \leq i_1 < i_2 < \cdots < i_m \leq n\} \tag{12.2.5}$$

span the algebra. This is because generators anti-commute and square to ± 1. So any concievable product of the generators can be first reordered and then simplified down to an ordered product of distinct generators.

In fact with a little more work one can prove that this set is a basis for the Clifford algebra.

Since each generator is either present or absent from a given basis element, there are 2^n basis vectors for $\mathrm{Cl}(\eta)$.

There is a structure theorem for these Clifford algebras which says that when n is even, any representation is faithful which means the smallest possible matrix algebra that could contain it would have to be matrices of size $N \times N$, where $N = 2^{n/2}$, because the dimension of that space of matrices is exactly 2^n. And indeed such a representation exists, and is unique up to equivalence. (It is necessarily irreducible.)

When n is odd, things become a bit more annoying. There is now a faithful representation of the Clifford algebra using block diagonal matrices with two blocks, each of size $N \times N$ where $N = 2^{(n-1)/2}$. Up to equivalence one can arrange things so that $\widetilde{\boldsymbol{\gamma}} = \begin{bmatrix} \mathbf{1}_N & \\ & -\mathbf{1}_N \end{bmatrix}$ with off-diagonal entries zero. In this representation, the first N basis vector span an irreducible subrepresentation in which $\widetilde{\boldsymbol{\gamma}} = \mathbf{1}_N$. Similarly, the last N basis vectors span a *different (inequivalent)* irreducible representation. This all is related to the fact that when n is odd, the chirality operator $\widetilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^1\boldsymbol{\gamma}^2\cdots\boldsymbol{\gamma}^n$ commutes with all the generators forcing it to be constant on an irreducible representation (by Schur's Lemma from representation theory).

The main take-away here is that such matrices exist, and are more or less unique. For $n = 4$ they have size $N \times N$ where $N = 2^{\lfloor n/2 \rfloor} = 4$ and are known as *Dirac gamma matrices*.

12.3 Dirac Spinors

We can now write down the *Dirac equation*

$$i\boldsymbol{\gamma}^\mu \partial_\mu \psi + m\psi = 0. \tag{12.3.1}$$

Here $\boldsymbol{\gamma}^\mu$ is an n-tuple of $N \times N$-matrices (for appropriate N) and ψ is a column vector

$$\psi = \psi^\alpha = \begin{bmatrix} \psi^1 \\ \psi^2 \\ \vdots \\ \psi^N \end{bmatrix} \tag{12.3.2}$$

of complex-valued functions on M (or, equivalently, defined on the image of the coordinate system).

We call α a *spinor index*. It runs from 1 to N. We will reserve the beginning of the greek alphabet $\alpha, \beta, \ldots$ for spinor indices, and continue to use the middle of the alphabet $\mu, \nu, \ldots$ for spacetime indices.

Supposing that ψ is a solution to the Dirac equation, and acting on (12.3.1) from the left by the operator $(-i\boldsymbol{\gamma}^\mu \partial_\mu + m\mathbf{1}_N)$ and using the Clifford algebra relations (12.2.2), we obtain that each component ψ^α satisfies the Klein-Gordon equation $\partial_\mu \partial^\mu \psi = -m^2 \psi$. Consequently, the parameter m should be identified with mass.

In order for the Dirac equation to be coordinate independent, the components ψ^α must transform in a certain way. Consider a Lorentz coordinate transformation

$$x^\mu \mapsto x^{\mu'} = \Lambda_\nu{}^{\mu'} x^\nu \tag{12.3.3}$$

Let $\psi' = \psi^{\alpha'}$ be the components expressed in the new coordinate system. We assume that

$$\psi' = S(\Lambda)\psi \tag{12.3.4a}$$

where $S(\Lambda)$ is some invertible $N \times N$-matrix associated to the Lorentz transformation Λ (which is an invertible $n \times n$-matrix). In component form we may write this as

$$\psi^{\alpha'} = S_\beta{}^{\alpha'}(\Lambda)\psi^\beta \tag{12.3.4b}$$

where α, β run over $1, 2, \ldots, N$.

