Not so classical Mechanics

James T. Wheeler

August 26, 2018

Contents

Part I
Preliminaries

Geometry is physics; physics is geometry. It is human nature to unify our experience, and one of the important specializations of humans to develop language to describe our world. Thus, we find the unifying power of geometric description a powerful tool for spotting patterns in our experience, while at the same time, the patterns we find lead us to create new terms, concepts, and most importantly, pictures. Applying geometric abstractions to model things in the world, we discover new physics. Creating new pictures, we invent new mathematics.

Our ability to make predictions based on perceived patterns is what makes physics such an important subject. The more mathematical tools we have at hand, the more patterns we can model, and therefore the more predictions we can make. We therefore start this adventure with some new tools, Lie groups and the calculus of variations. We will also cast some old familiar tools in a new form.

[DESCRIBE EACH]

With this extension of the familiar calculus, along with some new ways to look at curves and spaces, we will be able to demonstrate the naturalness and advantages of the most elegant formulation of Newton's laws of mechanics: the phase space formulation due to Euler, Lagrange, Hamilton and many others, in which the position and the momentum of each particle are treated as independent variables coupled by a system of first-order differential equations.

We will take a gradual approach to this formulation. Our starting point centers on the idea of space, an abstraction which started in ancient Greece. Basing our assumptions in our direct experience of the world, we provisionally work in a 3-dimensional space, together with time. Because we want to associate physical properties with objects which move in this space rather than with the space itself, we demand that the space be homogeneous and isotropic. This leads us to the construction of (among other possibilities) Euclidean 3-space. In order to make measurements in this space, we introduce a metric, which allows us to characterize what is meant by uniform motion. We also treat the description of matter from a fundamental approach, carefully defining the mathematical structures that might be used to model the stuff of the world. Finally, we develop techniques to describe the motion of matter.

Each of these developments involves the introduction of new mathematics. The description of uniform motion leads to the calculus of variations, the description of matter leads to a discussion of vectors and tensors, and our analysis of motion requires techniques of differential forms, connections on manifolds, and gauge theory.

Once these tools are in place, we derive various well-known and not so well-known techniques of classical (and quantum) mechanics.

Numerous examples and exercises are scattered throughout.

[MORE HERE]

Enjoy the ride!

1 The physical arena

We have presented general arguments that we can reconcile our visual and tactile experiences of the world by choosing a 3-dim model, together with time. We still need to specify what we mean by a space. Returning to Aristotle's question, we observe that we can maintain the idea of the "space" where an object was as independent of the body if we insist that "space" contain no absolute information. Thus, the orientation of a body is to be a property of the body, not of the space. Moreover, it should not matter whether a body is at this or that location in space. This notion of space lets us specify, for example, the relative nearness or farness of two bodies without any dependence on the absolute positions of the bodies. These properties are simply expressed by saying that space should have no preferred position, direction or scale. We therefore demand a 3-dim space which is *homogeneous, isotropic* and *scale invariant.*

1.1 Symmetry and groups

Mathematically, it is possible to construct a space with any desired symmetry using standard techniques. We begin with a simple case, reserving more involved examples for later chapters. To begin, we first define a mathematical object capable of representing symmetry. We may think of a symmetry as a collection of transformations that leave some essential properties of a system unchanged. Such a collection, G, of transformations must have certain properties:

1. We may always define an identity transformation, e, which leaves the system unchanged: $\exists\ e \in G$.
2. For every transformation, g, taking the system from description A to another equivalent description A', there must be another transformation, denoted g^{-1}, that reverses this, taking A' to A. The combined effect of the two transformations is therefore $g^{-1}g = e$. We may write: $\forall g \in G, \exists g^{-1} \in G \ni: g^{-1}g = e$.
3. Any two transformations must give a result which is also achievable by a transformation. That is, $\forall g_1, g_2 \in G,\ \exists g_3 \in G \ni: g_1 g_2 = g_3$.
4. Applying three transformations in a given order has the same effect if we replace either consecutive pair by their combined result. Thus, we have associativity: $\forall g_1, g_2, g_3 \in G,\ g_1 (g_2 g_3) = (g_1 g_2) g_3$.

These are the defining properties of a mathematical *group*. Precisely, a group is a set, S, of objects together with a binary operation satisfying properties $1 - 4$. We provide some simple examples.

The binary, or Boolean, group, B, consists of the pair $B = \{\{1, -1\}, \times\}$ where $\times$ is ordinary multiplication. The multiplication table is therefore

$$\begin{bmatrix} \times & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$

Naturally, 1 is the identity, while each element is its own inverse. Closure is evident by looking at the table, while associativity is checked by tabulating all triple products:

$$\begin{aligned} 1 \times (1 \times (-1)) &= -1 &&= (1 \times 1) \times (-1) \\ 1 \times (-1 \times (-1)) &= &&1 = (1 \times (-1)) \times (-1) \\ & && \textit{etc.} \end{aligned}$$

The pair B is therefore a group.

There is another way to write the Boolean group, involving modular addition. We define:

Let S be the set of n consecutive integers beginning zero, $S = \{0, 1, 2, \ldots, n-1\}$. Addition modulo n (or mod n), $\oplus_n$, is cyclic addition on S. That is, for all $a, b \in S$

$$a \oplus_n b = \begin{cases} a + b & a + b < n \\ a + b - n & a + b \geq n \end{cases}$$

where $+$ is the usual addition of real numbers.

Addition mod n always produces a group with n elements:

The pair $G_n = (S, \oplus_n)$ is a group.

For proof, we see immediately that multiplication modulo n is closed, because if

$$a + b < n$$

then $a + b \in S$ while if

$$a + b \geq n$$

then $a+b-n \in S$. Zero is the additive identity, while the inverse of $a$is $n-a$. Finally, to check associativity, we have three cases:

$$\begin{aligned} a+b &< n,\ b+c < n \\ a+b &< n,\ b+c \geq n \\ a+b &\geq n,\ b+c \geq n \end{aligned}$$

The first case is immediate because

$$\begin{aligned} (a \oplus_n b) \oplus_n c &= (a+b) \oplus_n c = \begin{cases} a+b+c & a+b+c < n \\ a+b+c-n & a+b+c \geq n \end{cases} \\ a \oplus_n (b \oplus_n c) &= a \oplus_n (b+c) = \begin{cases} a+b+c & a+b+c < n \\ a+b+c-n & a+b+c \geq n \end{cases} \end{aligned}$$

In the second case, we note that since $a+b<n$, and $b+c>n$, we must have $n < a+b+c < 2n$. Therefore,

$$\begin{aligned} (a \oplus_n b) \oplus_n c &= (a+b) \oplus_n c = a+b+c-n \\ a \oplus_n (b \oplus_n c) &= a \oplus_n (b+c-n) \\ &= \begin{cases} a+b+c-n & a+b+c < 2n \\ a+b+c-2n & a+b+c \geq 2n \end{cases} \\ &= a+b+c-n \end{aligned}$$

For the final case, we have two subcases:

$$\begin{aligned} n &< a+b+c < 2n \\ 2n &\leq a+b+c < 3n \end{aligned}$$

In the first subcase,

$$\begin{aligned} (a \oplus_n b) \oplus_n c &= (a+b-n) \oplus_n c = a+b+c-n \\ a \oplus_n (b \oplus_n c) &= a \oplus_n (b+c-n) = a+b+c-n \end{aligned}$$

while in the second subcase,

$$\begin{aligned} (a \oplus_n b) \oplus_n c &= (a+b-n) \oplus_n c = a+b+c-2n \\ a \oplus_n (b \oplus_n c) &= a \oplus_n (b+c-n) = a+b+c-2n \end{aligned}$$

Therefore, G_n is an n-element group.

Now, returning to our discussion of the Boolean group, consider addition mod 2. The multiplication table is

$\oplus_2$	0	1
0	0	1
1	1	0

Notice that if we rename the elements, replacing $0 \to 1$ and $1 \to -1$, we reproduce the multiplication of the Boolean group. When this is the case, we say we have two representations of the same group. We make this idea more precise below. First, we prove that, while there may be different representations, there is only group *one* with two elements. To prove this, suppose we pick any set with two elements, $S = \{a, b\}$ and write the general form of the multiplication table:

$\circ$	a	b
a		
b		

One of these must be the identity; without loss of generality we choose $a = e$. Then

$$\begin{array}{ccc} \circ & a & b \\ a & a & b \\ b & b & \end{array}$$

Finally, since b must have an inverse, and its inverse cannot be a, we must fill in the final spot with the identity, thereby making b its own inverse:

$$\begin{array}{ccc} \circ & a & b \\ a & a & b \\ b & b & a \end{array}$$

Comparing to the boolean table, we see that a simple renaming, $a \to 1, b \to -1$ reproduces the boolean group. The relationship between different representations of a given group is made precise by the idea of an *isomorphism*.

Let $G = (S, \oplus)$ and $H = (T, \otimes)$ be two groups and let ϕ be a one-to-one, onto mapping, ϕ, between G and H. Then ϕ is an *isomorphism* if it preserves the group product in the sense that for all g_1, g_2 in G,

$$\phi(g_1) \otimes \phi(g_2) = \phi(g_1 \oplus g_2) \tag{1}$$

When there exists an isomporphism between G and H, then G and H are said to be *isomorphic* to one another.

The definition essentially means that ϕ provides a renaming of the elements of the group. Thus, suppose $g_1 \oplus g_2 = g_3$. Thinking of $h = \phi(g)$ as the new name for g, and setting

$$\begin{aligned} h_1 &= \phi(g_1) \\ h_2 &= \phi(g_2) \\ h_3 &= \phi(g_3) \end{aligned}$$

eq.(1) becomes

$$h_1 \otimes h_2 = h_3$$

Applying the group product may be done before or after applying ϕ with the same result. In the Boolean case, for example, setting $\phi(a) = 0$ and $\phi(b) = 1$ shows that $G = (\circ, \{a, b\})$ and $H = (\oplus_2, \{0, 1\})$ are isomorphic.

Now consider a slightly bigger group. We may find all groups with three elements as follows. Let $G = \{\{a, b, c\}, \otimes\}$, where the group operation, $\otimes$, remains to be defined by its multiplication table. In order for G to be a group, one of the elements must be the identity. Without loss of generality, we pick $a = e$. Then the multiplication table becomes

$$\begin{array}{cccc} \otimes & e & b & c \\ e & e & b & c \\ b & b & & \\ c & c & & \end{array}$$

Next, we show that no element of the group may occur more than once in any given row or column. To prove this, suppose some element, z, occurs in the c column twice. Then there are two distinct elements (say, for generality, x and y) such that

$$\begin{aligned} x \otimes c &= z \\ y \otimes c &= z \end{aligned}$$

From this we may write

$$x \otimes c = y \otimes c$$

But since c must be invertible, we may multiply both sides of this equation by c^{-1} and use associativity:

$$\begin{aligned}(x \otimes c) \otimes c^{-1} &= (y \otimes c) \otimes c^{-1} \\ x \otimes \left(c \otimes c^{-1}\right) &= y \otimes \left(c \otimes c^{-1}\right) \\ x \otimes e &= y \otimes e \\ x &= y\end{aligned}$$

in contradiction with x and y being distinct elements. The argument for an element occurring twice in the same row is similar.

Returning to the three-element multiplication table, we see that we have no choice but to fill in the remaining squares as

$\otimes$	e	b	c
e	e	b	c
b	b	c	e
c	c	e	b

thereby showing that there is exactly one three element group.

Many groups are already familiar:

Let $G = \{Z, +\}$, the integers under addition. For all integers a, b, c we have $a + b \in R$ (closure); $0 + a = a + 0 = a$ (identity); $a + (-a) = 0$ (inverse); $a + (b + c) = (a + b) + c$ (associativity). Therefore, G is a group. The integers also form a group under addition mod$p$, where p is any integer (Recall that $a = b$ mod$p$ if there exists an integer n such that $a = b + np$).

Let $G = \{R, +\}$, the real numbers under addition. For all real numbers a, b, c we have $a+b \in R$ (closure); $0 + a = a + 0 = a$ (identity); $a + (-a) = 0$ (inverse); $a + (b + c) = (a + b) + c$ (associativity). Therefore, G is a group. Notice that the rationals, Q, are not a group under addition because they do not close under addition:

$$\pi = 3 + .1 + .04 + .001 + .0005 + .00009 + \ldots$$

Of course, the real numbers form a field, which is a much nicer object than a group.

In working with groups it is generally convenient to omit the multiplication sign, writing simply ab in place of $a \otimes b$.

A subgroup of a group $G = \{S, \otimes\}$ is a group $G' = \{S', \otimes\}$, with the same product, $\otimes$ such that S' is a subset of S. Prove that a group with $n + 1$ elements has no subgroup with n elements. (Hint: write the multiplication table for the n element subgroup and try adding one row and column.)

Find all groups (up to isomorphism) with four elements.

Show that the three reflections of the plane

$$\begin{aligned}R_x &: (x, y) \to (-x, y) \\ R_y &: (x, y) \to (x, -y) \\ R_{xy} &: (x, y) \to (-x, -y)\end{aligned}$$

together with the identity transformation, form a group. Write out the multiplication table.

	e	R_x	R_y	R_{xy}
e	e	R_x	R_y	R_{xy}
R_x	R_x	e	R_{xy}	R_y
R_y	R_y	R_{xy}	e	R_x
R_{xy}	R_{xy}	R_y	R_x	e

Find the 8-element group built from the three dimensional reflections and their products.

1.2 Lie groups

While groups having a finite number of elements are entertaining, and even find use in crystalography, most groups encountered in physics have infinitely many elements. To specify these elements requires one or more continuous parameters. We begin with some familiar examples.

1. The real numbers under addition, $G = \{R, +\}$, form a Lie group because each element of R provides its own label. Since only one label is required, R is a 1-dimensional Lie group.

2. The real, n-dim vector space V^n under vector addition is an n-dim Lie group, since each element of the group may be labeled by n real numbers.

3. Rotations in the plane. Consider a rotation of vectors (x, y) through an angle θ :

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \end{aligned}$$

which we may write as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

The transformation matrices

$$R = \left\{ \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \theta \in [0, 2\pi) \right\}, \times \right\}$$

where $\times$ is normal matrix multiplication, form a group. To see this, consider the product of two elements,

$$\begin{aligned} R(\theta) R(\varphi) &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta\cos\varphi - \sin\theta\sin\varphi & -\cos\theta\sin\varphi - \sin\theta\cos\varphi \\ \sin\theta\cos\varphi + \cos\theta\sin\varphi & -\sin\theta\sin\varphi + \cos\theta\cos\varphi \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta+\varphi) & -\sin(\theta+\varphi) \\ \sin(\theta+\varphi) & \cos(\theta+\varphi) \end{pmatrix} \end{aligned}$$

so the set is closed under multiplication as long as we consider the addition of angles to be addition modulo 2π. We immediately see that the inverse to any rotation $R(\theta)$ is a rotation by $R(2\pi - \theta)$, and the associativity of (modular) addition guarantees associativity of the product. Notice that rotations in the plane commute. A group in which all products commute is called *Abelian*. Show that the 2-dim rotation group $R(\theta)$ preserves the Euclidean length, $l^2 = x^2 + y^2$.

4. Rotations in 3-dim. These depend on three parameters, which may be defined in various ways. We could take an angle of rotation about each of the three Cartesian coordinate axes. A more systematic approach is to use the Euler angles. The simplest parameterization, however, is to specify a unit vector giving the axis of rotation, and an angle representing the amount of the rotation around that axis. The latter method involves the elegant use of quaternions, pioneered by Klein in the 19th century. We will give a complete treatment of this Lie group when we begin our study of Lie algebras.

5. Translations in 3-dim. Consider the translations of 3-space, given by vector addition:

$$T_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} + \mathbf{a}$$

where $\mathbf{a}$ is any 3-vector,

$$\mathbf{a} \in V^3$$

The pair $T = \left\{ \left\{ T_{\mathbf{a}}, \mathbf{a} \in V^3 \right\}, T_{\mathbf{a}} T_{\mathbf{b}} = T_{\mathbf{a}+\mathbf{b}} \right\}$ is a Lie group.

1.2.1 Topological spaces and manifolds

We can give a precise definition of a Lie group if we first define the useful class of spaces called *manifolds*. The definition relies on the idea of a *topological space*. This section is intended only to give the reader an intuitive grasp of these terms. Thus, while the definitions are rigorous, the proofs appeal to familiar properties of lines and planes. For example, a rigorous form of the proof of the equality of two topologies for the plane given below requires detailed arguments involving terms we do not introduce here such as limit points and metrics. A complete treatment takes us too far afield, but a few of the many excellent presentations of these topics are listed in the references.

A topological space, S, it a set S, for which we have a collection, τ, of subsets of S satisfying the following properties

1. For all A, B in τ, their intersection $A \cap B$ is also in τ.
2. For all A_α in τ, their union $\bigcup_\alpha A_\alpha$ is also in τ.
3. The empty set, ϕ, and S itself are in τ.

This definition is quite abstract, and it is amazing that it embodies many important properties we associate with spaces. The subsets in the collection are called *open sets*, while the complement of an open set with respect to S is called *closed*. To see how the definition works, consider the example of the real line. The open sets are just the usual open intervals, (a, b), finite unions of open intervals, and open infinite intervals, $(a, \infty), (-\infty, b)$ and $(-\infty, \infty)$. Closed sets are sets containing their endpoints, $[a, b]$. Notice that we can often define a topology by specifying a collection such as the open intervals, then extending the set by taking all possible finite intersections and arbitrary unions. It is worth mentioning that we require finite intersections because infinite intersections produce too many sets. For example, let A_n be the open interval

$$A_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$$

Then the infinite intersection over all n is the set containing a single point, $\{0\}$.

Another familiar example is the plane, R^2. Let τ_0 the collection of open disks,

$$\begin{aligned} \tau_0 &= \{U_\varepsilon(a, b) \,|\, a, b \in R, \varepsilon > 0\} \\ U_\varepsilon(a, b) &= \left\{(a, b) + \varepsilon(x, y) \,|\, -1 < \sqrt{x^2 + y^2} < 1\right\} \end{aligned}$$

and let τ be the collection of all unions and finite intersections of sets in τ_0. This is the usual topology of the plane. If we have any "open" region V of the plane – that is, an arbitrarily shaped region of contiguous points without any boundary points – then it is in τ. To see this, pick any point in V. Around this point we can find an open disk, $U_{\varepsilon(P)}(P)$, that is small enough that it lies entirely within V. Repeating this for every point in V, we see that V is the union of these open disks,

$$V = \bigcup_P U_{\varepsilon(P)}(P)$$

so that V is in τ.

Two topologies on a given set are equal if they contain the same open sets. Typically, we can define more than one distinct topology for any given set S, and there is more than one way to specify a given topology. To see the first, return to the real line but define the open sets to be the half-open intervals,

$$[a, b)$$

together with their unions and finite intersections. No half-open set is included in the usual topology because such an interval is not the union or finite intersection of open intervals. To see the second, we need a way to compare two topologies. The technique we used above for the plane works in general, for suppose we want

to show that two topologies, τ and τ', are equal. Then for an arbitrary open set U in τ and any point P in U, find an open set V_P in τ' which contains P and lies entirely in U. Since the union of the sets V_P is U, the set U must be in τ'. Then we repeat the argument to show that sets in τ' also lie in τ.

As an example of the method, consider a second topology for the plane consisting of unions and finite intersections of open squares,

$$V_\varepsilon(a,b) = \{(a,b) + \varepsilon(x,y) \mid -1 < x < 1, -1 < y < 1\}$$

Picking any point P, in any $V_\varepsilon(a,b)$, we can find an open disk centered on P and lying entirely within $V_\varepsilon(a,b)$. Conversely, for any point of any open disk, we can find a rectangle containing the point and lying entirely within the disk. Therefore, any set that can be built as a union or finite intersection of open rectangles may also be built as a union or finite intersection of disks, and vice versa. The two topologies are therefore equal.

This concludes our brief foray into topology. The essential idea is enough to proceed with the definition of a differentiable manifold.

An n-dim differentiable manifold is a topological space, $\mathcal{M}$, together with a set of 1-1 mappings, φ_α, from open sets $U_\alpha \subset \mathcal{M}$ onto open subsets V_α of real n-space, $V_\alpha \subset R^n$,

$$\varphi_\alpha : U_\alpha \to V_\alpha$$

for some fixed dimension n. Here α is simply a label for convenience. These mappings must satisfy the following properties:

1. For each point p of $\mathcal{M}$ there is some open open set, U_α containing p. Therefore, the union of the sets U_α is $\mathcal{M}$.

2. Let
$$U_\gamma = U_\alpha \cap U_\beta$$
be the intersection of any two of the open sets U_α, U_β, and consider the restriction of the mappings φ_α and φ_β to U_γ. Denoting these restricted mappings by
$$\begin{aligned} \varphi_{\alpha|\gamma} &: U_\gamma \to V_{\alpha|\gamma} \subset V_\alpha \\ \varphi_{\beta|\gamma} &: U_\gamma \to V_{\beta|\gamma} \subset V_\beta \end{aligned}$$
we require the real-valued map given by the composition
$$\varphi_{\alpha|\gamma} \circ \varphi_{\beta|\gamma}^{-1} : V_{\beta|\gamma} \to V_{\alpha|\gamma}$$
to be differentiable.

The basic idea here is that we have a correspondence between small regions of the manifold and regions of real n-space: the manifold is essentially R^n if you look closely enough. The overlap condition allows us to carry this correspondence to larger regions, but it is weak enough to allow $\mathcal{M}$ to be distinct from R^n. For example, consider the circle, S^1. The usual angle θ maps points on S^1 in a 1-1 way to points in the interval $[0, 2\pi)$, but this interval is not open in R^1. Nonetheless, S^1 is a manifold because we can choose *two* mappings from all but one point of S^1.

$$\begin{aligned} \theta &: S^1 - \{\pi\} \to (-\pi, +\pi) \\ \phi &: S^1 - \{0\} \to (0, 2\pi) \end{aligned}$$

Every point of the circle lies in at least one of the sets $S^1 - \{\pi\}$ or $S^1 - \{0\}$, each angle maps open set to open sets, and on the overlap region $S^1 - \{\pi\} - \{0\}$, the mapping

$$\theta \circ \phi^{-1}$$

is just

$$\theta \circ \phi^{-1}(x) = x - \pi$$

Prove that the 2-sphere, $S^2 = \left\{(x,y,z) \mid x^2+y^2+z^2 = 1\right\}$ is a manifold.

Prove that the 2-torus, $T^2 = \{(x,y) \mid 0 \le x < 1, 0 \le y < 1\}$ with the periodic boundary conditions, $(0,y) = (1,y)$ and $(x,0) = (x,1)$, is a manifold.

Show that a cylinder is a manifold.

Show that a Möbius strip is a manifold.

We may now define a Lie group.

A Lie group is a group $G = \{S, \otimes\}$ for which the set S is a manifold.

1.3 The Euclidean group

We now return to the problem of defining a homogeneous, isotropic, 3-dim space. (The result will also be scale invariant, but we postpone systematic treatment of this property). If we think of the properties of homogeneity and isotropy as transformations, we can use them to define a Lie group called the Euclidean group. The Euclidean group characterizes the properties we associate with the Newtonian arena, and may even be used to construct that arena.

Homogeneity of space means that there is no essential difference between distinct points. If we label points by triples of numbers, (x,y,z) then homogeneity is equivalent to invariance under translations:

$$T_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} + \mathbf{a}$$

These transformations may be written in matrix form,

$$\begin{pmatrix} 1 & & & a_1 \\ & 1 & & a_2 \\ & & 1 & a_3 \\ & & & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} x+a_1 \\ y+a_2 \\ z+a_3 \\ 1 \end{pmatrix}$$

In this form it is easy to check the group multiplication property,

$$\begin{pmatrix} 1 & & & a_1 \\ & 1 & & a_2 \\ & & 1 & a_3 \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & b_1 \\ & 1 & & b_2 \\ & & 1 & b_3 \\ & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & & & a_1+b_1 \\ & 1 & & a_2+b_2 \\ & & 1 & a_3+c_2 \\ & & & 1 \end{pmatrix}$$

which shows that $T_{\mathbf{a}} T_{\mathbf{b}} = T_{\mathbf{a}+\mathbf{b}}$ and $T_{\mathbf{a}}^{-1} = T_{-\mathbf{a}}$. Closure is guaranteed as long as each of the three parameters ranges over all real numbers, and associativity follows from the associativity of addition (or, equivalently, the associativity of matrix multiplication).

Isotropy is independence of direction in space, equivalent to invariance under rotations. The simplest way to characterize rotations is as transformations preserving lengths. Therefore, any rotation,

$$\mathbf{x}' = \mathbf{R}(\mathbf{x})$$

must satisfy

$$\mathbf{x}' \cdot \mathbf{x}' = \mathbf{R}(\mathbf{x}) \cdot \mathbf{R}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{x}$$

for all vectors $\mathbf{x}$. Such relations are often most usefully expressed in terms of index notation:

$$\begin{aligned} x_i' &= \sum_{j=1}^{3} R_{ij} x_j \\ \sum_{i=1}^{3} x_i' x_i' &= \sum_{i=1}^{3} \left(\sum_{j=1}^{3} R_{ij} x_j \right) \left(\sum_{k=1}^{3} R_{ik} x_k \right) = \sum_{i=1}^{3} x_i x_i \end{aligned}$$

Unfortunately, this notation is cumbersome. Expressions like these are greatly simplified if we adopt the Einstein summation convention, which simply omits the summation symbol $\sum$. Sums are then performed whenever an index is repeated in any given term:

$$\begin{aligned} x'_i &= R_{ij}x_j \\ x'_i x'_i &= R_{ij}x_j R_{ik}x_k = x_i x_i \end{aligned}$$

Since j is repeated in a single term on the right side of the first equation, we sum on j. Notice that the remaining, unpaired index i must be present on both sides of the equation. In the second equation, i is repeated in a single term on the left, while i, j and k are each repeated on the right. The summed indices are called *dummy* indices, and single indices are called $free$ indices. In every term of every equation the free indices must balance. We have a different equation for each value of each $free\ index$. The dummy indices can be changed for our convenience as long as we don't change their positions or repeat the same dummy index more than the required two times in any one term. Thus, we can write

$$T_{ijk}v_j w_k + \omega_i = S_{ij}u_j$$

or

$$T_{imn}v_m w_n + \omega_i = S_{ij}u_j$$

but *not*

$$T_{ijj}v_j w_j + \omega_i = S_{im}u_m$$

because using the same dummy index twice in the term $T_{ijj}v^j w^j$ means we don't know whether to sum v^j with the second or the third index of T_{ijk}. We will employ the Einstein summation convention throughout the remainder of the book, with one further modification occurring in Section 5.

Defining the transpose of R_{ij} by

$$R^t_{ij} = R_{ji}$$

we may write the defining equation as

$$R^t_{ji}x_j R_{ik}x_k = x_i x_i$$

Index notation has the advantage that we can rearrange terms as long as we don't change the index relations. We may therefore write

$$R^t_{ji}R_{ik}x_j x_k = x_i x_i$$

Finally, we introduce the identity matrix,

$$\delta_{ij} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$$

so that we can write the right hand side as

$$x_i x_i = x_j \delta_{jk} x_k = \delta_{jk} x_j x_k$$

Since there are no unpaired, or "free" indices, it doesn't matter that we have i on the left and j, k on the right. These "dummy" indices can be anything we like, as long as we always have exactly two of each kind in each term. In fact, we can change the names of dummy indices any time it is convenient. For example, it is correct to write

$$x_i x_i = x_j x_j$$

or

$$x'_i = R_{ij}x_j = R_{ik}x_k$$

We now have

$$R^t_{ji}R_{ik}x_j x_k = \delta_{jk}x_j x_k$$

This expression must hold for all x_j and x_k. But notice that the product $x_j x_k$ is really a symmetric matrix, the outer product of the vector with itself,

$$x_j x_k = \begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix}$$

Beyond symmetry, this matrix is arbitrary.

Now, we see that the double sum of $R^t_{ji}R_{ik} - \delta_{jk}$ with an arbitrary symmetric matrix

$$\left(R^t_{ji}R_{ik} - \delta_{jk}\right) x_j x_k = 0$$

must vanish. This is the case for any purely antisymmetric matrix, so that the symmetric part of the matrix in parentheses must vanish:

$$\left(R^t_{ji}R_{ik} - \delta_{jk}\right) + \left(R^t_{ki}R_{ij} - \delta_{kj}\right) = 0$$

This simplifies immediately because the identity matrix is symmetric, $\delta_{jk} = \delta_{kj}$ and so is the product of R_{ij} with its transpose:

$$\begin{aligned} R^t_{ki}R_{ij} &= R_{ik}R_{ij} \\ &= R_{ik}R^t_{ji} \\ &= R^t_{ji}R_{ik} \end{aligned}$$

In formal matrix notation, this is equivalent to simply

$$\left(R^t R\right)^t = R^t R$$

Combining terms and dividing by two we finally reach the conclusion that R times its transpose must be the identity:

$$\begin{aligned} R^t_{ji}R_{ik} &= \delta_{jk} \\ R^t R &= 1 \end{aligned}$$

Notice that we may use the formal notation for matrix multiplication whenever the indices are in the correct positions for normal matrix multiplication, i.e., with adjacent indices summed. Thus,

$$M_{ij}N_{jk} = S_{ik}$$

is equivalent to $MN = S$, while

$$M_{ji}N_{jk} = S_{ik}$$

is equivalent to $M^t N = S$.

Returning to the problem of rotations, we have shown that the transformation matrix R is a rotation if and only if

$$R^t R = 1$$

Since R must be invertible, we may multiply by R^{-1} on the right to show that

$$R^t = R^{-1}$$

This characterizes rotations. Any transformation whose transpose equals its inverse is called *orthogonal*. We will show later that rotations can be labeled by three parameters. The rotations in themselves form a 3-dim Lie group.

Show that the orthogonal transformations form a group. You may use the fact that matrix multiplication is associative.

Show that the orthogonal transformations form a Lie group.

We need to show that the elements of the rotation group form a manifold. This requires a 1-1, onto mapping between a neighborhood of any given group element and a region of R^3. To begin, find a neighborhood of the identity. Let R be a 3-dim matrix satisfying $R^t R = 1$ such that

$$R = 1 + A$$

where A is another matrix with all of its elements small, $|A_{ij}| << 1$ for all $i, j = 1, 2, 3$. Then we may expand

$$\begin{aligned} 1 &= R^t R = \left(1 + A^t\right)(1 + A) \\ &= 1 + A^t + A + A^t A \\ &\approx 1 + A^t + A \end{aligned}$$

We conclude that A is antisymmetric

$$A^t = -A$$

and may therefore be written as

$$A_{ij} = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}$$

To infinitesimal order, the rotations

$$R = \begin{pmatrix} 1 & a & b \\ -a & 1 & c \\ -b & -c & 1 \end{pmatrix}$$

form a neighborhood of the identity transformation and the mapping $\phi : R \leftrightarrow (a, b, c)$ is 1-1 and onto an open region of R^3. To generalize the result to a neighborhood of an arbitrary group element, R_0, we simply take the product

$$\begin{aligned} RR_0 &= (1 + A) R_0 \\ &= R_0 + AR_0 \end{aligned}$$

Since the components of R_0 are bounded, the components of AR_0 are both bounded and proportional to one or more of a, b, c. We may therefore make the components of A sufficiently small that $\left|(AR_0)_{ij}\right| << 1$, providing the required neighborhood.

The rotations and translations may immediately be combined into a single class of transformations comprising the Euclidean group:

$$\begin{pmatrix} R & \mathbf{a} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \begin{pmatrix} R\mathbf{x} + \mathbf{a} \\ 1 \end{pmatrix}$$

where R is an orthogonal matrix. The product of two such transformations is then

$$\begin{pmatrix} R & \mathbf{a} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R' & \mathbf{b} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} RR' & R\mathbf{b} + \mathbf{a} \\ & 1 \end{pmatrix}$$

This demonstrates closure because the product RR' is orthogonal

$$(RR')^t = (R')^t R^t = (R')^{-1} R^{-1} = (RR')^{-1}$$

and $R\mathbf{b} + \mathbf{a}$ is a 3-vector.

1.4 The construction of Euclidean 3-space from the Euclidean group

The Euclidean group is a Lie group, and therefore a manifold. This means that starting from the dimensionality of space and its local properties of homogeneity and isotropy, we have constructed a manifold. Unfortunately, points on the Euclidean group manifold are specified by six parameters, not three – the Euclidean group is six dimensional. There is a simple way to reduce this dimension, however, and there are important ways of elaborating the procedure to arrive at more complicated objects, including the curved spaces of general relativity and the higher dimensional spaces of Hamiltonian mechanics.

To recover 3-dim space, we first identify the *isotropy subgroup* of the Euclidean group. In principle this could be almost any subgroup, but we want it to be a group that leaves points fixed. The rotation group does this, but the translation group does not. The idea is to identify all points of the group manifold that differ by isotropy. That is, any two points of the 6-dimensional Euclidean group manifold that differ by a pure rotation will be regarded as identical.

As we show below, the result in the present case simply Euclidean 3-space. Why go to all this trouble to construct a space which was an obvious choice from the start? The answer is that the world is *not* described by Euclidean 3-space, but the technique we use here generalizes to more complicated symmetry. When we carefully analyze the assumptions of our measurement theory, we will find additional symmetries which should be taken into account. What we are doing here is illustrating a technique for moving from local symmetries of measurement to possible arenas for dynamics. Ultimately, we will find the most relevant arenas to be curved and of more than three dimensions. Still, the procedure we outline here will let us deduce the correct spaces.

Returning to the problem of rotationally equivalent points, it is easy to see which points these are. The points of the Euclidean group are labeled by the six parameters in matrices of the form

$$p\left(R,\mathbf{a}\right)=\left(\begin{array}{cc} R & \mathbf{a} \\ 0 & 1 \end{array}\right)$$

We define a point in Euclidean space to be the set of all $p\left(R,\ \mathbf{a}\right)$ that are related by a rotation. Thus, two points $p\left(R,\mathbf{a}\right)$ and $p'\left(R',\mathbf{a}'\right)$ are regarded as equivalent if there exists a rotation

$$R''=\left(\begin{array}{cc} R'' & 0 \\ 0 & 1 \end{array}\right)$$

such that

$$p'\left(R',\mathbf{a}'\right)=R''p\left(R,\mathbf{a}\right)$$

To see what Euclidean space looks like, we choose one representative from each equivalence class. To do this, start with an arbitrary point, p, and apply an arbitrary rotation to get

$$\begin{aligned} R''p &= \left(\begin{array}{cc} R'' & 0 \\ 0 & 1 \end{array}\right)\left(\begin{array}{cc} R & \mathbf{a} \\ 0 & 1 \end{array}\right) \\ &= \left(\begin{array}{cc} R''R & R''\mathbf{a} \\ 0 & 1 \end{array}\right) \end{aligned}$$

Any point of the form

$$\left(\begin{array}{cc} R''R & R''\mathbf{a} \\ 0 & 1 \end{array}\right)$$

is therefore in the same class as the original point, $p\left(R,\mathbf{a}\right)$, for any choice of R''. In particular, we may choose $R''=R^{-1}$ so that the point

$$p_0\left(1,R^{-1}\mathbf{a}\right)=\left(\begin{array}{cc} 1 & R^{-1}\mathbf{a} \\ 0 & 1 \end{array}\right)$$

is in the same equivlance class as p. We can simplify this further by writing

$$R^{-1}\mathbf{a} = \mathbf{x}$$

Since R is invertible, there is exactly one $\mathbf{x}$ for each $\mathbf{a}$ – each is simply some triple of numbers. This means that distinct equivalence classes are labeled by matrices of the form

$$\begin{pmatrix} 1 & \mathbf{x} \\ 0 & 1 \end{pmatrix}$$

which are obviously in 1-1 relationship with triples of numbers, $\mathbf{x} = (x_1, x_2, x_3)$. The points of our homogeneous, isotropic 3-dim space are now in 1-1 correspondence with triples of numbers, called coordinates. It is important to distinguish between a manifold and its coordinate charts, and a vector space. Thus, we speak of R^3 when we mean a space of points and V^3 when we wish to talk of vectors. It is the source of much confusion that these two spaces are essentially identical, because most of the generalizations of classical mechanics require us to keep close track of the difference between manifolds and vector spaces. When we want to discuss the 2-sphere, S^2, as a spherical space of points labeled by two angles (θ, φ), it is clear that (θ, φ) is a coordinate chart and not a vector. Thus, assigning vectors to points on S^2 must be done separately from specifying the space itself. As we treat further examples for which points and vectors differ, the distinction will become clearer, but it is best to be mindful of the difference from the start. When in doubt, remember that the 2-sphere is a homogeneous, isotropic space!

2 Measurement in Euclidean 3-space

By abstracting from our experience of the world, we postulate a 3-dim model for the world together with ever-progressing time. By abstracting from objects the idea of a homogeneous, isotropic arena for their motion, we arrive at the 6-dim Euclidean group. Finally, by regarding as equivalent those points of the Euclidean group whose action leaves some point fixed, we arrive at Euclidean 3-space as the arena for our physical description. We next return to our elements of a physical theory: dynamical laws and a measurement theory. Before we can say anything about dynamical evolution, we must have some idea of the types of object we wish to consider, and the ways we have of extracting information from them. This will lead us to a surprising array of new tools.

2.1 Newtonian measurement theory

One principal assumption of Newtonian mechanics is that we can assign positions in space to a particle for each instant of time. Such an assignment is called a *curve*, and the study of curves will make up the next part of our investigation. However, taken by themselves, one curve is as good as another. We need, in addition, a notion of *distance*. The assignment of distances along curves or between nearby points adds a richness that will lead us quickly to the calculus of variations.

But curves alone are insufficient to describe most physical systems, for at each point of a particle's curve in space we wish to assign properties such as velocity, linear and angular momentum, the moment of inertia tensor, and so on. Our second foray will therefore be the study of vectors and their generalization to tensors. In order to compare vectors at different locations (remember S^2!) we will introduce the *connection*.

2.2 Curves, lengths and extrema

So far, all we know about our Euclidean space is that we can place its points in 1-1 correspondence with triples of numbers,

$$\mathbf{x} = (x_1, x_2, x_3)$$

and that the properties of the space are invariant under translations and rotations. We proceed the definition of a curve.

2.2.1 Curves

Intuitively, we imagine a curve drawn in the plane, winding peacefully from point A to point B. Now assign a monotonic parameter to points along the curve so that as the parameter increases, the specified point moves steadily along the curve. In Cartesian coordinates in the plane, this amounts to assigning a number λ, to each point (x, y) of the curve, giving two functions $(x(\lambda), y(\lambda))$. We may write:

$$C(\lambda) = (x(\lambda), y(\lambda))$$

Sometimes it is useful to think of C as a map. If the parameter λ is chosen so that $(x(0), y(0))$ are the coordinates of the initial point A, and $(x(1), y(1))$ are the coordinates of the point B, the λ lies in the interval $[0, 1]$ and the curve C is a map from $[0, 1]$ into the plane:

$$C : [0, 1] \to R^2$$

Notice that a parameter λ easily specify the entire curve even if the it loops around and crosses itself. The alternative procedure of specifying, say, y as a function of x, breaks down for many curves.

We therefore make the brief definition:

A curve is a map from the reals into a manifold,

$$C : R \to \mathcal{M}$$

This suffices. Suppose we have a curve that passes though some point p of our manifold. Around p there is some open set that is mapped by a set of coordinates into a region of R^3,

$$\begin{aligned} p &\in U \\ \varphi &: U \to V \subset R^3 \end{aligned}$$

More concretely, $\varphi(p)$ is some triple of numbers in R^3,

$$\varphi(p) = (x_1, x_2, x_3)$$

The composition of φ with C therefore assigns a triple of numbers in R^3 to each value of the parameter λ,

$$\begin{aligned} C(\lambda) &= p(\lambda) \\ \varphi \circ C(\lambda) &= \varphi(p(\lambda)) = (x_1(\lambda), x_2(\lambda), x_3(\lambda)) \end{aligned}$$

so the mapping ultimately gives us a parameterized set of coordinates.

A further definition gives us reason to prefer some curves to others:

A differentiable curve is one for which the functions $x_i(\lambda)$ are differentiable. A smooth curve is one for which the functions $x_i(\lambda)$ are infinitely differentiable.

2.2.2 Lengths

In keeping with Newtonian measurement theory, we need a means of choosing particular curves as the path of motion of a particle. Taking our cue from Galileo, we want to describe "uniform motion." This should mean something like covering equal distances in equal times, and this brings us to a notion of distance.

The simplest idea of distance is given by the Pythagorean theorem for triangles,

$$a^2 + b^2 = c^2$$

Applying this to pairs of points, if two points in the plane are at positions (x_1, y_1) and (x_2, y_2) then the distance between them is

$$l = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

We may generalize this immediately to three dimensions.

a) Use the Pythagorean theorem in two dimensions to prove show that in three dimensions

$$l = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

b) Generalize to n-dimensions.

Perhaps more important for our discussion of curves is the specialization of this formula to infinitesimal separation of the points. For two infinitesimally nearby points, the proper distance between them is

$$ds = \sqrt{dx^2 + dy^2 + dz^2}$$

This form of the Pythagorean theorem allows us to compute the length of an arbitrary curve by adding up the lengths of infinitesimal bits of the curve.

Consider a curve, $C(\lambda)$ with $\varphi \circ C(\lambda) = x_i(\lambda)$. In going from λ to $\lambda + d\lambda$, the change in coordinates is

$$dx_i = \frac{dx_i}{d\lambda} d\lambda$$

so the length of the curve from $\lambda = 0$ to $\lambda = 1$ is

$$\begin{aligned} l_{01} &= \int_0^1 ds(\lambda) \\ &= \int_0^1 \sqrt{g_{ij} \frac{dx_i}{d\lambda} \frac{dx_j}{d\lambda}} d\lambda \end{aligned}$$

which is an ordinary integral. This integral gives us the means to assign an unbiased meaning to Galileo's idea of uniform motion. We have, in principal, a positive number l_{01} for each curve from $p(0)$ to $p(1)$. Since these numbers are bounded below, there exist one or more curves of shortest length. This shortest length is the infimum of the numbers $l_{01}(C)$ over all curves C.

For the next several sections we will address the following question: Which curve C has the shortest length?

To answer this question, we begin with a simplified case. Consider the class of curves in the plane, given in Cartesian coordinates:

$$C(\lambda) = (x(\lambda), y(\lambda))$$

The length of this curve between $\lambda = 0$ and $\lambda = 1$ is

$$\begin{aligned} s &= \int_0^1 ds \\ &= \int_0^1 \sqrt{dx^2 + dy^2} \\ &= \int_0^1 \sqrt{\left(\frac{dx}{d\lambda}\right)^2 + \left(\frac{dy}{d\lambda}\right)^2} d\lambda \end{aligned}$$

If the curve always has finite slope with respect to x, so that $\frac{dx}{d\lambda}$ never vanishes, we can choose $\lambda = x$ as the parameter. Then the curve may be written in terms of a single function, $y(\lambda)$,

$$C(x) = (\lambda, y(\lambda))$$

with length

$$s = \int_0^1 \sqrt{1 + \left(\frac{dy}{d\lambda}\right)^2} d\lambda$$

We begin by studying this example.

Compute the length of the curve

$$C(x) = (x, \sin x)$$

from the point $(0, 0)$ to the point $(\pi, 0)$.

2.3 The Functional Derivative

2.3.1 An intuitive approach

The question of the shortest length curve is strikingly different from the usual extremum problem encountered in calculus. The problem is that the length of a curve is not a function. Rather, it is an example of a *functional*. Whereas a function returns a single number in response to a single number input, a functional requires an entire function as input. Thus, the length of a curve requires the specification of an entire curve, consisting of one or more real valued functions, before a length can be computed. To indicate the difference notationally, we replace the usual $f(x)$ for a function and write $f[x]$ or $f[x(\lambda)]$ for a functional.