After the Lorentz transformation, the Dirac equation becomes (abbreviating $S(\Lambda)$ by S):

$$\begin{aligned} i\boldsymbol{\gamma}^{\mu'} \partial_{\mu'} \psi' + m\psi' &= i\boldsymbol{\gamma}^{\mu'} \Lambda_{\mu'}{}^\nu \partial_\nu S + mS\psi \\ &= S\big(iS^{-1}\boldsymbol{\gamma}^{\mu'} S \Lambda_{\mu'}{}^\nu \partial_\nu + m\psi\big) \end{aligned}$$

where $\Lambda_{\mu'}{}^\nu$ denotes the entries of the inverse transformation. Thus, we see that the Dirac equation itself transforms by a factor of S (and therefore is zero in one coordinate system if and only if it is zero in any other) provided that

$$S(\Lambda)^{-1} \cdot \boldsymbol{\gamma}^{\mu'} \cdot S(\Lambda) = \Lambda_\nu{}^{\mu'} \boldsymbol{\gamma}^\nu \tag{12.3.5}$$

We emphasize here that *the matrices $\boldsymbol{\gamma}^\mu$ should be the same in all coordinate systems*. This is becauase the metric $\eta_{\mu\nu}$ is the same in all coordinate systems: The coordinate transformations by definition have to preserve the metric.

There is a technical issue with solving the above equation for $S(\Lambda)$: it is defined only up to sign. One can show that as Λ traverses a loop inside the Lorentz group, $S(\Lambda)$ might have changed sign.

One can ignore this problem and only consider small (infinitesimal) Lorentz transformations. Or, we are forced to abandon the whole structure of Lorentzian spacetime in which coordinate transformations give rise to representations of $\mathrm{SO}(\eta)$ and instead start from scratch, replacing $\mathrm{SO}(\eta)$ by its double cover denoted $\mathrm{Spin}(\eta)$ called the spin group.

It is possible to do this in a framework of principal $\mathrm{Spin}(\eta)$-bundles and a version of the Vielbein to help facilitate the relation to the old Lorentzian tangent vectors.

We may return to this issue in a future lecture.

For now, we take the first route, and define spinor fields as follows.

Definition 12.3.1. Given a representation of the Clifford algebra by $N \times N$-matrices and given a representation $S(\Lambda)$ of the Lorentz group (near identity) satisfying (12.3.5), a *Dirac spinor field* is an N-tuple ψ^α of complex functions on M, given in each coordinate system, which transform according to (12.3.4) under Lorentz transformations (12.3.3).

Spinors are a new kind of field we have not encountered before. We can add this to our previous arsenal of (relative) tensor fields and consider mixed objects such as ψ^α_μ which carries both a Lorentz index μ and a spinor index α. We can have any number of upper and lower spinor indices.

The Dirac equation, like every field equation we consider, arises as the Euler-Lagrange equation of a Lagrangian:

$$\mathscr{L} = \overline{\psi}(i\boldsymbol{\gamma}^\mu \partial_\mu + m)\psi \tag{12.3.6}$$

where $\overline{\psi} = \psi^\dagger \boldsymbol{\gamma}^0$. One verifies that the expression $\overline{\psi}\phi$ is sesquilinear and Lorentz-invariant, while $\overline{\psi}\boldsymbol{\gamma}^\mu\phi$ is a Lorentz vector, whenever ψ, ϕ are Dirac spinor fields. Therefore the Dirac lagrangian is a Lorentz scalar (density) as required. Note that, just like for complex scalars, ψ and $\overline{\psi}$ should be considered independent fields. It is trivial to check that the Euler-Lagrange equation for $\overline{\psi}$ is the Dirac equation. (The field equation for ψ is equivalent.)

All matter fields in nature are actually spinors: They carry a spinor index in addition to any other index. For example, in electrodynamics, the electron field $\psi = \psi^\alpha$ is a Dirac spinor field, in addition to being a $U(1)$-matter field.

12.4 Weyl Spinors

To discuss the electroweak sector of the standard model we first need to address the notion of chirality.

[To be added.....]

12.5 Homework #12

1. Show that infinitesimal version of the condition (12.3.5) obtained by substituting $S = 1 + s\varepsilon + \mathcal{O}(\varepsilon^2)$, and $\Lambda = 1 + \lambda\varepsilon + \mathcal{O}(\varepsilon^2)$ is $[\boldsymbol{\gamma}^{\mu'}, s] = \lambda_\nu{}^{\mu'}\boldsymbol{\gamma}^\nu$. Then show that $s = s(\lambda) = \frac{-1}{8}[\boldsymbol{\gamma}^\mu, \boldsymbol{\gamma}^\nu]\lambda_{\mu\nu}$ provides a solution to that equation.