Formally, a function is a mapping from the reals to the reals,

$$f(x) : R \to R$$

while a functional is a maping from a space of functions to the reals. Let $\mathcal{F}$ be a function space, for example, the set of all bounded, real-valued functions $x(\lambda)$, on the unit interval,

$$\mathcal{F} = \{x(\lambda) \,|\lambda \in [0,1], |x(\lambda)| < \infty\}$$

Then a functional is a mapping from the function space to the reals

$$f[x] : \mathcal{F} \to R$$

Integrals are linear functions on curves, in the sense that they are additive in the parameter. Nonlinear functions of a curve are annoying, nonlocal beasts such as

$$f[x] = \int x(\lambda)\, x'\left(\lambda^2\right) d\lambda$$

While such integrals are uncommon, nonlocal functions are not. For example, we will later consider the well-known iterative map of a continuous varible n, satisfying

$$x_{n+1} = a\left(1 + bx_n^2\right)$$

which may be extended to a function X,

$$X(n+1) = a\left(1 + bX^2(n)\right)$$

Most functionals that arise in physics take the form of integrals over some function of the curve

$$f[x] = \int_0^1 L\left(x, \frac{dx}{d\lambda}, \frac{d^2x}{d\lambda^2}, \ldots \frac{d^nx}{d\lambda^n}\right) d\lambda$$

where L is a known function of n variables. Our simple example of the length of a curve given above takes this form with

$$\begin{aligned} L(y') &= \sqrt{1 + (y')^2} \\ \frac{df}{dx} &= \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon} \end{aligned}$$

The problem is that for a functional the denominator is not well defined. If we were to set

$$\frac{\delta f[x(\lambda)]}{\delta x(\lambda)} = \lim_{h(\lambda) \to 0} \frac{f[x(\lambda) + h(\lambda)] - f[x(\lambda)]}{h(\lambda)}$$

To find the function at which a given functional is extremal, we expect to impose the condition of vanishing derivative of the functional with respect to the function. But we need to define the derivative of a functional. Denoting the *functional deravitive* by

$$\frac{\delta f\left[x\left(\lambda\right)\right]}{\delta x(\lambda)}$$

we first try a direct analogy with the usual limit of calculus,

$$\frac{df}{dx}=\lim_{\varepsilon\to 0}\frac{f\left(x+\varepsilon\right)-f\left(x\right)}{\varepsilon}$$

The problem is that for a functional the denominator is not well defined. If we were to set

$$\frac{\delta f\left[x\left(\lambda\right)\right]}{\delta x(\lambda)}=\lim_{h(\lambda)\to 0}\frac{f\left[x\left(\lambda\right)+h\left(\lambda\right)\right]-f\left[x\left(\lambda\right)\right]}{h\left(\lambda\right)}$$

we would have to restrict $h\left(\lambda\right)$ to be everywhere nonzero and carefully specify its uniform approach to zero. To avoid this, we change the argument a little bit and look at just the numerator. Let $h(x)$ be any continuous, bounded function that vanishes at the endpoints, $x\left(A\right)$ and $x\left(B\right)$, but is otherwise arbitrary and set

$$y(x)=y_0(x)+\alpha h(x)$$

where $\mathrm{y}_0(x)$ is some fixed function. For the moment, let $h\left(x\right)$ be considered small. Then we can look at the difference $f\left[y\left(x\right)\right]-f\left[y_0\left(x\right)\right]$as a power series in $h\left(x\right)$. At each value of x, expand f$[y\left(x\right)]$ in a Taylor series around $f\left[y_0\left(x\right)\right]$, and identify the result with teh formal expression,

$$f\left[y\left(x\right)\right]=f\left[y_0\left(x\right)\right]+\frac{\delta f\left[y\left(x\right)\right]}{\delta y(x)}h(x)+\ldots$$

Let's see how this works with our length functional, $s\left[y\left(\lambda\right)\right]=s=\int_0^1 d\lambda\sqrt{1+\left(y'\right)^2}$ Setting $y\left(\lambda\right)=y_0\left(\lambda\right)+h\left(\lambda\right)$, we define the variation of s,

$$\begin{aligned}\delta s &\equiv s\left[y\left(x\right)\right]-s\left[y_0\left(x\right)\right]\\ &= \int_0^1 d\lambda\sqrt{1+\left(y_0'+h'\right)^2}-\int_0^1 d\lambda\sqrt{1+\left(y_0'\right)^2}\end{aligned}\tag{2}$$

Since the integrand is an ordinary function, we can expand the square root. This works as long as we treat h' as small

$$\begin{aligned}\sqrt{1+\left(y_0'+h'\right)^2} &= \sqrt{1+\left(y_0'\right)^2+2y_0'h'+\left(h'\right)^2}\\ &= \sqrt{1+\left(y_0'\right)^2}\sqrt{1+\frac{2y_0'h'+\left(h'\right)^2}{1+\left(y_0'\right)^2}}\\ &= \sqrt{1+\left(y_0'\right)^2}\left(1+\frac{y_0'h'}{1+\left(y_0'\right)^2}+\cdots\right)\end{aligned}$$

where we omit terms of order $\left(h'\right)^2$ and higher. At this order,

$$\begin{aligned}\delta s &= \int_0^1 d\lambda\sqrt{1+\left(y_0'\right)^2}\left(1+\frac{y_0'h'}{1+\left(y_0'\right)^2}-1\right)\\ &= \int_0^1 d\lambda\frac{y_0'h'}{\sqrt{1+\left(y_0'\right)^2}}\end{aligned}$$

The expression depends on the derivative of the arbitrary function h, but there is no reason not to integrate by parts:

$$\begin{aligned}\delta s_{AB} &= \int_0^1 d\lambda \left(\frac{d}{d\lambda} \left(\frac{y_0' h}{\sqrt{1+(y_0')^2}} \right) - h \frac{d}{d\lambda} \left(\frac{y_0'}{\sqrt{1+(y_0')^2}} \right) \right) \\ &= \left. \frac{y_0' h}{\sqrt{1+(y_0')^2}} \right|_{\lambda=1} - \left. \frac{y_0' h}{\sqrt{1+(y_0')^2}} \right|_{\lambda=0} \\ &\quad - \int_0^1 d\lambda h(\lambda) \frac{d}{d\lambda} \left(\frac{y_0'}{\sqrt{1+(y_0')^2}} \right)\end{aligned}$$

Since $h(0) = h(1) = 0$, the boundary terms vanish, so,

$$\delta s = - \int_0^1 d\lambda \; h(\lambda) \frac{d}{d\lambda} \left(\frac{y_0'}{\sqrt{1+(y_0')^2}} \right)$$

This relation must hold for any sufficiently small function $h(x)$. Since $h(x)$ represents a small change, δy in the function $y(\lambda)$, we are justified in writing,

$$\delta s = - \int_0^1 d\lambda \; \delta y \frac{d}{dx} \left(\frac{y_0'}{\sqrt{1+(y_0')^2}} \right)$$

so the functional derivative we want,

$$\frac{\delta s}{\delta y}$$

seems to be athand. But how do we get at it? We still have

$$\delta s$$

written as an integral and it contains an arbitrary function, $y = h()$.

How can we extract information about $y(\lambda)$? In particular, what is $\frac{\delta s}{\delta y}$? First, consider the problem of curves of extremal length. For such curves we expect $\frac{\delta s}{\delta y}$ and therefore δs to vanish for any variation δy. Therefore, demand

$$- \int_0^1 d\lambda \; \delta y \frac{d}{d\lambda} \left(\frac{y_0'}{\sqrt{1+(y_0')^2}} \right) = 0$$

for all (sufficiently small) functions $\delta y = h(\lambda)$ Now let $\lambda_0 \in \{0, 1\}$ and choose $h(\lambda)$ of the form

$$h(\lambda) = \begin{cases} h(\lambda) > 0 & \lambda \in (\lambda_0 - \varepsilon/2, \lambda_0 + \varepsilon/2) \\ h(\lambda) = 0 & otherwise \end{cases}$$

Then the integral becomes

$$\begin{aligned} 0 &= - \int_0^1 d\lambda \; h(\lambda) \frac{d}{d\lambda} \left(\frac{y_0'}{\sqrt{1+(y_0')^2}} \right) \\ &= - \int_{\lambda_0 - \varepsilon/2}^{\lambda_0 + \varepsilon/2} d\lambda \; h(\lambda) \frac{d}{d\lambda} \left(\frac{y_0'}{\sqrt{1+(y_0')^2}} \right)\end{aligned}$$

In the limit as $\varepsilon \to 0$ then integral approaches

$$\lim_{\varepsilon\to 0}\int_{\lambda_0-\varepsilon/2}^{\lambda_0+\varepsilon/2} d\lambda\ h(\lambda)\frac{d}{d\lambda}\left(\frac{y_0'}{\sqrt{1+(y_0')^2}}\right) = \varepsilon h(\lambda_0)\left[\frac{d}{d\lambda}\left(\frac{y_0'}{\sqrt{1+(y_0')^2}}\right)\right]_{\lambda=\lambda_0}$$

so vanishing δs requires

$$0 = \varepsilon h(\lambda_0)\left[\frac{d}{d\lambda}\left(\frac{y_0'}{\sqrt{1+(y_0')^2}}\right)\right]_{\lambda=\lambda_0}$$

Since $\varepsilon h\left(\lambda_0\right) > 0$, we may divide it out, leaving

$$\left[\frac{d}{d\lambda}\left(\frac{y_0'}{\sqrt{1+(y_0')^2}}\right)\right]_{\lambda=\lambda_0} = 0$$

But the point $\lambda_0$was arbitrary, so we can drop the $\lambda = \lambda_0$ condition. The argument holds at every point of the interval. Therefore, the function $y_0\left(\lambda\right)$ that makes the length extremal must satisfy

$$\frac{d}{d\lambda}\left(\frac{y_0'}{\sqrt{1+(y_0')^2}}\right) = 0$$

This is just what we need, if it really gives the condition we want! Integrating once gives

$$\frac{y_0'}{\sqrt{1+(y_0')^2}} = c = const.$$

so that

$$\begin{aligned} y_0' &= \sqrt{\frac{c}{1-c}} \equiv b = const. \\ y_0(\lambda) &= a + b\lambda \end{aligned}$$

Finally, with initial conditions $y_0\left(0\right) = y\left(A\right)$ and $y_0\left(1\right) = B$,

$$\begin{aligned} y_A &= a \\ y_B &= a + b \end{aligned}$$

we find

$$y_0(\lambda) = y_A + \left(y_B - y_A\right)\lambda$$

This is the unique straight line connection $(0, y_A)$ and $(1, y_B)$.

Find the form of the function $x\left(\lambda\right)$ that makes the variation, δf, of each of the following functionals vanish:

1. $f[x] = \int_0^\lambda \dot{x}^2 d\lambda$ where $\dot{x} = \frac{dx}{d\lambda}$.
2. $f[x] = \int_0^\lambda \left(\dot{x}^2 + x^2\right) d\lambda$
3. $f[x] = \int_0^\lambda \left(\dot{x}^n - x^m\right) d\lambda$
4. $f[x] = \int_0^\lambda \left(\dot{x}^2(\lambda) - V(x(\lambda))\right) d\lambda$ where $V(x)$ is an arbitrary function of $x(\lambda)$.
5. $f[x] = \int_0^\lambda \dot{\mathbf{x}}^2 d\lambda$ where $\dot{\mathbf{x}} = (\dot{x}, \dot{y}, \dot{z})$ and $\dot{\mathbf{x}}^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2$. Notice that there are *three* functions to be determined here: $x(\lambda)$, $y(\lambda)$ and $z(\lambda)$.

2.3.2 Formal definition of functional differentiation

There are several reasons to go beyond the simple variation of the previous section. First, we have only extracted a meaningful result in the case of vanishing variation. We would like to define a functional derivative that exists even when it does not vanish! Second, we would like a procedure that can be applied iteratively to define higher functional derivatives. Third, the procedure can become difficult or impossible to apply for functionals more general than curve length. It is especially important to have a definition that will generalize to field theory. Finally, the added rigor of a formal definition allows careful proofs of the properties of functionals, including generalizations to a complete functional calculus.

In this section we look closely at the calculation of the previous section to formally develop the *functional derivative*, the generalization of differentiation to functionals. We will give a general definition sufficient to find $\frac{\delta f}{\delta x}$ for any functional of the form

$$f[x] = \int_0^1 L\left(x, \frac{dx}{d\lambda}, \frac{d^2x}{d\lambda^2}, \ldots \frac{d^nx}{d\lambda^n}\right) d\lambda$$

One advantage of treating variations in this more formal way is that we can equally well apply the technique to relativistic systems and classical field theory.

There are two principal points to be clarified:

1. We would like to allow fully arbitrary variations, $h(\lambda)$.
2. We require a rigorous way to remove the integral, even when the variation, δf, does not vanish.

We treat these points in turn.

Arbitrary variations: one parameter families of functions We would like the functional derivative to formalize finding the extremum of a functional. Suppose we have a functional, $f[x(\lambda)]$. We argued that we can look at the difference between $f[x(\lambda) + h(\lambda)]$ and $f[x(\lambda)]$, demanding that for an extremum,

$$\delta f \equiv f[x+h] - f[x] = 0 \tag{3}$$

to first order in $h(\lambda)$. We want to generalize this to a definition of a functional derivative,

$$\frac{\delta f[x(\lambda)]}{\delta x(\lambda)} \tag{4}$$

such that at extrema, $\frac{\delta f[x(\lambda)]}{\delta x(\lambda)} = 0$ is equivalent to vanishing variation of $f[x]$.

Let f be given by

$$f[x(\lambda)] = \int_0^1 L(x(\lambda), x^{(1)}(\lambda), \ldots, x^{(n)}(\lambda))d\lambda \tag{5}$$

where L is a *function* of $x(\lambda)$ and its derivatives up to some maximum order, n. Because L is a function, we found δf by expanding $f[x(\lambda) + h(\lambda)]$ to first order in $h(\lambda)$. We can get the same result without making any assumptions about $h(x)$ by introducing a parameter α along with h. Thus, we write

$$f[x(\lambda, \alpha)] \equiv f[x(\lambda) + \alpha h(\lambda)] \tag{6}$$

and think of $x(\lambda, \alpha)$ as a function of two variables for any given $h(\lambda)$. Equivalently, we may think of $x(\lambda, \alpha)$ as a 1-parameter family of functions, one for each value of α. Then, although $f[x(\lambda) + \alpha h(\lambda)]$ is a functional of $x(\lambda)$ and $h(\lambda)$, it is a simple *function* of α:

$$\begin{aligned} f[x(\lambda, \alpha)] &= \int L\left(x(\lambda, \alpha), x^{(1)}(\lambda, \alpha), \ldots, x^{(n)}(\lambda, \alpha)\right) d\lambda \\ &= \int L\left(x(\lambda, \alpha) + \alpha h(\lambda, \alpha), x^{(1)} + \alpha h^{(1)}, \ldots\right) d\lambda \end{aligned} \tag{7}$$

Moreover, taking a regular derivative with respect to α then setting $\alpha = 0$ gives the same expression as the first term of a Taylor series in small $h(x)$. This happens because the only dependence f has on α is through x :

$$\begin{aligned}
\frac{d}{d\alpha} f\left[x(\lambda,\alpha)\right] &= \frac{d}{d\alpha}\int L\left(x\left(\lambda,\alpha\right), x^{(1)}\left(\lambda,\alpha\right), \ldots\right) d\lambda \\
&= \int\left(\frac{\partial L}{\partial x}\frac{\partial x}{\partial \alpha} + \frac{\partial L}{\partial x^{(1)}}\frac{\partial x^{(1)}}{\partial \alpha} + \ldots + \frac{\partial L}{\partial x^{(n)}}\frac{\partial x^{(n)}}{\partial \alpha}\right) d\lambda \qquad (8) \\
&= \int\left(\frac{\partial L}{\partial x}h + \frac{\partial L}{\partial x^{(1)}}h^{(1)} + \ldots + \frac{\partial L}{\partial x^{(n)}}h^{(n)}\right) d\lambda \qquad (9)
\end{aligned}$$

so that when we set $\alpha = 0$, all dependence on $x\left(\lambda\right) + \alpha h\left(\lambda\right)$ reduces to dependence on $x\left(\lambda\right)$ only. We therefore define the variation of $f\left[x\right]$ as

$$\begin{aligned}
\delta f\left[x\left(\lambda\right)\right] &\equiv \left(\frac{d}{d\alpha} f\left[x(\lambda,\alpha)\right]\right)\Big|_{\alpha=0} \\
&= \int\left(\frac{\partial L\left(x\left(\lambda,\alpha\right)\right)}{\partial x}h + \ldots + \frac{\partial L\left(x\left(\lambda,\alpha\right)\right)}{\partial x^{(n)}}h^{(n)}\right)\Big|_{\alpha=0} d\lambda \\
&= \int\left(\frac{\partial L\left(x\left(\lambda\right)\right)}{\partial x}h + \frac{\partial L\left(x\left(\lambda\right)\right)}{\partial x^{(1)}}h^{(1)} + \ldots + \frac{\partial L\left(x\left(\lambda\right)\right)}{\partial x^{(n)}}h^{n}\right) d\lambda
\end{aligned}$$

Notice that all dependence of L on h has dropped out, so the expression is linear in $h\left(\lambda\right)$ and its derivatives.

Now continue as before. Integrate the $h^{(1)}$ and higher derivative terms by parts. We assume that $h\left(\lambda\right)$ and its first $n-1$ derivatives all vanish at the endpoints. The k^{th} term becomes

$$\int_0^1 \frac{\partial L}{\partial x^{(k)}}\Big|_{x(\lambda,\alpha)=x(\lambda)} h^{(k)}\left(\lambda\right) d\lambda = \left(-1\right)^k \int_0^1 d\lambda \frac{d^k}{d\lambda^k}\left(\frac{\partial L}{\partial x^{(k)}}\Big|_{x=x(\lambda)}\right) h\left(\lambda\right)$$

so we have

$$\begin{aligned}
\delta f\left[x\left(\lambda\right)\right] &= \left(\frac{d}{d\alpha} f\left[x(\lambda,\alpha)\right]\right)\Big|_{\alpha=0} \\
&= \int_0^1 \left(\frac{\partial L\left(x\left(\lambda\right)\right)}{\partial x} - \frac{d}{d\lambda}\frac{\partial L\left(x\left(\lambda\right)\right)}{\partial x^{(1)}}\right. \\
&\quad \left. + \ldots + \left(-1\right)^n \frac{d^n}{d\lambda^n}\frac{\partial L\left(x\left(\lambda\right)\right)}{\partial x^{(n)}} h\left(\lambda\right)\right) d\lambda \qquad (10)
\end{aligned}$$

where now, $h\left(\lambda\right)$ is fully arbitrary.

Fill in the details of the preceeding derivation, paying particular attention to the integrations by parts.

The functional derivative We finish the procedure by generalizing the way we extract the term in parentheses from the integrand of eq.(10), without demanding that the functional derivative vanish. Recall that for the straight line, we argued that we can choose a sequence of functions $h(x)$ that vanish everywhere except in a small region around a point x_0. It is this procedure that we formalize. The required rigor is provided by the Dirac delta function.

The Dirac delta function To do this correctly, we need yet another class of new objects: distributions. Distributions are objects whose *integrals* are wel-defined though their values may not be. They may be identified with infinite sequences of functions. Therefore, distributions include functions, but include other objects as well.

The development of the theory of distributions is similar in some ways to the definition of real numbers as limits of Cauchy sequences of rational numbers. For example, we can think of π as the limit of the sequence of rationals

$$3, \frac{31}{10}, \frac{314}{100}, \frac{3141}{1000}, \frac{31415}{10000}, \frac{314159}{100000}, \ldots$$

By considering all sequences of rationals, we move to the larger class of numbers called the reals. Similarly, we can consider sequences of functions,

$$f_1(x), f_2(x), \ldots$$

The limit of such a sequence is not necessarily a function. For example, the sequence of functions

$$\frac{1}{2}e^{-|x|}, e^{-2|x|}, \frac{3}{2}e^{-3|x|}, \ldots, \frac{m}{2}e^{-m|x|}, \ldots$$

vanishes at nonzero x and diverges at $x = 0$ as $m \to \infty$. Nonetheless, the integrals of these functions, on the interval $[-\infty, \infty]$ is the independent of n,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{m}{2}e^{-m|x|} &= \int_0^{\infty} me^{-mx} \\ &= -e^{-mx}|_0^{\infty} \\ &= 1 \end{aligned}$$

so the integral of the limit function is well defined.

Another particularly useful example is a sequence of Gaussians which, like the previous example, becomes narrower and narrower while getting taller and taller:

$$f_m(x) = \sqrt{\frac{m}{2\pi}}e^{-\frac{m^2x^2}{2}}$$

As m increases, the width of the Gaussian decreases to $\frac{1}{m}$ while the maximum increases to $\sqrt{\frac{m}{2\pi}}$. Notice that the area under the curve is always 1, regardless of how large we make m. In the limit as $m \to \infty$, the width decreases to zero while the maximum becomes infinite – clearly not a function. However, the limit still defines a useful object – the distribution known as the Dirac delta function.

Let $f(x)$ be any function. The Dirac delta function, $\delta(x)$ is defined by the property

$$\int f(x)\delta(x)dx = f(0)$$

In particular, this means that

$$\int \delta(x)dx = 1$$

Heuristically, we can think of $\delta(x)$ as being zero everywhere but where its argument (in this case, simply x) vanishes. At the point $x = 0$, its value is sufficiently divergent that the integral of $\delta(x)$ is still one.

Formally we can define $\delta(x)$ as the limit of a sequence of functions such as $f_n(x)$, but note that there are many different sequences that give the same properties. To prove that the sequence f_n works we must show that

$$\lim_{m\to\infty} \int f(x)\sqrt{\frac{m}{2\pi}}e^{-\frac{m^2x^2}{2}}dx = f(0)$$

for an arbitrary function $f(x)$. Then it is correct to write

$$\delta(x) = \lim_{m\to\infty} \sqrt{\frac{m}{2\pi}}e^{-\frac{m^2x^2}{2}}$$

This proof is left as an exercise.

We don't usually need a specific sequence to solve problems using delta functions, because they are only meaningful under integral signs and they are very easy to integrate using their defining property. In physics they are extremely useful for making "continuous" distributions out of discrete ones. For example, suppose we have a point charge Q located at $\mathbf{x}_0 = (x_0, y_0, z_0)$. Then we can write a corresponding charge density as

$$\rho(\mathbf{x}) = Q\delta(\mathbf{x} - \mathbf{x}_0) = Q\delta(x - x_0)\,\delta(y - y_0)\,\delta(z - z_0)$$

Perform the following integrals over Dirac delta functions:

1. $\int f(x)\delta(x - x_0)\,dx$
2. $\int f(x)\delta(ax)\,dx$ (Hint: By integrating over both expressions, show that $\delta(ax) = \frac{1}{a}\delta(x)$. Note that to integrate $\delta(ax)$ you need to do a change of variables.)
3. Evaluate $\int f(x)\delta^{(n)}(x)\,dx$ where
$$\delta^{(n)}(x) = \frac{d^n}{dx^n}\delta(x)$$
4. $\int f(x)\delta\left(x^2 - x_0^2\right)dx$ (This is tricky!)

Show that

$$\delta(x) = \lim_{m\to\infty}\sqrt{\frac{m}{2\pi}}e^{-\frac{m^2x^2}{2}}$$

Show that

$$\delta(x) = \lim_{m\to\infty}\frac{m}{2}e^{-m|x|}$$

Let $f(x)$ be a differentiable function with zeros at $x_1, \ldots, x_n$. Assume that at any of these zeros, the derivative $f'(x_i)$ is nonzero. Show that

$$\delta(f(x)) = \sum_{i=1}^{n}\frac{1}{|f'(x_i)|}\delta(x - x_i)$$

The mass density of an infinitesimally thin, spherical shell of radius R may be written as

$$\rho(r, \theta, \varphi) = \frac{M}{4\pi R^2}\delta(r - R)$$

By integrating ρ over all space, find the total mass of the shell.

Write an expression for the mass density of an infinitesimally thin ring of matter of radius R, lying in the xy plane.

The definition at last! Using the idea of a sequence of functions and the Dirac delta function, we can now extract the variation. So far, we have defined the variation as

$$\delta f[x(\lambda)] \equiv \left(\frac{d}{d\alpha}f[x(\lambda, \alpha)]\right)\Bigg|_{\alpha=0} \tag{11}$$

where $x(\lambda, \alpha) = x(\lambda) + \alpha h(\lambda)$, and shown that the derivative on the right reduces to

$$\begin{aligned}\delta f[x(\lambda)] &= \int_0^1 \left(\frac{\partial L(x(\lambda))}{\partial x} - \frac{d}{d\lambda}\frac{\partial L(x(\lambda))}{\partial x^{(1)}}\right.\\ &\quad \left. + \ldots + (-1)^n\,\frac{d^n}{d\lambda^n}\frac{\partial L(x(\lambda))}{\partial x^{(n)}}\right)h(\lambda)\,d\lambda\end{aligned} \tag{12}$$

where $h(\lambda)$ is arbitrary. In particular, we may choose for $h(\lambda)$ a sequence of functions, $h_m(\lambda-\alpha)$, such that

$$\lim_{m\to\infty} h_m(\lambda-\beta) = \delta(\lambda-\beta)$$

where β is fixed. Then defining

$$x_m(\lambda,\alpha,\beta) \equiv x(\lambda) + \alpha h_m(\lambda-\beta)$$

the variation takes the form

$$\begin{aligned}
\delta f[x_m(\lambda,\beta)] &= \left[\frac{d}{d\alpha} f[x_m(\lambda,\alpha,\beta)]\right]_{\alpha=0} \\
&= \int_0^1 \left(\frac{\partial L(x(\lambda))}{\partial x} - \dots \right. \\
&\quad \left. + (-1)^n \frac{d^n}{d\lambda^n}\frac{\partial L(x(\lambda))}{\partial x^{(n)}}\right) h_m(\lambda-\beta)\, d\lambda
\end{aligned} \tag{13}$$

The functional derivative is now defined as follows.

The functional derivative of $f[x(\beta)]$ is defined as

$$\frac{\delta f[x(\beta)]}{\delta x(\beta)} \equiv \lim_{m\to\infty}\left[\frac{d}{d\alpha} f[x_m(\lambda,\alpha,\beta)]\right]_{\alpha=0} \tag{14}$$

where

$$\begin{aligned}
x_m(\lambda,\alpha,\beta) &\equiv x(\lambda) + \alpha h_m(\lambda-\beta) \\
\lim_{m\to\infty} h_m(\lambda-\beta) &= \delta(\lambda-\beta)
\end{aligned}$$

Since we have chosen $h_m(\lambda-\beta)$ to approach a Dirac delta function, and since nothing except $h_m(\lambda-\beta)$ on the right hand side of eq.(13) depends on m, we have

$$\begin{aligned}
\frac{\delta f[x(\beta)]}{\delta x(\beta)} &\equiv \lim_{m\to\infty}\left[\frac{d}{d\alpha} f[x_m(\lambda,\alpha,\beta)]\right]_{\alpha=0} \\
&= \int_0^1 \left(\frac{\partial L(x(\lambda))}{\partial x}\right. \\
&\quad \left. - \dots + (-1)^n \frac{d^n}{d\lambda^n}\left(\frac{\partial L(x(\lambda))}{\partial x^{(n)}}\right)\right) \lim_{m\to\infty} h_m(\lambda-\beta)\, d\lambda \\
&= \int_0^1 \left(\frac{\partial L(x(\lambda))}{\partial x} - \dots + (-1)^n \frac{d^n}{d\lambda^n}\left(\frac{\partial L(x(\lambda))}{\partial x^{(n)}}\right)\right) \delta(\lambda-\beta)\, d\lambda \\
&= \frac{\partial L(x(\beta))}{\partial x} - \dots + (-1)^n \frac{d^n}{d\lambda^n}\left(\frac{\partial L(x(\lambda))}{\partial x^{(n)}}\right)
\end{aligned}$$

This expression holds for any value of β. If we are seeking an extremum of $f[x(\lambda)]$, then we set the functional derivative to zero:

$$\frac{\delta f[x(\lambda)]}{\delta x(\lambda)} = 0$$

and recover the same result as we got from vanishing variation. The condition for the extremum is the *generalized Euler-Lagrange equation*,

$$\left.\frac{\partial L}{\partial x}\right|_{x(\lambda)} - \frac{d}{d\lambda}\left.\frac{\partial L}{\partial x^{(1)}}\right|_{x(\lambda)} + \dots + (-1)^n \frac{d^n}{d\lambda^n}\left.\frac{\partial L}{\partial x^{(n)}}\right|_{x(\lambda)} = 0 \tag{15}$$

This equation is the central result this chapter. Consider, for a moment, what we have accomplished. Any functional of the form

$$S\left[\mathbf{x}\left(t\right)\right]=\int_{C}L\left(\mathbf{x},\dot{\mathbf{x}},\ldots\right)dt$$

is a mapping that assigns a real number to any curve. Viewing motion as described by a set of uniform curves, we now have a way to specify what is meant by uniform – a uniform curve is one which makes the functional S extremal. We have seen that this makes sense if we take S to be the length of the curve. The extremals are the straight lines.

The integrand L of the action functional is called the *Lagrangian*. In general, once we choose a Lagrangian, $L\left(\mathbf{x},\dot{\mathbf{x}},\ldots\right)$, eq.(15) determines the preferred paths of motion associated with L. What we now must do is to find a way to associate particular forms of L with particular physical systems in such a way that the extremal paths are the paths the physical system actually follows.

In subsequent chapters we will devote considerable attention to the generalized Euler-Lagrange equation. But before trying to associate any particular form for L with particular physical systems, we will prove a number of general results. Most of these have to do with symmetries and conservation laws. We shall see that many physical laws hold by virtue of very general properties of S, and not the particular form. Indeed, there are cases where it possible to write infinitely many distinct functionals which all have the same extremals! Until we reach our discussion of gauge theory, we will have no reason to prefer any one of these descriptions over another. Until then we take the pragmatic view that any functional whose extremals give the correct equation of motion is satisfactory.

Using the formal definition, eq.(14) find the functional derivative of each of the following functionals.

1. $f\left[x\right]=\int_{0}^{t}\dot{x}^{2}dt$ where $\dot{x}=\frac{dx}{dt}$.
2. $f\left[x\right]=\int_{0}^{t}\left(\dot{x}^{2}(t)-V(x(t))\right)dt$ where $V(x)$ is an arbitrary function of $x\left(t\right)$.

Define the functional Taylor series by

$$\begin{aligned}\delta f &= \int d\lambda\frac{\delta f}{\delta x}\delta x \\ &\quad +\frac{1}{2}\int\int d\lambda_{1}d\lambda_{2}\frac{\delta^{2}f}{\delta x\left(\lambda_{2}\right)\delta x\left(\lambda_{1}\right)}\delta x\left(\lambda_{1}\right)\delta x\left(\lambda_{2}\right)+\cdots\end{aligned}$$

and let $f\left[x\left(\lambda\right)\right]=\int L\left(x,x'\right)d\lambda$. Using the techniques of this section, show that the first and second functional derivatives are given by

$$\begin{aligned}\frac{\delta f}{\delta x} &= \frac{\partial L}{\partial x}-\frac{d}{d\lambda}\frac{\partial L}{\partial\dot{x}} \\ \frac{\delta^{2}f}{\delta x\left(\lambda_{1}\right)\delta x\left(\lambda_{2}\right)} &= \left(\frac{\partial^{2}L}{\partial x\partial x}-\frac{d}{d\lambda_{1}}\frac{\partial^{2}L}{\partial x\partial x'}\right)\delta\left(\lambda_{1}-\lambda_{2}\right) \\ &\quad -\frac{d}{dt}\left(\frac{\partial^{2}L}{\partial\dot{x}\partial\dot{x}}\right)\frac{\partial\delta\left(\lambda_{1}-\lambda_{2}\right)}{\partial\lambda_{1}} \\ &\quad -\frac{\partial^{2}L}{\partial x'\partial x'}\frac{\partial^{2}\delta\left(\lambda_{1}-\lambda_{2}\right)}{\partial\lambda_{1}^{2}}\end{aligned}$$

To define higher order derivatives, we can simply apply the definition recursively. For example, the second functional derivative is given by applying the definition twice. The first functional derivative is

$$\frac{\delta f\left[x\left(\beta\right)\right]}{\delta x\left(\beta\right)} \equiv \lim_{m\to\infty}\left[\frac{d}{d\alpha}f\left[x_{m}\left(\lambda,\alpha,\beta\right)\right]\right]_{\alpha=0}$$

Now let $f\left[x\left(\beta\right)\right]=\frac{\delta g[x(\beta)]}{\delta x}$,

$$\frac{\delta^{2}g\left[x\left(\beta\right)\right]}{\delta x^{2}} \equiv \lim_{m\to\infty}\left[\frac{d}{d\alpha}\left(\lim_{n\to\infty}\left[\frac{d}{d\gamma}g\left[x_{n}\left(\lambda,\gamma,\beta\right)\right]\right]_{\gamma=0}\right)\right]_{\alpha=0}$$

Suppose $g(x(\beta)) = \int h(x, x')\, dx$. Then

$$f[x(\beta)] = \frac{\delta g[x(\beta)]}{\delta x} = \frac{\partial h}{\partial x} - \frac{d}{d\beta}\frac{\partial h}{\partial x'}$$

and the second derivative gives

$$\frac{\delta f[x(\beta)]}{\delta x(\beta)} \equiv \lim_{m\to\infty}\left[\frac{d}{d\alpha} f[x_m(\lambda, \alpha, \beta)]\right]_{\alpha=0}$$

where

$$\begin{aligned} x_m(\lambda, \alpha, \beta) &\equiv x(\lambda) + \alpha h_m(\lambda - \beta) \\ \lim_{m\to\infty} h_m(\lambda - \beta) &= \delta(\lambda - \beta) \end{aligned}$$

Substituting,

$$\begin{aligned} \frac{\delta^2 g[x(\beta)]}{\delta x^2} &= \lim_{m\to\infty}\left[\frac{d}{d\alpha} f[x_m(\lambda, \alpha, \beta)]\right]_{\alpha=0} \\ &= \lim_{m\to\infty}\left[\frac{d}{d\alpha}\left(\frac{\partial h}{\partial x} - \frac{d}{d\beta}\frac{\partial h}{\partial x'}\right)\right]_{\alpha=0} \\ &= \lim_{m\to\infty}\left[\frac{d}{d\alpha}\left(\frac{\partial^2 h}{\partial x^2}h_m + \frac{\partial^2 h}{\partial x \partial x'}h'_m - h_m\frac{\partial}{\partial x}\left(\frac{d}{d\beta}\frac{\partial h}{\partial x'}\right) - h'_m\frac{\partial}{\partial x'}\left(\frac{d}{d\beta}\frac{\partial h}{\partial x'}\right)\right)\right]_{\alpha=0} \end{aligned}$$

We need to expand

$$\frac{d}{d\beta}\frac{\partial h}{\partial x'} = x"\frac{\partial^2 h}{\partial x'^2} + x'\frac{\partial^2 h}{\partial x' \partial x}$$

Then

$$\begin{aligned} \frac{\partial}{\partial x}\left(\frac{d}{d\beta}\frac{\partial h}{\partial x'}\right) &= \frac{\partial}{\partial x}\left(x"\frac{\partial^2 h}{\partial x'^2} + x'\frac{\partial^2 h}{\partial x' \partial x}\right) \\ &= x"\frac{\partial^3 h}{\partial x \partial x'^2} + x'\frac{\partial^3 h}{\partial x' \partial x^2} \end{aligned}$$

and similarly,

$$\begin{aligned} \frac{\partial}{\partial x'}\left(\frac{d}{d\beta}\frac{\partial h}{\partial x'}\right) &= \frac{\partial}{\partial x'}\left(x"\frac{\partial^2 h}{\partial x'^2} + x'\frac{\partial^2 h}{\partial x' \partial x}\right) \\ &= x"\frac{\partial^3 h}{\partial x'^3} + x'\frac{\partial^3 h}{\partial x'^2 \partial x} + \frac{\partial^2 h}{\partial x' \partial x} \end{aligned}$$

Substituting,

$$\begin{aligned} \frac{\delta^2 g_m[\alpha, x(\beta)]}{\delta x^2} &= \frac{d}{d\alpha}\frac{\partial^2 h}{\partial x^2}h_m + \frac{d}{d\alpha}\frac{\partial^2 h}{\partial x \partial x'}h'_m - h_m\frac{d}{d\alpha}\left(x"\frac{\partial^3 h}{\partial x \partial x'^2} + x'\frac{\partial^3 h}{\partial x' \partial x^2}\right) \\ &\quad -h'_m\frac{d}{d\alpha}\left(x"\frac{\partial^3 h}{\partial x'^3} + x'\frac{\partial^3 h}{\partial x'^2 \partial x} + \frac{\partial^2 h}{\partial x' \partial x}\right) + \frac{\partial^2 h}{\partial x^2}h'_m + \frac{\partial^2 h}{\partial x \partial x'}h"_m - h'_m\left(x"\frac{\partial^3 h}{\partial x \partial x'^2} + x'\frac{\partial^3 h}{\partial x' \partial x^2}\right) \\ &\quad -h"_m\left(x"\frac{\partial^3 h}{\partial x'^3} + x'\frac{\partial^3 h}{\partial x'^2 \partial x} + \frac{\partial^2 h}{\partial x' \partial x}\right) \end{aligned}$$

To define higher order derivatives, we can simply carry the expansion of L to higher order.

$$S = \int L(x, x')\, d\lambda$$

$$\begin{aligned}
\delta S &= \int L\left(x+\delta x, x'+\delta x'\right) d\lambda - \int L\left(x, x'\right) d\lambda \\
&= \int L\left(x, x'\right) + \frac{\partial L}{\partial x}\delta x + \frac{\partial L}{\partial \dot{x}}\delta x' + \frac{1}{2!}\frac{\partial^2 L}{\partial x \partial x}\left(\delta x\right)^2 \\
&\quad + \frac{\partial^2 L}{\partial x \partial \dot{x}}\delta x \delta x' + \frac{1}{2!}\frac{\partial^2 L}{\partial x' \partial x'}\left(\delta x'\right)^2 - L\left(x, x'\right) dt \\
&= \int \frac{\partial L}{\partial x}\delta x + \frac{\partial L}{\partial x'}\delta x' + \frac{1}{2!}\frac{\partial^2 L}{\partial x\left(\lambda\right)\partial x\left(\lambda\right)}\left(\delta x\right)^2 \\
&\quad + \frac{\partial^2 L}{\partial x\left(\lambda\right)\partial x'\left(\lambda\right)}\delta x \delta x' + \frac{1}{2!}\frac{\partial^2 L}{\partial x'\left(\lambda\right)\partial x'\left(\lambda\right)}\left(\delta x'\right)^2
\end{aligned}$$

Now, to integrate each of the various terms by parts, we need to insert some delta functions. Look at one term at a time:

$$\begin{aligned}
\int \frac{\partial L}{\partial x'}\delta x' &= -\int \frac{d}{d\lambda}\frac{\partial L}{\partial x'}\delta x \\
I_2 &= \frac{1}{2!}\int d\lambda \frac{\partial^2 L}{\partial x \partial x}\left(\delta x\right)^2 \\
&= \frac{1}{2}\int d\lambda_1 \int d\lambda_2 \delta\left(\lambda_1 - \lambda_2\right)\frac{\partial^2 L}{\partial x \partial x}\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right) \\
I_3 &= \int d\lambda \frac{\partial^2 L}{\partial x \partial x'}\delta x \delta x' \\
&= \frac{1}{2}\int\int d\lambda_1 d\lambda_2 \delta\left(\lambda_1 - \lambda_2\right) \\
&\quad \times \frac{\partial^2 L}{\partial x \partial x'}\left(\delta x\left(\lambda_1\right)\delta x'\left(\lambda_2\right) + \delta x\left(\lambda_2\right)\delta x'\left(\lambda_1\right)\right) \\
&= -\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(\frac{d}{d\lambda_2}\left(\delta\left(\lambda_1 - \lambda_2\right)\frac{\partial^2 L}{\partial x \partial x'}\right)\right. \\
&\quad \left. + \frac{d}{d\lambda_1}\left(\delta\left(\lambda_1 - \lambda_2\right)\frac{\partial^2 L}{\partial x \partial x'}\right)\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right) \\
&= -\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(\frac{\partial}{\partial \lambda_2}\delta\left(\lambda_1 - \lambda_2\right)\frac{\partial^2 L}{\partial x \partial x'}\right. \\
&\quad + \frac{\partial}{\partial \lambda 1}\delta\left(\lambda_1 - \lambda_2\right)\frac{\partial^2 L}{\partial x \partial x'} \\
&\quad \left. + \delta\left(\lambda_1 - \lambda_2\right)\frac{d}{d\lambda_1}\frac{\partial^2 L}{\partial x \partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right) \\
&= -\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(\delta\left(\lambda_1 - \lambda_2\right)\frac{d}{d\lambda_1}\frac{\partial^2 L}{\partial x \partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)
\end{aligned}$$

and finally,

$$\begin{aligned}
I_4 &= \frac{1}{2!}\int d\lambda \frac{\partial^2 L}{\partial x'\left(\lambda\right)\partial x'\left(\lambda\right)}\left(\delta x'\right)^2 \\
&= \frac{1}{2}\int d\lambda_1 \int d\lambda_2 \delta\left(\lambda_1 - \lambda_2\right)\frac{\partial^2 L}{\partial x'\left(\lambda_2\right)\partial x'\left(\lambda_1\right)}\delta x'\left(\lambda_1\right)\delta x'\left(\lambda_2\right) \\
&= \frac{1}{2}\int d\lambda_1 \int d\lambda_2 \frac{\partial}{\partial \lambda_1}\frac{\partial}{\partial \lambda_2}\left(\delta\left(\lambda_1 - \lambda_2\right)\frac{\partial^2 L}{\partial x'\left(\lambda_1\right)\partial x'\left(\lambda_2\right)}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right) \\
&= \frac{1}{2}\int d\lambda_1 \int d\lambda_2 \frac{\partial}{\partial \lambda_1}\left(\frac{\partial}{\partial \lambda_2}\delta\left(\lambda_1 - \lambda_2\right)\frac{\partial^2 L}{\partial x' \partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(\frac{\partial^2}{\partial\lambda_1\partial\lambda_2}\delta\left(\lambda_1-\lambda_2\right)\frac{\partial^2 L}{\partial x'\partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)\\
&\quad +\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(\frac{\partial}{\partial\lambda_2}\delta\left(\lambda_1-\lambda_2\right)\frac{d}{d\lambda_1}\frac{\partial^2 L}{\partial x'\partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)\\
&= \frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(-\frac{\partial^2}{\partial\lambda_1^2}\delta\left(\lambda_1-\lambda_2\right)\frac{\partial^2 L}{\partial x'\partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)\\
&\quad +\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(-\frac{\partial}{\partial\lambda_1}\delta\left(\lambda_1-\lambda_2\right)\frac{d}{d\lambda_1}\frac{\partial^2 L}{\partial x'\partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)
\end{aligned}
$$

Combining,

$$
\begin{aligned}
\delta S &= \int d\lambda\left(\frac{\partial L}{\partial x}-\frac{d}{d\lambda}\frac{\partial L}{\partial x'}\right)\delta x\\
&\quad +\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \delta\left(\lambda_1-\lambda_2\right)\frac{\partial^2 L}{\partial x\partial x}\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)\\
&\quad -\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(\delta\left(\lambda_1-\lambda_2\right)\frac{d}{d\lambda_1}\frac{\partial^2 L}{\partial x\partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)\\
&\quad +\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(-\frac{\partial^2}{\partial\lambda_1^2}\delta\left(\lambda_1-\lambda_2\right)\frac{\partial^2 L}{\partial x'\partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)\\
&\quad +\frac{1}{2}\int d\lambda_1 \int d\lambda_2 \left(-\frac{\partial}{\partial\lambda_1}\delta\left(\lambda_1-\lambda_2\right)\frac{d}{d\lambda_1}\frac{\partial^2 L}{\partial x'\partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)\\
&= \int d\lambda\left(\frac{\partial L}{\partial x}-\frac{d}{d\lambda}\frac{\partial L}{\partial x'}\right)\delta x\\
&\quad +\int d\lambda_1 \int d\lambda_2 \frac{1}{2}\left(\delta\left(\lambda_1-\lambda_2\right)\left(\frac{\partial^2 L}{\partial x\partial x}-\frac{d}{d\lambda_1}\frac{\partial^2 L}{\partial x\partial x'}\right)\right.\\
&\quad -\frac{\partial}{\partial\lambda_1}\delta\left(\lambda_1-\lambda_2\right)\frac{d}{d\lambda_1}\left(\frac{\partial^2 L}{\partial x'\partial x'}\right)\\
&\quad \left.-\frac{\partial^2}{\partial\lambda_1^2}\delta\left(\lambda_1-\lambda_2\right)\frac{\partial^2 L}{\partial x'\partial x'}\right)\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)
\end{aligned}
$$

This agrees with DeWitt when there is only one function x. Setting

$$
\delta S = \int d\lambda \frac{\delta S}{\delta x}\delta x + \frac{1}{2}\int\int d\lambda_1 d\lambda_2 \frac{\delta^2 S}{\delta x(\lambda_1)\delta x(\lambda_2)}\delta x\left(\lambda_1\right)\delta x\left(\lambda_2\right)+\cdots
$$

we identify:

$$
\begin{aligned}
\frac{\delta S}{\delta x} &= \frac{\partial L}{\partial x}-\frac{d}{d\lambda}\frac{\partial L}{\partial x'}\\
\frac{\delta^2 S}{\delta x(\lambda_1)\delta x(\lambda_2)} &= \left(\frac{\partial^2 L}{\partial x\partial x}-\frac{d}{d\lambda_1}\frac{\partial^2 L}{\partial x\partial x'}\right)\delta\left(\lambda_1-\lambda_2\right)\\
&\quad -\frac{d}{d\lambda_1}\left(\frac{\partial^2 L}{\partial x'\partial x'}\right)\frac{\partial\delta\left(\lambda_1-\lambda_2\right)}{\partial\lambda_1}\\
&\quad -\frac{\partial^2 L}{\partial x'\partial x'}\frac{\partial^2\delta\left(\lambda_1-\lambda_2\right)}{\partial\lambda_1^2}
\end{aligned}
$$

Third and higher order derivatives may be defined by extending this procedure. The result may also be found by taking two independent variations from the start.

2.4 Functional integration

It is also possible to integrate functionals. Since a functional has an entire function as its argument, the functional integral is a sum over all functions in some well-defined function space. Such a sum is horribly uncountable, but it is still possible to perform some functional integrals exactly by taking the infinite limit of the product of finitely many normal integrals. For more difficult functional integrals, there are many approximation methods. Since functional integrals have not apppeared widely in classical mechanics, we do not treat them further here. However, they do provide one approach to quantum mechanics, and play an important role in quantum field theory.

3 Physical theories

Within any theory of matter and motion we may distinguish two conceptually different features: dynamical laws and measurement theory. We discuss each in turn.

By *dynamical laws*, we mean the description of various motions of objects, both singly and in combination. The central feature of our description is generally some set of dynamical equations. In classical mechanics, the dynamical equation is Newton's second law,

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt}$$

or its relativistic generalization, while in classical electrodynamics two of the Maxwell equations serve the same function:

$$\begin{aligned} \frac{1}{c}\frac{d\mathbf{E}}{dt} - \nabla \times \mathbf{B} &= 0 \\ \frac{1}{c}\frac{d\mathbf{B}}{dt} + \nabla \times \mathbf{E} &= \frac{4\pi}{c}\mathbf{J} \end{aligned}$$

The remaining two Maxwell equations may be regarded as constraints on the initial field configuration. In general relativity the Einstein equation gives the time evolution of the metric. Finally, in quantum mechanics the dynamical law is the Schrödinger equation

$$\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}$$

which governs the time evolution of the wave function, ψ.

Several important features are implicit in these descriptions. Of course there are different objects – particles, fields or probability amplitudes – that must be specified. But perhaps the most important feature is the existence of some arena within which the motion occurs. In Newtonian mechanics the arena is Euclidean 3-space, and the motion is assumed to be parameterized by universal time. Relativity modified this to a 4 -dimensional spacetime, which in general relativity becomes a curved Riemannian manifold. In quantum mechanics the arena is phase space, comprised of both position and momentum variables, and again having a universal time. Given this diverse collection of spaces for dynamical laws, you may well ask if there is any principle that determines a preferred space. As we shall see, the answer is a qualified yes. It turns out that symmetry gives us an important guide to choosing the dynamical arena.

A *measurement theory* is what establishes the correspondence between calculations and measurable numbers. For example, in Newtonian mechanics the primary dynamical variable for a particle is the position vector, $\mathbf{x}$. While the dynamical law predicts this vector as a function of time, we never measure a vector directly. In order to extract measurable magnitudes we use the Euclidean inner product,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v}$$

If we want to know the position, we specify a vector basis ($\left(\hat{\imath}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\right)$ say) for comparison, then specify the numbers

$$\begin{aligned} x &= \hat{\imath} \cdot \mathbf{x} \\ y &= \hat{\mathbf{j}} \cdot \mathbf{x} \\ z &= \hat{\mathbf{k}} \cdot \mathbf{x} \end{aligned}$$

These numbers are then expressed as dimensionless ratios by choosing a length standard, l. If l is chosen as the meter, then in saying the position of a particle is a $(x, y, z) = (3m, 4m, 5m)$ we are specifying the dimensionless ratios

$$\begin{aligned} \frac{x}{l} &= 3 \\ \frac{y}{l} &= 4 \\ \frac{z}{l} &= 5 \end{aligned}$$

A further assumption of Newtonian measurement theory is that particles move along unique, well-defined curves. This is macroscopically sound epistemology, since we can see a body such as a ball move smoothly through an arc. However, when matter is not continuously monitored the assumption becomes suspect. Indeed, the occasional measurements we are able to make on fundamental particles do not allow us to claim a unique path is knowable, and quantum experiments show that it is incorrect to assume that unique paths exist.

Thus, quantum mechanics provides a distinct example of a measurement theory – we do not assume unique evolution. Perhaps the chief elements of quantum measurement theory is the Hermitian inner product on Hilbert space:

$$\langle \psi | \psi \rangle = \int_V \psi^* \psi d^3x$$

and its interpretation as the probability of finding the particle related to ψ in the volume V. As noted in the preceeding paragraph, it is incorrect to assume a unique path of motion. The relationship between expectation values of operators and measurement probabilities is a further element of quantum measurement theory.

The importance of the distinction between dynamical laws and measurement theories will become clear when we introduce the additional element of symmetry in the final sections of the book. In particular, we will see that the techniques of *gauge theory* allow us to reconcile differences between the symmetry of a dynamical law and the symmetry of the associated measurement theory. We shall show how different applications of gauge theory lead to the Lagrangian and Hamiltonian formulations of classical mechanics, and eventually, how a small change in the measurement theory leads to quantum mechanics.

4 The objects of measurement

In this section we develop the mathematical tools required to describe physically measurable properties. Which properties we can measure depends on the nature of the physical theory, and in particular, on the symmetry of our physical laws and measurement theory. This idea of symmetry is one of the most important concepts in modern physics. All of the known fundamental interactions, and even supersymmetry relating matter and force, are described in terms of definite symmetries. As we shall see, even the Lagrangian and Hamiltonian formulations of classical mechanics may be derived from symmetry considerations.

As discussed in Chapter 2, a symmetry may typically be represented by the action of a group.

Suppose we have a physical model based on a given symmetry. The measurable properties of the physical model will be *invariant* under the action of the symmetry. The physical properties associated with that model are called *scalars*. These scalars are the only objects that we can measure. However, there are

many objects besides scalars that are useful, including many that are invariant. For example, a cylindrically symmetric system may be characterized by an invariant vector along the symmetry axis, but measurement of that vector relies on forming scalars from that vector.

The most important class of non-scalars are the *tensors*. There are two principal reasons that tensors are useful. First, their transformation properties are so simple that it is easy to construct scalars from them. Second, the form of tensor equations is unchanged by transformation. Specifically, tensors are those objects which transform linearly and homogeneously under the action of a group symmetry (Λ), or the under inverse action of the group symmetry (Λ^{-1}). This linear, homogeneous transformation property is called *covariance*. If we write, schematically,

$$\begin{aligned} T' &= \Lambda T \\ S' &= S\Lambda^{-1} \end{aligned}$$

for some tensors of each type, then it is immediate that combining such a pair gives a scalar, or invariant quantity,

$$S'T' = S\Lambda^{-1}\Lambda T = ST$$

It is also immediate that tensor equations are covariant. This means that the *form* of tensor equations does not change when the system is transformed. Thus, if we arrange any tensor equation to have the form

$$T = 0$$

where T may be an arbitrarily complicated tensor expression, we immediately have the same equation after transformation, since

$$T' = \Lambda T = 0$$

Knowing the symmetry and associated tensors of a physical system we can quickly go beyond the dynamical law in making predictions by asking what *other* objects besides the dynamical law are preserved by the transformations. Relations between these covariant objects express possible physical relationships, while relationships among other, non-covariant quantities, will not.

4.1 Examples of tensors

Before proceeding to a formal treatment of tensors, we provide some concrete examples of scalars, of vector transformations, and of some familiar second rank tensors.

4.1.1 Scalars and non-scalars

If we want to describe a rod, its length is a relevant feature because its length is independent of what coordinate transformations we perform. However, it isn't reasonable to associate the change, Δx, in the x coordinate between the ends of the rod with anything physical because as the rod moves around Δx changes arbitrarily. Tensors allow us to separate the properties like length,

$$L = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$$

from properties like Δz; invariant quantities like L can be physical but coordinate dependent quantities like Δz cannot.

There are many kinds of objects that contain physical information. You are probably most familiar with numbers, functions and vectors. A function of position such as temperature has a certain value at each point, regardless of how we label the point. Similarly, we can characterize vectors by their magnitude and direction relative to other vectors, and this information is independent of coordinates. But there are other objects that share these properties.

4.1.2 Vector transformations

To help cement these ideas, suppose we have three vectors which transform according to

$$\begin{aligned}\tilde{A}_i &= M_{ij}A_j \\ \tilde{B}_i &= M_{ij}B_j \\ \tilde{C}_i &= N_{ij}C_j + a_i\end{aligned}$$

Notice that C_i, that does *not* transform in the same way as A_i and B_i, nor does it transform homogeneously. It makes mathematical sense to postulate a relationship between A_i and B_i such as

$$A_i = \lambda B_i$$

because if we transform the system we still have

$$\tilde{A}_i = \lambda \tilde{B}_i$$

The relationship between A_i and B_i is consistent with the symmetry of the physical system. However, we cannot this type of physical relationship between A_i and C_i, because the two expressions

$$A_i = \beta C_i$$

and

$$\tilde{A}_i = \beta \tilde{C}_i$$

are not equivalent. Indeed, the second is equivalent to

$$M_{ij}A_j = \beta \left(N_{ij}C_j + a_i\right)$$

or, multiplying both sides by M^{-1}_{mi},

$$\begin{aligned}M^{-1}_{mi}M_{ij}A_j &= \beta M^{-1}_{mi}\left(N_{ij}C_j + a_i\right) \\ A_m &= \beta M^{-1}_{mi}N_{ij}C_j + \beta M^{-1}_{mi}a_i\end{aligned}$$

so that unless $N = M^{-1}$ and $a = 0$, the two expression are quite different.

4.1.3 The Levi-Civita tensor

One of the most useful tensors is the totally antisymmetric Levi-Civita tensor, e_{ijk}. To define e_{ijk}, we first define the totally antisymmetric symbol ε_{ijk} by setting

$$\varepsilon_{123} = 1$$

All remaining components of ε_{ijk} follow by using the antisymmetry. Thus, if any two of the indices of a component are the same the component must vanish (e.g., $\varepsilon_{112} = -\varepsilon_{112} = 0$), while the nonvanishing components are

$$\begin{aligned}\varepsilon_{123} &= \varepsilon_{231} = \varepsilon_{312} = 1 \\ \varepsilon_{132} &= \varepsilon_{213} = \varepsilon_{321} = -1\end{aligned}$$

Note that the triple sum

$$\varepsilon_{ijk}\varepsilon_{ijk} = 3! = 6$$

is easily found by summing of squares of all the components.

As we show below, one way to define a tensor is to specify its components in one coordinate system, and then demand that it transform as a tensor. Once we have described the transformations of tensors, we will apply this technique to the Levi-Civita tensor. In the meantime, we define the components of the Levi-Civita tensor in Cartesian coordinates to be the same as the antisymmetric symbol. Thus, in Cartesian coordinates,

$$e_{ijk} = \varepsilon_{ijk}$$

There are a number of useful identities satisfied by these tensors, most of which follow from

$$\begin{aligned} \varepsilon_{ijk}\varepsilon_{lmn} &= \delta_{il}\delta_{jm}\delta_{kn} + \delta_{im}\delta_{jn}\delta_{kl} + \delta_{in}\delta_{jl}\delta_{km} \\ &\quad -\delta_{il}\delta_{jn}\delta_{km} - \delta_{in}\delta_{jm}\delta_{kl} - \delta_{im}\delta_{jl}\delta_{kn} \end{aligned} \tag{16}$$

It is easy to check that this expression is correct by noting that in order for the left hand side to be nonzero, each of the sets (ijk) and (lmn) must take a permutation of the values (123). Therefore, since i must be 1 or 2 or 3, it must be equal to exactly one of l, m or n. Similarly, j must equal one of the remaining two indices, and k the third. The right hand side is therefore a list of the possible ways this can happen, with appropriate signs. Alternatively, noting that specifying a single component of a totally antisymmetric object determines all of the remaining components, we can argue that any two totally antisymmetric tensors must be proportional. It is then easy to establish that the proportionality constant is 1.

Prove that eq.(16) is correct by the second method. First, show that the right side of eq.(16) is antisymmetric under any pair exchange of the indices (ijk) and also under any pairwise exchange of (lmn). This allows us to write

$$\begin{aligned} \varepsilon_{ijk}\varepsilon_{lmn} &= \lambda\,(\delta_{il}\delta_{jm}\delta_{kn} + \delta_{im}\delta_{jn}\delta_{kl} + \delta_{in}\delta_{jl}\delta_{km} \\ &\quad -\delta_{il}\delta_{jn}\delta_{km} - \delta_{in}\delta_{jm}\delta_{kl} - \delta_{im}\delta_{jl}\delta_{kn}) \end{aligned}$$

for some constant λ. Show that $\lambda = 1$ by setting $l = i, m = j$ and $n = k$ on both sides and finding the resulting triple sum on each side.

We can now produce further useful results. If, instead of summing over all three pairs of indices, we only sum over one, we find that

$$\begin{aligned} \varepsilon_{ijk}\varepsilon_{imn} &= \delta_{ii}\delta_{jm}\delta_{kn} + \delta_{im}\delta_{jn}\delta_{ki} + \delta_{in}\delta_{ji}\delta_{km} \\ &\quad -\delta_{ii}\delta_{jn}\delta_{km} - \delta_{in}\delta_{jm}\delta_{ki} - \delta_{im}\delta_{ji}\delta_{kn} \\ &= 3\delta_{jm}\delta_{kn} + \delta_{km}\delta_{jn} + \delta_{jn}\delta_{km} - 3\delta_{jn}\delta_{km} - \delta_{kn}\delta_{jm} - \delta_{jm}\delta_{kn} \\ &= \delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km} \end{aligned} \tag{17}$$

Since the cross product may be written in terms of the Levi-Civita tensor as

$$[\mathbf{u} \times \mathbf{v}]_i = \varepsilon_{ijk} u_j v_k$$

this identity gives a simple way to reduce multiple cross products.

Prove that the components of the cross product may be written as

$$[\mathbf{u} \times \mathbf{v}]_i = \varepsilon_{ijk} u_j v_k$$

then use the identity of eq.(17) to prove the "*bac-cab*" rule:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}\,(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}\,(\mathbf{a} \cdot \mathbf{b})$$

Prove that

$$\varepsilon_{ijk}\varepsilon_{ijn} = 2\delta_{kn}$$

When we generalize these results to arbitrary coordinates, we will find that the Levi-Civita tensor e_{ijk} requires a multiplicative function. When we generalize the Levi-Civita tensor to higher dimensions it is always a *maximally* antisymmetric tensor, and therefore unique up to an overall multiple. Thus, in d-dimensions, the Levi-Civita tensor has d indices, $e_{i_1 i_2 \ldots i_d}$, and is antisymmetric under all pair interchanges. In any dimension, a pair of Levi-Civita tensors may be rewritten in terms of antisymmetrized products of Kronecker deltas.

4.1.4 Some second rank tensors

Moment of intertia tensor For example, consider the angular momentum of a rigid body . If the force on a small piece of the body is given by

$$\begin{aligned} d\mathbf{F} &= dm\frac{d\mathbf{v}}{dt} \\ &= dm\frac{d}{dt}\left(\omega\times\mathbf{r}\right) \end{aligned}$$

then the contribution of the mass element $dm=\rho d^3x$ to the torque is

$$\begin{aligned} d\mathbf{N} &= \mathbf{r}\times d\mathbf{F} \\ &= dm\mathbf{r}\times\frac{d\mathbf{v}}{dt} \end{aligned}$$

where $\mathbf{r}$ is the position of the mass element relative to the center of mass. Integrating over the rigid body, we get the total torque on the left and therefore

$$\begin{aligned} \mathbf{N} &= \int \mathbf{r}\times d\mathbf{F} \\ &= \int \rho\left(\mathbf{r}\times\frac{d\mathbf{v}}{dt}\right)d^3x \end{aligned}$$

Now notice that

$$\frac{d}{dt}\left(\mathbf{r}\times\mathbf{v}\right)=\mathbf{v}\times\mathbf{v}+\mathbf{r}\times\frac{d\mathbf{v}}{dt}=\mathbf{r}\times\frac{d\mathbf{v}}{dt}$$

so we can pull out the time derivative. Let the body rotate with angular velocity ω, where the direction of ω is along the axis of rotation according to the right-hand rule. Then, since the density is independent of time,

$$\begin{aligned} \mathbf{N} &= \frac{d}{dt}\int\rho\left(\mathbf{r}\times\mathbf{v}\right)d^3x \\ &= \frac{d}{dt}\int\rho\left(\mathbf{r}\times\left(\omega\times\mathbf{r}\right)\right)d^3x \\ &= \frac{d}{dt}\int\rho\left(\omega r^2-\mathbf{r}\left(\omega\cdot\mathbf{r}\right)\right)d^3x \end{aligned}$$

Here's the important part: we can separate the dynamics from physical properties of the rigid body if we can get omega out of the integral, because then the integral will depend only on intrinsic properties of the rigid body. We can accomplish the separation if we write the torque in components:

$$\begin{aligned} N_i &= \frac{d}{dt}\int\rho\left(\mathbf{r}\times\mathbf{v}\right)_i d^3x \\ &= \frac{d}{dt}\int\rho\left(\omega_i r^2-r_i\left(\omega_j r_j\right)\right)d^3x \\ &= \frac{d}{dt}\int\rho\left(\delta_{ij}\omega_j r^2-r^i\left(\omega_j r_j\right)\right)d^3x \\ &= \frac{d}{dt}\omega_j\int\rho\left(\delta_{ij}r^2-r_i r_j\right)d^3x \end{aligned}$$

Notice how we inserted an identity matrix, using $\omega_i=\delta_{ij}\omega_j$, to get the indices on both factors of ω_j to be the same. Now defined the *moment of inertia tensor*,

$$I_{ij}\equiv\int\rho\left(\delta_{ij}r^2-r_i r_j\right)d^3x$$

This is nine separate equations, one for each value of i and each value of j. However, since the moment of inertia tensor is symmetric, $I_{ij} = I_{ji}$, only six of these numbers are independent. Once we have computed each of them, we can work with the much simpler expression,

$$N_i = \frac{d}{dt} I_{ij} \omega_j$$

The sum on the right is the normal way of taking the product of a matrix on a vector. The product is the angular momentum vector,

$$L_i = I_{ij} \omega_j$$

and we have, simply,

$$N_i = \frac{dL_i}{dt}$$

We can also write the rotational kinetic energy of a rigid body in terms of the moment of inertia tensor,

$$\begin{aligned} T &= \frac{1}{2} \int \mathbf{v}^2 dm \\ &= \frac{1}{2} \int (\omega \times \mathbf{r}) \cdot (\omega \times \mathbf{r}) \, \rho(x) d^3x \end{aligned}$$

Working out the messy product,

$$\begin{aligned} (\omega \times \mathbf{r}) \cdot (\omega \times \mathbf{r}) &= \varepsilon_{ijk} \varepsilon_{imn} \omega_j r_k \omega_m r_n \\ &= (\delta_{jm} \delta_{kn} - \delta_{jn} \delta_{km}) \, \omega_j r_k \omega_m r_n \\ &= \omega_i \omega_j \left(\delta_{ij} r^2 - r_i r_j \right) \end{aligned}$$

Therefore,

$$\begin{aligned} T &= \frac{1}{2} \int \omega \cdot (\mathbf{r} \times (\omega \times \mathbf{r})) \, \rho(x) d^3x \\ &= \frac{1}{2} \omega_i \omega_j \int \left(\delta_{ij} r^2 - r_i r_j \right) \rho(x) d^3x \\ &= \frac{1}{2} I_{ij} \omega_i \omega_j \end{aligned}$$

The metric tensor We have already used the Pythagorean formulat to specify the length of a curve, but the usual formula works only in Cartesian coordinates. However, it is not difficult to find an expression valid in any coordinate system – indeed, we already know the squared separation of points in Cartesian, polar and spherical coordinates,

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 \\ ds^2 &= d\rho^2 + \rho^2 d\phi^2 + dz^2 \\ ds^2 &= dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta \; d\phi^2 \end{aligned}$$

respectively. Notice that all of these forms are quadratic in coordinate differentials. In fact, this must be the case for any coordinates. For suppose we have coordinates y_i given as arbitrary invertible functions of a set of Cartesian coordinates x_i,

$$\begin{aligned} y_i &= y_i (x_j) \\ x_i &= x_i (y_j) \end{aligned}$$

Then the differentials are related by

$$dx_i = \frac{\partial x_i}{\partial y_j} dy_j$$

The squared separation is therefore

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 \\ &= dx_i dx_i \\ &= \left(\frac{\partial x_i}{\partial y_j} dy_j\right)\left(\frac{\partial x_i}{\partial y_k} dy_k\right) \\ &= \frac{\partial x_i}{\partial y_j}\frac{\partial x_i}{\partial y_k} dy_j dy_k \\ &= g_{jk} dy_j dy_k \end{aligned}$$

where in the last step we define the *metric tensor*, g_{jk}, by

$$g_{jk} \equiv \frac{\partial x_i}{\partial y_j}\frac{\partial x_i}{\partial y_k}$$

Using the metric tensor, we can now write the distance between nearby points regardless of the coordinate system. For two points with coordinate separations dx_i, the distance between them is

$$ds = \sqrt{g_{ij} dx_i dx_j}$$

We have already introduced the metric as a way of writing the infinitesimal line element in arbitrary coordinates,

$$ds^2 = g_{ij} dx_i dx_j$$

We show below that the metric also characterizes the inner product of two vectors:

$$\mathbf{u} \cdot \mathbf{v} = g_{ij} u_i v_j$$

The metric is a *rank two covariant tensor*.

Find the metric tensor for:

1. polar coordinates, and
2. spherical coordinates.

The stress tensor We can write an infinitesimal area element as

$$dS_i = n_i d^2x$$

where n_i is orthogonal to the surface element d^2x. Now imagine such a surface element imersed in a continuous medium. In general, there will be a force, dF_i, acting across this surface area, but it is not generally in the same direction as dS_i. However, we do expect its magnitude to be proportional to the area, so we may write a linear equation,

$$dF_i = P_{ij} dS_j$$

The coefficients in this expression comprise the *stress tensor*. If P_{ij} is diagonal,

$$P_{ij} = \begin{pmatrix} p_1 & & \\ & p_2 & \\ & & p_3 \end{pmatrix}$$

then the numbers p_i are just the forces per unit area – i.e., pressures – in each of the three independent directions. Any off-diagonal elements of P_{ij} are called *stresses*. These are due to components of the force that are parallel rather than perpendicular to the surface, which therefore tend to produce shear, shifting parallel area elements along one another.

Thus, when P_{12} is nonzero, there is an x-component to the force on a surface whose normal is in the y direction. But now consider P_{21}, which gives the y-component of the force on a similar surface, with normal in the x direction. The z-component of the torque produced on a cube of side $2a$ by these two forces together, about the center of the cube is

$$\begin{aligned} N_3 &= r_2 dF_1 - r_1 dF_2 = r_2 P_{1j} dS_j - r_1 P_{2j} dS_j \\ &= r_2 P_{12} dS_2 - r_1 P_{21} dS_1 \\ &= \frac{1}{2} a^3 \left(P_{12} - P_{21} \right) \end{aligned}$$

The moment of inertia of the infinitesimal cube, taking the density constant, is

$$I_{ij} = \frac{1}{12} \rho a^5 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The equation of motion becomes

$$\begin{aligned} N_3 &= I_{33} \frac{d\omega}{dt} \\ \frac{1}{2} a^3 \left(P_{12} - P_{21} \right) &= \frac{1}{12} \rho a^5 \frac{d\omega}{dt} \end{aligned}$$

so the angular acceleration is given by

$$\frac{d\omega}{dt} = \frac{6}{\rho a^2} \left(P_{12} - P_{21} \right)$$

Since the angular acceleration must remain finite as the side of the cube tends to zero, we must have $P_{12} = P_{21}$. A similar argument applies to the other off diagonal terms, so the stress tensor must be symmetric.

4.2 Vectors

The simplest tensors are scalars, which are the measurable quantities of a theory, left invariant by symmetry transformations. By far the most common non-scalars are the vectors, also called rank-1 tensors. Vectors hold a distinguished position among tensors – indeed, tensors must be defined in terms of vectors. The reason for their importance is that, while tensors are those objects that transform linearly and homogeneously under a given set of transformations, we require vectors in order to define the action of the symmetry in the first place. Thus, vectors cannot be defined in terms of their transformations.

In the next subsection, we provide an axiomatic, algebraic definition of vectors. Then we show how to associate two distinct vector spaces with points of a manifold. Somewhat paradoxically, one of these vector spaces is called the space of vectors while the other is called the space of 1 -forms. Fortunately, the existence of a metric on the manifold allows us to relate these two spaces in a 1-1, onto way. Moreover, the metric allows us to define an inner product on each of the two vectors spaces. Therefore, we discuss the properties of metrics in some detail.

After the geometric description of vectors and forms, we turn to transformations of vectors. Using the action of a group on a vector space to define a linear representation of the group, we are finally able to define outer products of vectors and give a general definition of tensors in terms of their transformation properties.

4.2.1 Vectors as algebraic objects

Alternatively, we can define vectors algebraically. Briefly, a vector space is defined as a set of objects, $V = \{\mathbf{v}\}$, together with a field $\mathcal{F}$ of numbers (general R or C) which form a commutative group under addition and permit scalar multiplication. The scalar multiplication must satisfy distributive laws.

More concretely, being a group under addition guarantees the following:

1. V is closed under addition. If $\mathbf{u}, \mathbf{v}$ are any two elements of V, then $\mathbf{u} + \mathbf{v}$ is also an element of V.
2. There exists an additive identity, which we call the zero vector, $\mathbf{0}$.
3. For each element $\mathbf{v}$ of V there is an additive inverse to $\mathbf{v}$. We call this element $(-\mathbf{v})$.
4. Vector addition is associative, $\mathbf{w} + (\mathbf{u} + \mathbf{v}) = (\mathbf{w} + \mathbf{u}) + \mathbf{v}$

In addition, addition is commutative, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

The scalar multiplication satisfies:

1. Closure: $a\mathbf{v}$ is in V whenever $\mathbf{v}$ is in V and a is in $\mathcal{F}$.
2. Scalar identity: $1\mathbf{v} = \mathbf{v}$
3. Scalar and vector zero: $0\mathbf{v} = \mathbf{0}$ for all $\mathbf{v}$ in V and $a\mathbf{0} = \mathbf{0}$ for all a in $\mathcal{F}$.
4. Distributive 1: $(a + b)\,\mathbf{v} = a\mathbf{v} + b\mathbf{v}$
5. Distributive 2: $a\,(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
6. Associativity: $(ab)\,\mathbf{v} = a\,(b\mathbf{v})$

All of the familiar properties of vectors follow from these. An important example is the existence of a basis for any finite dimensional vectors space. We prove this in several steps as follows.

First, define linear dependence. A set of vectors $\{\mathbf{v}_i \mid i = 1, \ldots, n\}$ is *linearly dependent* if there exist numbers $\{a_i \mid i = 1, \ldots, n\}$, not all of which are zero, such that the sum $a_i\mathbf{v}_i$ vanishes,

$$a_i\mathbf{v}_i = 0$$

As set of vectors is *linearly independent* if it is not dependent. Now suppose there exists a maximal linearly independent set of vectors. By this we mean that there exists some finite number n, such that we can find one or more linearly independent sets containing n vectors, but there do not exist any linearly independent sets containing $n + 1$ vectors. Then we say that n is the *dimension* of the vector space.

In an n-dimensional vector space, and collection of n independent vectors is called a *basis*. Suppose we have a basis,

$$B = \{\mathbf{v}_i \mid i = 1, \ldots, n\}$$

Then, since every set with $n + 1$ elements is linearly dependent, the set

$$\{\mathbf{u}\} \cup B = \{\mathbf{u}, \mathbf{v}_i \mid i = 1, \ldots, n\}$$

is dependent, where $\mathbf{u}$ is any nonzero vector in V. Therefore, there exist numbers a_i, b, not all zero, such that

$$b\mathbf{u} + a_i\mathbf{v}_i = 0$$

Now suppose $b = 0$. Then we have a linear combination of the $\mathbf{v}_i$ that vanishes, $a_i\mathbf{v}_i = 0$, contrary to our assumption that they form a basis. Therefore, b is nonzero, and we can divide by it. Adding the inverse to the sum $a_i\mathbf{v}_i$ we can write

$$\mathbf{u} = -\frac{1}{b}a_i\mathbf{v}_i$$

This shows that every vector in a finite dimensional vector space V can be written as a linear combination of the vectors in any basis. The numbers $u_i = -\frac{a_i}{b}$ are called the components of the vector $\mathbf{u}$ in the basis B.

Prove that two vectors are equal if and only if their components are equal.

Notice that we have chosen to write the labels on the basis vectors as subscripts, while we write the components of a vector as superscripts. This choice is arbitrary, but leads to considerable convenience later. Therefore, we will carefully maintain these positions in what follows.

Often vector spaces are given an inner product. An inner product on a vector space is a symmetric bilinear mapping from pairs of vectors to the relevant field, $\mathcal{F}$,

$$g : V \times V \to \mathcal{F}$$

Here the Cartesian product $V \times V$ means the set of all ordered pairs of vectors, $(\mathbf{u}, \mathbf{v})$, and bilinear means that g is linear in each of its two arguments. Symmetric means that $g(\mathbf{u}, \mathbf{v}) = g(\mathbf{v}, \mathbf{u})$.

There are a number of important consequences of inner products.

Suppose we have an inner product which gives a nonnegative real number whenever the two vectors it acts on are identical:

$$g(\mathbf{v}, \mathbf{v}) = s^2 \geq 0$$

where the equal sign holds if and only if $\mathbf{v}$ is the zero vector. Then g is a *norm* or *metric* on V – it provides a notion of length for each vector. If the inner product satisfies the triangle inequality,

$$g(\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) \leq g(\mathbf{u}, \mathbf{u}) + g(\mathbf{v}, \mathbf{v})$$

then we can also define angles between vectors, via

$$\cos\theta = \frac{g(\mathbf{u}, \mathbf{v})}{\sqrt{g(\mathbf{u}, \mathbf{u})\, g(\mathbf{v}, \mathbf{v})}}$$

If the number s is real, but not necessarily positive, then g is called a *pseudo-norm* or a *pseudo-metric*. We will need to use a pseudo-metric when we study relativity.

If $\{\mathbf{v}_i\}$ is a basis, then we can write the inner product of any two vectors as

$$\begin{aligned} g(\mathbf{u}, \mathbf{v}) &= g\left(a^i \mathbf{v}_i, b^j \mathbf{v}_j\right) \\ &= a^i b^j g(\mathbf{v}_i, \mathbf{v}_j) \end{aligned}$$

so if we know how g acts on the basis vectors, we know how it acts on any pair of vectors. We can summarize this knowledge by defining the matrix

$$g_{ij} \equiv g(\mathbf{v}_i, \mathbf{v}_j)$$

Now, we can write the inner product of any two vectors as

$$g(\mathbf{u}, \mathbf{v}) = a^i g_{ij} b^j = g_{ij} a^i b^j$$

It's OK to think of this as sandwiching the metric, g_{ij}, between a row vector a^i on the left and a column vector b^j on the right. However, index notation is more powerful than the notions of row and column vectors, and in the long run it is more convenient to just note which sums are required. A great deal of computation can be accomplished without actually carrying out sums. We will discuss inner products in more detail in later Sections.

4.2.2 Vectors in space

In order to work with vectors in physics, it is most useful to think of them as geometric objects. This approach allows us to associate one or more vector spaces with each point of a manifold, which in turn will allow us to discuss motion and the time evolution of physical properties.

Since there are spaces – for example, the spacetimes of general relativity or on the surface of a sphere – where we want to talk about vectors but can't draw them as arrows because the space is curved, we need a more general, abstract definition. To define vectors, we need three things:

1. A manifold, $\mathcal{M}$, that is, a topological space which in a small enough region looks like a small piece of R^n. Manifolds include the Euclidean spaces, R^n, but also things like the 2-dimensional surface of a sphere or a doughnut or a saddle.

2. Functions on $\mathcal{M}$. A function on a manifold assigns a number to each point of the space.

3. Curves on the space. A curve is a mapping from the reals into the space. Such maps don't have to be continuous or differentiable in general, but we are interested in the case where they are. If M is our space and we have a curve $C : R \to M$, then for any real number λ, $C(\lambda)$ is a point of M. Changing λ moves us smoothly along the curve in M. If we have coordinates x^i for M we can specify the curve by giving $x^i(\lambda)$. For example, $(\theta(\lambda), \varphi(\lambda))$ describes a curve on the surface of a sphere.

Given a space together with functions and curves, there are *two* ways to associated a space of vectors with each point. We will call these two spaces $forms$ and $vectors$, respectively, even though both are vector spaces in the algebraic sense. The existence of two distinct vector spaces associated with a manifold leads us to introduce some new notation. From now on, vectors from the space of vectors will have components written with a raised index, v^i, while the components of forms will be written with the index lowered, ω_i. The convention is natural if we begin by writing the indices on coordinates in the raised position and think of a derivative with respect to the coordinates as having the index in the lowered position. The benefits of this convention will quickly become evident. As an additional aid to keeping track of which vector space we mean, whenever practical we will name forms with Greek letters and vectors with Latin letters.

The definitions are as follows:

A $form$ is defined for each function as a linear mapping from curves into the reals. The vector space of forms is denoted V_*.

A *vector* is defined for each curve as a linear mapping from functions into the reals. The vector space of vectors is denoted V^*.

Here's how it works. For a form, start with a function and think of the form, ω_f, as the differential of the function, $\omega_f = df$. Thus, for each function we have a form. The form is defined as a linear mapping on curves, $\omega_f : f \to R$. We can think of the linear mapping as integration along the curve C, so

$$\omega_f(C) = \int_C df = f(C(1)) - f(C(0))$$

In coordinates, we know that df is just

$$df = \frac{\partial f}{\partial x^i} dx^i$$

If we restrict the differentials dx^i to lie along the curve C, we have

$$df = \frac{\partial f}{\partial x^i} \frac{dx^i}{d\lambda} d\lambda$$

We can think of the coordinate differentials dx^i as a basis, and the partial derivatives $\frac{\partial f}{\partial x^i}$ as the components of the vector ω_f.Formal definition of forms.

This argument shows that integrals of the differentials of functions are forms, but the converse is also true – any linear mapping from curves to the reals may be written as the integral of the differential of a function. The proof is as follows. Let ϕ be a linear map from differentiable curves to the reals, and let the curve C be parameterized by $s \in [0, 1]$. Break C into N pieces, C_i, parameterized by $s \in \left[\frac{i-1}{N}, \frac{i}{N}\right]$, for $i = 1, 2, ...N$. By linearity, $\phi(C)$ is given by

$$\phi(C) = \sum_{i=1}^{N} \phi(C_i)$$

By the differentiability (hence continuity) of ϕ, we know that $\phi(C_i)$ maps C_i to a bounded interval in R, say, (a_i, b_i), of length $|b_i - a_i|$. As we increase N, each of the numbers $|b_i - a_i|$ tends monotonically to zero so that the value of $\phi(C_i)$ approaches arbitrarily close to the value a_i. We may therefore express $\phi(C)$ by

$$\phi(C) \quad = \quad \lim_{N \to \infty} \sum_{i=1}^{N} \phi(C_i)$$

$$= \lim_{N\to\infty} \sum_{i=1}^{N} \frac{a_i}{|b_i - a_i|} ds$$

where ds replaces $\lim_{N\to\infty} \frac{1}{N}$. Notice that as N becomes large, a_i becomes small, since the average value of $\phi(C_i)$ is $\frac{\phi(C)}{N}$. The normalized expression

$$\frac{a_i}{|b_i - a_i|}$$

therefore remains of order $\phi(C)$ and we may define a function $f(s)$ as the piecewise continuous function

$$f(s_i) = \frac{a_i}{|b_i - a_i|}, \quad s_i \in \left[\frac{i-1}{N}, \frac{i}{N}\right]$$

Then $f(s)$ becomes smooth in the limit as $N \to \infty$, and $\phi(C)$ is given by

$$\phi(C) = \int f(s)\, ds$$

The fundamental theorem of calculus now show that if we let $F = \int_C f(s)\, ds$, then

$$\phi(C) = \int_C dF$$

so that the linear map on curves is the integral of a differential.

For vectors, we start with the curve and think of the corresponding vector as the tangent to the curve. But this "tangent vector" isn't an intrinsic object – straight arrows don't fit into curved spaces. So for each curve we define a vector as a linear map from functions to the reals – the directional derivative of f along the curve C. The directional derivative can be defined just using $C(\lambda)$:

$$v(f) = \lim_{\delta\lambda\to 0} \frac{f(C(\lambda + \delta\lambda) - f(C(\lambda)))}{\delta\lambda}$$

It is straightforward to show that the set of directional derivatives forms a vector space. In coordinates, we're used to thinking of the directional derivative as just

$$\mathbf{v} \cdot \nabla f$$

and this is just right if we replace $\mathbf{v}$ by the tangent vector, $\frac{dx^i}{d\lambda}$:

$$v(f) = \frac{dx^i}{d\lambda} \frac{\partial f}{\partial x^i}$$

We can abstract v as the differential operator

$$v = \frac{dx^i}{d\lambda} \frac{\partial}{\partial x^i}$$

and think of $\frac{dx^i}{d\lambda}$ as the components of v and $\frac{\partial}{\partial x^i}$ as a set of basis vectors.

For both forms and vectors, the linear character of integration and differentiation guarantee the algebraic properties of vector spaces, while the usual chain rule applied to the basis vectors,

$$\begin{aligned} dx^i &= \frac{\partial x^i}{\partial y^k} dy^k \\ \frac{\partial}{\partial x^i} &= \frac{\partial y^k}{\partial x^i} \frac{\partial}{\partial y^k} \end{aligned}$$

together with the coordinate invariance of the formal symbols ω and v, shows that the components of ω and v transform to a new coordinate system $y^i\left(x^k\right)$ according to

$$\begin{aligned}\tilde{v}^k &= \frac{dy^k}{d\lambda} = \frac{\partial y^k}{\partial x^m}\frac{dx^m}{d\lambda} = \frac{\partial y^k}{\partial x^m}v^m \\ \tilde{\omega}_k &= \frac{\partial}{\partial y^k} = \frac{\partial x^k}{\partial y^i}\frac{\partial}{\partial x^k} = \frac{\partial x^k}{\partial y^i}\omega_k\end{aligned}$$

Since the Jacobian matrix, $J^k{}_m$, and its inverse are given by

$$\begin{aligned}J^k{}_m &= \frac{\partial y^k}{\partial x^m} \\ \bar{J}^k{}_m &= \frac{\partial x^k}{\partial y^m}\end{aligned}$$

we can write the transformation laws for vectors and forms as

$$\tilde{v}^k = J^k{}_m v^m \tag{18}$$

$$\tilde{\omega}_k = \bar{J}^m{}_k \omega_m \tag{19}$$

In general, any object which transforms according to eq.(18) is called *contravariant*, while any object which transforms according to eq.(19) is called *covariant*. There are two uses of the word covariant – here the term refers to transformation with the inverse Jacobian matrix, but the term is also used generically to refer to tensors. Thus, any object which transforms linearly under a group action may be said to transform covariantly under the group. The context usually makes clear which meaning is intended.

The geometric approach taken here shows that, corresponding to the two types of transformation there are two types of geometric object. Both of these are familiar from vector calculus – the vectors that are used in introductory physics and denoted by arrows, $\vec{\mathbf{v}}$ are vectors, while differentials of functions, df, are forms. We shall show below that we can pass from one type to the other whenever our space has one other bit of structure: a metric.

Prove that

$$\begin{aligned}J^k{}_m &= \frac{\partial y^k}{\partial x^m} \\ \bar{J}^k{}_m &= \frac{\partial x^k}{\partial y^m}\end{aligned}$$

are actually inverse to one another.

Prove that the set of directional derivatives satisfies the algebraic definition of a vector space.

Prove that the set of forms satisfies the algebraic definition of a vector space.

There is a natural duality between the coordinate basis for vectors and the coordinate basis for forms. We define the bracket between the respective basis vectors by

$$\left\langle \frac{\partial}{\partial x^j}, dx^i \right\rangle = \delta^i_j$$

This induces a linear map from $V^* \times V_*$ into the reals,

$$\langle , \rangle : V^* \times V_* \to R$$

given by

$$\langle v, \omega \rangle = \left\langle v^j \frac{\partial}{\partial x^j}, \omega_i dx^i \right\rangle$$

$$
\begin{aligned}
&= v^j \omega_i \left\langle \frac{\partial}{\partial x^j}, dx^i \right\rangle \\
&= v^j \omega_i \delta^i_j \\
&= v^i \omega_i
\end{aligned}
$$

If we pick a particular vector v, then $\langle v, \cdot \rangle$ is a linear mapping from forms to the reals. Since there is exactly one linear map for each vector v, it is possible to define the space of vectors V^* as the set of linear mappings on forms. The situation is symmetric – we might also choose to define forms as the linear maps on vectors. However, both vectors and forms have intrinsic geometric definitions, independently of one another.

Notice that all of the sums in this section involve one raised and one lowered index. This must always be the case, because this is the only type of "inner product" that is invariant. For example, notice that if we transform between $(\tilde{v}, \tilde{\omega})$ and (v, w) the bracket is invariant:

$$
\begin{aligned}
\langle \tilde{v}, \tilde{\omega} \rangle &= \tilde{v}^i \tilde{\omega}_i \\
&= \left(J^i{}_m v^m \right) \left(\bar{J}^n{}_i \omega_n \right) \\
&= \bar{J}^n{}_i J^i{}_m v^m \omega_n \\
&= \delta^n_m v^m \omega_n \\
&= v^m \omega_m \\
&= \langle v, \omega \rangle
\end{aligned}
$$

4.3 The metric

We have already introduced the metric as a line element, giving the distance between infinitesimally separated points on a manifold. This definition may be used to define an inner product on vectors and forms as well.