2. Let $E^i{}_j$ denote the $n \times n$-matrix with zeroes everywhere except in row i and column j where it is 1. The Lie algebra $\mathfrak{so}(\eta)$ is defined as the space of matrices $\lambda_i{}^j E^i{}_j$ which satisfy $\lambda_i{}^k\eta_{kj} + \lambda_j{}^k\eta_{ki} = 0$. Put $\omega_{ij} = E_{ij} - E_{ji}$ where $E_{ij} = E_i{}^k\eta_{kj}$. Show that there is a Lie algebra map from $\mathfrak{so}(\eta)$ to $\mathrm{Cl}(\eta)$, given by $\omega_{ij} \mapsto \frac{1}{2}[\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j]$.

3. (Charge conjugation) Find conditions on a matric C which would ensure that if ψ solves the Dirac equation then so does $C\psi^*$ where $(\psi^*)^\alpha = (\psi^\alpha)^*$ is the complex conjugate of each spinor component.

Lecture 13

The Standard Model

13.1 Toy Models

13.1.1 $(1+0)$-Dimensional Theory

In $1+0$ dimensions, the signature of η is $(-)$ and the real Clifford algebra $\mathrm{Cliff}(\eta)$ has one generators γ_0 with relations $\gamma_0^2 = -1$. In other words, $\mathrm{Cliff}(\eta) = \mathbb{C}$.

In this extremely simplified model, we consider a $U(1)$-gauge theory with a fermion χ and a gauge field $A_\mu = A_0 = A$. Note the field strengths $F_{00} = 0$. Letting $\bar{\chi} = -i\chi^*$:

$$\mathscr{L} = \bar{\chi}\gamma^0 D_0\chi = \bar{\chi}i(\partial_0 + qA_0)\chi = \chi^*\partial\chi + q\chi^*\chi A \tag{13.1.1}$$

This lagrangian is invariant under local gauge transformations $\chi \mapsto \exp(i\varepsilon)\chi$, $A \mapsto A - \frac{1}{q}i\partial\varepsilon$ where $\varepsilon = \varepsilon(x^0)$ is a function.

13.1.2 Charged Scalar Field Case

Suppose we have a charged scalar ϕ, i.e. a complex-valued scalar field $\phi = \phi(x) = \frac{1}{\sqrt{2}}(\varphi_1(x) + \varphi_2(x)i)$, where ϕ_1 and ϕ_2 are smooth real-valued functions defined on the spacetime manifold M, which we again take to be a pseudo-Riemannian manifold $(M, g_{\mu\nu})$ where $g = g_{\mu\nu}dx^\mu \wedge dx^\nu$ is the metric – a symmetric, non-degenerate smooth two-form; possibly indefinite.

The standard lagrangian for such a field reads[1]

$$\begin{aligned}\mathscr{L} &= \partial_\mu \phi^* \partial^\mu \phi - V(\phi^* \phi) \\ &= \frac{1}{2}\partial_\mu \varphi_1 \partial^\mu \varphi_1 + \frac{1}{2}\partial_\mu \varphi_2 \partial^\mu \varphi_2 - V(\phi^* \phi). \end{aligned} \tag{13.1.2}$$

Here we have included a potential term involving a smooth real-valued function $V : \mathbb{R} \to \mathbb{R}$, to be specified later. Note that, in general, we should be using the covariant partial derivative ∇_μ, to ensure the lagrangian is invariant under general coordinate transformations. However, in the case of a scalar field ϕ, $\partial_\mu \phi$ already transforms as a covariant vector (this is because ϕ takes values in the trivial representation of $\mathfrak{gl}_n$)

We observe that the lagrangian $\mathscr{L}$ is invariant under the global (meaning x-independent) transformation

$$\phi \mapsto e^{i\theta}\phi \tag{13.1.3}$$

for any real constant $\theta \in \mathbb{R}$.

As we have seen, in order to upgrade this global symmetry to a local gauge symmetry, we must introduce a gauge (connection) field A_μ counteracting the error term coming from the product rule for derivatives, and replace ∂_μ by the $U(1)$-covariant partial derivative $D_\mu = \partial_\mu + A_\mu$. The upgraded lagrangian then reads

$$\mathscr{L}' = D_\mu \phi^* D^\mu \phi - V(\phi^* \phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{13.1.4}$$

where we also included the kinetic term for the new gauge field A_μ. We remark again that substituting $D_\mu = \partial_\mu + A_\mu$ and expanding, we get terms that couple the charged scalar field ϕ to the gauge field A_μ:

$$\mathscr{L}' = \partial_\mu \phi^* \partial^\mu \phi + A_\mu \phi^* \partial_\mu \phi + \partial_\mu \phi^* A_\mu \phi + \phi^* \phi A_\mu A^\mu - V(\phi^* \phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{13.1.5}$$

A common choice for the function V is

$$V(\phi^* \phi) = \frac{1}{2} m^2 \phi^* \phi \tag{13.1.6}$$

where $m \geq 0$ is a real number. For this choice of V, m is interpreted as the *mass* of the scalar field ϕ. In this case, a small perturbation of the trivial solution $\phi \equiv 0$ yields a wave-like solution oscillating around the value 0, i.e. it has expectation value zero.