4.3.1 The inner product of vectors

Recall the geometric definition of a vector v as the differential operator

$$v = \frac{dx^i}{d\lambda}\frac{\partial}{\partial x^i}$$

with components $\frac{dx^i}{d\lambda}$ and $\frac{\partial}{\partial x^i}$ as a set of basis vectors. Then rewriting the form

$$ds^2 = g_{ij}dx^i dx^j$$

as

$$\begin{aligned} ds^2 &= g_{ij}\frac{dx^i}{d\lambda}\frac{dx^j}{d\lambda}d\lambda^2 \\ \left(\frac{ds}{d\lambda}\right)^2 &= g_{ij}\frac{dx^i}{d\lambda}\frac{dx^j}{d\lambda} \end{aligned} \tag{20}$$

shows us that g_{ij} provides a norm on vectors. Formally, g is linear map from a vector into the reals,

$$g : \mathbf{A} \to R$$

where

$$\mathbf{A} = \left.\frac{dx^i(\lambda)}{d\lambda}\frac{\partial}{\partial x^i}\right|_P$$

We generalize this to arbitrary pairs of vectors as follows.

Let $\mathbf{A}$,$\mathbf{B} \in V^*$ be vectors, defined as tangent operators on curves $x^i(\lambda), x^i(\sigma)$ at a point, P, on a manifold, so that

$$\begin{aligned} \mathbf{A} &= \left.\frac{dx^i(\lambda)}{d\lambda}\frac{\partial}{\partial x^i}\right|_P \\ \mathbf{B} &= \left.\frac{dx^i(\sigma)}{d\sigma}\frac{\partial}{\partial x^i}\right|_P \end{aligned}$$

An *inner product* on the space of vectors at a point on a manifold is a symmetric, bilinear mapping

$$g : (\mathbf{A}, \mathbf{B}) \to R$$

The linearity of g on each vector allows us to fully specify the mapping by its effect on the basis. Setting

$$\begin{aligned} A^i &= \left.\frac{dx^i(\lambda)}{d\lambda}\right|_P \\ B^i &= \left.\frac{dx^i(\sigma)}{d\sigma}\right|_P \end{aligned}$$

we have

$$\begin{aligned} g(\mathbf{A}, \mathbf{B}) &= g\left(A^i\frac{\partial}{\partial x^i}, B^j\frac{\partial}{\partial x^j}\right) \\ &= A^i B^j g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) \end{aligned}$$

Now, defining

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$$

the inner product becomes

$$g\left(\mathbf{A}, \mathbf{B}\right) = A^i B^j g_{ij}$$

When the two vectors are the same, we recover eq.(20).

4.3.2 Duality and linear maps on vectors

We may always define a duality relationship between vectors and forms. Starting with the coordinate basis, $\frac{\partial}{\partial x^i}$, for vectors and the corresponding coordinate basis, $\mathbf{d}x^i$, for forms, we define a linear bracket relation by

$$\left\langle \frac{\partial}{\partial x^j}, \mathbf{d}x^i \right\rangle = \delta^i_j$$

Now suppose the space of vectors, V^* is given an arbitrary set of basis vectors, e_i. We define the basis dual to e_i by

$$\left\langle e_j, \mathbf{e}^i \right\rangle = \delta^i_j$$

Since each basis set can be expanded in terms of a coordinate basis,

$$\left\langle e_j \ {}^m \frac{\partial}{\partial x^m}, e_n \ {}^i \mathbf{d}x^n \right\rangle = \delta^j_i$$

we may use linearity to find

$$\begin{aligned} \delta^i_j &= e_j \ {}^m e_n \ {}^i \left\langle \frac{\partial}{\partial x^m}, \mathbf{d}x^n \right\rangle \\ &= e_j \ {}^m e_n \ {}^i \delta^n_m \\ &= e_j \ {}^m e_m \ {}^i \end{aligned}$$

It follows that the matrix $e_n \ {}^i$ giving the form basis in terms of the coordinate differentials is inverse to the matrix giving the dual basis for vectors.

Now consider an arbitrary vector, $\mathbf{v} \in V^*$ and form, $\omega \in V_*$. The duality relation becomes a map,

$$\begin{aligned} \langle \mathbf{v}, \omega \rangle &= \left\langle v^j \mathbf{e}_j, \omega_i \mathbf{e}^i \right\rangle \\ &= v^j \omega_i \left\langle \mathbf{e}_j, \mathbf{e}^i \right\rangle \\ &= v^j \omega_i \delta^i_j \\ &= v^i \omega_i \end{aligned}$$

Using this map together with the metric, we define a unique 1-1 relationship between vectors and forms. Let $\mathbf{w}$ be an arbitrary vector. The 1 -form ω corresponding to $\mathbf{w}$ is defined by demanding

$$g\left(\mathbf{v}, \mathbf{w}\right) = \langle \mathbf{v}, \omega \rangle \tag{21}$$

for all vectors $\mathbf{v} \in V^*$. In components this relationship becomes

$$g_{ij} v^i w^j = v^i \omega_i$$

In order for this to hold for all vectors v^i, the components of the form ω must be related to those of $\mathbf{w}$ by

$$\omega_i = g_{ij} w^j$$

Formally, we are treating the map

$$g : (V^*, V^*) \to R$$

as a mapping from V^* to R by leaving one slot empty:

$$g(v, \cdot) : V^* \to R$$

The duality relation may therefore be viewed as a 1-1 correspondence between 1-forms and linear maps on vectors.

The mapping between vectors and forms allows us to apply the metric directly to forms. Let u and v be any two vectors, and μ and ν the corresponding forms. We then define

$$g(\mu, \nu) = g(\mathbf{u}, \mathbf{v})$$

Making use of linearity, we find components

$$\begin{aligned} g(\mu, \nu) &= g\left(\mu_i \mathbf{e}^i, v_j \mathbf{e}^j\right) = \mu_i v_j g\left(\mathbf{e}^i, \ \mathbf{e}^j\right) \\ g(\mathbf{u}, \mathbf{v}) &= g\left(u^i \mathbf{e}_i, v^j \mathbf{e}_j\right) = u^i v^j g_{ij} \end{aligned}$$

Now equating these and using the relationship between vectors and forms,

$$g_{ik} u^k g_{jl} v^l g\left(\mathbf{e}^i, \mathbf{e}^j\right) = u^k v^l g_{kl}$$

Since this must hold for arbitrary u^k and v^m, it must be that

$$g_{ik} g_{jl} g\left(\mathbf{e}^i, \mathbf{e}^j\right) = g_{kl}$$

We now define g^{ij} to be the inverse of g_{ij},

$$g^{ij} = \left(g^{-1}\right)_{ij}$$

This bit of notation will prove extremely useful. Notice that g_{ij} is the *only* matrix whose inverse is given this special notation. It allows us to write

$$g^{ij} g_{jk} = \delta^i_k = g_{kj} g^{ji}$$

and, as we shall see, makes the relationship between forms and vectors quite transparent. Continuing, we multiply both sides of our expression by two inverse metrics:

$$\begin{aligned} g^{mk} g^{nl} g_{ik} g_{jl} g\left(\mathbf{e}^i, \mathbf{e}^j\right) &= g^{mk} g^{nl} g_{kl} \\ \delta^m_i \delta^n_j g\left(\mathbf{e}^i, \mathbf{e}^j\right) &= g^{mk} \delta^n_k \\ g\left(\mathbf{e}^m, \mathbf{e}^n\right) &= g^{mn} \end{aligned}$$

This establishes the action of the metric on the basis. The inner product of two arbitrary 1-forms follows immediately:

$$g(\mu, \nu) = \mu_i v_j g\left(\mathbf{e}^i, \mathbf{e}^j\right) = g^{ij} \mu_i \nu_j$$

In summary, we have established a 1-1 relationship between vectors and forms,

$$g(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \alpha \rangle$$

and a corresponding correspondence of inner products,

$$g(\mu, \nu) = g(\mathbf{u}, \mathbf{v})$$

where the component form of the duality relation is

$$\langle \mathbf{v}, \omega \rangle = v^i \omega_i$$

In components, these relations imply

$$\begin{aligned} w^i &= g^{ij}\omega_j \\ \omega_i &= g_{ij}w^j \\ g(\mathbf{u},\mathbf{v}) &= u^i v^j g_{ij} \\ g(\mu,\nu) &= g^{ij}\mu_i\nu_j \end{aligned}$$

Show that

$$\begin{aligned} w^i &= g^{ij}\omega_j \\ \omega_i &= g_{ij}w^j \end{aligned}$$

are consistent by substituting one into the other and simplifying.

Suppose the components of a certain vector field in polar coordinates are given by $v^i = (\rho\sin\varphi, -\cos\varphi)$. Find the components of the corresponding form, ν_i. What is the duality invariant, $\langle v,\nu\rangle$? What is the norm, $g(v,v)$?

The relationship between vectors and forms leads us to a modification of the Einstein summation convention. We now modifiy it to say that whenever an index is repeated, *once up and once down*, we perform a sum. This still leads to equations with some doubled indices and some free indices, such as

$$T_{ij}v^j = \omega_i$$

but we will no longer write

$$T_{ij}v_j = \omega_i$$

The point is that the first expression is a relationship between vectors while the second is not. To see why, transform to a different coordinate system. In both cases the right hand side transforms as

$$\tilde{\omega}_m = \omega_n \frac{\partial x^n}{\partial y^m}$$

The left hand side of the first transforms in the same way, because

$$\begin{aligned} \tilde{T}_{mn}\tilde{v}^n &= \left(T_{ij}\frac{\partial x^i}{\partial y^m}\frac{\partial x^j}{\partial y^n}\right)\left(v^k\frac{\partial y^n}{\partial x^k}\right) \\ &= T_{ij}v^k\frac{\partial x^i}{\partial y^m}\left(\frac{\partial x^j}{\partial y^n}\frac{\partial y^n}{\partial x^k}\right) \\ &= T_{ij}v^k\frac{\partial x^i}{\partial y^m}\delta^j_k \\ &= \left(T_{ik}v^k\right)\frac{\partial x^i}{\partial y^m} \end{aligned}$$

However, the second expression is not tensorial because in the new system the left hand side is

$$\begin{aligned} \tilde{T}_{mn}\tilde{v}_n &= \left(T_{ij}\frac{\partial x^i}{\partial y^m}\frac{\partial x^j}{\partial y^n}\right)\left(v_k\frac{\partial x^k}{\partial y^n}\right) \\ &= T_{ij}v^k\frac{\partial x^i}{\partial y^m}\left(\frac{\partial x^j}{\partial y^n}\frac{\partial x^k}{\partial y^n}\right) \end{aligned}$$

which is not related in any simple way to $T_{ij}v_j$.

Since every form now corresponds to a vector and vice versa, we will usually give corresponding vectors and forms similar names:

$$\begin{aligned} v^i &= g^{ij}v_j \\ v_i &= g_{ij}v^j \end{aligned}$$

When we do this, the only distinction between forms and vectors is the position of the index. We can "raise" and "lower" free indices in any equation by using the inverse metric and the metric. Let's look at the torque equation for an example. The angular momentum is the inner product of the moment of inertia tensor with the angular velocity, so the equation requires a metric (in the previous section we assumed Cartesian coordinates, so the metric was just the identity matrix). Therefore we have

$$L^i = I^{ij} g_{jk} \omega^k = I^i{}_k \omega^k$$

where we have used the metric to lower the second index on I^{ij}. Then

$$\begin{aligned} N^i &= \frac{d}{dt}\left(I^i{}_k \omega^k\right) \\ &= I^i{}_k \frac{d\omega^k}{dt} \end{aligned}$$

where we assume the moment of inertia is constant in time. If we multiply the entire equation by g_{mi}, it becomes

$$\begin{aligned} g_{mi} N^i &= g_{mi} I^i{}_k \frac{d\omega^k}{dt} \\ N_m &= I_{mk} \frac{d\omega^k}{dt} \end{aligned}$$

The two forms of the equation are completely equivalent since we can always return the indices to their original positions. The same principles work with each index of any tensor, regardless of the rank. For example, if we have a tensor of type $\binom{4}{3}$ with components $T^{ij}{}_{kl}{}^{mn}{}_o$, we can convert it to a tensor of type $\binom{5}{2}$ by raising any of the indices k, l or o:

$$\begin{aligned} T^{ijs}{}_l{}^{mn}{}_o &= g^{sk} T^{ij}{}_{kl}{}^{mn}{}_o \\ T^{ij}{}_k{}^{smn}{}_o &= g^{sl} T^{ij}{}_{kl}{}^{mn}{}_o \\ T^{ij}{}_{kl}{}^{mns} &= g^{so} T^{ij}{}_{kl}{}^{mn}{}_o \end{aligned}$$

Notice that it is important to preserve the order of the indices, regardless of whether the position is raised or lowered.

Show that $v^i w_i = v_i w^i$.

The examples of Section (2.2) are now correctly written as

$$T_{ijk} v^j w^k + \omega_i = S_{ij} u^j$$

This now represents a relationship between certain 1-forms. Rewrite the expression as a relationship between vectors.

4.3.3 Orthonormal frames

When dealing with vectors or forms, it is almost always simpler to choose a basis which is orthonormal. Like the Cartesian $\left(\hat{\imath}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\right)$, the familiar reference frames of classical mechanics – spherical $\left(\hat{\mathbf{r}}, \hat{\theta}, \hat{\varphi}\right)$ or polar $\left(\hat{\rho}, \hat{\varphi}, \hat{\mathbf{k}}\right)$, for example – are orthonormal.

First, note that the matrices defining a given basis in terms of the coordinate basis are related to the form of the metric in that basis, for we have

$$\begin{aligned} \tilde{g}_{ij} &= g\left(e_i, e_j\right) \\ &= e_i{}^m e_j{}^n g\left(\frac{\partial}{\partial x^m}, \frac{\partial}{\partial x^n}\right) \\ &= e_i{}^m e_j{}^n g_{mn} \end{aligned}$$

This relationship is of particular importance when we define an orthornormal basis.

Now we establish that we can always pick a pseudo-orthonormal basis. Since g_{mn} is by definition a symmetric matrix, it is diagonalizable. Let E_m be a basis in which g_{mn} is diagonal, so that

$$g\left(E_m, E_n\right) = g_{mn} = diag\left(a_1, a_2, \ldots, a_d\right)$$

Then, since g_{mn} is invertible, none of the a_i vanish and we can define

$$e_i = \frac{1}{\sqrt{|a_i|}} E_i$$

Then the norm of each e_i is either $+1$ or -1, and the inner product takes the form

$$\eta_{ab} = g\left(e_a, e_b\right) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & -1 & & & \\ & & & & \ddots & & \\ & & & & & -1 & \end{pmatrix}$$

where there are p positive and q negative elements, with $p + q = d$. We will reserve the symbol η_{ab} for such orthonormal frames, and will distinguish two kinds of indices. Letters from the beginning of the alphabet, $a, b, c, \ldots$ will refer to an orthonormal basis, while letters from the middle of the alphabet, $i, j, k, \ldots$ will refer to a coordinate basis.

The relationship between an orthonormal frame (or pseudo-orthonormal if q is nonzero) and a coordinate frame therefore appear as

$$e_a = e_a{}^m \frac{\partial}{\partial x^m}$$

The form basis dual to an orthonormal vector basis is given by

$$\mathbf{e}^a = e_m{}^a \mathbf{d}x^m$$

Furthermore, we have the relations

$$\begin{aligned} g_{mn} &= e_m{}^a e_n{}^b \eta_{ab} \\ \eta_{ab} &= e_a{}^m e_b{}^n g_{mn} \end{aligned} \tag{22}$$

Prove eqs(22).

Let g^{mn} denote the inverse to g_{mn}, so that $g^{in} g_{nm} = \delta^i_m$. Similarly, let η^{ab} be inverse to η_{ab}. Show that

$$\begin{aligned} g^{mn} &= \eta^{ab} e_a{}^m e_b{}^n \\ \eta^{ab} &= e_m{}^a e_n{}^b g^{mn} \end{aligned}$$

The polar basis vectors $\left(\hat{\rho}, \hat{\varphi}, \hat{\mathbf{k}}\right)$ form an orthonormal set, and since the line element is

$$ds^2 = d\rho^2 + \rho^2 d\varphi^2 + dz^2$$

the metric in polar coordinates is

$$g_{mn} = \begin{pmatrix} 1 & & \\ & \rho^2 & \\ & & 1 \end{pmatrix}$$

1. Find the inverse metric.

2. Express the orthonormal basis $\mathbf{e}^a = \left(\hat{\rho}, \hat{\varphi}, \hat{\mathbf{k}}\right)$ in terms of the coordinate basis $\left(\frac{\partial}{\partial\rho}, \frac{\partial}{\partial\varphi}, \frac{\partial}{\partial z}\right)$.

The inverse metric is

$$g^{mn} = \begin{pmatrix} 1 & & \\ & \frac{1}{\rho^2} & \\ & & 1 \end{pmatrix}$$

The orthonormal basis, $\mathbf{e}_a$, is related to the orthonormal basis $\frac{\partial}{\partial x^m}$ by

$$\mathbf{e}_a = e_a{}^m \frac{\partial}{\partial x^m}$$

where

$$\begin{aligned} \eta^{ab} e_a{}^m e_b{}^n &= g^{mn} \\ \eta^{ab} &= \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \end{aligned}$$

Therefore, we may choose

$$e_a{}^m = \begin{pmatrix} 1 & & \\ & \frac{1}{\rho} & \\ & & 1 \end{pmatrix}$$

so that

$$\begin{aligned} \mathbf{e}_a &= e_a{}^m \frac{\partial}{\partial x^m} \\ \begin{pmatrix} \hat{\rho} \\ \hat{\varphi} \\ \hat{\mathbf{k}} \end{pmatrix} &= \begin{pmatrix} 1 & & \\ & \frac{1}{\rho} & \\ & & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial\rho} \\ \frac{\partial}{\partial\varphi} \\ \frac{\partial}{\partial z} \end{pmatrix} \end{aligned}$$

and therefore

$$\begin{aligned} \hat{\rho} &= \frac{\partial}{\partial\rho} \\ \hat{\varphi} &= \frac{1}{\rho}\frac{\partial}{\partial\varphi} \\ \hat{\mathbf{k}} &= \frac{\partial}{\partial z} \end{aligned}$$

The spherical basis vectors $\left(\hat{\mathbf{r}}, \hat{\theta}, \hat{\varphi}\right)$ form an orthonormal set, and since the line element is

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \; d\varphi^2$$

the metric in polar coordinates is

$$g_{mn} = \begin{pmatrix} 1 & & \\ & r^2 & \\ & & r^2\sin^2\theta \end{pmatrix}$$

1. Find the inverse metric.
2. Express the orthonormal basis $\mathbf{e}^a = \left(\hat{\mathbf{r}}, \hat{\theta}, \hat{\varphi}\right)$ in terms of the coordinate basis $\left(\frac{\partial}{\partial r}, \frac{\partial}{\partial\theta}, \frac{\partial}{\partial\varphi}\right)$.

4.4 Group representations

One important way to understand tensors is through their transformation properties. Indeed, we will define tensors to be the set of objects which transform linearly and homogeneously under a given set of transformations. Thus, to understand tensors, we need to understand transformations, and in particular, transformation groups. The most important transformation groups are the Lie groups. To specify how a Lie group acts on tensors, we first define a linear representation of a Lie group. Such a representation requires a vector space.

To understand the idea of a representation, note that there are many ways to write any given Lie group. For example, the group of translations in 1-dim may be written as

$$g(a) = \exp\left(a\frac{d}{dx}\right)$$

because when we act on an arbitrary function of x, $g(a)$ just gives the Taylor series for $f(x+a)$ expanded about $f(x)$:

$$\begin{aligned} g(a) f(x) &= \exp\left(a\frac{d}{dx}\right) f \\ &= \sum_{n=0}^{\infty} \frac{a^n}{n!} \frac{d^n f}{dx^n}(x) \\ &= f(x+a) \end{aligned}$$

In particular, the Taylor series for the action of g on x has only two terms,

$$\begin{aligned} g(a) x &= \exp\left(a\frac{d}{dx}\right) x \\ &= x + a \end{aligned}$$

But we have already seen a quite different representation for tranlations. Representing a point on the real line by a pair

$$\begin{pmatrix} x \\ 1 \end{pmatrix}$$

the group elements take the form

$$g(a) = \begin{pmatrix} 1 & a \\ & 1 \end{pmatrix}$$

so that

$$g(a) \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} x+a \\ 1 \end{pmatrix}$$

When considering tensors, we will use only *linear representations* of Lie groups. A linear representation of a group is a vector space upon which the group acts. Be aware that vector spaces can take surprising forms – both of the examples above for the translations are linear representations. The difference is that the first vector space is a function space, whereas the second is a 2-dim vector space. Thus, different vector space give rise to different representations.

Once we choose a vector space on which a group acts, the form of the group elements is fixed. Thus, if we choose a function space as the representation and demand that

$$g(a) x = x + a \tag{23}$$

then the form of $g(a)$ is determined. To see how, notice that we may rewrite the right side of eq.(23) as

$$\begin{aligned} x + a &= \left(1 + a\frac{d}{dx}\right) x \\ &= \exp\left(a\frac{d}{dx}\right) x \end{aligned}$$

By constrast, suppose we know that $g(a)$ is to act on a 2 -vector according to

$$g(a)\begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} x+a \\ 1 \end{pmatrix}$$

Once again the form of g is immediately determined. A systematic development of a technique for finding group elements from their representations will be given in a later section.

In general, suppose we have a finite dimensional linear representation, V. Then for any vector $\mathbf{v} \in V$, group elements will act linearly to give another element, $\mathbf{w}$, of V :

$$\mathbf{w} = g(\mathbf{v})$$

If we expand $\mathbf{v}$ and $\mathbf{w}$ in a basis $\mathbf{e}_i$, then using linearity

$$\begin{aligned} w^i \mathbf{e}_i &= g\left(v^i \mathbf{e}_i\right) \\ &= v^i g(\mathbf{e}_i) \end{aligned}$$

so the form of g is determined by its action on the basis vectors. By closure, this action must be another vector, and therefore expressible in terms of the basis vectors,

$$\begin{aligned} g(\mathbf{e}_i) &= \mathbf{u} \\ &= {u_{(i)}}^j \mathbf{e}_j \\ &\equiv \mathbf{e}_j {g^j}_i \end{aligned}$$

Now the action of g on a general vector is given by

$$\begin{aligned} w^i \mathbf{e}_i &= g\left(v^i \mathbf{e}_i\right) \\ &= v^i g(\mathbf{e}_i) \\ &= v^i \mathbf{e}_j {g^j}_i \end{aligned}$$

or simply the matrix transformation

$$w^j = {g^j}_i v^i$$

Finite linear representations of Lie groups therefore have matrix representations.

Notice that there are also nonlinear representations of Lie groups. For example, the set of fractional linear transformations of the complex numbers (including a point at infinity) have the nonlinear representation

$$w = \frac{az+b}{cz+d}$$

which includes, among other things, inversions.

4.5 Tensors

There is a great deal that is new in the notation we have just introduced, and the reader may wonder why we need these tools. Certainly there is a great deal of power in being able to use any coordinates, but we could probably figure out the expressions we need on a case by case basis. However, there are some deeper things going on. First, we are gaining access to new objects – we introduced the moment of inertia tensor and the metric tensor, and in time will introduce other tensors of even higher rank. Without these tools, these new objects won't make much sense, even though the objects directly describe physical properties of material bodies.

But there is a more important reason for learning these techniques. Over the last 50 years, symmetry has come to play a central role in understanding the fundamental interactions of nature and the most basic

constituents of matter. In order to study these particles and interactions, we need to work with objects that transform in a simple way *under the relevant symmetry.* Thus, if we want to use the group $SU(2)$ to study the weak interaction, we need to be able to write down $SU(2)$ invariant quantities in a systematic way. Similarly, if we want to study special relativity, we need to work with Lorentz tensors – that is, objects which transform linearly and homogeneously under Lorentz transformations. Knowing these, we can easily construct objects which are Lorentz invariant using the 4 -dimensional equivalent of the three dimensional dot product. Such invariants will be the same for all observers, so we won't need to worry about actually *doing* Lorentz transformations. We will have formulated the physics in terms of quantities that can be calculated in *any* frame of reference.

We are now in a position to develop tensors of arbitrary rank $(0, 1, 2, \ldots)$ and type (form, vector). We accomplish this by taking outer products of vectors and forms. Given two vectors $\mathbf{u}$ and $\mathbf{v}$ we can define their (linear) outer product,

$$\mathbf{u} \otimes \mathbf{v}$$

If we think of $\mathbf{u}$ and $\mathbf{v}$ as directional derivatives along curves parameterized by λ and τ respectively, then we can let the outer product act on a pair of functions (f, g) to get

$$(\mathbf{u} \otimes \mathbf{v})(f, g) = \frac{df}{d\lambda}\frac{dg}{d\tau}$$

so the product is a doubly linear operator,

$$\mathbf{u} \otimes \mathbf{v} = \frac{d}{d\lambda} \otimes \frac{d}{d\tau}$$

We can also expand in a basis and think of the product as a matrix,

$$\begin{aligned} \mathsf{M} &= \mathbf{u} \otimes \mathbf{v} \\ &= u^i v^j \mathbf{e}_i \otimes \mathbf{e}_j \end{aligned}$$

with components

$$[\mathsf{M}]^{ij} = u^i v^j$$

This is just what we would get if we took a column-row product:

$$\begin{pmatrix} u^1 \\ u^2 \\ u^3 \end{pmatrix} \left(v^1, v^2, v^3\right) = \begin{pmatrix} u^1v^1 & u^1v^2 & u^1v^3 \\ u^2v^1 & u^2v^2 & u^2v^3 \\ u^3v^1 & u^3v^2 & u^3v^3 \end{pmatrix}$$

Of course, the most general 3×3 matrix cannot be written as $u^i v^j$ since there are not enough degrees of freedom. We can fix this by taking linear combinations. To see what this means, we examine the basis elements.

Consider the basis elements, $\mathbf{e}_i \otimes \mathbf{e}_j$. These are regarded as formally independent, and they may be written as a basis for matrices. If we choose an orthonormal basis with

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

then we can think of the products as

$$\begin{aligned} \mathbf{e}_1 \otimes \mathbf{e}_1 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \mathbf{e}_1 \otimes \mathbf{e}_2 &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

and so on so that $\mathbf{e}_i \otimes \mathbf{e}_j$ has a 1 in the i^{th} row and the j^{th} column and zeros everywhere else. Linear combinations of these clearly give every possible 3×3 matrix. Because the outer product is linear, we can add two or more of these matrices together,

$$\begin{aligned} \mathsf{M} &= \mathbf{u} \otimes \mathbf{v} + \mathbf{w} \otimes \mathbf{s} + \ldots \\ &= \left(u^i v^j + w^i s^j + \ldots\right) \mathbf{e}_i \otimes \quad \mathbf{e}_j \\ &= [\mathsf{M}]^{ij}\, \mathbf{e}_i \otimes \mathbf{e}_j \end{aligned}$$

Since the nine matrices $\mathbf{e}_i \otimes \mathbf{e}_j$ form a basis for all matrices, it is clear that by adding together enough outer products of vectors we can construct any matrix we like.

Show that the set of all matrices, $\mathcal{M} = \{\mathsf{M}\}$, forms a vector space. What is the dimension of $\mathcal{M}$?

The vector space $\mathcal{M}$ is the space of second rank tensors. To prove this we must show that they transform linearly and homogeneously under coordinate transformations. They do because they are built linearly from vectors.

Let M be expanded in a coordinate basis,

$$\mathsf{M} = [\mathsf{M}]^{ij}\, \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j}$$

Treating M as invariant, show that the components of M transform with two factors of the Jacobian matrix, i.e;,

$$\left[\tilde{\mathsf{M}}\right]^{ij} = J^i{}_m J^j{}_n\, [\mathsf{M}]^{mn}$$

We can continue taking outer products to form rank-n tensors by taking sums of products of n vectors:

$$\mathsf{T} = \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \ldots \otimes \mathbf{u}_n + \mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \ldots \otimes \mathbf{v}_n + \ldots$$

In order to keep track of the rank of tensors, we can use *abstract index notation*, putting a lower case latin index on T for each rank. A rank-n tensor will have n *labels*

$$\mathsf{T}^{a_1 a_2 \ldots a_n}$$

Keep in mind that these are labels, not indices: $\mathsf{T}^{a_1 a_2 \ldots a_n}$ is the same (invariant) object as T, but with a bit of extra information attached. By looking at the labels we immediately know that $\mathsf{T}^{a_1 a_2 \ldots a_n}$ is n-times contravariant.

Alternatively, we can write the *components* of T in any basis we choose. These are written using letters from further down the alphabet, or a different alphabet altogether,

$$T^{m_1 m_2 \ldots m_n}$$

Since these objects are components in a particular basis, and therefore change when we change the basis. The relationship between the notations is evident if we write

$$\mathsf{T}^{a_1 a_2 \ldots a_n} = T^{m_1 m_2 \ldots m_n} \mathbf{e}_{m_1} \otimes \mathbf{e}_{m_2} \otimes \cdots \otimes \mathbf{e}_{m_n}$$

Usually one or the other form is being used at a time, so little confusion occurs. Each value of each m_i gives a different component of the tensor, so in 3-dimensions a rank-n tensor will have 3^n different components.

We will *not* use abstract index notation outside of this section of this book, for two reasons. First, it is unnecessarily formal for an introductory work. There are numerous differential geometry and general relativity books that use the convention throughout, so that the interested reader can easily learn to use it elsewhere. Second, we use early-alphabet and middle-alphabet letters to distinguish between orthonormal and coordinate bases – a different distinction that will be more important for our purposes. This convention will be introduced in our discussion of differential forms.

When working with vector and forms together, we can take arbitrary outer products of both:

$$\begin{aligned}\mathsf{T} &= \mathbf{u}_1 \otimes \omega_2 \otimes \ldots \otimes \mathbf{u}_n \\ &\quad + \mathbf{v}_1 \otimes \psi_2 \otimes \ldots \otimes \mathbf{v}_n + \ldots\end{aligned}$$

Notice that any alternation of vectors and forms must always occur in the same order. In this example, the second position is always a form. When forms are present, the corresponding labels or indices of T will be written as subscripts to indicate that they are in a form basis. We need to keep the horizontal spacing of the indices fixed so we don't lose track of which position the forms occur in. Thus,

$$\mathsf{T}^{a_1}{}_{a_2}{}^{a_3 \ldots a_n} = T^{m_1}{}_{m_2}{}^{m_3 \ldots m_n} \mathbf{e}_{m_1} \otimes \mathbf{e}^{m_2} \otimes \mathbf{e}_{m_3} \otimes \ \cdots \otimes \mathbf{e}_{m_n}$$

If we have a metric, so that forms and vectors correspond to the same object, we can raise and lower any indices we like on any tensor. For example, formally,

$$\begin{aligned}\mathsf{T}^a{}_b{}^{c \ldots d} &= \mathsf{g}_{be} \mathsf{T}^{aec \ldots d} \\ \mathsf{T}_{abc \ldots d} &= \mathsf{g}_{ae} \mathsf{g}_{bf} \mathsf{g}_{cg} \cdots \mathsf{g}_{dh} \mathsf{T}^{efg \ldots h}\end{aligned}$$

or in components

$$T^{m_1}{}_m{}^{m_3 \ldots m_n} = g_{mn} T^{m_1 n m_3 \ldots m_n}$$

In general, the abstract indices are moved about in the same way as coordinate indices. This is only confusing for a while.

The most general tensors have m contravariant and n covariant labels or indices. They are of type $\binom{m}{n}$, and may be regarded as multilinear mappings from n functions and m curves to the reals. They transform with m copies of the Jacobian matrix and n copies of its inverse.

The space of tensors of all ranks is generally large enough to encompass our needs for physical quantities. Tensors themselves are not measurable directly because their components change when we change coordinates. But, like the length of a vector, we can form truly invariant combinations. Making invariant combinations is easy because tensors transform *covariantly.* This is a different use of the word covariant! All we mean here is that tensors, of any type and rank, transform linearly and homogeneously under coordinate transformations. Because of this, whenever we form a sum between a form-type tensor index and a vector-type index, the "dummy" indices no longer transform – the Jacobian of one dummy index cancels the inverse Jacobian of the other. Therefore, any inner products of tensors transform according to the number of *free* indices. To form an invariant quantity – one capable of physical measurement – we only need to produce an expression with no free indices. For example, the rotational kinetic energy

$$T = \frac{1}{2} I_{ij} \omega^i \omega^j$$

is coordinate invariant because it has no free indices. Any exotic combination will do. Thus,

$$T^{ijk}{}_{mn} v^m I_{jk} R^n{}_i$$

is coordinate invariant, and, in principal, measurable.

We also consider $\binom{m}{n}$tensors which have m contravariant and n covariant indices, in some specified order. Notice how the convention for index placement corresponds to the tensor type. When dealing with mixed tensors, it is important to exactly maintain index order: $T^a{}_b{}^c$ is a different object than $T^{ac}{}_b$!

By having tensors at hand, it becomes easy to form quantities which, like the dot product, are invariant under a set of transformations. It is these invariant quantities that must constitute our physical theory, because physical quantities cannot depend on our choice of coordinates.Inner products and norms.

The tensors we have discussed here are covariant with respect to the diffeomorphism group in 3 dimensions. Evaluated at any point, such a transformation is a general linear transformation, hence an element of the Lie group $GL(3)$. However, we may define other objects where the relevant transformations are given

by a Lie group. In the next sections, for example, we will consider orthonormal bases for vectors and forms. By placing this restriction on the allowed bases, we restrict the allowed transformations to orthogonal group, $SO(3)$. Numerous other choices have physical application as well.

Part II
Motion: Lagrangian mechanics

Starting our investigation with our immediate perceptions of the world, we chose to model the world with a 3-dimensional space, with a universal time. In order to guarantee that the physical properties of an object do not depend on the absolute position or orientation of the object, we asked for the space to be homogeneous and isotropic. This led us to construct space from the Euclidean group of translations and rotations. To be able to describe uniform motion in the resulting space, we developed the tools of variational calculus. The result was the generalized Euler-Lagrange equation. We now turn to a systematic study of the generalized Euler-Lagrange equation, eq.(15), as a description of motion in Euclidean 3-space. In this chapter we explore some of the properties of Lagrangian systems which depend only on certain general properties of the action functional.

According to the claims of the previous chapters, in order for the Euler-Lagrangian equation to have physical content, it must be a tensor. It may seem to be sufficient for it to be a tensor with respect to the Euclidean group, since that is the symmetry of our chosen arena. But remember that we also need to avoid any dependence on our choice of coordinates – the laws of motion should not depend on how we label points. For this reason, it is very important that our description of motion be covariant with respect to the full diffeomorphism group.

Once we have established the tensor character of the Euler-Lagrange equation, we turn to the idea of symmetry. Defining symmetry as an invariance of the action, we prove the Noether theorem – for every continuous symmetry of a Lagrangian system, we can find a corresponding conserved quantity. We then study a number of common symmetries and their corresponding conservation laws. Because the generalized Euler-Lagrange equation conceals as much as it reveals about these conservation laws, we begin with a restricted class of Lagrangians – those which depend only on the path, $x^i(t)$, and its first time derivative, the velocity $v^i = \dot{x}^i$. Then, for completeness, we extend some of the results to general Lagrangians.

5 Covariance of the Euler-Lagrangian equation

We begin with the symmetry of sets of variational equations. Suppose we have a functional S, expressed as an integral along a curve, $C = x^i(t)$:

$$S[x(t)] = \int_C L(\mathbf{x}, \dot{\mathbf{x}}, \ldots)\, dt$$

The function L is called the *Lagrangian*. Suppose further that the Lagrangian L depends only on $\mathbf{x}$, $\dot{\mathbf{x}}$ and t, but not higher derivatives. Then the generalized Euler-Lagrange equation, eq.(15), reduces to the *Euler-Lagrange equation*,

$$\frac{\partial L}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = 0 \tag{24}$$

Now consider a general change of coordinates, $q^i = q^i(\mathbf{x})$ and its inverse, $x^i = x^i(\mathbf{q})$. How does the equation of motion change? We find

$$\begin{aligned} S[\mathbf{x}(\mathbf{q}(t), t)] &= \int_C L(\mathbf{x}(\mathbf{q}, t), \dot{\mathbf{x}}(\mathbf{q}, \dot{\mathbf{q}}, t))\, dt \\ &= \int_C L(\mathbf{q}, \dot{\mathbf{q}}, t)\, dt \end{aligned}$$

so the action may now be treated as a functional of the new coordinate. We wish to find the relationship between the Euler-Lagrange equation for $x^i(t)$ and the Euler-Lagrange equation for $q^i(t)$.

Consider the variational equation for q^i, computed in two ways. First, we may immediately write the Euler-Lagrange equation, eq.(15) by varying $S[\mathbf{q}(t)]$. Following the usual steps,

$$\begin{aligned}\delta S &= \delta \int_C L(\mathbf{q}, \dot{\mathbf{q}}, t)\, dt \\ &= \int_C \left(\frac{\partial L}{\partial q^k} \delta q^k + \frac{\partial L}{\partial \dot{q}^k} \delta \dot{q}^k \right) dt \\ &= \int_C \left(\frac{\partial L}{\partial q^k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^k} \right) \right) \delta q^i\end{aligned}$$

dropping the surface term in the final step, since the variation is taken to vanish at the endpoints. Now compare what we get by varying $S\left[\, \mathbf{x}\left(q^i(t), t\right)\right]$ with respect to $q^i(t)$:

$$\begin{aligned}0 &= \delta S \\ &= \delta \int_C L(\mathbf{x}(\mathbf{q}, t), \dot{\mathbf{x}}(\mathbf{q}, \dot{\mathbf{q}}, t))\, dt \\ &= \int_C \left(\frac{\partial L}{\partial x^k} \frac{\partial x^k}{\partial q^i} \delta q^i + \frac{\partial L}{\partial \dot{x}^k} \frac{\partial \dot{x}^k}{\partial q^i} \delta q^i + \frac{\partial L}{\partial \dot{x}^k} \frac{\partial \dot{x}^k}{\partial \dot{q}^i} \delta \dot{q}^i + \frac{\partial L}{\partial x^k} \frac{\partial x^k}{\partial \dot{q}^i} \delta \dot{q}^i \right) dt\end{aligned}$$

Since x^i is a function of q^j and t only, $\frac{\partial x^k}{\partial \dot{q}^i} = 0$ and the last term vanishes. Expanding the velocity, $\dot{x}$, explicitly, we have:

$$\begin{aligned}\dot{x}^k &= \frac{dx^k}{dt} \\ &= \frac{d}{dt} x^k\left(q^i(t), t\right) \\ &= \frac{\partial x^k}{\partial q^i} \dot{q}^i + \frac{\partial x^k}{\partial t}\end{aligned} \tag{25}$$

so that, differentiating,

$$\frac{\partial \dot{x}^k}{\partial \dot{q}^i} = \frac{\partial x^k}{\partial q^i}$$

Finally, we differentiate eq.(25) for the velocity with respect to q^i :

$$\begin{aligned}\frac{\partial \dot{x}^k}{\partial q^i} &= \frac{\partial^2 x^k}{\partial q^i \partial q^j} \dot{q}^j + \frac{\partial^2 x^k}{\partial q^i \partial t} \\ &= \frac{\partial}{\partial q^j}\left(\frac{\partial x^k}{\partial q^i} \right) \dot{q}^j + \frac{\partial}{\partial t} \frac{\partial x^k}{\partial q^i} \\ &= \frac{d}{dt} \frac{\partial x^k}{\partial q^i}\end{aligned}$$

Substituting, the variation now reduces to

$$\begin{aligned}0 &= \delta S \\ &= \int_C \left(\frac{\partial L}{\partial x^k} \frac{\partial x^k}{\partial q^i} \delta q^i + \frac{\partial L}{\partial \dot{x}^k} \frac{d}{dt}\left(\frac{\partial x^k}{\partial q^i} \right) \delta q^i + \frac{\partial L}{\partial \dot{x}^k} \frac{\partial x^k}{\partial q^i} \delta \dot{q}^i \right) dt \\ &= \int_C \left(\frac{\partial L}{\partial x^k} \frac{\partial x^k}{\partial q^i} + \frac{\partial L}{\partial \dot{x}^k} \frac{d}{dt}\left(\frac{\partial x^k}{\partial q^i} \right) \delta q^i - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k} \frac{\partial x^k}{\partial q^i} \right) \right) \delta q^i + surface\ term\end{aligned}$$

$$
\begin{aligned}
&= \int_C \left(\frac{\partial L}{\partial x^k} \frac{\partial x^k}{\partial q^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^k} \right) \frac{\partial x^k}{\partial q^i} \right) \delta q^i \\
&= \int_C \left(\frac{\partial L}{\partial x^k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^k} \right) \right) \frac{\partial x^k}{\partial q^i} \delta q^i
\end{aligned}
$$

We can now write δS in either of two ways:

$$
\delta S = \int_C \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) \right) \delta q^i dt
$$

or

$$
\delta S = \int_C \left(\frac{\partial L}{\partial x^k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^k} \right) \right) \frac{\partial x^k}{\partial q^i} \delta q^i dt
$$

Since the coefficient of δq^i must be the same whichever way we write the variation, this means that the Euler-Lagrange equation transforms as

$$
\begin{aligned}
\frac{\partial L}{\partial q^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}^i} \right) &= \left(\frac{\partial L}{\partial x^k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^k} \right) \right) \frac{\partial x^k}{\partial q^i} \\
&= \left(\frac{\partial L}{\partial x^k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^k} \right) \right) J_i^k
\end{aligned}
$$

where the matrix $J_i^k = \frac{\partial x^k}{\partial q^i}$ is just the Jacobian matrix of a general coordinate transformation. The Euler-Lagrange equation is therefore covariant with respect to the diffeomorphism group.

Let L be a function of $x^i(t)$ and its first and second derivatives and consider an arbitrary diffeomorphism, $x^i = x^i(\mathbf{q}(t), t)$ Repeat the preceeding calculation, considering variations, δq^i, which vanish at the endpoints of the motion, together with their first time derivative,

$$
\delta q^i = \delta \dot{q}^i = 0
$$

to show that the generalized Euler-Lagrange equation is a diffeomorphism group tensor.

Hint. Differentiate

$$
\begin{aligned}
\dot{x}^k &= \frac{dx^k}{dt} \\
&= \frac{\partial x^k}{\partial q^i} \dot{q}^i + \frac{\partial x^k}{\partial t}
\end{aligned} \tag{26}
$$

again to find

$$
\begin{aligned}
\ddot{x}^k &= \frac{d\dot{x}^k}{dt} \\
&= \frac{d}{dt} \left(\frac{\partial x^k}{\partial q^i} \dot{q}^i + \frac{\partial x^k}{\partial t} \right) \\
&= \frac{\partial x^k}{\partial q^i} \ddot{q}^i + \frac{\partial^2 x^k}{\partial q^j \partial q^i} \dot{q}^j \dot{q}^i + \frac{\partial^2 x^k}{\partial q^i \partial t} \dot{q}^i + \frac{\partial^2 x^k}{\partial t^2}
\end{aligned}
$$

from which we can easily find the partial derivatives, of $x^k, \dot{x}^k, \ddot{x}^k$ with respect to $q^m, \dot{q}^m$ and $\ddot{q}^m$. It is also useful to show by expanding in partial derivatives that

$$
\frac{d}{dt} \frac{\partial x^k}{\partial q^m} = \frac{\partial \dot{x}^k}{\partial q^m}
$$

The remainder of the problem is straightforward, but challenging.

6 Symmetries and the Euler-Lagrange equation

There are important general properties of Euler-Lagrange systems based on the symmetry of the Lagrangian. The most important result based on symmetry is Noether's Theorem, which we prove for the generalized Euler-Lagrange system. Then, to introduce the ideas in a more accessible way, we present applications of Noether's Theorem for Lagrangians which depend only on the coordinates and their first derivatives. Finally, we generalize several of the main results for general Lagrangians.

6.1 Noether's theorem for the generalized Euler-Lagrange equation

We have shown that the action

$$S\left[\mathbf{x}\left(t\right)\right]=\int_{C}L\left(x^{i},x_{(1)}^{i},x_{(2)}^{i},\ldots x_{(n)}^{i}\right)dt$$

where $x_{(k)}^{i}$ denotes the kth derivative of $x^{i}\left(t\right)$, is extremal when $x^{i}\left(t\right)$ satisfies the generalized Euler-Lagrange equation,

$$\left.\frac{\partial L}{\partial x}\right|_{x(t)}-\left.\frac{d}{dt}\frac{\partial L}{\partial x^{(1)}}\right|_{x(t)}+\cdots+\left(-1\right)^{n}\left.\frac{d^{n}}{dt^{n}}\frac{\partial L}{\partial x^{(n)}}\right|_{x(t)}=0 \tag{27}$$

This condition guarantees that δS vanishes for *all* variations, $x\left(t\right)\rightarrow x\left(t\right)+\delta x\left(t\right)$ which vanish at the endpoints of the motion.

Sometimes it is the case that δS vanishes for certain limited variations of the path without imposing any condition at all. When this happens, we say that S has a symmetry:

A *symmetry* of an action functional $S\left[x\right]$ is a transformation of the path, $x^{i}\left(t\right)\rightarrow\lambda^{i}\left(x^{j}\left(t\right),t\right)$ that leaves the action invariant,

$$S\left[x^{i}\left(t\right)\right]=S\left[\lambda^{i}\left(x^{j}\left(t\right),t\right)\right]$$

In particular, when $\lambda^{i}\left(x\right)$ represents the action of a Lie group on x, we may expand the transformation infinitesimally, so that

$$\begin{aligned}x^{i} &\rightarrow x'^{i}=x^{i}+\varepsilon^{i}\left(x\right)\\ \delta x^{i} &= x'^{i}-x^{i}=\varepsilon^{i}\left(x\right)\end{aligned}$$

Since the infinitesimal transformation must leave $S\left[x\right]$ invariant, we have

$$\delta S=S\left[x^{i}+\varepsilon^{i}\left(x\right)\right]-S\left[x^{i}\right]=0$$

Conversely, if the infinitesimal variation vanishes, we may apply arbitrarily many infinitesimal transformations to recover the invariance of S under finite transformations.

Here $\lambda(x)$ is a particular function of the coordinates. This is quite different from performing a general variation – we are not placing any new demand on the action, just noticing that particular transformations don't change it. Notice that neither λ^{i} nor ε^{i} is required to vanish at the endpoints of the motion.

Let $x^{i}\left(t\right)$ be a solution to the generalized Euler-Lagrange equations of motion. Then a function of $x^{i}\left(t\right)$ and its time derivatives,

$$f\left(x^{i}\left(t\right),\ldots,x_{(n)}^{i}\left(t\right)\right)$$

is *conserved* if it is constant along the paths of motion,

$$\frac{df}{dt}=0$$

We now show that when an action has a symmetry, we can derive a conserved quantity.

(Noether's Theorem) Suppose an action has a Lie symmetry so that it is invariant under

$$\delta x^{i}=x'^{i}-x^{i}=\varepsilon^{i}\left(x\right)$$

where $\varepsilon^i(x)$ is a fixed function of $x^i(t)$. Then the quantity

$$I = \sum_{k=1}^{n} \sum_{m=1}^{k} (-1)^{m-1} \frac{d^{m-1}}{dt^{m-1}} \frac{\partial L(x(\lambda))}{\partial x^i_{(k)}} \frac{d^{k-m}}{dt^{k-m}} \varepsilon^i(x)$$

is conserved.

We prove the theorem for $n = 1$. The proof for arbitrary n is left as an exercise. When $n = 1$, variation of the action gives

$$0 = \delta S[x(t)] \equiv \int_{t_1}^{t_2} \left(\frac{\partial L(x(t))}{\partial x^i} \varepsilon^i(x) + \left(\frac{\partial L(x(t))}{\partial \dot{x}^i_{(n)}} \right) \frac{d\varepsilon^i(x)}{dt} \right) dt$$

Notice that δS vanishes identically because the action has a symmetry. No equation of motion has been used. Integrating by parts we have

$$\begin{aligned} 0 &= \int \left(\frac{\partial L}{\partial x^i} \varepsilon^i(x) + \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^i} \varepsilon^i(x) \right) - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^i} \right) \varepsilon^i(x) \right) dt \\ &= \frac{\partial L}{\partial \dot{x}^i} \varepsilon^i(x) \Big|_{t_1}^{t_2} + \int \left(\frac{\partial L}{\partial x^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^i} \right) \right) \varepsilon^i(x) dt \end{aligned}$$

This expression must vanish for *every* path. Now suppose $x^i(t)$ is an actual classical path of the motion, that is, one that satisfies the Euler-Lagrange equation,

$$\frac{\partial L}{\partial x^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^i} \right) = 0$$

Then *for that path*, the integrand vanishes and it follows that

$$0 = \delta S[\mathbf{x}] = \frac{\partial L}{\partial \dot{x}^i} \varepsilon^i(x(t)) \Big|_{t_1}^{t_2}$$

or

$$\frac{\partial L(\mathbf{x}(t_2), \dot{\mathbf{x}}(t_2))}{\partial \dot{x}^i} \varepsilon^i(x(t_2)) = \frac{\partial L(\mathbf{x}(t_1), \dot{\mathbf{x}}(t_1))}{\partial \dot{x}^i} \varepsilon^i(x(t_1))$$

Since t_1 and t_2 are arbitrary, the function

$$\frac{\partial L(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{x}^i} \varepsilon^i(x)$$

is a constant of the motion.

Prove Noether's theorem when the Lagrangian depends on the first n time derivatives of the position $x^i(t)$. That is, show that the quantity

$$I = \sum_{k=1}^{n} \sum_{m=1}^{k} (-1)^{m-1} \frac{d^{m-1}}{dt^{m-1}} \frac{\partial L(x(\lambda))}{\partial x^i_{(k)}} \frac{d^{k-m}}{dt^{k-m}} \varepsilon^i(x)$$

is conserved.

Hint: For the general case the variation of the action is

$$\delta S[x(t)] \equiv \int_{t_1}^{t_2} \left(\frac{\partial L(x(t))}{\partial x^i} \varepsilon^i(x) + \ldots + \left(\frac{\partial L(x(t))}{\partial x^i_{(n)}} \right) \frac{d^n}{dt^n} \varepsilon^i(x) \right) d\lambda$$

where the k^{th} term is:

$$I_k = \int_{t_1}^{t_2} \frac{\partial L(x(\lambda))}{\partial x^i_{(k)}} \frac{d^k}{d\lambda^k} \varepsilon^i(x) \, d\lambda$$

Integrate this term by parts k times, keeping careful track of the surface terms. After writing the surface term for the k^{th} integral as a sum over m, sum over all k.

We conclude this subsection with a definition.

A coordinate, q, is *cyclic* if it does not occur in the Lagrangian, i.e.,

$$\frac{\partial L}{\partial q} = 0$$

In the following section, we consider the application of Noether's Theorem to restricted Euler-Lagrange systems, that is, those for which the Lagrangian depends only on x and its first time derivative. In the subsequent section, we generalize some of the results to arbitrary Lagrangians.

6.2 Conserved quantities in restricted Euler-Lagrange systems

For restricted Euler-Lagrange systems, the Lagrangian take the form

$$L = L\left(x^i, \dot{x}^i = x^i_{(1)}\right)$$

so that the Euler-Lagrange equation reduces to

$$\frac{\partial L}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right) = 0$$

In such cases, Noether's theorem states that for an infinitesimal symmetry

$$\delta x^i = x'^i - x^i = \varepsilon^i(x)$$

the quantity

$$Q = \frac{\partial L(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{x}^i}\varepsilon^i(\mathbf{x})$$

is conserved. We make one further definition for restricted Euler-Lagrange systems:

The *conjugate momentum*, p, to any coordinate q is defined to be

$$p = \frac{\partial L}{\partial \dot{q}}$$

6.2.1 Cyclic coordinates and conserved momentum

We have the following consequences of a cyclic coordinate:

If a coordinate q is cyclic then

1. The system has translational symmetry, since the action is invariant under the translation

$$q \to q + a$$

2. The momentum conjugate to q is conserved.

To prove the first result, simply notice that if

$$\frac{\partial L}{\partial q} = 0$$

then L has no dependence on q at all. Therefore, replacing q by $q + a$ does nothing to L, hence nothing to the action. Equivalently, the variation of the action with respect to the infinitesimal symmetry $(a \to \varepsilon)$,

$$\begin{aligned} \delta q &= \varepsilon \\ \delta \dot{q} &= 0 \end{aligned}$$

is

$$\begin{aligned}\delta S &= \int \left(\frac{\partial L}{\partial q} \delta q + \frac{\partial L}{\partial \dot{q}} \delta \dot{q} \right) dt \\ &= \int \left(0 \cdot \delta q + \frac{\partial L}{\partial \dot{q}} \cdot 0 \right) dt \\ &= 0\end{aligned}$$

so the translation is a symmetry of the action.

For the second result, the Euler-Lagrange equation for the coordinate q immediately gives

$$\begin{aligned}0 &= \frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}} \right) \\ &= -\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}} \right)\end{aligned}$$

so that

$$p = \frac{\partial L}{\partial \dot{q}}$$

is conserved.

We now generalize this result to general translational invariance of the action. Suppose the action for a 1-particle system is invariant under arbitrary translations,

$$\bar{x}^i = x^i + a^i$$

or infinitesimally,

$$\delta x^i = \varepsilon^i$$

We may express the invariance of S under $\delta x^i = \varepsilon^i$ explicitly,

$$\begin{aligned}0 &= \delta S \\ &= \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x^i} \delta x^i + \frac{\partial L}{\partial \dot{x}} \delta \dot{x} \right) dt \\ &= \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x^i} \delta x^i + \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}} \delta x \right) - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}} \right) \delta x \right) dt \\ &= \frac{\partial L}{\partial \dot{x}^i} \varepsilon^i \Big|_{t_1}^{t_2} + \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x^i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i} \right) \right) \varepsilon^i dt\end{aligned}$$

For a particle which satisfies the Euler-Lagrange equation, the integral vanishes. Then, since t_1 and t_2 are arbitrary we must have

$$\frac{\partial L}{\partial \dot{x}^i} \varepsilon^i = p_i \varepsilon^i$$

conserved for all constants ε^i. Therefore, the momentum p_i conjugate to x^i is conserved as a result of translational invariance.

6.2.2 Rotational symmetry and conservation of angular momentum

Now consider a simple 2-dimensional example. Suppose the Lagrangian takes the form

$$L(x, y) = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2 \right)$$

Then the transformation

$$\begin{aligned} x &\rightarrow x' = x\cos\theta - y\sin\theta \\ y &\rightarrow y' = x\sin\theta + y\cos\theta \end{aligned}$$

for any fixed value of θ leaves

$$S\left[\mathbf{x}\right] = \int L dt$$

invariant; the change in x^i is

$$\begin{aligned} \lambda^1 &= \delta x = x' - x = x\cos\theta - y\sin\theta - x \\ \lambda^2 &= \delta y = y' - y = x\sin\theta + y\cos\theta - y \end{aligned}$$

The variation $\varepsilon^i(x)$ will be infinitesimal if the angle θ is infinitesimal, so to first order in θ we have

$$\begin{aligned} \varepsilon^1 &= x\cos\theta - y\sin\theta - x \\ &= x\left(1 - \frac{1}{2}\theta^2 + \cdots\right) - y\left(\theta - \frac{1}{6}\theta^3 + \cdots\right) - x \\ &= -y\theta \\ \varepsilon^2 &= x\theta \end{aligned}$$

Therefore, we have the conserved quantity,

$$\begin{aligned} \frac{\partial L}{\partial \dot{x}^i}\varepsilon^i\left(x\right) &= m\dot{x}\varepsilon^1 + m\dot{y}\varepsilon^2 \\ &= m\dot{x}\left(-y\theta\right) + m\dot{y}\left(x\theta\right) \\ &= 2\theta m\left(\dot{y}x - \dot{x}y\right) \end{aligned}$$

as long as x and y satisfy the equations of motion. Since 2θ is just an arbitrary constant to begin with, we can identify the angular momentum,

$$J = m\left(\dot{y}x - \dot{x}y\right)$$

as the conserved quantity.

We can easily generalize this result. Suppose $S$is rotationally invariant. Then S is invariant under the relacement of x^i by

$$\bar{x}^i = R^i{}_j x^j$$

or, infinitesimally,

$$\begin{aligned} \delta x^i &= \bar{x}^i - x^i \\ &= \left(\delta^i{}_j + A^i{}_j\right)x^j - x^i \\ &= A^i{}_j x^j \end{aligned}$$

where, according to our discussion of the Euclidean group, the infinitesimal matrix $A^i{}_j$ is antisymmetric,

$$A^j{}_k = g^{ji}A_{ik} = -g^{ji}A_{ki} = -A_k{}^j$$

Equivalently, we may write

$$A^{ij} = -A^{ji}$$

Now consider the (vanishing) variation of S under a rotation. We have

$$\begin{aligned} 0 &= \delta S \\ &= \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x^i} \delta x^i + \frac{\partial L}{\partial \dot{x}} \delta \dot{x} \right) dt \\ &= \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x^i} \delta x^i + \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \delta x \right) - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) \delta x \right) dt \\ &= \frac{\partial L}{\partial \dot{x}^i} \delta x^i \Big|_{t_1}^{t_2} + \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x^i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^i} \right) \right) \delta x^i dt \end{aligned}$$

For any particle obeying the Euler-Lagrange equation, the final integral vanishes. Since t_1 and t_2 are arbitrary times, we find the conserved quantity,

$$\begin{aligned} M &= \frac{\partial L}{\partial \dot{x}^i} \delta x^i \\ &= \frac{\partial L}{\partial \dot{x}^i} A^i{}_j x^j \\ &= p_i A^i{}_j x^j \\ &= p_i A^{ij} g_{jk} x^k \\ &= A^{ij} p_i x_j \\ &= \frac{1}{2} A^{ij} \left(p_i x_j - p_j x_i \right) \end{aligned}$$

Since we may write a general antisymmetric matrix using the Levi-Civita tensor as

$$A^{ij} = w_k \varepsilon^{ijk}$$

where

$$w^m = \frac{1}{2} A_{ij} \varepsilon^{ijm}$$

we have

$$M = \frac{1}{2} w_k \varepsilon^{ijk} \left(p_i x_j - p_j x_i \right) \equiv -\frac{1}{2} w_k M^k$$

Since w_k is an arbitrary constant vector and we may drop an overall constant $-\frac{1}{2}$, the vector

$$\mathbf{M} = \mathbf{x} \times \mathbf{p}$$

must be conserved. Thus, conservation of angular momentum is a consequence of rotational symmetry.

Conservation of angular monementum is a property of a number of important physical systems; moreover, the total angular momentum of any isolated system is conserved. The following theorems illustrate the usefulness of this conservation law in the case of a central potential, $V(r)$.

The angular momentum of a particle moving in an arbitrary central force, in any dimension $d \geq 2$, is conserved.

Proof The action may be written as

$$S = \int dt \left(\frac{1}{2} m \delta_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} - V(r) \right)$$

where the x^i are Cartesian coordinates and $r = \sqrt{\delta_{ij} x^i x^j}$. It follows that

$$m \frac{d^2 x^i}{dt^2} = -V' \frac{x^i}{r}$$

The total angular momentum

$$\begin{aligned} M_{ij} &= x_i p_j - x_j p_i \\ &= m\left(x_i \dot{x}_j - x_j \dot{x}_i\right) \end{aligned}$$

is conserved, since

$$\begin{aligned} \frac{d}{dt} M_{ij} &= m \frac{d}{dt}\left(x_i \dot{x}_j - x_j \dot{x}_i\right) \\ &= m\left(x_j \frac{d^2 x_k}{dt^2} - x_k \frac{d^2 x_j}{dt^2}\right) \\ &= -\frac{V'}{r}\left(x_j x_k - x_k x_j\right) \\ &= 0 \end{aligned}$$

We next prove that central forces always lead to planar motion.

The motion of a particle in a central potential always lies in a plane.

Proof: Let $\mathbf{x}_0$ and $\mathbf{v}_0$ be the initial postion and velocity, with $\mathbf{x}_0$ measured from the center of force. Then the angular momentum is

$$M_{ij} = x_{0i} v_{0j} - x_{0j} v_{0i}$$

For M_{ij} nonzero, let $\mathbf{w}_{(a)}$, $a = 1, \ldots, n-2$, be a collection of vectors perpendicular to the initial plane

$$\begin{aligned} P &= \left\{\mathbf{v} = \alpha \mathbf{x}_0 + \beta \mathbf{v}_0 \,|\forall \alpha, \beta\right\} \\ \mathbf{w}_{(a)} \cdot \mathbf{v} &= 0 \end{aligned}$$

so that the set $\left\{\mathbf{x}_0, \mathbf{v}_0, \mathbf{w}_{(a)}\right\}$ forms a basis. We consider the $M_{ij} = 0$ case below. For M_{ij} nonzero, for all a,

$$w^i_{(a)} M_{ij} = 0$$

At any time t, M_{ij} is given by

$$M_{ij} = m\left(x_i v_j - x_j v_i\right)$$

and since M_{ij} is constant we still have

$$\begin{aligned} 0 &= w^i_{(a)} m\left(x_i v_j - x_j v_i\right) \\ 0 &= \left(\mathbf{w}_{(a)} \cdot \mathbf{x}\right)\mathbf{v} - \mathbf{x}\left(\mathbf{w}_{(a)} \cdot \mathbf{v}\right) \end{aligned}$$

Suppose, for some a_0, that

$$\mathbf{w}_{(a_0)} \cdot \mathbf{x} \neq 0$$

Then

$$\mathbf{v} = \mathbf{x}\left(\frac{\mathbf{w}_{a_0} \cdot \mathbf{v}}{\mathbf{w}_{a_0} \cdot \mathbf{x}}\right)$$

and M_{ij} is identically zero, in contradiction to its constancy. Therefore, we conclude

$$\mathbf{w}_{(a)} \cdot \mathbf{x} = 0$$

for all a. A parallel argument shows that

$$\mathbf{w}_{(a)} \cdot \mathbf{v} = 0$$

for all a, so the motion continues to lie in the original plane. Finally, if $M_{ij} = 0$ then at any time t,

$$M_{ij} = m\left(x_i v_j - x_j v_i\right) = 0$$

so x_i and v_i are always parallel and we can write

$$x^i = \lambda v^i$$

for some $\lambda\left(x^j, t\right)$and all t. Then at any time

$$v^i = \frac{d}{dt}x^i = \frac{d\lambda}{dt}v^i + \lambda\frac{dv^i}{dt}$$

so any change in velocity is parallel to the velocity:

$$\frac{dv^i}{dt} = \frac{1}{\lambda}\left(1 - \frac{d\lambda}{dt}\right)v^i$$

and the motion remains along the initial line.

6.2.3 Conservation of energy

Conservation of energy is related to *time translation invariance*. However, this invariance is more subtle than simply replacing $t \to t + \tau$ in the action, since this transformation is simply a reparameterization of the action integral. Instead, the conservation law holds whenever the Lagrangian does not depend explicitly on time so that

$$\frac{\partial L}{\partial t} = 0$$

We can then write the total time derivative of L as

$$\frac{dL}{dt} = \frac{\partial L}{\partial x_i}\dot{x}_i + \frac{\partial L}{\partial \dot{x}_i}\ddot{q}_i$$

Using the Lagrange equations to replace

$$\frac{\partial L}{\partial x^i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}$$

in the first term, we get

$$\begin{aligned}\frac{dL}{dt} &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\right)\dot{x}^i + \frac{\partial L}{\partial \dot{x}^i}\ddot{x}^i \\ &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\dot{x}^i\right)\end{aligned}$$

Bringing both terms to the same side, we have

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^i}\dot{x}^i - L\right) = 0$$

so that the quantity

$$E \equiv \frac{\partial L}{\partial \dot{x}^i}\dot{x}^i - L$$

is conserved. The quantity E is called the *energy*.

6.2.4 Scale Invariance

As we have noted, physical measurements are always relative to our choice of unit. The resulting dilatational symmetry will be examined in detail when we study Hamiltonian dynamics. However, there are other forms

of rescaling a problem that lead to physical results. These results typically depend on the fact that the Euler-Lagrange equation is unchanged by an overall constant, so that the actions

$$\begin{aligned} S &= \int L dt \\ S' &= \alpha \int L dt \end{aligned}$$

have the same extremal curves.

Now suppose we have a Lagrangian which depends on some constant parameters $(a_1, \ldots, a_n)$ in addition to the arbitrary coordinates,

$$L = L\left(x^i, \dot{x}^i, a_1, \ldots, a_n, t\right)$$

These parameters might include masses, lengths, spring constants and so on. Further, suppose that each of these variables may be rescaled by some factor in such a way that L changes by only an overall factor. That is, when we make the replacements

$$\begin{aligned} x^i &\rightarrow \alpha x^i \\ t &\rightarrow \beta t \\ \dot{x}^i &\rightarrow \frac{\alpha}{\beta}\dot{x}^i \\ a_i &\rightarrow \gamma_i a_i \end{aligned}$$

for certain constants $(\alpha, \beta, \gamma_1, \ldots, \gamma_n)$ we find that

$$L\left(\alpha x^i, \frac{\alpha}{\beta}\dot{x}^i, \gamma_1 a_1, \ldots, \gamma_n a_n, \beta t\right) = \sigma L\left(x^i, \dot{x}^i, a_1, \ldots, a_n, t\right)$$

for some constant σ which depends on the scaling constants. Then the Euler-Lagrange equations for the system described by $L\left(\alpha x^i, \frac{\alpha}{\beta}\dot{x}^i, \gamma_1 a_1, \ldots, \gamma_n a_n, \beta t\right)$ are the same as for the original Lagrangian, and we may make the replacements in the solution.

Consider the 1-dimensional simple harmonic oscillator. The motion of the oscillator may be described by the Lagrangian

$$L = \frac{1}{12}m^2\dot{x}^4 + \frac{1}{2}km\dot{x}^2x^2 - \frac{1}{4}k^2x^4$$

since the restricted Euler-Lagrange equation gives

$$\begin{aligned} 0 &= \frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) \\ &= km\dot{x}^2x - k^2x^3 - \frac{d}{dt}\left(\frac{1}{3}m^2\dot{x}^3 + km\dot{x}x^2\right) \\ &= km\dot{x}^2x - k^2x^3 - \left(m^2\dot{x}^2\ddot{x} + km\ddot{x}x^2 + 2km\dot{x}^2x\right) \\ &= -\left(kx^2 + m\dot{x}^2\right)kx - \left(m\dot{x}^2 + kx^2\right)m\ddot{x} \end{aligned}$$

or simply

$$\left(m\dot{x}^2 + kx^2\right)\left(kx + m\ddot{x}\right) = 0$$

Assuming m and k are positive, the factor $\left(m\dot{x}^2 + kx^2\right)$ is positive definite except for the case of degenerate motion, $x(t) = 0$. Dividing by this factor, we have the usual equation of motion,

$$m\ddot{x} + kx = 0$$

Using scalings of the variables, we do not need to solve this equation to gain insight into the solution. For example, suppose we know that the motion is periodic, with period T. Now we may make any or all of the replacements

$$\begin{aligned} x &\rightarrow \alpha x \\ t &\rightarrow \beta t \\ \dot{x} &\rightarrow \frac{\alpha}{\beta}\dot{x} \\ m &\rightarrow \gamma m \\ k &\rightarrow \delta k \end{aligned}$$

for constants $(\alpha, \beta, \gamma, \delta)$. The Lagrangian becomes

$$L = \frac{1}{12}\gamma^2\frac{\alpha^4}{\beta^4}m^2\dot{x}^4 + \frac{1}{2}\gamma\delta\frac{\alpha^4}{\beta^2}km\dot{x}^2x^2 - \frac{1}{4}\delta^2\alpha^4k^2x^4$$

This is a multiple, σL, of the original Lagrangian if

$$\sigma = \gamma^2\frac{\alpha^4}{\beta^4} = \gamma\delta\frac{\alpha^4}{\beta^2} = \delta^2\alpha^4$$

The value of α is arbitrary, while the remaining constants must satisfy

$$\frac{\gamma^2}{\beta^4} = \frac{\gamma\delta}{\beta^2} = \delta^2$$

Both conditions are satisfied by the single condition,

$$\gamma = \delta\beta^2$$

Returning to the periodicity of the oscillator, we now know that if we change the mass by a factor γ and the spring constant k by a factor δ then the period changes by a factor $\beta = \sqrt{\frac{\gamma}{\delta}}$. Now suppose we start with a system with $m_0 = k_0 = 1$ and period T_0. Then with

$$\begin{aligned} m &= \gamma m_0 = \gamma \\ k &= \delta k_0 = \delta \end{aligned}$$

the period is

$$\begin{aligned} T &= \beta T_0 \\ &= \sqrt{\frac{\gamma}{\delta}}T_0 \\ &= \sqrt{\frac{m}{k}}T_0 \end{aligned}$$

We therefore learn that the frequency is proportional to $\sqrt{\frac{k}{m}}$ without solving for the motion.

6.3 Consequences of Newtonian dynamical and measurement theories

One of our goals is to develop a systematic approach to finding dynamical laws and measurement theories. This will require us to examine some mathematical techniques, including functional analysis, group theory and gauge theory. Nonetheless, some features of our ultimate methods may be employed immediately, with a more sophisticated treatmenr to follow. With this in mind, we now turn to a development of Newton's law from certain prior ideas.

Our starting point is geometry. Over two thousand years ago, Aristotle asked whether the space occupied by an object follows the object or remains where it was after the object moves away. This is the conceptual beginning of abstract space, independent of the objects in it. The idea is clearly an abstraction, and physicists have returned again and again to the inescapable fact that we only know space through the relationships between objects. Still, the idea of a continuum in which objects move may be made rigorous by considering the full set of possible positions of objects. We will reconsider the idea in light of some more contemporary philosophy.

Before beginning our agruments concerning spacde, we define another abstraction: the particle. By a particle, we mean an object sufficiently small and uncomplicated that its behavior may be accurately captured by specifying its position only. This is our physical model for a mathematical point. Naturally, the smallness of size required depends on the fineness of the description. For macroscopic purposes a small, smooth marble may serve as a model particle, but for the description of atoms it becomes questionable whether such a model even exists. For the present, we assume the existence of effectively point particles, and proceed to examine space. It is possible (and desirable if we take seriously the arguements of such philosophers as Popper and Berkeley[1]), to begin with our immediate experience.

Curiously enough, the most directly accessible geometric feature of the world is time. Our experience is a near-continuum of events. This is an immediate consequence of the richness of our experience. In fact, we might define *any* continuous or nearly continuous element of our experience – a succession of colors or a smooth variation of tones – as a direction for time. The fact that we do not rely on any one particular experience for this is probably because we choose to label time in a way that makes sense of the most possible experiences we can. This leads us to rely on correlations between many different experiences, optimizing over apparently causal relationships to identify time. Henceforward, we assume that our experience unfolds in an ordered sequence.

A simple experiment can convince us that a 3-dim model is *convenient* for describing that experience. First, I note that my sense of touch allows me trace a line down my arm with my finger. This establishes the existence of a single continuum of points which I can distinguish by placing them in 1-1 correspondence with successive times. Further, I can lay my hand flat on my arm, experiencing an entire 2-dim region of my skin. Finally, still holding my hand against my arm, I cup it so that the planar surface of my arm and the planar surface of my hand are not in contact, although they still maintain a continuous border. This establishes the usefulness of a third dimension.

A second, similar experiment makes use of vision. Reflecting the 2-dim nature of our retina, visual images appear planar – I can draw lines between pairs of objects in such a way that the lines intersect in a single intermediate point. This cannot be done in one dimension. Furthermore, I am not presented with a single image, but perceive a succession in time. As time progresses, I see some images pass out of view as they approach others, then reemerge later. Such an occultation is easily explained by a third dimension. The vanishing object has passed on the far side of the second object. In this way we rationalize the difference between our (at least) 2-dim visual experience and our (at least) 3-dim tactile experience.

What idea of spatial relation can we gain from the senses of smell and taste?

What is the dimension of the world of touch?

As we all know, these three spatial dimensions together with time provide a useful model for the physical world. Still, a simple mathematical proof will demonstrate the arbitrariness of this choice. Suppose we have a predictive physical model in 3-dim that adequately accounts for various phenomena. Then there exists a completely equivalent predictive model in any dimension. The proof follows from the proof that there exist 1-1 onto maps between dimensions, which we present first.

For simplicity we focus on a unit cube. For any point in the three dimensional unit cube, let the decimal expansions for the Cartesian coordinates be

$$\begin{aligned} x &= .a_1a_2a_3\ldots \\ x &= .b_1b_2b_3\ldots \\ z &= .c_1c_2c_3\ldots \end{aligned}$$

[1] Brief statement about Popper and Berkeley.

We map this point into a 1-dim continuum, w, by setting

$$w = .a_1b_1c_1a_2b_2c_2a_3b_3c_3\ldots$$

This mapping is clearly 1-1 and onto. To map to a higher dimension, we take any given w,

$$w = .d_1d_2d_3d_4\ldots$$

and partition the decimal expansion,

$$\begin{aligned} x_1 &= .d_1d_{n+1}d_{2n+1}\ldots \\ x_2 &= .d_2d_{n+2}d_{2n+2}\ldots \\ x_3 &= .d_3d_{n+3}d_{2n+3}\ldots \\ &\vdots \\ x_n &= .d_nd_{2n}d_{3n}\ldots \end{aligned}$$

Now suppose we have a physical model making, say, a prediction of the position of a particle 3-dim as a function of time,

$$x^i(t)$$

Applying the mapping gives a 1-dim sequence,

$$w(t)$$

containing all the same information.

Thus, any argument for a three dimensional model must be a pragmatic one. In fact, even though this mapping is never continuous, there might exist models in other dimensions that display useful properties more effectively than our usual 3-dim models.

Here is an example that shows how descriptions in different dimensions can reveal different physical information about a body. Consider the description of an extended body. In a three dimensional representation, we might specify a 3 parameter family of positions, $x^i(\alpha,\beta,\gamma)$ together with suitable ranges for the parameters. Alternative, we may represent this as a single number as follows. Divide a region of 3 -space into 1 meter cubes; divide each cube into 1000 smaller cubes, each one decimeter on a side, and so on. Number the 1 meter cubes from 0 to 999; number the decimeter cubes within each meter cube from 0 to 999, and so on. Then a specific location in space may be expressed as a sequence of numbers between 000 and 999,

$$(999, 345, 801, \ldots)$$

which we may concatenate to give

$$w = .999345801\ldots$$

This is clearly a 1-1, onto map. Now, for an extended body, choose a point in the body. About this point there will be a smallest cube contained entirely within the body. The specification of this cube is a *finite* decimal expansion,

$$w = .999345801\ldots 274$$

Additional cubes which together fill the body may be specified. Disuss the optimization of this list of numbers, and argue that an examination of a suitably defined list can quickly give information about the total size and shape of the body.

Devise a scheme for mapping arbitrary points in R^3 to a single real number. Hint: The essential problem here is that the decimal expansion may be arbitrarily long in both directions:

$$x = a_1a_2\ldots a_n.b_1b_2b_3\ldots$$

Try starting at the decimal point.

Devise a 1-1, onto mapping from the 3-dim position of a particle to a 1-dim representation in such a way that the number $w(t)$ is always within 10^{-n} of the x component of $(x(t), y(t), z(t))$.

6.4 Conserved quantities in generalized Euler-Lagrange systems

Most of the results of the preceeding sections have forms that hold for generalized Euler-Lagrange systems. Recall that if $L = L\left(x, \dot{x}, \ddot{x}, \ldots, x^{(n)}, t\right)$ the resulting variation leads to

$$\sum_{k=0}^{n} (-)^k \frac{d^k}{dt^k} \frac{\partial L}{\partial x^{(k)}} = 0$$

This generalized Euler-Lagrange equation is generically of order $2n$.

6.4.1 Conserved momenta

Suppose the action constructed from L is translationally invariant. The infinitesimal variation which leaves S invariant is again the constant vector

$$\delta x^i = \varepsilon^i$$

so the variation gives

$$\begin{aligned} 0 &= \delta S \\ &= \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x^i} \delta x^i + \frac{\partial L}{\partial \dot{x}^i} \delta \dot{x}^i + \cdots + \frac{\partial L}{\partial x^i_{(n)}} \delta x^i_{(n)} \right) dt \\ &= \sum_{k=0}^{n} \int_{t_1}^{t_2} \frac{\partial L}{\partial x^i_{(k)}} \delta x^i_{(k)} dt \end{aligned}$$

Integrating the k^{th} term by parts and noting that $\delta x^i_{(k)} = 0$ unless $k = 0$, gives

$$\begin{aligned} \int_{t_1}^{t_2} \frac{\partial L}{\partial x^i_{(k)}} \delta x^i_{(k)} dt &= \frac{\partial L}{\partial x^i_{(k)}} \delta x^i_{(k-1)} \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{d}{dt} \frac{\partial L}{\partial x^i_{(k)}} \delta x^i_{(k-1)} dt \\ &= -\int_{t_1}^{t_2} \frac{d}{dt} \frac{\partial L}{\partial x^i_{(k)}} \delta x^i_{(k-1)} dt \\ &= (-1)^2 \int_{t_1}^{t_2} \frac{d^2}{dt^2} \left(\frac{\partial L}{\partial x^i_{(k)}} \right) \delta x^i_{(k-2)} dt \\ &= (-1)^{k-1} \frac{d^{k-1}}{dt^{k-1}} \left(\frac{\partial L}{\partial x^i_{(k)}} \right) \delta x^i \Big|_{t_1}^{t_2} \\ &\quad + (-1)^k \int_{t_1}^{t_2} \frac{d^k}{dt^k} \left(\frac{\partial L}{\partial x^i_{(k)}} \right) \delta x^i dt \end{aligned}$$

so the variation gives

$$\begin{aligned} 0 &= \delta S \\ &= \sum_{k=1}^{n} (-1)^{k-1} \frac{d^{k-1}}{dt^{k-1}} \left(\frac{\partial L}{\partial x^i_{(k)}} \right) \Big|_{t_1}^{t_2} \varepsilon^i + \sum_{k=0}^{n} \int_{t_1}^{t_2} (-1)^k \frac{d^k}{dt^k} \left(\frac{\partial L}{\partial x^i_{(k)}} \right) \delta x^i dt \end{aligned}$$

The coefficient of δx^i in the integrand of the final integral is the Euler-Lagrange equation, so when the Euler-Lagrange equation is satisfied we must have

$$p_i = \sum_{k=1}^{n} (-1)^{k-1} \frac{d^{k-1}}{dt^{k-1}} \left(\frac{\partial L}{\partial x^i_{(k)}} \right) \tag{28}$$

for all δx. This p_i is the generalized conjugate momentum to x^i. Notice that when $n = 1$, p_i reduces to the previous expression. When any of the coordinates x^i is cyclic, the corresponding momentum is conserved.

Prove this claim directly from the generalized Euler-Lagrange equation. Suppose that some coordinate q is cyclic, so that $\frac{\partial L}{\partial q} = 0$. Show that

$$p_i = \sum_{k=1}^{n} (-1)^{k-1} \frac{d^{k-1}}{dt^{k-1}} \left(\frac{\partial L}{\partial q_{(k)}} \right)$$

is conserved.

Because p_i depends on time derivatives, there may be further conserved quantities. From Euler-Lagrange equation

$$\sum_{k=0}^{n} (-)^k \frac{d^k}{dt^k} \frac{\partial L}{\partial x^{(k)}} = 0$$

is is immediate that if the first m partial derivatives of L with respect to $x, x_{(1)}, \ldots, x_{(m-1)}$ vanish,

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial x^{(1)}} = \cdots = \frac{\partial L}{\partial x^{(m-1)}} = 0$$

then the Euler-Lagrange equation reduces to

$$\sum_{k=m}^{n} (-)^k \frac{d^k}{dt^k} \frac{\partial L}{\partial x^{(k)}} = 0$$

Rewriting the sum we have

$$\begin{aligned} 0 &= \frac{d^m}{dt^m} \sum_{k=m}^{n} (-)^k \frac{d^{k-m}}{dt^{k-m}} \frac{\partial L}{\partial x^{(k)}} \\ &= \frac{d^m}{dt^m} \sum_{k=0}^{n-m} (-)^{k+m} \frac{d^k}{dt^k} \frac{\partial L}{\partial x^{(k+m)}} \end{aligned}$$

showing that the first m time derivatives of

$$f(t) = \sum_{k=0}^{n-m} (-)^{k+m} \frac{d^k}{dt^k} \frac{\partial L}{\partial x^{(k+m)}}$$

vanish. Thus, $\frac{d^{m-1} f}{dt^{m-1}}$ is conserved. Moreover, we may immediately integrate m times, introducing $m - 1$ additional constants,

$$f(t) = \sum_{k=0}^{m-1} \frac{1}{k!} p_k t^k$$

The constants p_k are all conserved quantities.

6.4.2 Angular momentum

Suppose a general action is invariant under infinitesimal (and hence finite) rotations,

$$\begin{aligned} \delta x^i &= A^i{}_j x^j \\ A^{ij} &= -A^{ji} \end{aligned}$$

Notice that higher order variations no longer vanish. Instead, we have

$$\delta x^i_{(k)} = A^i{}_j x^j_{(k)}$$

Then applying this limited variation,

$$\begin{aligned} 0 &= \delta S \\ &= \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x^i}\delta x^i + \frac{\partial L}{\partial \dot{x}^i}\delta \dot{x}^i + \cdots + \frac{\partial L}{\partial x^i_{(n)}}\delta x^i_{(n)} \right) dt \\ &= \sum_{k=0}^{n} \int_{t_1}^{t_2} \frac{\partial L}{\partial x^i_{(k)}}\delta x^i_{(k)} dt \end{aligned}$$

This time we must keep all of the surface terms. For the k^{th} term,

$$\begin{aligned} \int_{t_1}^{t_2} \frac{\partial L}{\partial x^i_{(k)}}\delta x^i_{(k)} dt &= \frac{\partial L}{\partial x^i_{(k)}}\delta x^i_{(k-1)}\bigg|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{d}{dt}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i_{(k-1)} dt \\ &= \frac{\partial L}{\partial x^i_{(k)}}\delta x^i_{(k-1)}\bigg|_{t_1}^{t_2} - \frac{d}{dt}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i_{(k-2)}\bigg|_{t_1}^{t_2} \\ &\quad + \int_{t_1}^{t_2} \frac{d^2}{dt^2}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i_{(k-2)} dt \\ &\quad \vdots \\ &= \sum_{m=1}^{l} (-1)^{m-1} \frac{d^{m-1}}{dt^{m-1}}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i_{(k-m)}\bigg|_{t_1}^{t_2} \\ &\quad + (-1)^m \int_{t_1}^{t_2} \frac{d^m}{dt^m}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i_{(k-m)} dt \\ &\quad \vdots \\ &= \sum_{m=1}^{k} (-1)^{m-1} \frac{d^{m-1}}{dt^{m-1}}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i_{(k-m)}\bigg|_{t_1}^{t_2} \\ &\quad + (-1)^k \int_{t_1}^{t_2} \frac{d^k}{dt^k}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i dt \end{aligned}$$

Summing over k the variation becomes

$$\begin{aligned} 0 &= \delta S \\ &= \sum_{k=1}^{n}\sum_{m=1}^{k} (-1)^{m-1} \frac{d^{m-1}}{dt^{m-1}}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i_{(k-m)}\bigg|_{t_1}^{t_2} \\ &\quad + \sum_{k=0}^{n} (-1)^k \int_{t_1}^{t_2} \frac{d^k}{dt^k}\left(\frac{\partial L}{\partial x^i_{(k)}}\right)\delta x^i dt \end{aligned}$$

As usual, the final integral vanishes when we apply the equation of motion, so the quantity

$$\begin{aligned} M &= \sum_{k=1}^{n}\sum_{m=1}^{k} (-1)^{m-1} \frac{d^{m-1}}{dt^{m-1}}\left(\frac{\partial L}{\partial x^i_{(k)}}\right) A^i{}_j x^j_{(k-m)} \\ &= A^{ij} \sum_{k=1}^{n}\sum_{m=1}^{k} (-1)^{m-1} \frac{d^{m-1}}{dt^{m-1}}\left(\frac{\partial L}{\partial x^i_{(k)}}\right) x_{j(k-m)} \end{aligned}$$

is constant in time for all antisymmetric A^{ij}. Therefore, we may write the conserved angular momentum vector as

$$M^s = \sum_{k=1}^{n} \sum_{m=1}^{k} (-1)^m \, \varepsilon^{ijs} x_{i(k-m)} \frac{d^{m-1}}{dt^{m-1}} \left(\frac{\partial L}{\partial x^j_{(k)}} \right)$$

Notice that the relationship to the momentum conjugate to x^i is not simple, since the conjugate momentum

$$p_i = \sum_{k=1}^{n} (-1)^{k-1} \, \frac{d^{k-1}}{dt^{k-1}} \left(\frac{\partial L}{\partial x^i_{(k)}} \right)$$

contains only the $k-1$ derivative of $\frac{\partial L}{\partial x^i_{(k)}}$, whereas M_{ij} depends on all derivatives $\frac{d^{m-1}}{dt^{m-1}} \left(\frac{\partial L}{\partial x^j_{(k)}} \right)$ up to and including this one.

6.4.3 Energy

Finally we consider energy. Suppose L is independent of time. Then

$$\frac{dL}{dt} = \sum_{k=0}^{n} x^{(k+1)} \frac{\partial L}{\partial x^{(k)}}$$

But

$$\begin{aligned} x^{(k+1)} \frac{\partial L}{\partial x^{(k)}} &= \frac{d}{dt} \left(x^{(k)} \frac{\partial L}{\partial x^{(k)}} \right) - x^{(k)} \frac{d}{dt} \frac{\partial L}{\partial x^{(k)}} \\ &= \frac{d}{dt} \left(x^{(k)} \frac{\partial L}{\partial x^{(k)}} \right) - \frac{d}{dt} \left(x^{(k-1)} \frac{d}{dt} \frac{\partial L}{\partial x^{(k)}} \right) \\ &\quad + x^{(k-1)} \frac{d^2}{dt^2} \frac{\partial L}{\partial x^{(k)}} \\ &\quad \vdots \\ &= \sum_{m=0}^{k-1} (-)^m \, \frac{d}{dt} \left(x^{(k-m)} \frac{d^m}{dt^m} \frac{\partial L}{\partial x^{(k)}} \right) - x^{(1)} \, (-)^{k-1} \, \frac{d^k}{dt^k} \frac{\partial L}{\partial x^{(k)}} \end{aligned}$$

so

$$\frac{dL}{dt} = \frac{d}{dt} \sum_{k=0}^{n} \sum_{m=0}^{k-1} (-)^m \left(x^{(k-m)} \frac{d^m}{dt^m} \frac{\partial L}{\partial x^{(k)}} \right) + x^{(1)} \sum_{k=0}^{n} (-)^k \, \frac{d^k}{dt^k} \frac{\partial L}{\partial x^{(k)}}$$

Using the equation of motion,

$$\sum_{k=0}^{n} (-)^k \, \frac{d^k}{dt^k} \frac{\partial L}{\partial x^{(k)}} = 0$$

the final sum vanishes and we have the conserved energy

$$E = \sum_{k=0}^{n} \sum_{m=0}^{k-1} (-)^m \left(x^{(k-m)} \frac{d^m}{dt^m} \frac{\partial L}{\partial x^{(k)}} \right) - L$$

The $n = 3$ case of this result is given in [43] and elsewhere.

6.4.4 Scale invariance

Rescaling of the variables and constants works the same way for higher order Lagrangians as it does in the restricted case.

6.5 Exercises

Find the Euler-Lagrange equation for the following action functionals:

1. $S[x] = \int \exp\left(\alpha \mathbf{x}^2 + \beta \mathbf{v}^2\right) dt$ for constants α and β.
2. $S[x] = \int f\left(\mathbf{x}^2 \mathbf{v}^2\right) dt$ for any given function, f.
3. $S[x] = \frac{1}{\int \mathbf{x}\cdot\mathbf{a} dt} + \int \mathbf{x}\cdot\mathbf{a} dt$, where $\mathbf{a} = \ddot{\mathbf{x}}$.

Apply the techniques for generalized Euler-Lagrange systems to the following fourth-order action:

$$\begin{aligned} S &= \int L dt \\ &= \int \left(\frac{1}{2} k m \dot{x}^2 x^2 - \frac{1}{4} k^2 x^4 + \frac{1}{4} m^2 x \dot{x}^2 \ddot{x} + \frac{1}{4} m^2 x^2 \ddot{x}^2 + \frac{1}{4} m^2 x^2 \dot{x} x^{(3)}\right) dt \end{aligned}$$

Find the equation of motion and the conserved energy.