[1] With these conventions we have $\partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi^*)} = \partial_\mu \partial^\mu \phi$ while $\partial_\mu \frac{\partial \mathscr{L}}{\partial(\partial_\mu \varphi_j)} = \partial_\mu \partial^\mu \varphi_j$ for $j = 1, 2$.

Also, the energy is minimized for the solution $\phi = 0$ (constant).

The point is that, when nothing is going on, $\phi = 0$. By pluggin in $\phi = 0$ into the lagrangian, we delete all terms involving ϕ and we obtain a theory with a single photon-field A_μ and its equations of motions are Maxwell's equations in vacuo. As we know, this theory still has local $U(1)$-gauge invariance.

To obtain the phenomenon of spontaneous symmetry-breaking, we take the potential V to be

$$V(r^2) = \frac{\lambda^2}{2}(r^2 + \frac{\mu^2}{2\lambda^2})^2 = \frac{\lambda^2}{2}(r^2 - \frac{v^2}{2})^2 = \frac{\lambda^2}{2}r^4 + \frac{\mu^2}{2}r^2 + \text{const} \tag{13.1.7}$$

where $\lambda^2 > 0$ but $\mu^2 < 0$, and the constant may be neglected (it disappears in the Euler-Lagrange equations), and $v = \sqrt{\frac{-\mu^2}{\lambda^2}} \in \mathbb{R}$. Here λ is dimensionless, while v has mass-dimension 1, same as ϕ, φ_1 and φ_2.

Here the situation is different. A vacuum (lowest energy) state is now a constant function ϕ_0 satisfying $\phi_0^*\phi_0 = v^2/2$. Thus, up to a constant phase, $\phi_0 = v/\sqrt{2}$.

The key observation here is that ϕ_0 *is not invariant under* $U(1)$*-transformations even though the lagrangian is.* Substituting $\phi = \phi_0 = \frac{v}{\sqrt{2}}$ (constant) into the lagrangian (13.1.5), we obtain:

$$\mathscr{L}_0 = \frac{1}{2}v^2 A_\mu A^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{13.1.8}$$

We see that the gauge field A_μ has acquired a mass of v.

We can do a more detailed analysis by instead of forcing ϕ to be in a vacuum state, expand ϕ around this state:

$$\phi = \frac{v}{\sqrt{2}} + \varphi \tag{13.1.9}$$

Due to local $U(1)$-gauge invariance, we may replace ϕ by $\widetilde{\phi}(x) = e^{i\epsilon(x)}\phi(x)$ where $\epsilon(x)$ is an arbitrary smooth real-valued function on M, as long as we also replace A_μ by $A_\mu + \partial_\mu\epsilon(x)$. Let us choose $\epsilon(x)$ so that $\widetilde{\phi}(x)$ is real-valued for all $x \in M$.

(Recall that we already chose a gauge in which $\varphi_2 = 0$ so that φ is real-valued.) Substituting this into the lagrangian (13.1.5) we get

13.2 Standard Model of Particle Physics

In the standard model (SM) of particle physics,[2] there are 18 massless, chiral (i.e. each is either left- or right-handed), Dirac spinor fields denoted

$$\ell_L^a, \nu_R^a, e_R^a; \quad q_L^a, u_R^a, d_R^a, \qquad (a = 1, 2, 3)$$

where subscript L/R indicates handedness. The first 9 are called *leptons* and the last 9 are called *quarks*. The a labels the *generation*. In addition, there are three gauge fields

$$B_\mu, W_\mu, G_\mu$$

respectively taking values in the three irreducible summands of the adjoint representation of G_{SM} on its Lie algebra: $\mathfrak{u}(1)$, $\mathfrak{su}(2)$, $\mathfrak{su}(3)$ of dimensions $1, 3, 8$. The interactions between these 3 gauge fields and the matter fields give rise to what we perceive as "forces", particle attraction and repulsion and various decay processes.

Last, but not least, there is one scalar boson field denoted ϕ, called the *Higgs doublet*. It enters the SM lagrangian with a very special potential term and Yukawa couplings.