Consider the 3-dimensional action

$$S = \int \left(\frac{1}{2} m \dot{\mathbf{x}}^2 - mgz\right) dt$$

where $\mathbf{x} = (x, y, z)$.

1. Show that there are four symmetries of S.
2. Find the four conserved quantities.

Consider the 2-dimensional action functional

$$S = \frac{1}{2} \int \left(m l^2 \dot{\varphi}^2 + mgl\varphi^2\right) dt$$

Find all rescalings of the parameters and coordinates (m, g, l, φ, t) which leave S changed by no more than an overall constant. Use these rescalings to show that the period of the motion is proportional to $\sqrt{\frac{l}{g}}$.

The action functional

$$S = \int_{t_1}^{t_2} \left(\frac{1}{2} m \dot{\mathbf{x}}^2 - \frac{K}{\sqrt{\mathbf{x}^2}}\right) dt$$

Use a scaling argument to derive Kepler's law relating the period and a characteristic length of the orbit.

7 The physical Lagrangian

We have seen that the extrema of functionals give preferred paths of motion among the class of curves. These extremal paths appear to be smooth curves. Indeed, when the action is taken to be path length, the extremal curves are straight lines. Now we return to the problem of physical motion stated in Chapter III: can we specify a functional for specific physical problems in such a way that the extremals are the physical paths of motion?

We have seen examples where this is the case. For example, the action functional

$$L = \frac{1}{12}m^2\dot{x}^4 + \frac{1}{2}km\dot{x}^2x^2 - \frac{1}{4}k^2x^4 \tag{29}$$

describes the simple harmonic oscillator in 1-dimension. Now we take a more systematic approach, based on the principles of *gauge theory*. Our gauging of Newton's second law begins with the observation that the symmetry which preserves Newton's law is different than the symmetry of the Euler-Lagrange equation. Specifically, Newton's second law is invariant under a special set of constant transformations called Galilean transformations. By contrast, the Euler-Lagrange equation is invariant under the diffeomorphism group, that is, the group of all coordinate transformations.

It follows that, if we are to develop a systematic way of writing physical equations of motion using the Euler-Lagrange equation, then it must involve a generalization of Newton's second law to arbitrary coordinate systems. This is exactly what gauging accomplishes – the systematic extension of a symmetry. There is nothing surprising in such a generalization. In fact, we would find it odd not to be able to work out a form of the second law valid in any coordinate system of our choosing. The key point here is to do it systematically, rather than just substituting a specific transformation of the coordinates.

There are several steps to our generalization. First, we derive the most general set of transformations under which Newton's law is covariant – the Galilean group. This group has important subgroups. In particular, we are interested in the subgroup that also preserves Newtonian measurement theory, the Euclidean group.

Next, we derive the geodesic equation. Geodesics are the straightest possible lines in arbitrary, even curved, spaces. In our Euclidean 3-space, the geodesic equation simply provides a diffeomorphism-invariant way of describing a straight line – the condition of vanishing acceleration. As a result, we will have expressed the acceleration in arbitrary coordinates.

Finally, we gauge Newton's law by writing it in terms of coordinate covariant expressions. We conclude by showing that this generalized form of Newton's law indeed follows as the extremum of a functional, and derive a general expression for the action.

7.1 Galilean symmetry and the invariance of Newton's Law

Newton's second law, written as

$$F^i = \frac{d}{dt}\left(mv^i\right)$$

is a relationship between vectors. Therefore, we write the law in terms of the velocity because the coordinates, x^i, are not vectors, but the velocity – being the tangent vector to a curve – is. We will assume that the force is a vector (see the section on tensors), so that it changes according to:

$$F^i(\mathbf{x}) = \frac{\partial x^i}{\partial q^j}F^j(\mathbf{q}) \tag{30}$$

where $\frac{\partial x^i}{\partial q^j}$ is the Jacobian matrix of the coordinate transformation. Covariance of the second law requires this same linear, homogeneous transformation on each term,

$$\frac{\partial x^i}{\partial q^j}F^j(\mathbf{q}) = m\frac{\partial x^i}{\partial q^j}\ddot{q}^j \tag{31}$$

but this relationship cannot hold form arbitrary changes of coordinate,

$$\begin{aligned} q^i &= q^i(\mathbf{x}, t) \\ t' &= t'(t) \end{aligned} \tag{32}$$

To find what transformations are allowed, we apply the general coordinate change of eq.(32) to Newton's law, and examine the conditions required to make it transform according to eq.(31). Since the coordinate transformation of eq.(32) must be invertible, we may also write

$$\begin{aligned} x^i &= x^i(\mathbf{q}, t) \\ t &= t(t') \end{aligned}$$

We can immediately set

$$t' = t - t_0$$

since the Newtonian assumption of universal time requires equal intervals, $dt = dt'$.

The real limitation on covariance comes from the time derivatives in the acceleration term. Along any path specified by $x^i(t) = x^i\left(q^j(t), t\right)$ the acceleration $\ddot{x}^i$ may be written in terms of $\ddot{q}^j$ and $\dot{q}^k$ as

$$\begin{aligned} \ddot{x}^i &= \frac{d}{dt}\left(\frac{\partial x^i}{\partial q^j}\dot{q}^j + \frac{\partial x^i}{\partial t}\right) \\ &= \left(\frac{\partial x^i}{\partial q^j}\ddot{q}^j + \frac{\partial^2 x^i}{\partial q^k \partial q^j}\dot{q}^k\dot{q}^j + \frac{\partial^2 x^i}{\partial \dot{q}^k \partial t}\dot{q}^k\right) + \left(\frac{\partial^2 x^i}{\partial t \partial q^j}\dot{q}^j + \frac{\partial^2 x^i}{\partial t^2}\right) \\ &= \frac{\partial x^i}{\partial q^j}\ddot{q}^j + \frac{\partial^2 x^i}{\partial q^k \partial q^j}\dot{q}^k\dot{q}^j + 2\frac{\partial^2 x^i}{\partial \dot{q}^k \partial t}\dot{q}^k + \frac{\partial^2 x^i}{\partial t^2} \end{aligned}$$

The first term is proportional to the second time derivative of q^i, but the remaining terms are not. Comparing to the actual transformation of the acceleration, covariance therefore requires

$$\begin{aligned} \frac{\partial x^i}{\partial q^j}\ddot{q}^j &= \frac{\partial x^i}{\partial q^j}\ddot{q}^j + \frac{\partial^2 x^i}{\partial q^k \partial q^j}\dot{q}^k\dot{q}^j + 2\frac{\partial^2 x^i}{\partial \dot{q}^k \partial t}\dot{q}^k + \frac{\partial^2 x^i}{\partial t^2} \\ 0 &= \frac{\partial^2 x^i}{\partial q^k \partial q^j}\dot{q}^k\dot{q}^j + 2\frac{\partial^2 x^i}{\partial \dot{q}^k \partial t}\dot{q}^k + \frac{\partial^2 x^i}{\partial t^2} \end{aligned}$$

Since this must hold for all velocities $\dot{q}^j$, the coefficient of each order in velocity must vanish. Thus,

$$\begin{aligned} 0 &= \frac{\partial^2 x^i}{\partial t^2} \\ 0 &= \frac{\partial^2 x^i}{\partial \dot{q}^k \partial t} \\ 0 &= \frac{\partial^2 x^i}{\partial q^k \partial q^j} \end{aligned}$$

Integrating the first equation shows that x^i must be linear in the time

$$x^i = x_0^i + \dot{x}_0^i t$$

The second equation then requires $\dot{x}_0^i = v^i = \text{constant}$. Finally, we see that x_0^i must be linear in q^j :

$$\begin{aligned} \frac{\partial^2 x^i}{\partial q^k \partial q^j} &= 0 \\ \frac{\partial x^i}{\partial q^j} &= M^i{}_j \\ x^i &= M^i{}_j q^j + a^i \end{aligned}$$

where $M^i{}_j$ and a^i are constant. Therefore, the most general coordinate transformation for which Newton's second law is covariant is

$$x^i(\mathbf{q},t) = M^i{}_j q^j + a^i + v^i t \tag{33}$$

and the velocity and acceleration transform as

$$\begin{aligned} \dot{x}^i &= M^i{}_j \dot{q}^j + v^i \\ &= \frac{\partial x^i}{\partial q^j}\dot{q}^j + v^i \\ \ddot{x}^i &= \frac{\partial x^i}{\partial q^j}\ddot{q}^j \end{aligned}$$

Now consider each transformation. The arbitrary constant matrix $M^i{}_j$ need only be invertible, and is therefore an element of the general linear group, $GL(3)$. The constant a^i is an arbitrary translation, while the time-dependent translation $v^i t$ is called a *boost*. The full group of transformations of eq.(33), called the Galilean group, describes the set of *inertial frames*.

There are important subgroups of the Galilean group. First, we would like the velocity to transform as a vector. This requires $v^i = 0$. Furthermore, Newtonian measurement theory requires an inner product of vectors, so we may restrict to an orthonormal basis. The subgroup of transformations which preserves an orthonormal basis is the orthogonal group, with $M^i{}_j = R^i{}_j$ a rotation. The resulting coordinate transformations

$$x^i(\mathbf{q}) = R^i{}_j q^j + a^i \tag{34}$$

show that we have recovered the Euclidean symmetry of the background space.

We now derive a covariant expression for the acceleration.

7.2 Galileo, Lagrange and inertia

Recall the Galileo-Newton law of inertia:

- A body remains in uniform motion unless acted upon by an external force.

With this in mind, we define

Uniform motion is motion at a constant rate along extremals of length. The extremal paths are called *geodesics*.

We have seen that extremals of length in the Euclidean plane are straight lines, and this coincides with our notion of uniform, or unaccelerated motion,

$$\frac{d^2x^i}{dt^2} = 0$$

By using a variational principle to find the covariant description of a straight line, we will have found a covariant expression for the acceleration.

We proceed to find the general expression for geodesics. In arbitrary coordinates, y^i, the infinitesimal separation between two points is given by

$$ds = \sqrt{g_{ij}dy^i dy^j} \tag{35}$$

where the Euclidean metric g_{ij} is given by

$$g_{ij} = \delta_{mn}\frac{\partial x^m}{\partial y^i}\frac{\partial x^n}{\partial y^j} \tag{36}$$

and x^i are Cartesian coordinates.

If we paramaterize a curve by λ,

$$\phi\left(C\left(\lambda\right)\right) = y^i\left(\lambda\right)$$

then uniform motion is described by the extremals of length, where length is given by the functional

$$\begin{aligned} S\left[y^i\left(\lambda\right)\right] &= \int_0^1 \sqrt{g_{ij}\frac{dy^i}{d\lambda}\frac{dy^j}{d\lambda}}d\lambda \\ &= \int_0^1 \sqrt{g_{ij}\dot{y}^i\dot{y}^j}d\lambda \end{aligned}$$

where we set $\frac{dy^i}{d\lambda} = \dot{y}^n$. Varying the path, and remembering that g_{ij} is a function of the coordinates and using $\delta\dot{y}^n = \frac{d\delta y^i}{d\lambda}$, we have

$$\begin{aligned} 0 &= \delta S\left[y\right] \\ &= \int_0^1 \frac{1}{2\sqrt{g_{mn}\dot{y}^m\dot{y}^n}}\left(\frac{\partial g_{ij}}{\partial y^k}\delta y^k\dot{y}^i\dot{y}^j + g_{ij}\frac{d\delta y^i}{d\lambda}\dot{y}^j + g_{ij}\dot{y}^i\frac{d\delta y^j}{d\lambda}\right)d\lambda \\ &= \int_0^1 \frac{1}{2\sqrt{g_{mn}\dot{y}^m\dot{y}^n}}\left(\frac{\partial g_{ij}}{\partial y^k}\delta y^k\dot{y}^i\dot{y}^j\right)d\lambda \\ &\quad -\int_0^1 \frac{1}{2}\frac{d}{d\lambda}\left(g_{ij}\dot{y}^j\left(g_{mn}\dot{y}^m\dot{y}^n\right)^{-1/2} + g_{ij}\dot{y}^i\left(g_{mn}\dot{y}^m\dot{y}^n\right)^{-1/2}\right)\delta y^j d\lambda \end{aligned}$$

where we have set the surface terms to zero as usual. The geodesic equation is therefore

$$\begin{aligned} 0 &= \frac{1}{2\sqrt{g_{mn}\dot{y}^m\dot{y}^n}}\left(\frac{\partial g_{ij}}{\partial y^k}\dot{y}^i\dot{y}^j\right) \\ &\quad -\frac{1}{2}\frac{d}{d\lambda}\left(g_{kj}\dot{y}^j\left(g_{mn}\dot{y}^m\dot{y}^n\right)^{-1/2}\right) \\ &\quad -\frac{1}{2}\frac{d}{d\lambda}\left(g_{ik}\dot{y}^i\left(g_{mn}\dot{y}^m\dot{y}^n\right)^{-1/2}\right) \end{aligned}$$

A considerable simplification is achieved if we choose the parameter λ along curves to be path length, s itself. Then, from eq.(35), we have

$$g_{mn}\dot{y}^m\dot{y}^n = 1$$

and the geodesic equation reduces to

$$0 = \frac{1}{2}\left(\frac{\partial g_{ij}}{\partial y^k}\dot{y}^i\dot{y}^j - \frac{d}{ds}\left(g_{kj}\dot{y}^j\right) - \frac{d}{ds}\left(g_{ik}\dot{y}^i\right)\right)$$

Since

$$\frac{d}{ds}\left(g_{ik}\dot{y}^i\right) = \frac{\partial g_{ik}}{\partial y^m}\dot{y}^m\dot{y}^i + g_{ik}\ddot{y}^i$$

this becomes

$$\begin{aligned} 0 &= \frac{1}{2}\left(\frac{\partial g_{ij}}{\partial y^k}\dot{y}^i\dot{y}^j - \frac{\partial g_{kj}}{\partial y^m}\dot{y}^m\dot{y}^j - g_{kj}\ddot{y}^j\right) \\ &\quad -\frac{1}{2}\left(\frac{\partial g_{ik}}{\partial y^m}\dot{y}^m\dot{y}^i + g_{ik}\ddot{y}^i\right) \\ &= \frac{1}{2}\left(\frac{\partial g_{mn}}{\partial y^k} - \frac{\partial g_{kn}}{\partial y^m} - \frac{\partial g_{km}}{\partial y^n}\right)\dot{y}^m\dot{y}^n - g_{jk}\ddot{y}^j \end{aligned}$$

where we have used the symmetry, $g_{mn} = g_{nm}$, and the symmetry

$$\dot{y}^m\dot{y}^n = \dot{y}^n\dot{y}^m$$

in the last step. We can give this a more compact appearance with some new notation. First, let partial derivatives be denoted by

$$\partial_k = \frac{\partial}{\partial y^k}$$

Next, we define the convenient symbol

$$\Gamma_{kmn} = \frac{1}{2}\left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right)$$

The object, Γ_{kmn}, is called the Christoffel connection for reasons that will become clear later. Notice that Γ_{kmn} is symmetric in the last two indices, $\Gamma_{kmn} = \Gamma_{knm}$. Substituting, and using the inverse metric g^{ik}, we have

$$0 = g^{ik}\Gamma_{kmn}\frac{dy^m}{ds}\frac{dy^n}{ds} + g^{ik}g_{jk}\frac{d^2y^j}{ds^2}$$

or the final form of the geodesic equation,

$$\frac{d^2y^i}{ds^2} + \Gamma^i{}_{mn}\frac{dy^m}{ds}\frac{dy^n}{ds} = 0$$

This equation may be regarded as either a second order equation for a curve $y^i(s)$, or as a first order equation for the tangent vector $u^i = \frac{dy^i}{ds}$,

$$\frac{du^i}{ds} + \Gamma^i{}_{mn}u^m u^n = 0$$

For the vector equation, we may also write

$$u^m\left(\partial_m u^i + u^n\Gamma^i{}_{nm}\right) = 0$$

The term in parentheses will reappear when we discuss gauge theory in more detail.

We consider two examples.

First, if the coordinates are Cartesian, the metric is simply $g_{ij} = \delta_{ij}$. Then all derivatives of the metric vanish, $\partial_k\delta_{ij} = 0$, and the Christoffel symbols, Γ_{kmn} and $\Gamma^i{}_{mn}$ vanish. Therefore,

$$\begin{aligned}\frac{d^2x^i}{ds^2} &= 0\\ x^i &= x_0^i + v_0^i s\end{aligned}$$

and we recover the straight lines of Chapter 3.

For a non-trivial example, consider motion constrained to the surface of a sphere of radius a. Since the metric in spherical coordinates is

$$g_{ij} = \begin{pmatrix} 1 & & \\ & r^2 & \\ & & r^2\sin^2\theta \end{pmatrix}$$

the surface is described by holding $r = a$ and dropping dr in the line element. We therefore set

$$g_{ij} = \begin{pmatrix} a^2 & \\ & a^2\sin^2\theta \end{pmatrix}$$

and compute

$$\Gamma_{kmn} = \frac{1}{2}\left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right)$$

for each combination of indices. Since all components of g_{ij} are constant except $g_{\varphi\varphi}$, the only nonvanishing components are:

$$\Gamma_{\varphi\varphi\theta} = \Gamma_{\varphi\theta\varphi} = -\Gamma_{\theta\varphi\varphi} = a^2\sin\theta\cos\theta$$

Raising an index with the inverse metric

$$g^{ij} = \frac{1}{a^2}\begin{pmatrix} 1 & \\ & \frac{1}{\sin^2\theta} \end{pmatrix}$$

we find

$$\begin{aligned} \Gamma^{\varphi}{}_{\varphi\theta} &= \Gamma^{\varphi}{}_{\theta\varphi} = \frac{\cos\theta}{\sin\theta} \\ \Gamma^{\theta}{}_{\varphi\varphi} &= -\sin\theta\cos\theta \end{aligned}$$

and the equations for geodesics become

$$\begin{aligned} 0 &= \frac{d^2\theta}{ds^2} + \Gamma^{\theta}{}_{\varphi\varphi}\frac{d\varphi}{ds}\frac{d\varphi}{ds} \\ 0 &= \frac{d^2\varphi}{ds^2} + \Gamma^{\varphi}{}_{\theta\varphi}\frac{d\theta}{ds}\frac{d\varphi}{ds} + \Gamma^{\varphi}{}_{\varphi\theta}\frac{d\varphi}{ds}\frac{d\theta}{ds} \end{aligned}$$

and therefore,

$$\begin{aligned} \frac{d^2\theta}{ds^2} - \sin\theta\cos\theta\frac{d\varphi}{ds}\frac{d\varphi}{ds} &= 0 \\ \frac{d^2\varphi}{ds^2} + 2\frac{\cos\theta}{\sin\theta}\frac{d\varphi}{ds}\frac{d\theta}{ds} &= 0 \end{aligned} \tag{37}$$

We can see immediately that if the motion starts with $\frac{d\varphi}{ds} = 0$, then initially

$$\begin{aligned} \frac{d^2\theta}{ds^2} &= 0 \\ \frac{d^2\varphi}{ds^2} &= 0 \end{aligned}$$

The second of these shows that $\frac{d\varphi}{ds}$ remains zero, so this form of the equations continues to hold for all s. The solution is therefore

$$\begin{aligned} \theta &= a + bs \\ \varphi &= c \end{aligned}$$

Thus, φ is constant and θ increases linearly. This describes a great circle on the 2-sphere, that is, the intersection of the 2-sphere with a plane through the origin. Since any solution may be rotated to have these initial conditions, this is the general solution.

Using eq.(36) for the metric, find an expression for $\Gamma_{kmn} = \frac{1}{2}\left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right)$ in terms of first and second derivatives of x^i with respect to y^j.

Show that the geodesic equation describes a straight line by writing it in Cartesian coordinates.

Find the equation for a straight line in the plane using polar coordinates, using two methods. First, find the extremum of the integral of the line element

$$\begin{aligned} S\left[q^i(t)\right] &= \int ds \\ &= \int \sqrt{\dot\rho^2 + \rho^2\dot\varphi^2}\, dt \end{aligned}$$

Then, find the Christoffel symbols and write out the geodesic equation using the same metric, $g_{ij} = \begin{pmatrix} 1 & \\ & \rho^2 \end{pmatrix}$.

7.3 Gauging Newton's law

Newton's second law of motion now provides the starting point for seeking a functional whose extrema are physical paths. But in order to make the connection between

$$F^i = \frac{d}{dt}\left(mv^i\right)$$

and a variational equation, we must first generalize Newton's law to a new expression valid in arbitrary coordinates. We can now do this in two easy steps.

First, we already know how to write a covariant expression for a straight line, or geodesic

$$\begin{aligned} 0 &= \frac{du^i}{ds} + \Gamma^i{}_{mn} u^m u^n \\ u^m &= \frac{dx^m}{ds} \end{aligned}$$

This equation expresses zero change of direction in arbitrary coordinates. We only need to alter the parameter s to time. Recall our definition of uniform motion as motion at a constant rate along geodesics. This means that uniform motion is described by the geodesic equation together with

$$\frac{ds}{dt} = v_0 = const.$$

Therefore, to change the parameterization of the geodesic equation to time, we simply multiply by v_0^2. Then, since v_0 is constant,

$$\begin{aligned} 0 &= v_0 \frac{d}{ds} v_0 \frac{dx^i}{ds} + \Gamma^i{}_{mn} v_0 \frac{dx^m}{ds} v_0 \frac{dx^n}{ds} = 0 \\ 0 &= \frac{ds}{dt}\frac{d}{ds}\left(\frac{ds}{dt}\frac{dx^i}{ds}\right) + \Gamma^i{}_{mn}\left(\frac{ds}{dt}\frac{dx^m}{ds}\right)\left(\frac{ds}{dt}\frac{dx^n}{ds}\right) \end{aligned}$$

and therefore motion in a straight line at constant velocity may be written in any coordinate system as

$$\frac{d^2x^i}{dt^2} + \Gamma^i{}_{mn}\frac{dx^m}{dt}\frac{dx^n}{dt} = 0$$

Notice that in Cartesian coordinates, where $\Gamma^i{}_{mn} = 0$, this equation just expresses vanishing acceleration, $\frac{d^2x^i}{dt^2} = 0$. Since the equation is covariant, it must express vanishing acceleration in any coordinates and the acceleration is covariantly described by.

$$a^i \equiv \frac{d^2x^i}{dt^2} + \Gamma^i{}_{mn}\frac{dx^m}{dt}\frac{dx^n}{dt}$$

whether it vanishes or not.

To rewrite Newton's second law in arbitrary coordinates, we may simply multiply a^i by the mass and equate to the force:

$$F^i = m\frac{dv^i}{dt} + m\Gamma^i{}_{mn} v^m v^n$$

For the force, recall that it is often possible to write F^i as minus the gradient of a potential. In general coordinates, however, the gradient requires a metric:

$$[\nabla f]^i = g^{ij}\frac{\partial f}{\partial y^j}$$

Newton's law may therefore be written as

$$-g^{ij}\frac{\partial V}{\partial y^j} = m\frac{dv^i}{dt} + m\Gamma^i{}_{mn} v^m v^n \tag{38}$$

where $V(x)$ is the potential for a force and

$$\begin{aligned} g_{ij} &= \sum_i \delta_{mn} \frac{\partial x^m}{\partial y^i} \frac{\partial x^n}{\partial y^j} \\ g^{ij} &= \left[g^{-1}\right]_{ij} \\ \Gamma^i{}_{mn} &= \frac{1}{2} g^{ik} \left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right) \end{aligned}$$

Eq.(38) holds in any coordinate system if Newton's second law holds in Cartesian coordinates.

Now we manipulate eq.(38). Substituting the expression for the connection,

$$\begin{aligned} -g^{ij} \frac{\partial V}{\partial y^j} &= m \frac{dv^i}{dt} + \frac{m}{2} g^{ik} \left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right) v^m v^n \\ -\frac{\partial V}{\partial y^k} &= m g_{ki} \frac{dv^i}{dt} + \frac{m}{2} \left(\left(v^m \partial_m g_{kn}\right) v^n + \left(v^n \partial_n g_{km}\right) v^m - \partial_k g_{mn} v^m v^n\right) \\ -\frac{\partial V}{\partial y^k} &= m g_{ki} \frac{dv^i}{dt} + \frac{m}{2} \left(\frac{dg_{kn}}{dt} v^n + \frac{dg_{km}}{dt} v^m - \left(\partial_k g_{mn}\right) v^m v^n\right) \end{aligned}$$

where we have used the chain rule to write $v^m \partial_m g_{kn} = \frac{dg_{kn}}{dt}$. Collecting the gradient terms on the right,

$$\begin{aligned} 0 &= m g_{ki} \frac{dv^i}{dt} + m \frac{dg_{kn}}{dt} v^n - \frac{\partial}{\partial y^k} \left(\frac{m}{2} g_{mn} v^m v^n - V\right) \\ &= \frac{d}{dt} \left(m g_{kn} v^n\right) - \frac{\partial}{\partial y^k} \left(\frac{m}{2} g_{mn} v^m v^n - V\right) \end{aligned}$$

Now, observing that

$$\frac{\partial}{\partial v^k} \left(\frac{m}{2} g_{mn} v^n v^m\right) = m g_{kn} v^n$$

we substitute to get

$$0 = \frac{d}{dt} \left(\frac{\partial}{\partial v^k} \left(\frac{m}{2} g_{mn} v^n v^m\right)\right) - \frac{\partial}{\partial y^k} \left(\frac{m}{2} g_{mn} v^m v^n - V\right)$$

Since V depends only on position and not velocity we can put it into the first term as well as the second,

$$0 = \frac{d}{dt} \left(\frac{\partial}{\partial v^k} \left(\frac{m}{2} g_{mn} v^n v^m - V\right)\right) - \frac{\partial}{\partial y^k} \left(\frac{m}{2} g_{mn} v^m v^n - V\right)$$

Finally, we recognize the Euler-Lagrange equation,

$$0 = \frac{d}{dt} \left(\frac{\partial L}{\partial v^k}\right) - \frac{\partial L}{\partial y^k}$$

with the Lagrangian

$$L = \frac{1}{2} m g_{mn} v^n v^m - V$$

We identify the first term on the right as the kinetic energy,

$$T = \frac{1}{2} m g_{mn} v^n v^m$$

since it reduces to $\frac{1}{2} m \mathbf{v}^2$ in Cartesian coordinates.

We have successfully gauged Newton's second law, and shown that the new diffeomorphism invariant version is given by extrema of the action functional,

$$S\left[x^i(t)\right] = \int (T - V)\, dt$$

There are distinct advantages to this formulation of classical mechanics. First, as we have seen, Noether's theorem lets us take advantage of symmetries of the system in a direct way. Second, the diffeomorphism invariance allows us to write the problem in the fewest number of independent variables easily. Finally, starting with an action permits easy inclusion of constraints.

As an example, we consider the motion in 3-dim of two masses, both of mass m, connected by a light, rigid rod of length L. If we describe the system in Cartesian coordinates, each mass requires 3 coordinates, for a total of 6 coordinates. However, using symmetries this is considerably reduced. First, since the system is isolated, angular momentum must be conserved. Consequently, the motion lies in the plane orthogonal to the angular momentum vector. Let the origin, at position R^i, coincide with the midpoint of the rod. Picking x and y axes in the plane of motion, let the rod make an angle φ with the x-axis. Then the positions of the two masses are

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{R} + \frac{L}{2}\left(\mathbf{i}\cos\varphi + \mathbf{j}\sin\varphi\right) \\ \mathbf{x}_2 &= \mathbf{R} - \frac{L}{2}\left(\mathbf{i}\cos\varphi + \mathbf{j}\sin\varphi\right) \end{aligned}$$

and the corresponding velocites are

$$\begin{aligned} \dot{\mathbf{x}}_1 &= \dot{\mathbf{R}} + \frac{L\dot{\varphi}}{2}\left(-\mathbf{i}\sin\varphi + \mathbf{j}\cos\varphi\right) \\ \dot{\mathbf{x}}_2 &= \dot{\mathbf{R}} - \frac{L\dot{\varphi}}{2}\left(-\mathbf{i}\sin\varphi + \mathbf{j}\cos\varphi\right) \end{aligned}$$

The kinetic energy is therefore

$$\begin{aligned} T &= \frac{1}{2}m\dot{\mathbf{x}}_1\cdot\dot{\mathbf{x}}_1 + \frac{1}{2}m\dot{\mathbf{x}}_2\cdot\dot{\mathbf{x}}_2 \\ &= \frac{1}{2}m\left(\dot{\mathbf{R}} + \frac{L\dot{\varphi}}{2}\left(-\mathbf{i}\sin\varphi + \mathbf{j}\cos\varphi\right)\right)^2 \\ &\quad + \frac{1}{2}m\left(\dot{\mathbf{R}} + \frac{L\dot{\varphi}}{2}\left(-\mathbf{i}\sin\varphi + \mathbf{j}\cos\varphi\right)\right)^2 \\ &= \frac{1}{2}m\left(2\dot{R}^2 + \frac{L^2\dot{\varphi}^2}{2}\right)^2 \end{aligned}$$

Since there is no potential energy, the action is simply

$$S = \int m\left(\dot{R}^2 + \frac{L^2\dot{\varphi}^2}{4}\right)dt$$

Since both $\mathbf{R}$ and φ are cyclic we immediately have two conservation laws,

$$\begin{aligned} P_i &= \frac{\partial L}{\partial \dot{R}^i} = 2m\dot{R}_i = const. \\ J &= \frac{1}{2}mL^2\dot{\varphi} = const. \end{aligned}$$

Integrating, we have the complete solution,

$$\begin{aligned} R^i &= R^i_0 + P^i t \\ \varphi &= \varphi_0 + \frac{2J}{mL^2}t \end{aligned}$$

Write the action for the Kepler problem. The Kepler problem describes motion in the gravitational potential $V = -\frac{GM}{r}$. To formulate the problem in spherical coordinates we first write the kinetic energy.

The easiest way to find v^2 is to divide the squared infinitesimal line element by dt^2:

$$\begin{aligned} ds^2 &= dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2 \\ v^2 &= \left(\frac{ds}{dt}\right)^2 = \left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2 + r^2\sin^2\theta\left(\frac{d\varphi}{dt}\right)^2 \end{aligned}$$

The kinetic energy is therefore just

$$\begin{aligned} T &= \frac{1}{2}mv^2 \\ &= \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\dot{\varphi}^2\sin^2\theta\right) \end{aligned}$$

and the Lagrangian is

$$L = T - V = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\dot{\varphi}^2\sin^2\theta\right) + \frac{\alpha}{r}$$

Thus, the Kepler action is

$$S = \int\left(\frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\dot{\varphi}^2\sin^2\theta\right) + \frac{\alpha}{r}\right)dt$$

Suppose the Lagrangian for a physical problem, $L = T - V$ has no explicit time dependence. Then the energy

$$E = \dot{x}^i\frac{\partial L}{\partial\dot{x}^i} - L$$

is conserved. Prove that $E = T + V$.

We showed that the action of eq.(29) describes the simple harmonic oscillator, but according to our new physical correspondence, extrema of the simpler action functional

$$\begin{aligned} S &= \int (T - V)\,dt \\ &= \frac{1}{2}\int\left(m\dot{x}^2 - kx^2\right)dt \end{aligned}$$

should describe this same motion. Show that vanishing variation of this simpler action does give the correct equation of motion.

For each of the following, give the number of Cartesian coordinates required to describe the system, and give the actual number of degrees of freedom of the problem. Then write the Lagrangian, $L = T - V$, and find the equations of motion for each of the following physical systems:

1. A particle moving in 3-dim under the influence of a gravitational force $-mg\mathbf{k}$.
2. A pendulum of length l and mass m is suspended from a second pendulum of length L and mass M. Both are constrained to move in the xz plane.
3. A ball moves frictionlessly on a horizontal tabletop. The ball of mass m is connected to a string of length L which passes through a hole in the tabletop and is fastened to a pendulum of mass M. The string is free to slide through the hole in either direction.
4. The isotropic oscillator is a particle moving in the spherically symmetric potential $V = \frac{1}{2}kr^2$.

Use scaling arguments to show how the frequency of small oscillations depends on the amplitude for any potential of the form

$$V = ax^n$$

In the next fews sections we study various applications and properties of the new techniques we have developed. One the most important applications of classical mechanics is the problem of central forces, and we begin with this example.

8 Motion in central forces

While the study of gravity has gone far beyond classical potentials, our approach models the treatment of any physical theory in that we study an entire class of theories containing the one we think the best. This gives a theory-independent way to design experiments.

For example, by studying arbitrary power-law potentials we learn that most cannot produce closed orbits, and this might provide a sensitive test of the theory. Or, by examining the dependence of period on eccentricity of orbits in power-law potentials, we gain another test. In this way, designing experiments that explore an entire class of possible models, we both identify new tests and quantify our confidence in a $\frac{1}{r^2}$-force law.

Consider any potential which depends only on the radial distance of a particle from a center of force,

$$V = V(r)$$

This class of potentials includes important subclasses. In order of increasing specificity, we will study

1. Monotonically increasing potentials, $V' > 0$.
2. Power law potentials, $V = ar^n$.
3. Potentials with perturbatively closed orbits
4. Bertrand's theorem: potentials with non-perturbatively closed orbits
   (a) Kepler/Coulomb potential, $V = -\frac{\kappa}{r}$
   (b) Isotropic oscillator, $V = ar^2$
5. Newtonian gravity

The two potentials described by Bertrand's theorem – for Newtonian gravity and the isotropic oscillator – are extremely important problems. The Kepler/Coulomb potential, in particular, has led to the most striking confirmations of Newton's gravity theory and is still widely applicable in astronomical applications.

Corrections to Newton's law of gravity become necessary when the escape velocity, $\sqrt{\frac{GM}{r}}$, becomes a substantial fraction of the speed of light. Suppose the escape velocity from a star with the density of water, $1gm/cc$, is half the speed of light. What is the radius of the star?

In the empty space surrounding an isolated black hole, general relativity must be used to correctly describe gravitational effects, since the escape velocity reaches the speed of light at the event horizon. However, sufficiently far from the hole, the Newtonian theory is approximately correct. If the horizon of a solar mass black hole has radius $1km$, approximately how far from the hole will the Newtonian approximation give answers correct to within one percent. In other words, at what distance is the escape velocity $.01c$, where c is the speed of light?

Suppose two particles of masses m_1 and m_2 move in a potential that depends only on the separation of the two particles, $V = V(|\mathbf{x}_1 - \mathbf{x}_2|)$ so the action is

$$S = \int \frac{1}{2}m_1\mathbf{v}_1^2 + \frac{1}{2}m_2\mathbf{v}_2^2 - V(\mathbf{x}_1 - \mathbf{x}_2)$$

Reformulate the problem in terms of the motion of the center of mass,

$$\mathbf{R} = \frac{1}{m_1 + m_2}(m_1\mathbf{x}_1 + m_2\mathbf{x}_2)$$

plus the motion of a single particle of effective mass

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

and position $\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2$ in a central potential, $V(r)$, showing that the action may be rewritten in the form

$$S = \int \frac{1}{2}(m_1 + m_2)\dot{\mathbf{R}}^2 + \frac{1}{2}\mu\dot{\mathbf{r}}^2 - V(\mathbf{x}_1 - \mathbf{x}_2)$$

Solve the Kepler problem completely. Begin with the Lagrangian $L = T - V$ in spherical coordinates, with the potential

$$V = -\frac{\kappa}{r}$$

Use any techniques of the preceeding sections to simplify the problem. Solve for both bounded and unbounded motions.

Study the motion of an isotropic harmonic oscillator, that is, a particle moving in 3-dimensions in the central potential

$$V = \frac{1}{2}kr^2$$

where r is the radial coordinate.

There is a great deal that can be said about central forces without specifying the force law further. These results are largely consequences of the conservation of angular momentum. Recalling theorem 6.7, the total angular momentum of a particle moving in a central potential is always conserved, and by theorem 6.8 the resulting motion is confined to a plane. We may therefore always reduce the action to the form

$$S = \int \left(\frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\varphi}^2\right) - V(r)\right)dt$$

Moreover, we immediately have two conserved quantities. Since $\frac{\partial L}{\partial t} = 0$, energy is conserved,

$$E = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\varphi}^2\right) + V(r)$$

and because φ is cyclic, the magnitude of the angular momentum is conserved,

$$J = mr^2\dot{\varphi}$$

Using J to eliminate $\dot{\varphi}$ from the energy,

$$E = \frac{1}{2}m\dot{r}^2 + \frac{J^2}{2mr^2} + V(r)$$

the problem is equivalent to a particle in 1-dimension moving in the effective potential

$$U(r) = V(r) + \frac{J^2}{2mr^2}$$

In general, orbits will therefore be of three types. Near any minimum, r_0, of the effective potential U, there will be bound states with some maximum and minimum values of r given by the nearest roots of

$$E - U(r) = 0$$

If the energy is high enough that this equation has only one root, the motion will have (at most) one turning point then move off to infinity. Finally, if there are no roots to $E - U = 0$, the motion is unbounded with no turning points.

From the conserved energy, we may reduce the problem to quadratures. Solving for $\frac{dr}{dt}$, we find

$$\sqrt{\frac{2}{m}}t = \int \frac{dr}{\sqrt{E - U(r)}}$$

Alternatively, after solving for $\dot{r}$ we may divide by $J = mr^2\dot{\varphi}$, converting the solution to one for the orbit equation, $r(\varphi)$:

$$\begin{aligned}\frac{dr}{d\varphi} &= \frac{dr/dt}{d\varphi/dt} = \frac{mr^2}{J}\sqrt{\frac{2}{m}(E-U)} \\ \frac{\sqrt{2m}}{J}\varphi &= \int \frac{dr}{r^2\sqrt{(E-U(r))}}\end{aligned}$$

In keeping with the reduction to quadratures and our understanding of the general types of motion possible, we now show that a time-dependent transformation of variables can regularize the central force problem by turning it into an isotropic oscillator.

8.1 Regularization

We can regularize a physical problem if we can transform it smoothly into another problem with well-behaved solutions. In the case of central potentials, it is always possible to transform the problem of bound orbits into the isotropic oscillator. For certain of these transformations, notably the Kepler problem, there are no singular points of the transformation.

8.1.1 Euler's regularization

Essential features of the regularizing transformation are evident even in the 1-dim case. The Euler solution uses the substitutions

$$\begin{aligned}x &= -u^{-2} \\ \frac{d}{dt} &= u^3\frac{d}{d\tau}\end{aligned}$$

to turn the 1-dim Kepler equation of motion into the 1-dim harmonic oscillator. Thus,

$$m\frac{d^2x}{dt^2} = -\frac{\alpha}{x^2}$$

becomes simply

$$\frac{d^2u}{d\tau^2} = -\frac{\alpha}{2m}u$$

Before moving to a proof for the general n-dim case, we note that more general transformations are possible in the 1-dim case. Suppose we begin with an arbitrary potential, $V(x)$

$$m\frac{d^2x}{dt^2} = -\frac{dV}{dx}$$

Then substituting

$$\begin{aligned}x &= f(u) \\ \frac{d}{dt} &= \frac{1}{f'}\frac{d}{d\tau}\end{aligned}$$

we have

$$\begin{aligned}m\frac{1}{f'}\frac{d}{d\tau}\frac{1}{f'}\frac{d}{d\tau}f(u) &= -\frac{dV(f(u))}{dx} \\ m\frac{1}{f'}\frac{d}{d\tau}\frac{du}{d\tau} &= -\frac{dV(f(u))}{df} \\ m\frac{d}{d\tau}\frac{du}{d\tau} &= -\frac{dV(f(u))}{df}\frac{df}{du} \\ m\frac{d^2u}{d\tau^2} &= -\frac{dV(f(u))}{du}\end{aligned}$$

If we choose $f = V^{-1}$ so that $V(f(u)) = \frac{1}{2}u^2$ then the right side is just $-u$ and we have

$$u = u_0 \sin\tau$$

Thus, we may turn any one-dimensional problem into a harmonic oscillator! The catch, of course, is that we have changed the time from t to τ, and transforming back requires a rather troublesome integral. For instance, suppose V is a power law, $V = ax^n$ for any n. Then we choose

$$x = f(u) = u^{\frac{2}{n}}$$

so that the time transforms as

$$\begin{aligned} d\tau &= \frac{n}{2} dt u^{\frac{n-2}{n}} \\ \frac{nt}{2} &= \int u^{\frac{2-n}{n}} d\tau \\ &= A^{\frac{2-n}{n}} \int \sin^{\frac{2-n}{n}} \omega\tau d\tau \end{aligned}$$

The integral involves an integer power of $\sin\omega\tau$ only if $n = -2, -1, 1, 2$.

In higher dimensions the regularizing transformation is complicated by the presence of angular momentum. Still, the general proof is similar, involving a change of both the radial coordinate and the time. Once again, more general potentials can be treated. To begin, we eliminate the angular momentum variables to reduce the problem to a single independent variable. The only remaining difficulty is to handle the angular momentum term in the radial equation.

Use Euler's regularization to solve the 1-dim Kepler problem. First, carry out the Euler substitution to derive the simple harmonic oscillator equation above. Then, to keep the calculation simple, take the solution to the harmonic oscillator equation to be $u = Ae^{i\omega t}$, where $\omega = \sqrt{\frac{\alpha}{2m}}$, and invert the regularizing transformation to solve the 1-dim Kepler problem. Check the answer you get by integrating the 1 -dim Kepler equation directly.

8.1.2 Higher dimensions

Consider the general central force motion in any dimension $d \geq 2$. We begin from the action

$$S = \int dt \left(\frac{1}{2} m \delta_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} - V(r) \right)$$

where the x^i are Cartesian coordinates and $r = \sqrt{\delta_{ij} x^i x^j}$. We have shown that angular momentum is conserved, and that the motion lies in a plane. Therefore, choosing polar coordinates in the plane of motion, the problem reduces to two dimensions. With coordinates $x^{(a)}$ in the directions orthogonal to the plane of motion, the central force equations of motion are

$$\begin{aligned} m\frac{d^2 x^{(a)}}{dt^2} &= 0 \\ m\left(\frac{d^2 r}{dt^2} - r\frac{d\varphi}{dt}\frac{d\varphi}{dt} \right) &= -V'(r) \\ \frac{d\left(mr^2\dot{\varphi}\right)}{dt} &= 0 \end{aligned} \tag{39}$$

We choose $x^{(a)} = 0$, and set $J = mr^2\dot{\varphi} =$ constant. Eliminating $\dot{\varphi}$, these reduce to the single equation

$$m\frac{d^2 r}{dt^2} - \frac{J^2}{mr^3} = -V'(r) \tag{40}$$

Notice that now any transform of r will change the required form of the angular momentum term. What works to avoid this is to recombine the angular momentum and force terms. We again start with

$$\begin{aligned} r &= f(u) \\ \frac{d}{dt} &= \frac{1}{f'}\frac{d}{d\tau} \end{aligned}$$

Then eq.(40) becomes

$$\frac{1}{f'}\frac{d}{d\tau}\left(\frac{1}{f'}f'\frac{du}{d\tau}\right) - \frac{J^2}{m^2 f^3} = -\frac{1}{m}\frac{dV}{df}\left(f\left(u\right)\right)$$

Rearranging, we have

$$\begin{aligned} \frac{d^2u}{d\tau^2} &= \frac{J^2 f'}{m^2 f^3} - f'\frac{dV}{df} \\ &= \frac{J^2 f'}{m^2 f^3} - \frac{df}{du}\frac{dV}{df} \\ &= \frac{J^2}{m^2 f^3}\frac{df}{du} - \frac{dV}{du} \end{aligned}$$

To obtain the isotropic harmonic oscillator we require the combination of terms on the right to give both the angular momentum and force terms of the oscillator:

$$\frac{J^2}{m^2 f^3}\frac{df}{du} - \frac{d}{du}V\left(f\left(u\right)\right) = \frac{\tilde{J}^2}{m^2 u^3} - ku$$

Integrating,

$$\frac{J^2}{2m^2 f^2} + V\left(f\left(u\right)\right) = \frac{\tilde{J}^2}{2m^2 u^2} + \frac{1}{2}ku^2 + \frac{c}{2} \tag{41}$$

If we define

$$g\left(f\right) \equiv \frac{J^2}{2m^2 f^2} + V\left(f\right)$$

the required function f is

$$f = g^{-1}\left(\frac{\tilde{J}^2}{2m^2 u^2} + \frac{1}{2}ku^2 + \frac{c}{2}\right)$$

Substituting this solution into the equation of motion, we obtain the equation for the isotropic oscillator,

$$m\frac{d^2u}{dt^2} - \frac{\tilde{J}^2}{mu^3} = -ku$$

Therefore, every central force problem is locally equivalent to the isotropic harmonic oscillator. We shall see that the same result follows from Hamilton-Jacobi theory, since *every* pair of classical systems with the same number of degrees of freedom are related by some time-dependent canonical transformation.