Local excitations in each of these 18+3+1=23 fields correspond to "particles".

At first, the description given so far seems far removed from what we see in particle accelerators.

Usually, the value of a field at a point is zero, signifying that nothing is going on there. However, the Higgs doublet ϕ, due to a special potential term in the Lagrangian, is locked into a state in which the value of the field at almost every point is *nonzero*. $\phi = 0$ is still a valid solution to the field equations, but it is an unstable solution, and shortly after the Big Bang, ϕ "fell down" down to a nonzero value. Replacing ϕ by this constant value in the Lagrangian breaks the gauge symmetry. This means we pass from the standard model gauge group G_{SM} to a certain subgroup G_{broken} isomorphic to $U(1) \times SU(3)$. This process is called *spontaneous electroweak symmetry breaking*. At high enough temperatures, such as in particle accelerators, one can observe the restoration of the full gauge group symmetry.

This symmetry breaking has far-reaching consequences and explains why leptons and quarks appear to us as if they have mass, why the three gauge bosons W^-, Z^0, W^+ (mediating the weak interaction) appear to have mass, thereby making the weak nuclear force a short-ranged force (as opposed to electromagnetism which is long-ranged due to the massless photon).

[2] For aesthetic reasons we include right-handed neutrinos, even though they cannot be (and have not been) detected since they take values in the trivial representation of the gauge group.

Let us write down the irreducible representation of G_{SM} for each of the 18 fermionic matter fields:

$$\ell_L^a : \quad \mathbb{C}_{-1} \otimes \mathbb{C}^2 \otimes \mathbb{C} \tag{13.2.1a}$$
$$\nu_R^a : \quad \mathbb{C}_0 \otimes \mathbb{C} \otimes \mathbb{C} \quad \text{(trivial representation)} \tag{13.2.1b}$$
$$e_R^a : \quad \mathbb{C}_{-2} \otimes \mathbb{C} \otimes \mathbb{C} \tag{13.2.1c}$$
$$q_L^a : \quad \mathbb{C}_{1/3} \otimes \mathbb{C}^2 \otimes \mathbb{C}^3 \tag{13.2.1d}$$
$$u_R^a : \quad \mathbb{C}_{4/3} \otimes \mathbb{C} \otimes \mathbb{C}^3 \tag{13.2.1e}$$
$$d_R^a : \quad \mathbb{C}_{-2/3} \otimes \mathbb{C} \otimes \mathbb{C}^3 \tag{13.2.1f}$$

and the single bosonic matter field (the Higgs doublet):

$$\phi : \quad \mathbb{C}_1 \otimes \mathbb{C}^2 \otimes \mathbb{C} \tag{13.2.1g}$$

Some things we can glean from this is that all the leptons are colorless, while quarks have color. In other words, leptons transform under the trivial representation of SU(3) which means they do not participate in the strong interaction. On the other hand, the quarks transform under the vector (aka standard, tautological, fundamental) representation $\mathbb{C}^3$ of SU(3) which means the quarks do participate in the strong interaction.

Another feature we see here is that the right-handed particles have no isospin because they transform under the trivial representation of SU(2). Only the Higgs doublet and the left-handed particles take part in the isospin interaction.

Lastly, we see that all particles, except the right-handed neutrinos, have nonzero hypercharge.

$$\begin{aligned}
\mathscr{L}_{\text{SM}} &= \bar{\Psi} i \boldsymbol{\gamma}^\mu D_\mu \Psi && (13.2.2)\\
&\quad - \frac{1}{4} F^k_{\mu\nu} F_k^{\mu\nu} && (13.2.3)\\
&\quad + \frac{1}{2} (D_\mu \phi)^\dagger D^\mu \phi && (13.2.4)\\
&\quad - V(\phi^\dagger \phi) && (13.2.5)\\
&\quad - \sum_{a=1}^{3} \Big(g_e^a \, \overline{e_R^a} \, \phi^\dagger \, \ell'^a_L + g_\nu^a \, \overline{\nu'^a_R} \, \phi^{\text{c}\dagger} \, \ell'^a_L \\
&\qquad + g_d^a \, \overline{d_R^a} \, \phi^\dagger \, q_L^a + g_u^a \, \overline{u_R^a} \, \phi^{\text{c}\dagger} \, q_L^a + \text{h.c.} \Big) && (13.2.6)
\end{aligned}$$