The solution takes a particularly simple form for the Kepler problem, $V = -\frac{\alpha}{r}$. In this case, eq.(41) becomes

$$\frac{J^2}{2m^2 f^2} - \frac{\alpha}{f} - \left(\frac{\tilde{J}^2}{2m^2 u^2} + \frac{1}{2}ku^2 + \frac{c}{2}\right) = 0$$

Solving the quadratic for $\frac{1}{f}$, we take the positive solution

$$\begin{aligned} \frac{1}{f} &= \frac{m^2}{J^2}\left(\alpha + \sqrt{\alpha^2 + \frac{J^2}{m^2}\left(\frac{\tilde{J}^2}{m^2 u^2} + ku^2 + c\right)}\right) \\ &= \frac{\alpha m^2}{J^2}\left(1 + \frac{J}{\alpha m u}\sqrt{ku^4 + \left(c + \frac{\alpha^2 m^2}{J^2}\right)u^2 + \frac{\tilde{J}^2}{m^2}}\right) \end{aligned}$$

There is also a negative solution.

We may choose c to complete the square under the radical and thereby simplify the solution. Setting

$$c = \frac{2\sqrt{k}J\tilde{J}}{m} - \frac{\alpha^2 m^2}{J^2}$$

the positive solution for f reduces to

$$\frac{1}{f} = \frac{\alpha m^2}{J^2} + m\sqrt{k}u + \frac{\tilde{J}}{Ju}$$

or

$$f = \frac{u}{m\sqrt{k}u^2 + \frac{\alpha m^2}{J^2}u + \frac{\tilde{J}}{J}}$$

The zeros of the denominator never occur for positive u, so the transformations f is regular in the Kepler case.

The regularity of the Kepler case is not typical – it is easy to see that the solution for f may have many branches. The singular points of the transformation in these cases should give information about the numbers of extrema of the orbits, the stability of orbits, and other global properties. The present calculation may provide a useful tool for studying these global properties in detail.

Consider the regularizing transformation when the radial potential is a power law, $V = ar^n$. Show that the solution for $f(u)$ is given by the polynomial equation

$$af^{n+2} - h(u) f^2 + b = 0$$

and find the form of the function $h(u)$. What values of n allow a simple solution?

8.2 General central potentials

We now examine various properties of orbits for different classes of potential. The next exercise illustrates one of the difficulties encountered in dealing with arbitrary potentials.

Consider the central potential

$$V = \alpha (r - r_0)^{2p}$$

Show that this leads to the effective potential

$$U = \frac{J^2}{2mr^2} + \alpha (r - r_0)^{2p}$$

1. Find the solution for circular orbits.
2. Add a small perturbation to the radial coordinate of a circular orbit, so that

   $$r = r_0 + x$$

   where $x << r_0$. Keeping the angular momentum M fixed, compute the frequency of radial oscillations.
3. Show that the frequency of small oscillations may be increased without bound by increasing p. Such closed paths will have arbitrarily many extrema per orbit.

The multiplicity of oscillations per orbit is just one of the things that can happen in arbitrary central potentials. In fact, general central potentials are just as general as arbitrary 1-dimensional potentials. If we choose

$$V(r) = -\frac{J_0^2}{2mr^2} + \tilde{V}(r)$$

then a particle having $J = J_0$ moves in the totally arbitrary effective potential $U(r) = \tilde{V}(r)$. In the next section, we explore some conditions which restrict the motion in two ways which are more in keeping with our expectations for planetary motion, demanding one or more of the following:

1. At any given value of the energy and angular momentum, there is exactly one stable orbit.
2. Circular orbits of increasing radius have increasing angular momentum and increasing energy.
3. Perturbations of circular orbits may precess, but do not wobble – that is, the frequency of small oscillations about circularity is no more than twice the orbital frequency.

Of course, there is no reason that orbits in nature must follow rules such as these, but it tells us a great deal about the structure of the theory to classify potentials by the shapes of the resulting orbits.

As shown in the previous problem, some central forces allow an orbiting particle to wobble arbitrarily many times while following an approximately circular orbit. On of the simplest constraints we might place on potentials is that they should be monotonic. However, a simple counterexample shows that orbits in monotonic potentials may have arbitrarily many extrema.

For a counterexample, consider motion in the potential

$$V(r) = -\frac{J_0^2}{2mr^2} + A\sin kr$$

Monotonicity requires the derivative to be positive definite, for all r,

$$V'(r) = \frac{J_0^2}{mr^3} + Ak\cos kr > 0$$

This is easy to arrange by taking Ak small or $\frac{J_0^2}{mr^3}$ large. Specifically, if we want many periods of kr per orbit, we allow kr to take values up to

$$kr \sim N\pi$$

for some large N. At the same time we require

$$\frac{J_0^2}{mr^3} - Ak > 0$$

Combining these, there will be up to N periods of oscillation per orbit if

$$\begin{aligned} \frac{J_0^2}{mr^2} &> AN\pi \\ r^2 &< \frac{J_0^2}{ANm\pi} \end{aligned}$$

which can always be satisfied by choosing A small. Now since the conserved energy for a central potential is

$$\begin{aligned} E &= \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\varphi}^2\right) + V(r) \\ &= \frac{1}{2}m\dot{r}^2 + \frac{J^2}{2mr^2} + V(r) \end{aligned}$$

the motion occurs in an effective, single particle potential

$$U = \frac{J^2}{2mr^2} + V(r)$$

Substituting for V, the effective potential has minima when

$$U' = -\frac{J_0^2}{mr^3} + \frac{J^2}{mr^3} + Ak\cos kr = 0$$

Therefore, when the angular momentum J is equal to J_0 the first terms cancel, leaving

$$Ak\cos kr = 0$$

Since kr may range up to $N\pi$, this has N solutions. Therefore, monotonic potentials may have arbitrarily many extrema at a given value of the energy and angular momentum. This means that a particle of given energy might be found orbiting at any of N distinct radii.

$$r_l = \left(2l + \frac{3}{2}\right)\frac{\pi}{k}$$

For the potential

$$-\frac{J_0^2}{2mr^2} + A\sin kr$$

with $J = J_0$, find the frequency of small oscillations about a circular orbit at r_l.

8.3 Energy, angular momentum and convexity

Let us now take a more systematic approach. Consider again the energy and effective potential for circular orbits:

$$\begin{aligned} 0 &= -\frac{J^2}{mr_0^3} + V'(r_0) \\ E &= \frac{J^2}{2mr_0^2} + V(r_0) \end{aligned}$$

Combining these results, and solving for E and J^2,

$$\begin{aligned} J^2 &= mr^3 V'(r) \\ E &= \frac{r}{2}V'(r) + V(r) \end{aligned}$$

Clearly we can find some values of E and J that give a circular orbit at any r. The problem of multiple orbits at a given energy or angular momentum may be avoided if we demand that, as functions of r, both J^2 and E increase monotonically. Then we have

$$\begin{aligned} \frac{dJ^2}{dr} &= 3mr^2V' + mr^3V'' \\ &= mr^2(3V' + rV'') > 0 \\ \frac{dE}{dr} &= \frac{1}{2}V' + \frac{r}{2}V'' + V' \\ &= \frac{1}{2}(3V' + rV'') > 0 \end{aligned}$$

In both cases we find the condition

$$V'' > -\frac{3}{r}V'$$

When this condition is met at each r, then there can be only one circular orbit of a given energy and angular momentum.

Check this for our oscillating example, $V(r) = -\frac{M_0^2}{2mr^2} + A\sin kr$. We have

$$\begin{aligned} -\frac{3}{r_0}V' &= -\frac{3M_0^2}{mr_0^4} - \frac{3}{r_0}Ak\cos kr_0 \\ V'' &= -\frac{3M_0^2}{mr_0^4} - Ak^2\sin kr_0 \end{aligned}$$

so the condition requires

$$\begin{aligned} -Ak^2 \sin kr &> -\frac{3}{r} Ak \cos kr \\ kr \sin kr &< 3 \cos kr \end{aligned}$$

While this condition is satisfied at each minimum, it cannot be true for all values of r, so the condition rules out this potential.

Now consider perturbations around circular orbits. For a circular orbit the orbital frequency is

$$\dot{\varphi} = \frac{J}{mr_0^2}$$

Let $r = r_0 + r_0 x$, with $x << 1$ so that expanding the energy to quadratic order,

$$\begin{aligned} E + \varepsilon &= \frac{1}{2} m r_0^2 \dot{x}^2 + \frac{J^2}{2mr_0^2 (1+x)^2} + V(r_0 + r_0 x) \\ &= \frac{1}{2} m r_0^2 \dot{x}^2 + \frac{J^2}{2mr_0^2} \left(1 - 2x + 3x^2\right) \\ &\quad + V(r_0) + V'(r_0) r_0 x + \frac{1}{2} V''(r_0) r_0^2 x^2 \\ \varepsilon &= \frac{1}{2} m r_0^2 \dot{x}^2 - \frac{J^2}{mr_0^2} x + V'(r_0) r_0 x + \frac{3J^2}{2mr_0^2} x^2 + \frac{1}{2} V''(r_0) r_0^2 x^2 \\ &= \frac{1}{2} m r_0^2 \dot{x}^2 + \frac{1}{2} \left(\frac{3J^2}{mr_0^2} + r_0^2 V''(r_0) \right) x^2 \end{aligned}$$

These perturbations therefore oscillate with effective mass mr_0^2 and constant $\left(\frac{3J^2}{mr_0^2} + r_0^2 V''(r_0) \right)$, so the frequency of oscillations is given by

$$\omega^2 = \frac{1}{mr_0^2} \left(\frac{3J^2}{m^2 r_0^2} + r_0^2 V''(r_0) \right)$$

Comparing to the orbital frequency, we have

$$\begin{aligned} \frac{\omega^2}{\dot{\varphi}^2} &= \frac{\frac{3J^2}{m^2 r_0^4} + \frac{1}{m} V''(r_0)}{\frac{J^2}{m^2 r_0^4}} \\ &= 3 + \frac{mr_0^4}{J^2} V''(r_0) \end{aligned}$$

Using the condition for minimal potential, $\frac{J^2}{mr_0^3} = V'(r_0)$, this becomes

$$\frac{\omega^2}{\dot{\varphi}^2} = 3 + \frac{r_0}{V'(r_0)} V''(r_0)$$

Finally, with the convexity condition

$$V''(r_0) > -\frac{3}{r_0} V'(r_0)$$

the ratio of frequencies is given by

$$\frac{\omega^2}{\dot{\varphi}^2} = 3 + \frac{r_0}{V'(r_0)} V''(r_0) > 0$$

The second derivative condition is therefore the necessary and sufficient condition for the existence of simple harmonic oscillations about circular orbits. To insure that these oscillations produce precession but not wobble, we further require

$$4 \geq \frac{\omega^2}{\dot{\varphi}^2} = 3 + \frac{r_0}{V'(r_0)} V''(r_0) > 0$$

or

$$-3 < \frac{r_0 V''}{V'} \leq 1$$

For a power law potential, $V = ar^n$, this gives

$$\begin{aligned} V' &= nar^{n-1} \\ V'' &= n(n-1)ar^{n-2} \\ -3 &< (n-1) \leq 1 \\ -2 &< n \leq 2 \end{aligned}$$

Not surprisingly, any power law potential with $n > -2$ satisfies the energy and momentum conditions, but there is an upper limit on n if we are to avoid wobble. As we shall see, this condition is closely related to the condition for nonperturbatively stable closed orbits.

8.4 Bertrand's theorem: closed orbits

We now consider a slightly different question: under what conditions do orbits close? We will restrict our attention to power law potentials, proceeding by finding circular orbits, then finding the frequency of small oscillations about those orbits. By comparing the frequency of small oscillations to the frequency of the overall orbital motion, we can tell whether the motion closes.

For circular orbits in power law potentials

$$V = ar^n$$

the energy and angular momentum given by,

$$\begin{aligned} E_0 &= \frac{J_0^2}{2mr_0^2} + ar_0^n \\ J_0 &= mr_0^2 \dot{\varphi}_0 \end{aligned}$$

where $\dot{\varphi}_0$ is the frequency of the orbit. We also know that the motion occurs at a minimum of the effective potential, so that

$$U'(r_0) = nar_0^{n-1} - \frac{J_0^2}{mr_0^3} = 0$$

Now consider small oscillations around the circular orbits. Let

$$\begin{aligned} r &= r_0(1+x) \\ \dot{\varphi} &= \dot{\varphi}_0\left(1 + \alpha x + \beta x^2\right) \\ E &= E_0 + \varepsilon \\ J &= J_0 + j \end{aligned}$$

where we assume

$$x << 1$$

First consider angular momentum. We have, to quadratic order in x,

$$\begin{aligned} J &= J_0 + j \\ &= mr_0^2 \dot{\varphi}_0 \left(1 + 2x + x^2\right)\left(1 + \alpha x + \beta x^2\right) \\ j &= mr_0^2 \dot{\varphi}_0 \left((2+\alpha)x + (1 + 2\alpha + \beta)x^2\right) \end{aligned}$$

Since j must be constant while $x = x(t)$, we must have

$$\begin{aligned} j &= 0 \\ \alpha &= -2 \\ \beta &= 3 \end{aligned}$$

The value of the angular momentum is unchanged,

$$J = J_0 = mr_0^2 \dot{\varphi}_0$$

while the angular velocity acquires small, time-dependent corrections

$$\dot{\varphi} = \dot{\varphi}_0 \left(1 - 2x + 3x^2\right)$$

Expanding the energy to second order in small quantities, we have

$$\begin{aligned} E_0 + \varepsilon &= \frac{1}{2} m\dot{x}^2 + \frac{J^2}{2mr_0^2 (1+x)^2} + ar_0^n (1+x)^n \\ &= \frac{1}{2} mr_0^2 \dot{x}^2 + \frac{J_0^2}{2mr_0^2} \left(1 - 2x + 3x^2\right) + ar_0^n \left(1 + nx + \frac{1}{2} n(n-1) x^2\right) \\ &= \frac{1}{2} mr_0^2 \dot{x}^2 + \frac{J_0^2}{2mr_0^2} + ar_0^n + \left(-\frac{J_0^2}{mr_0^2} + nar_0^n\right) x \\ &\quad + \left(\frac{3J_0^2}{2mr_0^2} + \frac{1}{2} n(n-1) ar_0^n\right) x^2 \end{aligned}$$

Using the value of E_0 and the condition for a minimum of the potential, the constant and linear terms cancel. Replacing $nar_0^n = \frac{J_0^2}{mr_0^2}$ leaves

$$\varepsilon = \frac{1}{2} mr_0^2 \dot{x}^2 + \frac{1}{2} \frac{J_0^2}{mr_0^2} (3 + (n-1)) x^2$$

This is clearly the energy for a simple harmonic oscillator with effective mass

$$\mu = mr_0^2$$

and constant

$$k = (n+2) \frac{J_0^2}{mr_0^2}$$

and therefore of squared frequency

$$\begin{aligned} \omega^2 &= \frac{k}{\mu} \\ &= (n+2) \frac{J_0^2}{m^2 r_0^4} \\ &= (n+2) \dot{\varphi}_0^2 \end{aligned}$$

The orbits will close after q circuits if ω is any rational multiple, $\frac{p}{q}$ of $\dot{\varphi}_0$, for non-negative integers p and q. This occurs when

$$\begin{aligned} \sqrt{n+2} &= \frac{p}{q} \\ n &= \frac{p^2}{q^2} - 2 \end{aligned}$$

When $q = 1$, each single orbital sweep is identical. In this case we have

$$\begin{aligned} n &= p^2 - 2 \\ &\in \{-2, -1, 2, 7, 14, \ldots\} \end{aligned}$$

Notice that we have computed the frequency of small oscillations only to lowest order, and that at the same order the orbital frequency is still $\dot{\varphi}_0$:

$$\dot{\varphi} = \dot{\varphi}_0 \left(1 - 2x + 3x^2\right) \simeq \dot{\varphi}_0$$

It can be shown that the $n = -2$ case allows no bound orbits at all, and second, that unless $n = 2$ or $n = -1$, the result holds only for orbits which are perturbatively close to circular, while orbits deviating nonperturbatively fail to close. The conclusion is therefore that generic orbits close only for the Kepler/Coulomb and Hooke's law potentials

$$\begin{aligned} V &= \frac{a}{r} \\ V &= ar^2 \end{aligned}$$

This result is *Bertrand's Theorem*. A complete proof requires keeping additional terms in the perturbative expansion.

Study motion in a central potential $V = -\frac{a}{r^2}$. Prove the following:

1. There are no bound orbits.
2. Orbits which initially increase in distance from the center of force continue to spiral to infinity. Find an expression for the angle as a function of time, and find the limiting angle as $t \to \infty$.
3. Orbits which initially decrease in distance from the center of force spiral inward, reaching the center in a finite amount of time. During this finite time, the angle increases without bound.

8.5 Symmetries of motion for the Kepler problem

Recent decades have seen new techniques and revivals of long-forgotten symmetries of the Kepler problem ([32],[33]). The best-known rediscovery concerning the Kepler problem is that in addition to the energy, E and angular momentum,

$$\begin{aligned} E &= \frac{1}{2} m \dot{\mathbf{x}}^2 - \frac{\alpha}{r} \\ \mathbf{L} &= \mathbf{r} \times \mathbf{p} \end{aligned} \tag{42}$$

the Laplace-Runge-Lenz vector ([34], [35], [36], [37]) is conserved.

Keplerian orbits can be described completely in terms of six initial conditions, and since one of these is the initial position on a given ellipse, only five remain among the seven degrees of freedom in the energy, angular momentum and Laplace-Runge-Lenz vector [38]. Two constraints – the orthogonality of $\mathbf{A}$ and $\mathbf{L}$, and a relationship between the magnitudes A, L and E – give the correct count. Of course, these three quantities are not the only set of constants we can choose. A number of fairly recent authors ([39], [40], [41], [3] have identified a simpler conserved vector quantity, which (lacking evidence for an earlier reference) we will call the Hamilton vector [42], which may be used to replace either A or $\mathbf{L}$.

To begin our investigation, consider the time rate of change of the angular unit vector $\hat{\varphi}$, given by

$$\begin{aligned} \hat{\varphi} &= -\hat{\imath} \sin\varphi + \hat{\mathbf{j}} \cos\varphi \\ \frac{d\hat{\varphi}}{dt} &= -\hat{\imath}\dot{\varphi} \cos\varphi - \hat{\mathbf{j}}\dot{\varphi} \sin\varphi \\ &= -\dot{\varphi}\hat{\mathbf{r}} \end{aligned}$$

Using the force law and the angular momentum, we can write this as

$$\begin{aligned}\frac{d\hat{\varphi}}{dt} &= -\dot{\varphi}\hat{\mathbf{r}} \\ &= -\frac{L}{mr^2}\hat{\mathbf{r}} \\ &= \frac{L}{m\alpha}\mathbf{f}(r)\end{aligned}$$

where

$$\mathbf{f}(r) = -\frac{\alpha}{r^2}\hat{\mathbf{r}}$$

is Newton's gravitational force. By the law of motion, we have

$$\frac{d\mathbf{p}}{dt} = \mathbf{f}(r)$$

so we may write

$$\frac{d\hat{\varphi}}{dt} = \frac{L}{m\alpha}\frac{d\mathbf{p}}{dt}$$

or simply

$$\frac{d}{dt}\left(\mathbf{p} - \frac{m\alpha}{L}\hat{\varphi}\right) = 0$$

This provides a conservation law for Kepler orbits. We define Hamilton's vector,

$$\mathbf{h} = \mathbf{p} - \frac{m\alpha}{L}\hat{\varphi}$$

as the conserved quantity.

An alternative conserved quantity, the Laplace-Runge-Lenz vector, is given by the cross product of $\mathbf{h}$ with the angular momentum,

$$\begin{aligned}\mathbf{A} &= \mathbf{h} \times \mathbf{L} \\ &= \left(\mathbf{p} - \frac{m\alpha}{L}\hat{\varphi}\right) \times (\mathbf{r} \times \mathbf{p}) \\ &= \mathbf{p} \times \mathbf{L} - \frac{m\alpha}{L}\left(\mathbf{r}\left(\hat{\varphi} \cdot \mathbf{p}\right) - \mathbf{p}\left(\mathbf{r} \cdot \hat{\varphi}\right)\right) \\ &= \mathbf{p} \times \mathbf{L} - \frac{m\alpha}{L}\left(mr^2\dot{\varphi}\right)\hat{\mathbf{r}} \\ &= \mathbf{p} \times \mathbf{L} - m\alpha\hat{\mathbf{r}}\end{aligned}$$

The Laplace-Runge-Lenz vector, the Hamilton vector and the angular momentum form a mutually orthogonal basis.

Show that

$$\mathbf{L} \times \mathbf{A}$$

is parallel to $\mathbf{h}$ and find the proportionality constant.

Check directly that $\frac{d\mathbf{A}}{dt} = 0$. Choose coordinates so that the motion lies in the xy plane with the perihelion on the x axis. Show that the Laplace-Runge-Lenz vector points in the direction of perihelion.

Hamilton's vector may be used to find the motion of the system. We follow a proof due to Muñoz [3]. Let motion be in the xy-plane and choose the perihelion of the orbit to occur at time $t = 0$ on the x-axis. Then the initial velocity is given by

$$\mathbf{v} = v_0\hat{\varphi} = v_0\hat{\mathbf{j}}$$

Then at $t = 0$,

$$\begin{aligned}\mathbf{h} &= m\mathbf{v} - \frac{m\alpha}{L}\hat{\varphi} \\ &= \left(mr_0\dot{\varphi}_0 - \frac{m\alpha}{L}\right)\hat{\mathbf{j}}\end{aligned}$$

At an arbitrary time, dotting $\mathbf{h}$ with $\hat{\varphi}$ gives

$$\begin{aligned}\mathbf{h}\cdot\hat{\varphi} &= \mathbf{p}\cdot\hat{\varphi} - \frac{\alpha m}{L} \\ \left(mr_0\dot{\varphi}_0 - \frac{m\alpha}{L}\right)\cos\phi &= mr\dot{\varphi} - \frac{\alpha m}{L}\end{aligned}$$

or replacing $\dot{\varphi} = \frac{L}{mr^2}$,

$$\begin{aligned}\left(\frac{L}{r_0} - \frac{m\alpha}{L}\right)\cos\phi &= \frac{L}{r} - \frac{m\alpha}{L} \\ \frac{m\alpha}{L^2} + \left(\frac{1}{r_0} - \frac{m\alpha}{L^2}\right)\cos\phi &= \frac{1}{r}\end{aligned}$$

or

$$r = \frac{L^2/m\alpha}{1 + \left(\frac{L^2}{m\alpha r_0} - 1\right)\cos\phi}$$

as usual. In terms of the initial energy

$$\begin{aligned}E &= \frac{1}{2}mr_0^2\dot{\varphi}_0^2 - \frac{\alpha}{r_0} \\ E &= \frac{L^2}{2mr_0^2} - \frac{\alpha}{r_0} \\ 0 &= \frac{L^2}{2mr_0^2} - \frac{\alpha}{r_0} - E \\ \frac{1}{r_0} &= \frac{\alpha + \sqrt{\alpha^2 + \frac{4EL^2}{2m}}}{\frac{2L^2}{2m}} \\ &= \frac{m\alpha}{L^2}\left(1 + \sqrt{1 + \frac{2EL^2}{m\alpha^2}}\right)\end{aligned}$$

Defining

$$\begin{aligned}r_m &= \frac{L^2}{m\alpha} \\ \varepsilon &= \sqrt{1 + \frac{2EL^2}{m\alpha^2}}\end{aligned}$$

The motion is therefore given in terms of the energy and angular momentum by

$$r = \frac{r_m}{1 + \varepsilon\cos\varphi} \tag{43}$$

These curves are hyperbolas, parabolas and ellipses, all of which are conic sections, that is, curves that arise from the intersection of a plane with a cone.

8.5.1 Conic sections

We can characterize the conic sections as follows. Consider a cone aligned along the z axis with its vertex at the origin,

$$z = \alpha\sqrt{x^2 + y^2}$$

A plane with closest approach to the origin given by a fixed vector $\mathbf{a}$ consists of all vectors, $\mathbf{x} - \mathbf{a}$, from $\mathbf{a}$ which are orthogonal to $\mathbf{a}$,

$$(\mathbf{x} - \mathbf{a}) \cdot \mathbf{a} = 0$$

Because the cone is rotationally symmetric, we can choose $\mathbf{a}$ to lie in the xz plane without any loss of generality. With $\mathbf{a} = (a, 0, b)$, the plane is given by

$$ax + bz = a^2 + b^2$$

Solving for z and substituting into the equation for the cone, we have

$$z^2 = \frac{1}{b^2}\left(a^2 + b^2 - ax\right)^2 = \alpha^2\left(x^2 + y^2\right)$$

Setting $d = a^2 + b^2$ and $c = \alpha^2 b^2$ and simplifying,

$$\left(c - a^2\right)x^2 + cy^2 + 2adx = d^2$$

When $c = a^2$, this is the equation of a parabola. When $c - a^2$ is nonzero we can set $e = +\sqrt{|c - a^2|}$ and $\beta = sign\left(c - a^2\right)$, and rewrite the equation as

$$\beta\left(ex + \frac{2\beta ad}{e}\right)^2 + cy^2 = d^2 + \left(\frac{2\beta ad}{e}\right)^2$$

In this form we recognize the equation for an hyperbola when $\beta = -1$ and an ellipse when $\beta = +1$.

When an ellipse is centered at the origin, we may write its equation in the simpler form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

To derive this form from the equation for a Kepler orbit we must convert eq.(43) from r and ϕ, which are measured from one focus of the ellipse, to Cartesian coordinates measured from the center. The full range of r from $r_{\min}$ at $\phi = 0$ to $r_{\max}$ at $\phi = \pi$ gives the semi-major axis $d = 2a$ as

$$a = \frac{r_m}{1 - \varepsilon^2}$$

while the maximum value of y gives the semiminor axis

$$b = \frac{r_m}{\sqrt{1 - \varepsilon^2}}$$

It is then not hard to see that the coordinates of any point are given by

$$\begin{aligned} y &= \frac{r_m \sin\phi}{1 + \varepsilon\cos\phi} \\ x &= \varepsilon a + \frac{r_m \cos\phi}{1 + \varepsilon\cos\phi} \end{aligned}$$

Then we compute:

$$\begin{aligned} \frac{x^2}{a^2} + \frac{y^2}{b^2} &= \left(\frac{1 - \varepsilon^2}{r_m}\right)^2\left(\varepsilon a + \frac{r_m\cos\phi}{1 + \varepsilon\cos\phi}\right)^2 + \frac{1 - \varepsilon^2}{r_m^2}\left(\frac{r_m\sin\phi}{1 + \varepsilon\cos\phi}\right)^2 \\ &= \frac{1}{\left(1 + \varepsilon\cos\phi\right)^2}\left(\left(\varepsilon\left(1 + \varepsilon\cos\phi\right) + \left(1 - \varepsilon^2\right)\cos\phi\right)^2 + \left(1 - \varepsilon^2\right)\sin^2\phi\right) \\ &= 1 \end{aligned}$$

so the orbit is an ellipse.

Show that the Laplace-Runge-Lenz vector points in the direction of perihelion and has magnitude Emr_0.

Show that when $\varepsilon > 1$, the orbit formula, eq.(43), describes hyperbolas.

8.6 Newtonian gravity

Newtonian gravity provides the simplest example of a field theory. We begin with the potential of an infinitesimal point mass, dm,

$$V = -\frac{G}{r}dm$$

Now, if we have a continuous distribution of mass, $\rho(\mathbf{x}')$ we can characterize the potential additively. For each volume d^3x' we have mass

$$dm = \rho(\mathbf{x}')\, d^3x'$$

The infinitesimal contribution to the potential at a point $\mathbf{x}$ due to dm at point $\mathbf{x}'$ is then

$$d\phi(\mathbf{x}) = -\frac{G\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d^3x'$$

and the total potential at $\mathbf{x}$ is

$$\phi(\mathbf{x}) = -\int \frac{G\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d^3x'$$

Applying the Laplacian to ϕ, we have

$$\begin{aligned}\nabla^2\phi(\mathbf{x}) &= -\nabla^2\int \frac{G\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d^3x' \\ &= -\int G\rho(\mathbf{x}')\left(\nabla^2\frac{1}{|\mathbf{x}-\mathbf{x}'|}\right)d^3x'\end{aligned}$$

We now show that the Laplacian gives a Dirac delta function,

$$\nabla^2\frac{1}{|\mathbf{x}-\mathbf{x}'|} = -4\pi\delta^3(\mathbf{x}-\mathbf{x}')$$

Without loss of generality, look at $\mathbf{x}' = 0$, and define the sequence of functions

$$\begin{aligned}f_a(\mathbf{x}-\mathbf{x}') &= \nabla^2\frac{1}{\sqrt{|\mathbf{x}-\mathbf{x}'|^2+a^2}} \\ f_a(\mathbf{x}) &= \nabla^2\frac{1}{\sqrt{r^2+a^2}}\end{aligned}$$

We want to show that

$$\lim_{a\to\infty} f_a(\mathbf{x}) = \lim_{a\to\infty}\nabla^2\frac{1}{\sqrt{r^2+a^2}} = -4\pi\delta(\mathbf{x})$$

First, for $r \neq 0$,

$$\begin{aligned}\nabla^2\frac{1}{\sqrt{r^2+a^2}} &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\left(\frac{1}{\sqrt{r^2+a^2}}\right)\right) \\ &= -\frac{1}{r^2}\frac{\partial}{\partial r}\left(\frac{r^3}{(r^2+a^2)^{3/2}}\right) \\ &= -\frac{1}{r^2}\left(\frac{3r^2}{(r^2+a^2)^{3/2}} - \frac{3}{2}\frac{2r^4}{(r^2+a^2)^{5/2}}\right) \\ &= -\frac{3a^2}{(r^2+a^2)^{5/2}}\end{aligned}$$

Now, for the integral

$$\begin{aligned} I &= \lim_{a\to\infty}\int g(\mathbf{x})\,\nabla^2\frac{1}{\sqrt{r^2+a^2}}d^3r \\ &= \lim_{a\to\infty}\int_0^R g(\mathbf{x})\,\nabla^2\frac{1}{\sqrt{r^2+a^2}}d^3r + \lim_{a\to\infty}\int_R^\infty g(\mathbf{x})\,\nabla^2\frac{1}{\sqrt{r^2+a^2}}d^3r \\ &= \lim_{a\to\infty}\int_0^R g(\mathbf{x})\,\nabla^2\frac{1}{\sqrt{r^2+a^2}}d^3r - \lim_{a\to\infty}3a^2\int_R^\infty\frac{g(\mathbf{x})}{\left(r^2+a^2\right)^{5/2}}r^2drd\Omega \end{aligned}$$

The second integral is bounded by

$$\begin{aligned} \lim_{a\to\infty}3a^2\int_R^\infty\frac{g(\mathbf{x})}{\left(r^2+a^2\right)^{5/2}}r^2drd\Omega &< \lim_{a\to\infty}3a^2\int_\varepsilon^\infty\frac{g(\mathbf{x})}{r^3}drd\Omega \\ &< \lim_{a\to\infty}\frac{3a^2}{R^3}\int_R^\infty g(\mathbf{x})\,drd\Omega \\ &= 0 \end{aligned}$$

since g, being a test function, has finite integral. For the first integral,

$$\begin{aligned} I &= \lim_{a\to\infty}\int_0^R g(\mathbf{x})\,\nabla^2\frac{1}{\sqrt{r^2+a^2}}d^3r \\ &= -g(0)\lim_{a\to\infty}\int_0^R\frac{3a^2}{\left(r^2+a^2\right)^{5/2}}d^3r \\ &= -4\pi g(0)\lim_{a\to\infty}\int_0^\varepsilon\frac{3a^2}{\left(r^2+a^2\right)^{5/2}}r^2dr \end{aligned}$$

and integrating directly,

$$\begin{aligned} \int_0^R\frac{3a^2}{\left(r^2+a^2\right)^{5/2}}r^2dr &= 3a^2\int_0^R\frac{r^2}{\left(r^2+a^2\right)^{5/2}}dr \\ &= 3a^2\int_0^R\frac{a^2\tan^2\theta}{\left(a^2\tan^2\theta+a^2\right)^{5/2}}\frac{ad\theta}{\cos^2\theta} \\ &= 3\int_0^{\tan^{-1}\frac{R}{a}}\sin^2\theta d\left(\sin\theta\right) \\ &= \sin^3\left(\tan^{-1}\frac{R}{a}\right) \\ &= \frac{R^3}{\left(R^2+a^2\right)^{3/2}} \end{aligned}$$

so that

$$\begin{aligned} \lim_{a\to\infty}\int g(\mathbf{x})\,\nabla^2\frac{1}{\sqrt{r^2+a^2}}d^3r &= -4\pi g(0)\lim_{a\to\infty}\int_0^R\frac{3a^2}{\left(r^2+a^2\right)^{5/2}}r^2dr \\ &= -4\pi g(0)\lim_{a\to\infty}\frac{R^3}{\left(R^2+a^2\right)^{3/2}} \\ &= -4\pi g(0) \end{aligned}$$

and finally

$$\delta(\mathbf{x}) = \lim_{a\to\infty}f_a(\mathbf{x}) = \lim_{a\to\infty}\nabla^2\frac{1}{\sqrt{r^2+a^2}}$$

Applying this to the gravitational potential,

$$\begin{aligned}\nabla^2\phi(\mathbf{x}) &= \nabla^2\int\frac{G\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d^3x' \\ &= -\int G\rho(\mathbf{x}')\,4\pi\delta(\mathbf{x}-\mathbf{x}')\,d^3x' \\ &= -4\pi G\rho(\mathbf{x})\end{aligned}$$

This is now a field equation for the gravitational potential ϕ :

$$\nabla^2\phi(\mathbf{x}) = -4\pi G\rho(\mathbf{x})$$

Defining the *gravitational field,*

$$\mathbf{g}(\mathbf{x}) = -\nabla\phi(\mathbf{x})$$

we may write

$$\nabla\cdot\mathbf{g}(\mathbf{x}) = 4\pi G\rho(\mathbf{x})$$

and using Stoke's theorem, we see that Gauss' law applies to the gravitational field:

$$\begin{aligned}\oint_S \mathbf{g}(\mathbf{x})\cdot\mathbf{n}\,d^2x &= \int_V \nabla^2\phi(\mathbf{x})\,d^3x \\ &= -4\pi G\int_V \rho(\mathbf{x})\,d^3x \\ &= -4\pi G M_V\end{aligned}$$

That is, the integral of the normal component of the field over the closed boundary surface S of a volume V is equal to $-4\pi G$ times the total mass contained in that volume.

Use Gauss' law to find the potential inside and outside a uniform spherical ball of total mass M. Prove that the gravitational field at any point outside the ball is the same as it would be if the mass M were concentrated at a point at the center.

Notice that the gravitational field equation,

$$\nabla\cdot\mathbf{g}(\mathbf{x}) = 4\pi G\rho(\mathbf{x})$$

is independent of time, so that changes in the field are felt instantaneously at arbitrary distances. We might try to fix the problem by including time derivatives in the equation for the potential,

$$-\frac{1}{c^2}\frac{\partial^2\phi(\mathbf{x},t)}{\partial t^2} + \nabla^2\phi(\mathbf{x},t) = -4\pi G\rho(\mathbf{x},t)$$

Show that, in empty space where $\rho = 0$ this equation permits gravitational wave solutions.

9 Constraints

We are often interested in problems which do not allow all particles a full range of motion, but instead restrict motion to some subspace. When constrained motion can be described in this way, there is a simple technique for formulating the problem.

Subspaces of constraint may be described by relationships between the coordinates,

$$f\left(x^i,t\right) = 0$$

The trick is to introduce f into the problem in such a way that it must vanish in the solution. Our understanding of the Euler-Lagrange equation as the covariant form of Newton's second law tells us how to

do this. Since the force that maintains the constraint must be orthogonal to the surface $f = 0$, it will be in the direction of the gradient of f and we can write

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} - \frac{\partial L}{\partial x^i} = \lambda \frac{\partial f}{\partial x^i}$$

where $\lambda = \lambda\left(x^i, \dot{x}^i, t\right)$ determines the amplitude of the gradient required to provide the constraining force. In addition, we need the constraint itself.

Remarkably, both the addition of $\lambda \frac{\partial f}{\partial x^i}$ and the constraint itself follow as a variation of a slightly altered form of the action. Since f itself is independent of the velocity, the simple replacement of the action by

$$S = \int (L + \lambda f)\, dt$$

means the the variation of S now gives

$$\delta S = \int \left(\left(-\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} + \frac{\partial L}{\partial x^i} + \lambda \frac{\partial f}{\partial x^i} \right) \delta x^i + f \delta \lambda \right) dt$$

where we treat λ as an independent degree of freedom. Thus, the variation $\delta\lambda$ is independent of the n coordinate variations δx^i and we get $n + 1$ equations,

$$\begin{aligned} 0 &= -\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} + \frac{\partial L}{\partial x^i} + \lambda \frac{\partial f}{\partial x^i} \\ 0 &= f \end{aligned}$$

These are exactly what we require – the extra equation gives just enough information to determine λ.

Thus, by increasing the number of degrees of freedom of the problem by one for each constraint, we include the constraint while allowing free variation of the action. In exchange for the added equation of motion, we learn that the force required to maintain the constraint is

$$F^i_{constraint} = \lambda g^{ij} \frac{\partial f}{\partial x^i}$$

The advantage of treating constraints in this way is that we now may carry out the variation of the coordinates freely, as if all motions were possible. The variation of λ, called a *Lagrange multiplier*, brings in the constraint automatically. In the end, we will have the N Euler-Lagrange equations we started with (assuming an initial N degrees of freedom), plus an additional equation for each Lagrange multiplier.

When the constraint surface is fixed in space the constraint force never does any work since there is never any motion in the direction of the gradient of f. If there is time dependence of the surface then work will be done. Because f remains zero its total time derivative vanishes so

$$\begin{aligned} 0 &= \frac{df}{dt} \\ &= \frac{\partial f}{\partial x^i}\frac{dx^i}{dt} + \frac{\partial f}{\partial t} \end{aligned}$$

or multiplying by λdt and integrating,

$$W = \int \lambda \frac{\partial f}{\partial x^i} dx^i = -\int \lambda \frac{\partial f}{\partial t} dt$$

Thus, the Lagrange multiplier allows us to compute the work done by a moving constraint surface.

As a simple example, consider the motion of a particle under the influence of gravity, $V = mgz$, constrained to the inclined plane

$$f(x, z) = z - x \tan\theta = 0$$

where θ is a fixed angle. We write the action as

$$S = \int \left(\frac{1}{2} m \dot{\mathbf{x}}^2 - mgz + \lambda \left(z - x \tan\theta \right) \right) dt$$

Because y is cyclic we immediately have

$$p_y = m\dot{y} = mv_{0y} = const.$$

so that

$$y = y_0 + v_{0y} t$$

We also have conservation of energy,

$$E = \frac{1}{2} m \dot{\mathbf{x}}^2 + mgz - \lambda \left(z - x \tan\theta \right)$$

Varying x, z and λ we have three further equations,

$$\begin{aligned} 0 &= m\ddot{x} + \lambda \tan\theta \\ 0 &= m\ddot{z} + mg - \lambda \\ 0 &= z - x \tan\theta \end{aligned}$$

We also have conservation of energy,

$$\begin{aligned} E &= \frac{1}{2} m \dot{\mathbf{x}}^2 + mgz - \lambda \left(z - x \tan\theta \right) \\ &= \frac{1}{2} m \dot{\mathbf{x}}^2 + mgz \end{aligned}$$

This shows that for this example the constraint contributes no energy.

To solve the x and z equations, we must eliminate λ. Differentiate the constraint equation twice. Then, subtracting m times the result from the weighted difference of the x and z equations,

$$\begin{aligned} 0 &= \tan\theta \left(m\ddot{x} + \lambda \tan\theta \right) - \left(m\ddot{z} + mg - \lambda \right) + m \left(\ddot{z} - \ddot{x} \tan\theta \right) \\ &= \lambda \tan^2\theta - mg + \lambda \\ &= \frac{1}{\cos^2\theta} \lambda - mg \end{aligned}$$

giving

$$\lambda = mg \cos^2\theta$$

so in this case, λ is constant. Replacing λ in the x and z equations gives

$$\begin{aligned} 0 &= \ddot{x} + g \cos\theta \sin\theta \\ 0 &= \ddot{z} + g \sin^2\theta \end{aligned}$$

From this we immediately find x and z separately:

$$\begin{aligned} x &= x_0 + v_{0x} t + \frac{1}{2} \cos\theta \sin\theta g t^2 \\ z &= z_0 + v_{0z} t - \frac{1}{2} \sin^2\theta g t^2 \end{aligned}$$

The constraint force is given by

$$\begin{aligned} F^i &= \lambda g^{ij} \frac{\partial f}{\partial x^j} \\ &= mg \cos^2\theta \left(-\tan\theta, 0, 1 \right) \\ &= mg \cos\theta \left(-\sin\theta, 0, \cos\theta \right) \end{aligned}$$

Notice that the magnitude is $mg\cos\theta$ as expected. Moreover, since the vector

$$w^i = (1, 0, \tan\theta)$$

is parallel to the surface of the plane, we check that

$$\begin{aligned} w_i F^i &= (1, 0, \tan\theta) \cdot mg\cos\theta\,(-\sin\theta, 0, \cos\theta) \\ &= 0 \end{aligned}$$

so the direction is perpendicular to the surface.

Work the following problems using Lagrange multipliers.

Work the inclined plane problem directly from Newton's second law and check explicitly that the force applied by the plane is

$$F^i = mg\cos\theta\,(-\sin\theta, 0, \cos\theta)$$

Repeat the inclined plane problem with a moving plane. Let the plane move in the direction

$$v^i = (v_1, 0, v_3)$$

Find the work done on the particle by the plane. For what velocities v^i does the particle stay in the same position on the plane?

A particle of mass m moves frictionlessly on the surface $z = k\rho^2$, where $\rho = \sqrt{x^2 + y^2}$ is the polar radius. Let gravity act in the $-z$ direction, $\mathbf{F} = -mg\mathbf{k}$. Find the motion of the system.

A ball moves frictionlessly on a horizontal tabletop. The ball of mass m is connected to a string of length L which passes through a hole in the tabletop and is fastened to a pendulum of mass M. The string is free to slide through the hole in either direction. Find the motion of the ball and pendulum.

Study the motion of a spherical pendulum: a negligably light rod of length L with a mass m attached to one end. The remaining end is fixed in space. The mass is therefore free to move anywhere on the surface of a sphere of radius L, under the influence of gravity, $-mg\mathbf{k}$.

1. Write the Lagrangian for the system.
2. Identify any conserved quantities.
3. Use the conservation laws and any needed equations of motion to solve for the motion. In particular study the following motions:
   (a) Motion confined to a vertical plane, of small amplitude.
   (b) Motion confined to a vertical plane, of arbitrary amplitude.
   (c) Motion confined to a horizontal plane.
4. Beginning with your solution for motion in a horizontal plane, study small oscillations away from the plane.