Here

- (13.2.2) is the kinetic term for the left- and right-handed Dirac spinors stacked together in a single multiplet

$$\Psi = \begin{pmatrix} \ell_L^1 \\ \nu_R^1 \\ e_R^1 \\ \vdots \\ q_L^1 \\ u_R^1 \\ d_R^1 \\ \vdots \end{pmatrix} \tag{13.2.7}$$

  Importantly, $D_\mu = \partial_\mu + \boldsymbol{A}_\mu$ is the covariant derivative, where

$$\boldsymbol{A}_\mu = ig_Y B_\mu 1 \oplus ig_L W_\mu^k T_k \oplus ig_c G_\mu^k \Lambda_k, \tag{13.2.8}$$

  and T_k (resp. Λ_k) a basis for the Lie algebra of $SU(2)$ (resp. $SU(3)$), and g_Y, g_L, g_c are coupling constants to hypercharge, isospin, and color charge respectively.

- (13.2.3) is the kinetic term for the gauge fields
- (13.2.6) is the *Yukawa coupling term* describing how the Higgs doublet couples to the lepton and quark matter fields. Here $\phi^c = \begin{pmatrix} -\phi_0^* \\ \phi_+^* \end{pmatrix}$ is the charge-conjugate of the Higgs doublet;

$$g_\nu^a,\ g_e^a,\ g_u^a,\ g_d^a, \quad a = 1, 2, 3 \tag{13.2.9}$$

  are 12 coupling constants describing the strengths of the interactions between the matter fields and the Higgs field. "+h.c." means we are to add the hermitian conjugate of the Yukawa terms (to ensure the lagrangian is real).

13.3 Spontaneous Symmetry Breaking

The so called U-gauge is the choice in which $\phi_+ = 0$ and ϕ_0 is real valued. Due to the potential, the vacuum expectation value (vev) is $v/\sqrt{2}$ and therefore we expand ϕ_0 around this (the $1/\sqrt{2}$ is just a common normalization) we have

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ h \end{pmatrix} \tag{13.3.1}$$

where h is a real-valued scalar field with zero vev. The higgs bosons are quanta in the field h. As a reminder, v is a constant (roughly 246 GeV).

Because ϕ has a nonzero vev, the field is constantly influencing all other fields (except gluons) through the various interaction terms. The U-gauge is just a convenient gauge to help us understand these interactions. As an example, let us consider the kinetic term for the higgs doublet after substituting (13.3.1) and ignoring terms involving h:

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \cdots\cdots = \frac{v^2 g_Y}{2} B_\mu B^\mu + v^2 g_L g_Y W^0_\mu B^\mu(-\frac{1}{2}) + \frac{g_L^2 v}{2}(W^-_\mu W^{\mu+} - W^0_\mu W^{\mu 0}) \tag{13.3.2}$$

where $W^\pm$, W^0 are the coefficients of the SU(2) generators $\sigma^\pm, \sigma^0 = \sigma^3$ (in mathematics known as $e, f, h/2$).

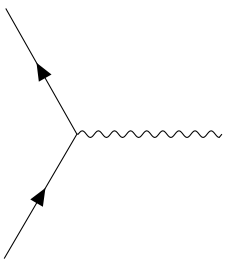
The point is that some of the $\mathrm{U(1)_Y} \times \mathrm{SU(2)_L}$ gauge bosons have become massive. More specifically, there is a unique way to diagonalize the quadratic form displayed. When doing so we find a mass terms given to three linear combinations called $W^+_\mu, W^-_\mu, Z^0_\mu = \cos\theta_\mathrm{W} B_\mu - \sin\theta_\mathrm{W} W^0_\mu$ (trig to be checked) and one orthogonal to this $A_\mu = \sin\theta_\mathrm{W} B_\mu + \cos\theta_\mathrm{W} W^0_\mu$ without a massterm. The angle θ_W is called the Weinberg angle and is given by $\cos\theta_\mathrm{W} = g_Y v/\sqrt{g_Y^2 + g_L^2}$ (if anyone is reading, please know all this is wrong in detail but true in spirit; these formulas are just placeholders).

The massless vector field A_μ that has appeared is now identified with the photon-field. Furthermore, rewriting the kinetic terms for the left-handed doublet fields ℓ and q so that the covariant derivative is expressed in terms of A_μ, Z^0_μ, $W^\pm_\mu$ we find that the bottom half of these doublets have the same em charge (couples to A_μ the same way) as the right-handed electron-type and down-type fermions. These fields therefore blend together into what we think of as the charged leptons (electron, muon, tau) and similarly on the quark side.

(.... insert example calculations here and/or exercises .....)

Lastly, the Yukawa terms are seen to give masses to the leptons and quarks.

(..... provide some details ....)



13.4 Homework #13

1. ...
2. ...

Lecture 14

Grand Unified Theories

14.1 A Cascade of Broken Dreams

14.2 $SU(5)$

Spontaneous symmetry breaking in the electroweak sector of the standard model enabled by the higgs mechanism has been spectacularly successful.

But three fundamental interactions (hypercharge, isospin, color) in the unbroken gauge group $G_{\rm SM} = {\rm SU}(3) \times {\rm SU}(2) \times U(1)$ and (three generations) of three leptons ($\ell_{\rm L}$, $e_{\rm R}$, ν_R) and three quarks (q_L, d_R, u_R) plus the higgs doublet still seem like a strange set of fields to serve as the fundamental nature of reality (ignoring gravity). Theoretically, it is very unsatisfying. Of course, nature does not care whether you are satisfied or not, and this might just be how the world works.

Nevertheless, given that spontaneous symmetry breaking happened once in the history of the universe, it is possible it has happened once before (and might happen again).

Is it possible that, just like the left-handed ${\rm SU}(2)_{\rm L} \times {\rm U}(1)_{\rm Y}$-doublets ℓ_L and q_L break into ${\rm U}(1)_{\rm em}$-singlets ν_L, e_L and u_L, d_L respecively, the set of six fermions are the broken down remnants of a smaller set of more fundamental fields, transforming under some larger symmetry group $G_{\rm GUT}$ containing $G_{\rm SM}$ as a subgroup?

In a 1974 paper by Georgi and Glashow [**?**], a tantalizing possibility was presented. Consider the group $G = {\rm SU}(5)$. It is the group of unitary 5×5-matrices of determinant 1. Consequently, it is clear that it contains ${\rm SU}(3) \times {\rm SU}(2)$ as a subgroup:

$$\begin{pmatrix} {\rm SU}(3) & \mathbf{0} \\ \mathbf{0} & {\rm SU}(2) \end{pmatrix} \subset {\rm SU}(5) \tag{14.2.1}$$

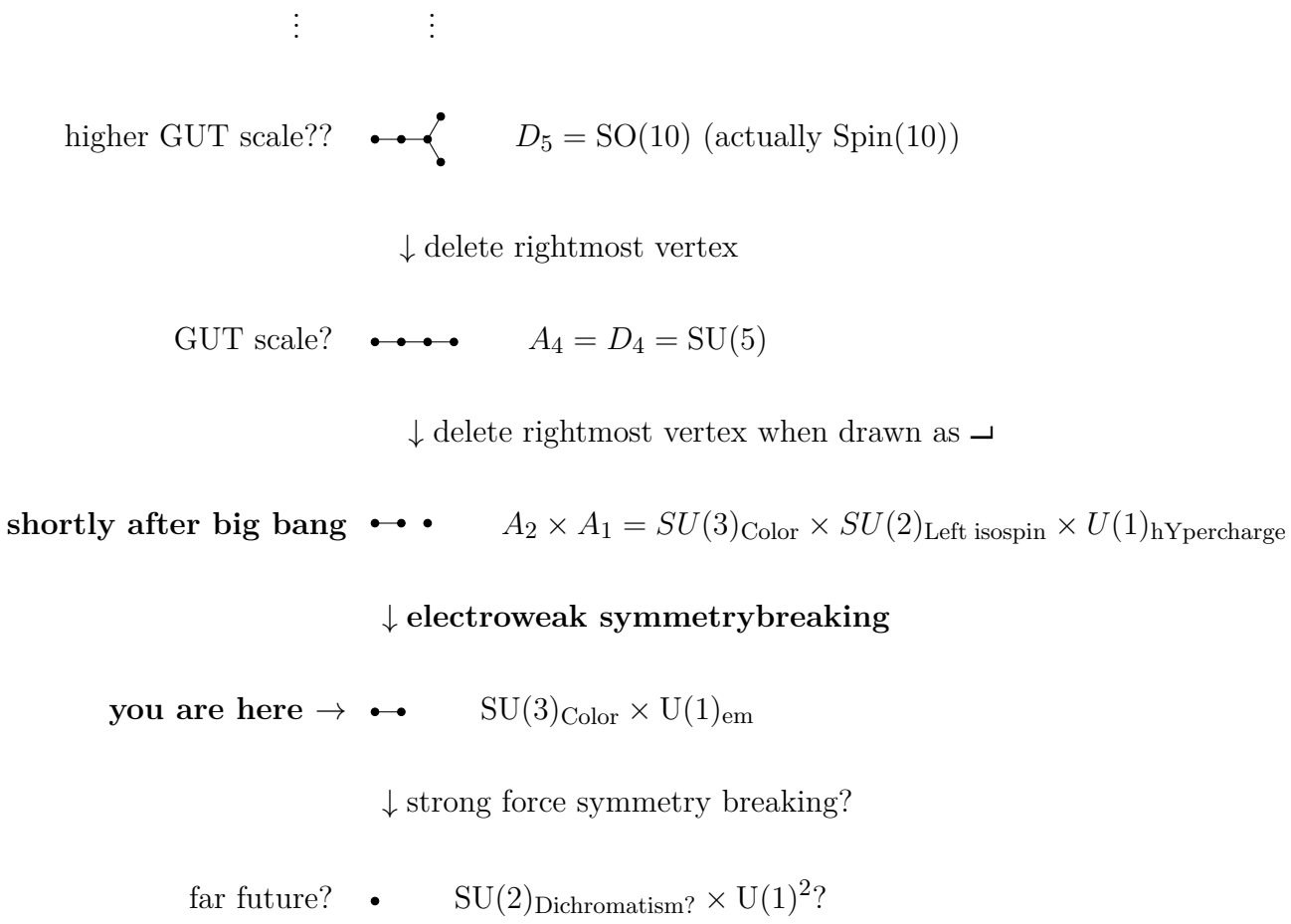


Figure 14.1: A cascade of symmetry breaking events. Bold parts are experimentally confirmed parts of the standard model. Non-bold parts are theoretical/speculative. The graphs are Dynkin diagrams associated with the gauge (Lie) group.

where $\mathbf{0}$ are rectangular zero matrices of appropriate sizes. In fact, there is enough room to simultaneously fit a copy of $U(1)$ in there as well:

$$U(1) \ni z \mapsto \begin{pmatrix} z^2\mathbf{1}_3 & \mathbf{0} \\ \mathbf{0} & z^{-3}\mathbf{1}_2 \end{pmatrix} \tag{14.2.2}$$

where $\mathbf{1}_n$ denotes the $n \times n$ identity matrix, and z is a complex number of modulus one.

Explicitly, there is a group homomorphism $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \to \mathrm{SU}(5)$ given by

$$(A, B, z) \mapsto \begin{pmatrix} z^2 A & \mathbf{0} \\ \mathbf{0} & z^{-3}B \end{pmatrix}, \qquad A \in \mathrm{SU}(3),\ B \in \mathrm{SU}(2),\ z \in \mathrm{U}(1). \tag{14.2.3}$$

Since SU(5) is a 24-dimensional Lie group ($\mathrm{SU}(n)$ has dimension $n^2 - 1$), and G_{SM} has dimension $8 + 3 + 1 = 12$, there are an additional 12 real components worth of gauge fields when SU(5) is used as a local gauge group. These are divided into two kinds, called X and Y, corresponding to block upper- and lower-triangular matrices in the Lie algebra:

$$\begin{pmatrix} \mathbf{0} & X_{ab} \\ Y_{ba} & \mathbf{0} \end{pmatrix} \tag{14.2.4}$$

where $a = 1, 2, 3$ and $b = 1, 2$. As we will see, these 12 new gauge bosons facilitate interactions which can turn quarks into leptons. Accordingly, X and Y are called *leptoquarks*. In particular, the up-quark can decay which implies a the proton is not stable.

The most natural representation of SU(5) is the fundamental representation denoted $\mathbb{C}^5$, or just $\mathbf{5}$, consisting of column matrices with 5 complex entries. The dual representation $\overline{\mathbf{5}}$ is the dual space $\mathrm{Hom}_{\mathbb{C}}(\mathbb{C}^5, \mathbb{C})$ with action $(A\xi)(v) = \xi(A^{-1}v)$ for $A \in \mathrm{SU}(5), \xi \in \overline{5}, v \in \mathbf{5}$. Let us consider how this representation breaks

14.3 $SO(10)$

Bibliography

[1] Georgi, H., Glashow, S.L., *Unity of All Elementary-Particle Forces*, Phys. Rev. Lett. 32(8) (1974) 438–441.

[2] Lovelock, D., Rund, H., *Tensors, Differential Forms, and Variational Principles*, Dover, New York, 1989.

[3] Soper, Davison E., *Classical Field Theory*, Wiley, New York, 1976. Also available as Dover edition, 2008.