10 Rotating coordinates

It frequently happens that we want to describe motion in a non-inertial frame. For example, the motion of a projectile relative to coordinates fixed to Earth is modified slightly by the rotation of Earth. The use of a non-inertial frame introduces additional terms into Newton's law. We have shown that the Euler-Lagrangian is covariant under time-dependent diffeomorphisms,

$$x^i = x^i\left(y^j, t\right)$$

To study this case, let x^i be a Cartesian coordinate in an inertial reference frame, and let y^i rotate relative to x^i. Let $R^i{}_j(t)$ be the rotation taking y^i to x^i, so that

$$\begin{aligned} y^i &= R^i{}_j(t)\, x^j \\ x^i &= \left[R^{-1}\right]^i{}_j\, y^j \end{aligned} \tag{44}$$

where $R^t R = 1$. Before proceeding, we need the explicit form of the rotation matrices $R^i{}_j(t)$.

10.1 Rotations

Let $R^i{}_j$ be an orthogonal transformation, so that $R^t R = 1$ and let the rotating coordinates y^i be related to inertial (and by definition non-rotating) coordinates $x^i(t) = x^i(0)$ by eqs.(44). After an infinitesimal time dt,

$$y^j(dt) = x^i + \delta x^i$$

where δx^i is proportional to dt. In Section 6.2.2 we showed that the change of x^i under such an infinitesimal rotation can be written as

$$\begin{aligned} \delta x^i &= -\varepsilon^i{}_{jk}\omega^j dt\; x^k \\ &= \beta^i{}_k\, x^k \end{aligned}$$

where the parameters ω^i are arbitrary. Setting

$$\omega^i dt = n^i \omega dt = n^i d\varphi$$

the magnitude ω and angle φ may be functions of time, but we assume the unit vector n^i is constant. Computing the effect of many infinitesimal rotations, we can find the effect of a finite rotation,

$$R(t) = \lim_{n\to\infty} (\delta + \beta d\varphi)^n$$

where

$$(\delta + \beta d\varphi)^n$$

means the n^{th} power of a matrix. Expanding this power using the binomial theorem,

$$\begin{aligned} R^i{}_j(t) &= \lim_{\substack{n\to\infty \\ nd\varphi\to\varphi}} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} (\delta)^{n-k} (\beta d\varphi)^k \\ &= \lim_{\substack{n\to\infty \\ nd\varphi\to\varphi}} \sum_{k=0}^{n} \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k k!} (\beta n d\varphi)^k \\ &\quad \lim_{\substack{n\to\infty \\ nd\varphi\to\varphi}} \sum_{k=0}^{n} \frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{k!} (\beta n d\varphi)^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} (\beta\varphi)^k \\ &= \exp(\beta\varphi) \end{aligned}$$

Finally, substituting the expression for the generator gives

$$R^i{}_k(t) = \exp\left(-\varepsilon^i{}_{jk} n^j \varphi\right) x^k$$

We can evaluate the exponential by finding powers of the Levi-Civita tensor.

$$\begin{aligned}\left[\left(\varepsilon^i_{jk} n^j \varphi\right)^2\right]^i{}_m &= \varphi^2 \varepsilon^i{}_{jk} n^j \varepsilon^k{}_{lm} n^l \\ &= \varphi^2 \left(\delta^i_l \delta_{jm} - \delta^i_m \delta_{jl}\right) n^j n^l \\ &= -\varphi^2 \left(\delta^i_m - n_m n^i\right)\end{aligned}$$

Then

$$\begin{aligned}\left[\left(\varepsilon^i_{jk} \omega^j\right)^3\right]^i{}_n &= -\varphi^3 \left(\delta^i_m - n_m n^i\right) \varepsilon^m{}_{jn} n^j \\ &= -\varphi^3 \left(\varepsilon^i{}_{jn} n^j - \varepsilon_{mjn} n^j n^m n^i\right) \\ &= -\varphi^3 \varepsilon^i{}_{jn} n^j\end{aligned}$$

This pattern repeats, so that we can immediately write the whole series. Separate the magnitude and direction of ω^i as $\omega^i = \omega n^i$. Then

$$\begin{aligned}R^i{}_k(t) &= \exp\left(-\varepsilon^i{}_{jk} n^j \varphi\right) \\ &= \sum_{n=0}^{\infty} \frac{1}{(2n)!} \left[\left(\varepsilon^i_{jk} n^j \varphi\right)^{2n}\right]^i{}_k \\ &\quad - \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} \left[\left(\varepsilon^i_{jk} n^j \varphi\right)^{2n+1}\right]^i{}_k \\ &= \delta^i_k + \sum_{n=1}^{\infty} \frac{1}{(2n)!} \left(-\varphi^2\right)^n \left(\delta^i_k - n^i n_k\right) \\ &\quad - \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} \left(-\varphi^2\right)^n \omega t \left(\varepsilon^i{}_{jk} n^j\right) \\ &= \delta^i_k - \left(\delta^i_k - n^i n_k\right) + \left(\delta^i_k - n^i n_k\right) \sum_{n=0}^{\infty} \frac{(-1)^n \varphi^{2n}}{(2n)!} \\ &\quad - \varepsilon^i{}_{jk} n^j \sum_{n=0}^{\infty} \frac{(-1)^n \varphi^{2n+1}}{(2n+1)!} \\ &= n^i n_k + \left(\delta^i_k - n^i n_k\right) \cos\varphi - \varepsilon^i{}_{jk} n^j \sin\varphi\end{aligned}$$

and the transformation of x^i is

$$\begin{aligned}y^i(t) &= R^i{}_k(t)\, x^k \\ &= n^i n_k x^k + \left(\delta^i_k - n^i n_k\right) x^k \cos\varphi - \varepsilon^i{}_{jk} n^j x^k \sin\varphi\end{aligned}$$

where φ is an arbitrary function of time. In vector notation,

$$\mathbf{y} = (\mathbf{n} \cdot \mathbf{x})\,\mathbf{n} + (\mathbf{x} - (\mathbf{n} \cdot \mathbf{x})\,\mathbf{n}) \cos\omega t - (\mathbf{n} \times \mathbf{x}) \sin\omega t$$

This is an entirely sensible result. The rotating vector has been decomposed into three mutually orthogonal directions,

$$\mathbf{n}, (\mathbf{x} - (\mathbf{n} \cdot \mathbf{x})\,\mathbf{n}), \mathbf{n} \times \mathbf{x}$$

which are, respectively, along $\mathbf{n}$, in the $\mathbf{nx}$ plane and orthogonal to $\mathbf{n}$, and orthogonal to both $\mathbf{x}$ and $\mathbf{n}$. The part of $\mathbf{y}$ parallel to $\mathbf{n}$ is equal to the projection of $\mathbf{x}$ parallel to $\mathbf{n}$ while the motion clearly rotates the plane orthogonal to $\mathbf{n}$. To see this more clearly, choose $\mathbf{n}$ in the z direction, $\mathbf{n} = \hat{\mathbf{k}}$. Then $\mathbf{y}$ is given by

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \cos\omega t & -\sin\omega t & 0 \\ \sin\omega t & \cos\omega t & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

corresponding to a right-handed rotation about the z-axis.

Show that for a general rotation $R^i{}_j(\omega, \mathbf{n}, t)$, $\mathbf{n}$ gives the axis of rotation while ωt is the angle of rotation at time t.

We will also need the rate of change of $R^i{}_j(\mathbf{n}, \varphi)$. From

$$R^i{}_k(t) = \exp\left(\varepsilon^i{}_{jk}\omega^j t\right)$$

we see that

$$\begin{aligned} \frac{d}{dt}R^i{}_k &= -\left[\exp\left(-\varepsilon^i{}_{jk}n^j\varphi\right)\right]^i{}_m \varepsilon^m{}_{jk}n^j \frac{d\varphi}{dt} \\ &= -R^i{}_m\varepsilon^m{}_{jk}n^j\omega \end{aligned}$$

Show that the velocity $\frac{d\mathbf{y}}{dt}$ is given by

$$\dot{\mathbf{y}} = R\left(\frac{d\mathbf{x}}{dt} - \omega \times \mathbf{x}\right)$$

Take another derivative to find $\ddot{\mathbf{y}}$.

Suppose both n^i and ω depend on time. Show that:

$$\begin{aligned} \frac{dy^i}{dt} &= R^i{}_m\left(\frac{dx^m}{dt} + \varepsilon^m{}_{jk}n^j\omega x^k\right) \\ &\quad + \left(\left(\frac{dn^i}{dt}n_k + n^i\frac{dn_k}{dt}\right)(1-\cos\varphi) + \varepsilon^i{}_{jk}\frac{dn^j}{dt}\sin\varphi\right)x^k \end{aligned}$$

10.2 The Coriolis theorem

Let's look at the metric and Christoffel symbol more closely. The metric term in the action is really the kinetic energy, and there may be more kinetic energy with y^i than meets the eye. For x^i we have

$$T = \frac{1}{2}m\delta_{ij}\dot{x}^i\dot{x}^j$$

Now, with y^i rotating, we get

$$\begin{aligned} \dot{x}^i &= \frac{dx}{dt} \\ &= \frac{d\left(\left[R^{-1}\right]^i{}_j y^j\right)}{dt} \\ &= \left[R^{-1}\right]^i{}_j\frac{dy^j}{dt} + \frac{d\left[R^{-1}\right]^i{}_j}{dt}y^j \end{aligned}$$

We show below that if the rotation is always about a fixed direction in space then

$$\begin{aligned} \frac{d}{dt}R^i{}_k &= -R^i{}_m\varepsilon^m{}_{jk}\omega^j \\ \frac{d\left[R^{-1}\right]^i{}_k}{dt} &= \left[R^{-1}\right]^i{}_m\varepsilon^m{}_{jk}\omega^j \end{aligned}$$

so

$$\begin{aligned}\dot{x}^i &= \left[R^{-1}\right]^i{}_j \dot{y}^j + \left[R^{-1}\right]^i{}_m \varepsilon^m{}_{jk}\omega^j y^k \\ &= \left[R^{-1}\right]^i{}_m \left(\dot{y}^m + \varepsilon^m{}_{jk}\omega^j y^k\right)\end{aligned}$$

The kinetic energy is therefore

$$\begin{aligned}T &= \frac{1}{2}m\delta_{ij}\dot{x}^i\dot{x}^j \\ &= \frac{1}{2}m\delta_{ij}\left[R^{-1}\right]^i{}_m\left[R^{-1}\right]^i{}_n\left(\dot{y}^m + \varepsilon^m{}_{jk}\omega^j y^k\right)\left(\dot{y}^n + \varepsilon^n{}_{jk}\omega^j y^k\right) \\ &= \frac{1}{2}m\delta_{mn}\left(\dot{y}^m + \varepsilon^m{}_{jk}\omega^j y^k\right)\left(\dot{y}^n + \varepsilon^n{}_{jk}\omega^j y^k\right)\end{aligned}$$

and this is *not* purely quadratic in the velocity! Nonetheless, the equation of motion is the Euler-Lagrange equation. Setting $L = T - V$, we have

$$\begin{aligned}\frac{d}{dt}m\delta_{mi}\left(\dot{y}^m + \varepsilon^m{}_{jk}\omega^j y^k\right) + \frac{\partial V}{\partial y^i} + m\delta_{mn}\left(\dot{y}^m + \varepsilon^m{}_{jk}\omega^j y^k\right)\left(\varepsilon^n{}_{li}\omega^l\right) &= 0 \\ m\left(\ddot{y}_i + \varepsilon_{ijk}\omega^j\dot{y}^k + \varepsilon_{ijk}\dot{\omega}^j y^k - \varepsilon_{mli}\omega^l\dot{y}^m - \varepsilon^m{}_{jk}\varepsilon_{mli}\omega^l\omega^j y^k\right) + \frac{\partial V}{\partial y^i} &= 0 \\ m\left(\ddot{y}_i + 2\varepsilon_{ijk}\omega^j\dot{y}^k + \varepsilon_{ijk}\dot{\omega}^j y^k - \varepsilon^m{}_{jk}\varepsilon_{mli}\omega^l\omega^j y^k\right) + \frac{\partial V}{\partial y^i} &= 0\end{aligned}$$

Writing the result in vector notation, we have

$$m\frac{d^2\mathbf{y}}{dt^2} + 2m\left(\omega\times\dot{\mathbf{y}}\right) + m\dot{\omega}\times\mathbf{y} + m\omega\times\left(\omega\times\mathbf{y}\right) = \mathbf{F}$$

Alternatively, we may write

$$m\frac{d^2\mathbf{y}}{dt^2} = \mathbf{F} - 2m\left(\omega\times\dot{\mathbf{y}}\right) - m\dot{\omega}\times\mathbf{y} - m\omega\times\left(\omega\times\mathbf{y}\right)$$

and think of the added terms on the right as effective forces. This is the *Coriolis theorem*.

The meaning of the various terms is easy to see. The last term on the right is the usual centrifugal force, acts radially away from the axis of rotation.

The second term on the right, called the *Coriolis force*, depends only on the velocity of the particle. It is largest for a particle moving radially toward or away from the axis of rotation. The particle experiences a change in radius, and at the new radius will be moving at the wrong speed for an orbit. For example, suppose we drop a particle from a height above the equator. At the initial moment, the particle is moving with the rotation of Earth, but as it falls, it is moving faster than the surface below it, and therefore overtakes the planet's surface. Since Earth rotates from west to east, the particle will fall to the east of a point directly below it.

The third term applies if the angular velocity is changing. Suppose it is increasing, $\dot{\omega}$ is in the same direction as ω. Then the particle will tend to be left behind in its rotational motion. If Earth were spinning up, this would give an acceleration to the west.

11 Inequivalent Lagrangians

One of the more startling influences of quantum physics on the study of classical mechanics is the realization that there exist inequivalent Lagrangians determining a given set of classical paths. Inequivalent Lagrangians for a given problem are those whose difference is *not* a total derivative. While it is not too surprising that a

given set of paths provides extremals for more than one functional, it is striking that some systems permit infinitely many Lagrangians for the same paths. There remain many open questions on this subject, with most of the results holding in only one dimension.

To begin our exploration of inequivalent Lagrangians, we describe classes of free particle Lagrangians and give some examples. Next we move to the theorems for 1-dim systems due to Yan, Kobussen and Leubner ([12], [13], [14], [15], [16]) including a simple example. Then we consider inequivalent Lagrangians in higher dimensions.

11.1 General free particle Lagrangians

There are distinct classes of Lagrangian even for free particle motion. We derive the classes and give an example of each, noting how Galilean invariance singles out the usual choice of Lagrangian.

The most general velocity dependent free particle Lagrangian is

$$S = \int f(v)dt$$

We assume the Cartesian form of the Euclidean metric, so that $v = \sqrt{\delta_{ij}v^i v^j}$. The equation of motion is

$$\frac{d}{dt}\frac{\partial f}{\partial v^i} = 0$$

so the conjugate momentum

$$p_i = \frac{\partial f}{\partial v^i} = f'\frac{v_i}{v}$$

is conserved. We need only solve this equation for the velocity. Separating the magnitude and direction, we have

$$\begin{aligned} \frac{v_i}{v} &= \frac{p_i}{p} \\ v &= g(p) \equiv [f']^{-1}(p) \end{aligned}$$

This solution is well-defined on any region in which the mapping between velocity and momentum is $1-1$. This means that velocity ranges may be any of four types: $v \in (0,\infty), (0,v_1), (v_1,v_2), (v_1,\infty)$. Which of the four types occurs depends on the singularities of $f'v^i/v$. Since v^i/v is a well-defined unit vector for all nonzero v_i, it is f' which determines the range. Requiring the map from v_i to p_i to be single valued and finite, we restrict to regions of f' which are monotonic. Independent physical ranges of velocity will then be determined by each zero or pole of f'. In general there will be $n+1$ such ranges:

$$v \in [0,v_1), (v_1,v_2), \ldots, (v_n,\infty)$$

if there are n singular points of f'. Of course it is possible that $v_1 = 0$ (so that on the lowest range, $(0,v_2)$, zero velocity is forbidden), or $v_1 = \infty$ so that the full range of velocities is allowed. Within any of these regions, the Hamiltonian formulation is well-defined and gives the same equations of motion as the Lagrangian formulation.

Thus, the motion for general f may be described as follows. Picture the space of all velocities divided into a number of spheres centered on the origin. The radii of these spheres are given by the roots and poles of f'. Between any pair of spheres, momentum and velocity are in $1-1$ correspondence and the motion is uniquely determined by the initial conditions. In these regions the velocity remains constant and the resulting motion is in a straight line. On spheres corresponding to zeros of f', the direction of motion is not determined by the equation of motion. On spheres corresponding to poles of f', no solutions exist. It is amusing to note that all three cases occur in practice. We now give an example of each.

First, consider the regular situation when f' is monotonic everywhere so the motion is uniquely determined to be straight lines for all possible initial velocities. The condition singles out the case of unconstrained

Newtonian mechanics. this is the only case that is Galilean invariant, since Galilean boosts require the full range of velocities, $v \in [0, \infty)$.

When f' has zeros, we have situations where a complete set of initial conditions is insufficient to determine the motion. Such a situation occurs in Lovelock, or extended, gravity, in which the action in d-dimensions (for d even) is a polynomial in the curvature tensor. The general Lovelock theory is the most general curved spacetime gravity theory in which the field equations depend on no higher than second derivatives of the metric [17]. In general, the field equations depend on powers of the second derivatives of the metric, whereas in general relativity this dependence is linear. Among the solutions are certain special cases called "geometrically free" [18]. These arise as follows. We may factor the action into a product which, schematically, takes the form

$$S = \int \prod_{k=0}^{d/2} \left(R^{a_k b_k}{}_{c_k d_k} - \frac{1}{2}\alpha_k \left(\delta^{a_k}_{c_k}\delta^{b_k}_{d_k} - \delta^{b_k}_{c_k}\delta^{a_k}_{d_k} \right) \right) \varepsilon_{a_1 b_1 \cdots a_{d/2} b_{d/2}} \varepsilon^{c_1 d_1 \cdots c_{d/2} d_{d/2}}$$

where $R^{ab}{}_{cd}$ is the curvature tensor and α_k are constants. Suppose that for all $k = 1, \ldots, n$ for some n in the range $2 < n < d/2$, we have

$$\alpha_k = \alpha$$

for some fixed value α. Then the variational equations all contain at least $n - 1$ factors of

$$R^{a_k b_k}{}_{c_k d_k} - \frac{1}{2}\alpha \left(\delta^{a_k}_{c_k}\delta^{b_k}_{d_k} - \delta^{b_k}_{c_k}\delta^{a_k}_{d_k} \right)$$

Therefore, if there is a subspace of dimension $m > d - n + 1$ of constant curvature

$$R^{a_k b_k}{}_{c_k d_k} = \frac{1}{2}\alpha \left(\delta^{a_k}_{c_k}\delta^{b_k}_{d_k} - \delta^{b_k}_{c_k}\delta^{a_k}_{d_k} \right)$$

for $a, b = 1, \ldots, m$, then the field equations are satisfied regardless of the metric on the complementary subspace. This is similar to the case of vanishing f', where the equation of motion is satisfied regardless of the direction of the velocity,

$$p_i = f' \frac{v_i}{v} \equiv 0$$

as long as v, but not v_i, is constant.

Finally, suppose f' has a pole at some value v_0. Then the momentum diverges and motion never occurs at velocity v_0. Of course, this is the case in special relativity, where the action of a free particle may be written as

$$\begin{aligned} S &= \int p_\alpha dx^\alpha \\ &= -\int E dt + p_i dx^i \\ &= -mc^2 \int \sqrt{1 - \frac{v^2}{c^2}}\, dt \end{aligned}$$

With $f(v) = -mc^2\sqrt{1 - \frac{v^2}{c^2}}$, we have

$$f' = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}}$$

with the well known pole in momentum at $v = c$.

11.2 Inequivalent Lagrangians

The existence of inequivalent Lagrangians for a given physical problem seems to trace back to Lie [19]. Dirac ([20],[21]) was certainly well aware of the ambiguities involved in passing between the Lagrangian and Hamiltonian formulations of classical mechanics. Later, others ([22],[23],[24],[25]), identified certain non-canonical transformations which nonetheless preserve certain Hamiltonians. A specific non-canonical transformation of the 2-dim harmonic oscillator is provided by Gelman and Saletan [26]. Bolza [27] showed that independent Lagrangians can give the same equations of motion, and a few years later, Kobussen [12], Yan ([13],[14]) and Okubo ([28],[29]) independently gave systematic developments showing that an infinite number of inequivalent Lagrangians exist for 2-dim mechanical systems. Shortly thereafter, Leubner [16] generalized and streamlined Yan's proof to include arbitrary functions of two constants of motion.

Leubner's result, the most general to date, may be stated as follows. Given any two constants of motion, (α, β), associated with the solution to a given 1-dim equation of motion, the solution set for any Lagrangian of the form

$$\begin{aligned} L(x, \dot{x}, t) &= \int_v^{\dot{x}} \frac{\dot{x} - v}{v} \left| \frac{\partial(\alpha, \beta)}{\partial(v, t)} \right| dv \\ &\quad + \int_{x_0}^{x} f(\tilde{x}, \nu, t) \frac{1}{v_0} \left| \frac{\partial(\alpha, \beta)}{\partial(v, t)} \right|_{\substack{\nu = \nu_0 \\ x = \tilde{x}}} d\tilde{x} + \frac{d\Omega}{dt} \end{aligned} \tag{45}$$

where $\left| \frac{\partial(\alpha,\beta)}{\partial(v,t)} \right|$ is the Jacobian, includes the same solutions locally. Notice that α and β are arbitrary constants of the motion – each may be an arbitrary function of simpler constants such as the Hamiltonian. We argue below that in 1-dim the solution sets are locally identical, though [16] provides no explicit proof. In higher dimensions there are easy counterexamples.

We illustrate a special case of this formula, of the form

$$L(x, v) = \dot{x} \int^{\dot{x}} \frac{K(x, v)}{v^2} dv \tag{46}$$

where K is any constant of the motion of the system. This expression is valid when the original Lagrangian has no explicit time dependence. Following Okubo [29], we prove that eq.(46) leads to the constancy of K. The result follows immediately from the Euler-Lagrange expression for L :

$$\begin{aligned} \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} &= \frac{d}{dt}\left(\int^{\dot{x}} \frac{K(x, v)}{v^2} dv + \dot{x} \frac{K(x, \dot{x})}{\dot{x}^2} \right) - \dot{x} \int^{\dot{x}} \frac{1}{v^2} \frac{\partial K(x, v)}{\partial x} dv \\ &= \frac{\ddot{x}}{\dot{x}} \frac{\partial K(x, \dot{x})}{\partial \dot{x}} + \frac{\partial K(x, \dot{x})}{\partial x} \\ &= \frac{1}{\dot{x}} \frac{dK(x, \dot{x})}{dt} \end{aligned}$$

Therefore, the Euler-Lagrange equation holds if and only if $K(x, \dot{x})$ is a constant of the motion.

The uniqueness in 1-dim follows from the fact that a single constant of the motion is sufficient to determine the solution curves up to the initial point. The uniqueness also depends on there being only a single Euler-Lagrange equation. These observations lead us to a higher dimensional result below.

It is interesting to notice that we can derive this form for L, but with K replaced by the Hamiltonian, by inverting the usual expression,

$$H = \dot{x} \frac{\partial L}{\partial \dot{x}} - L$$

for the Hamiltonian in terms of the Lagrangian. First, rewrite the right side as:

$$\begin{aligned} H &= \dot{x} \frac{\partial L}{\partial \dot{x}} - L \\ &= \dot{x}^2 \frac{\partial}{\partial \dot{x}} \left(\frac{L}{\dot{x}} \right) \end{aligned}$$

Now, dividing by $\dot{x}$ and integrating (regarding H as a function of the velocity) we find:

$$L = \dot{x} \int^{\dot{x}} \frac{H(x,v)}{v^2} dv$$

The remarkable fact is that the Hamiltonian may be replaced by any constant of the motion in this expression. Conversely, suppose we begin with the Lagrangian in terms of an arbitrary constant of motion, K, according to eq.(46),

$$L(x,v) = \dot{x} \int^{\dot{x}} \frac{K(x,v)}{v^2} dv$$

Then constructing the conserved energy,

$$\begin{aligned}
\tilde{E}(x,p) &= \dot{x}\frac{\partial L}{\partial \dot{x}} - L \\
&= \dot{x}\frac{\partial}{\partial \dot{x}}\left(\dot{x}\int^{\dot{x}} \frac{K(x,v)}{v^2} dv\right) - \dot{x}\int^{\dot{x}} \frac{K(x,v)}{v^2} dv \\
&= \dot{x}\left(\int^{\dot{x}} \frac{K(x,v)}{v^2} dv + \frac{K(x,\dot{x})}{\dot{x}}\right) - \dot{x}\int^{\dot{x}} \frac{K(x,v)}{v^2} dv \\
&= K(x,\dot{x})
\end{aligned}$$

we arrive at the chosen constant of motion! This proves the Gelman-Saletan-Currie conjecture [26]: any nontrivial time-independent constant of motion gives rise to a possible Hamiltonian. Proofs of the conjecture are due to Yan ([13],[14]) and Leubner [16].

The conjugate momentum to x constructed according to eq.(46) is

$$\begin{aligned}
\tilde{p} &= \frac{\partial L}{\partial \dot{x}} \\
&= \frac{\partial}{\partial \dot{x}}\left(\dot{x}\int^{\dot{x}} \frac{K(x,v)}{v^2} dv\right) \\
&= \int^{\dot{x}} \frac{K(x,v)}{v^2} dv + \frac{K(x,\dot{x})}{\dot{x}}
\end{aligned}$$

Of course, if $K = \frac{1}{2}m\dot{x}^2 + V$, both $\tilde{H}$ and $\tilde{p}$ reduce to the usual expressions.

The simple harmonic oscillator suffices to illustrate the method ([30],[31]). Since the energy, $E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2$, is a constant of the motion so is H^2, so we write

$$\begin{aligned}
L &= \frac{1}{4}\dot{x}\int^{\dot{x}} \frac{1}{v^2}\left(m^2v^4 + 2kmv^2x^2 + k^2x^4\right) dv \\
&= \frac{1}{12}m^2\dot{x}^4 + \frac{1}{2}km\dot{x}^2x^2 - \frac{1}{4}k^2x^4
\end{aligned}$$

The Euler-Lagrange equation resulting from L is

$$\begin{aligned}
0 &= \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} \\
&= \frac{d}{dt}\left(\frac{1}{3}m^2\dot{x}^3 + km\dot{x}x^2\right) - \left(km\dot{x}^2x - k^2x^3\right) \\
&= \left(m\ddot{x} + kx\right)\left(m\dot{x}^2 + kx^2\right)
\end{aligned}$$

Either of the two factors may be zero. Setting the first to zero is gives the usual equation for the oscillator, while setting the second to zero we find the same solutions in exponential form:

$$x = Ae^{i\omega t} + Be^{-i\omega t}$$

11.2.1 Are inequivalent Lagrangians equivalent?

Inequivalent Lagrangians have been defined as Lagrangians which lead to the same equations of motion but differ by more than a total derivative. For the simple case above, the cubic order equation of motion factors into the energy times the usual equation of motion, and setting either factor to zero gives the usual solution and only the usual solution. However, is this true in general? The Yan-Leubner proof shows that the new Lagrangian has the same solutions, but how do we know that none of the higher order Lagrangians introduces spurious solutions? The proofs do not address this question explicitly. If some of these Lagrangians introduce extra solutions, then they are not really describing the same motions.

Suppose we write,

$$L = v \int^{v} \frac{f(\alpha(x,\xi))}{\xi^2} d\xi$$

where α is any constant of the motion. Then we know that the Euler-Lagrange equation is satisfied by the usual equation of motion. But what *is* the Euler-Lagrange equation? We have shown that

$$\begin{aligned} \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} &= \frac{1}{\dot{x}} \frac{dK(x,\dot{x})}{dt} \\ &= \frac{1}{\dot{x}} f' \frac{d\alpha(x,\dot{x})}{dt} \end{aligned}$$

Setting this to zero, we have two types of solution:

$$\begin{aligned} f'(\alpha) &= 0 \\ \frac{d\alpha}{dt} &= 0 \end{aligned}$$

If spurious solutions could arise from motions with $f' = 0$, those motions would have to stay at the critical point, α_0 say, of f. But this means that $\alpha = \alpha_0$ remains constant. Therefore, the only way to introduce spurious solutions is if $\frac{d\alpha}{dt} = 0$ has solutions beyond the usual solutions. This may not be possible in one dimension. Finally, the inverse of the equation $\alpha(x,t) = \alpha_0$ may not exist at critical points, so the theorem must refer only to *local* equivalence of the solutions for inequivalent Lagrangians.

11.3 Inequivalent Lagrangians in higher dimensions

It is of interest to extend the results on inequivalent systems to higher dimension. Presumably, the theorems generalize in some way, but while one dimensional problems may be preferable "for simplicity" [16], this restricted case has many special properties that may not generalize. In any case, the method of proof of the Kobussen-Yan-Leubner theorem does not immediately generalize.

For 1-dim classical mechanics, there are only two independent constants of motion. The Kobussen-Yan-Leubner theorem, eq.(45), makes use of one or both to characterize the Lagrangian and, as noted above, one constant can completely determine the paths motion in 1-dim. The remaining constant is required only to specify the initial point of the motion. This leads to a simple conjecture for higher dimensions, namely, that the paths are in general determined by n of the $2n$ constants of motion. This is because n of the constants specify the initial position, while the remaining constants determine the paths.

We make these comments concrete with two examples. First, consider again the free particle in n-dim. The energy is

$$E = \frac{\mathbf{p}^2}{2m}$$

and we immediately find that a complete solution is characterized by the initial components of the momentum, p_{0i} and the initial position, x_{0i}. Clearly, knowledge of the momenta is necessary and sufficient to determine a set of flows. If we consider inequivalent Lagrangians,

$$L = v \int^{v} \frac{f(\xi)}{\xi^2} d\xi = F(v)$$

where

$$v = \sqrt{\mathbf{v}^2}$$

then the momenta

$$p_{i0} = \frac{\partial L}{\partial v^i} = F' \frac{v_i}{v}$$

comprise a set of first integrals of the motion. Inverting for the velocity

$$v^i = v^i \left(p_{i0}\right)$$

fixes the flow without fixing the initial point.

In general we will need at least this same set of relations, $v^i = v^i \left(p_{i0}\right)$, to determine the flow, though the generic case will involve n relations depending on $2n$ constants:

$$v^i = v^i \left(p_{i0}, x_0^i\right)$$

Notice that fewer relations do not determine the flow even for free motion in two dimensions. Thus, knowing only

$$v_x = \frac{p_{0x}}{m}$$

leaves the motion in the y direction fully arbitrary.

In an arbitrary number of dimensions, we find that expression for the energy in terms of the Lagrangian is still integrable as in the 1-dim case above, as long as $v = \sqrt{\mathbf{v}^2}$. If the Lagrangian does not depend explicitly on time, then energy is conserved. Then, letting $\hat{\theta}^i = \frac{\dot{x}^i}{v}$, we can still write the Lagrangian as an integral over Hamiltonian:

$$L\left(\mathbf{x}, v, \hat{\theta}_v\right) = v \int^v \frac{H\left(x, \xi, \hat{\theta}\right)}{\xi^2} d\xi + f\left(\mathbf{x}, \hat{\theta}_v\right)$$

where $f\left(\mathbf{x}, \vec{\theta}_v\right)$ is now necessary in order for L to satisfy the Euler-Lagrange equations. The integral term of this expression satisfies *one* of the Euler-Lagrange equations. If we now define a new Lagrangian by replacing H by an arbitrary, time-independent constant of the motion, $\alpha\left(x, v, \hat{\theta}\right)$,

$$\tilde{L} = v \int^v \frac{\alpha\left(x, \xi, \hat{\theta}\right)}{\xi^2} d\xi + f\left(x, \hat{\theta}\right)$$

then the new Lagrangian, $\tilde{L}$, still satisfies the same Euler-Lagrange equation,

$$\dot{x}^i \left(\frac{d}{dt} \frac{\partial \tilde{L}}{\partial \dot{x}^i} - \frac{\partial \tilde{L}}{\partial x^i} \right) = 0$$

We conjecture that for a suitable choice of f, $\tilde{L}$ provides an inequivalent Lagrangian, thereby providing *one* of the n relations required to specify the flow.

Part III
Conformal gauge theory

We now have the tools we need to describe the most elegant and powerful formulation of classical mechanics, as well as the starting point for quantum theory. While our treatment of Hamiltonian mechanics is non-relativistic, we begin with a relativistic treatment of the related conformal symmetry because it is from that perspective that the Hamiltonian itself most naturally arises. The development follows the steps we

took in the first Chapters of the book, constuction the arena and a law of motion from symmetry and variational principles. The new element that gives Hamiltonian dynamics its particular character is the choice of symmetry. Whereas Lagrangian theory arises as the gauge theory of Newton's second law when we generalize from Galilean symmetry to the diffeomorphism group, Hamiltonian mechanics arises by gauging the conformal group.

Because we want to gauge the conformal symmetry of spacetime, we begin with a discussion of special relativity. Then we proceed in the following stages. First, we return to contrast the Galilean symmetry of Newton's dynamical equation with the conformal symmetry of Newtonian measurement theory. Next, we build a new dynamical theory based on the full conformal symmetry, beginning with the construction of a new space to provide the arena, and continuing with the postulating of a dynamical law. Ultimately, we show the equivalence of the new formulation to the original second law and to Lagrangian dynamics.

After completing these constructions, we develop the properties of Hamilton mechanics.

12 Special Relativity

We begin our discussion of special relativity with a power point presentation, available on the website.

12.1 Spacetime

From the power point presentation, you know that spacetime is a four dimensional vector space with metric

$$\eta_{\alpha\beta} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

where the infinitesimal proper time τ and proper length s are given by

$$\begin{aligned} c^2 d\tau^2 &= c^2 dt^2 - dx^2 - dy^2 - dz^2 \\ ds^2 &= -c^2 dt^2 + dx^2 + dy^2 + dz^2 \end{aligned}$$

are agreed upon by all observers. The set of points at zero proper interval from a given point, P, is the light cone of that point. The light cone divides spacetime into regions. Points lying inside the light cone and having later time than P is the future. Points inside the cone with earlier times lie in the past of P. Points outside the cone are called elsewhere.

Timelike curves always lie inside the light cones of any of their points, while spacelike curves lie outside the lightcones of their points. The tangent vector at any point of a timelike curve point into the light cone and are timelike vectors, while the tangent to any spacelike curve is spacelike. The elapsed physical time experienced travelling along any timelike curve is given by integrating $d\tau$ along that curve. Similarly, the proper distance along any spacelike curve is found by integrating ds.Spacetime

We refer to the coordinates of an event in spacetime using the four coordinates

$$x^\alpha = (ct, x, y, z)$$

where $\alpha = 0, 1, 2, 3$. We may also write x^α in any of the following ways:

$$\begin{aligned} x^\alpha &= (ct, \mathbf{x}) \\ &= \left(ct, x^i\right) \\ &= \left(x^0, x^i\right) \end{aligned}$$

where $i = 1, 2, 3$. This neatly separates the familiar three spatial components of the vector x^α from the time component, allowing us to recognize familiar relations from non-relativistic mechanics.

Our most important tool is the invariant interval, expressed in either of the vector forms,

$$\begin{aligned} s^2 &= -c^2t^2 + x^2 + y^2 + z^2 \\ c^2\tau^2 &= c^2t^2 - \left(x^2 + y^2 + z^2\right) \end{aligned}$$

where s is proper distance and τ is proper time. These intervals are agreed upon by all observers.

These quadratic forms define a metric,

$$\eta_{\alpha\beta} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

so that we may write the invariant interval as

$$s^2 = \eta_{\alpha\beta} x^\alpha x^\beta$$

or infinitesimally

$$ds^2 = \eta_{\alpha\beta} dx^\alpha dx^\beta$$

A Lorentz transformation may be defined as any transformation of the coordinates that leaves the length-squared s^2 unchanged. It follows that

$$y^\alpha = \Lambda^\alpha{}_\beta x^\beta$$

is a Lorentz transformation if and only if

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

The set of points at zero proper interval, $s^2 = 0$, from a given point, P, is the light cone of that point. The light cone divides spacetime into regions. Points lying inside the light cone and having later time than P lie in the future of P. Points inside the cone with earlier times lie in the past of P. Points outside the cone are called elsewhere.

Timelike vectors from P connect P to past or future points. Timelike curves are curves whose tangent vector at any point $x^\alpha(\lambda)$ are timelike vectors at $x^\alpha(\lambda)$, while spacelike curves have tangents lying outside the lightcones of their points. The elapsed physical time experienced travelling along any timelike curve is given by integrating $d\tau$ along that curve. Similarly, the proper distance along any spacelike curve is found by integrating ds.

12.2 Relativistic dynamics

We now turn to look at motion in spacetime. Consider a particle moving along a world line. The path of the particle is a curve in spacetime, and we can write that curve parametrically:

$$x^\alpha = x^\alpha(\lambda)$$

Here λ can be any parameter that increases monotonically along the curve. Notice that two choices for λ that are sometimes convenient are the time coordinate, t, relative to our frame of reference, or the proper time τ experienced by the particle. The proper time is an excellent choice because it may be calculated once we know the coordinates of the particle in *any* frame of reference.

To calculate the proper time experienced along the world line of the particle between events A and B, just add up the infinitesimal displacements $d\tau$ along the path. Thus

$$\tau_{AB} = \int_A^B d\tau$$

$$
\begin{aligned}
&= \int_A^B \sqrt{dt^2 - \frac{1}{c^2}\left(dx^i\right)^2} \\
&= \int_{t_A}^{t_B} dt \sqrt{1 - \frac{1}{c^2}\left(\frac{dx^i}{dt}\right)^2} \\
&= \int_{t_A}^{t_B} dt \sqrt{1 - \frac{\mathbf{v}^2}{c^2}}
\end{aligned}
$$

where $\mathbf{v}^2$ is the usual squared magnitude of the 3-velocity. Notice that if $\mathbf{v}^2$ is ever different from zero, then τ_{AB} is *smaller* than the time difference $t_B - t_A$:

$$
\tau_{AB} = \int_{t_A}^{t_B} dt \sqrt{1 - \frac{\mathbf{v}^2}{c^2}} \leq \int_{t_A}^{t_B} dt = t_B - t_A
$$

Equality holds only if the particle remains at rest in the given frame of reference. This difference has been measured to high accuracy. One excellent test is to study the number of muons reaching the surface of the earth after being formed by cosmic ray impacts on the top of the atmosphere. These particles have a halflife on the order of 10^{-11} seconds, so they would normally travel only a few centimeters before decaying. However, because they are produced in a high energy collision that leaves them travelling toward the ground at nearly the speed of light, many of them are detected at the surface of the earth.

We next need a generalization of the velocity of the particle. We can get a direction in spacetime corresponding to the direction of the particle's motion by looking at the tangent vector to the curve,

$$
t^\alpha = \frac{dx^\alpha(\lambda)}{d\lambda}
$$

We can see that this tangent vector is closely related to the velocity by expanding with the chain rule,

$$
\begin{aligned}
t^\alpha &= \frac{dx^\alpha(\lambda)}{d\lambda} \\
&= \frac{dt}{d\lambda}\frac{dx^\alpha}{dt} \\
&= \frac{dt}{d\lambda}\frac{d}{dt}\left(ct, x^i\right) \\
&= \frac{dt}{d\lambda}\left(c, v^i\right)
\end{aligned}
$$

where v^i is the usual Newtonian 3-velocity. This is close to what we need, but since λ is arbitrary, so is $\frac{dt}{d\lambda}$. However, we can define a true vector by using the proper time as the parameter. Let the world line be parameterized by the elapsed proper time, τ, of the particle. Then define the *4-velocity,*

$$
u^\alpha = \frac{dx^\alpha(\tau)}{d\tau}
$$

Since the coordinates of the particle transform according to the Lorentz transformation

$$
\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & \frac{vx}{c^2}\gamma & & \\ \gamma v & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}
$$

or more simply,

$$
x'^\alpha = \Lambda^\alpha{}_\beta x^\beta
$$

and τ is invariant, we can find u'^{α} in a new frame of reference,

$$\begin{aligned} u'^{\alpha} &= \frac{dx'^{\alpha}(\tau')}{d\tau'} \\ &= \frac{d}{d\tau}\Lambda^{\alpha}{}_{\beta}x^{\beta}(\tau) \\ &= \Lambda^{\alpha}{}_{\beta}\frac{d}{d\tau}x^{\beta}(\tau) \\ &= \Lambda^{\alpha}{}_{\beta}u^{\beta} \end{aligned}$$

This shows that the 4-velocity is a 4-vector.

A very convenient form for the 4-velocity is given by our expansion of the tangent vector. Just as for general λ, we have

$$u^{\alpha} = \frac{dt}{d\tau}\left(c, v^{i}\right)$$

but now we know what the function in front is. Compute

$$\begin{aligned} d\tau &= \sqrt{dt^2 - \frac{1}{c^2}\left(dx^i\right)^2} \\ &= dt\sqrt{1 - \frac{\mathbf{v}^2}{c^2}} \end{aligned}$$

Then we see that

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \frac{\mathbf{v}^2}{c^2}}} = \gamma$$

Therefore,

$$u^{\alpha} = \gamma\left(c, v^{i}\right) \tag{47}$$

This is an extremely useful form for the 4-velocity. We shall use it frequently.

Since u^{α} is a 4-vector, its magnitude

$$\eta_{\alpha\beta}u^{\alpha}u^{\beta}$$

must be invariant! This means that the velocity of every particle in spacetime has the same particular value. Let's compute it:

$$\begin{aligned} \eta_{\alpha\beta}u^{\alpha}u^{\beta} &= -\left(u^0\right)^2 + \sum_i \left(u^i\right)^2 \\ &= -\gamma^2c^2 + \gamma^2\mathbf{v}^2 \\ &= \frac{-c^2 + \mathbf{v}^2}{1 - \frac{\mathbf{v}^2}{c^2}} \\ &= -c^2 \end{aligned}$$

This is indeed invariant! Our formalism is doing what it is supposed to do.

Now let's look at how the 4-velocity is related to the usual 3-velocity. If $\mathbf{v}^2 << c^2$, the components of the 4-velocity are just

$$u^{\alpha} = \gamma\left(c, v^{i}\right) \approx \left(c, v^{i}\right) \tag{48}$$

The speed of light, c, is just a constant, and the spatial components reduce to precisely the Newtonian velocity. This is just right. Moreover, it takes no new information to write the general form of u^{α} once we know v^{i} – there is no new information, just a different form.

From the 4-velocity it is natural to define the 4-momentum by multiplying by the mass,

$$\begin{aligned} p^\alpha &= mu^\alpha \\ &= m\gamma\left(c, v^i\right) \end{aligned}$$

As we might expect, the 3-momentum part of p^α is closely related to the Newtonian expression mv^i. In general it is

$$p^i = \frac{mv^i}{\sqrt{1-\frac{v^2}{c^2}}}$$

If $v << c$ we may expand the denominator to get

$$\begin{aligned} p^i &\approx mv^i\left(1+\frac{1}{2}\frac{v^2}{c^2}+\cdots\right) \\ &\approx mv^i \end{aligned}$$

Thus, while relativistic momentum differs from Newtonian momentum, they only differ at order $\frac{v^2}{c^2}$. Even for the 7 $mi/\sec$ velocity of a spacecraft which escapes Earth's gravity this ratio is only

$$\frac{v^2}{c^2} = 1.4 \times 10^{-9}$$

so the Newtonian momentum is correct to parts per billion. In particle accelerators, however, where near-light speeds are commonplace, the difference is substantial (see exercises).

Now consider the remaining component of the 4-momentum. Multiplying by c and expanding γ we find

$$\begin{aligned} p^0 c &= mc^2\gamma \\ &= mc^2\left(1+\frac{1}{2}\frac{v^2}{c^2}+\frac{3}{8}\frac{v^4}{c^4}\cdots\right) \\ &\approx mc^2+\frac{1}{2}mv^2+\frac{3}{8}mv^2\frac{v^2}{c^2} \end{aligned}$$

The third term is negligible at ordinary velocities, while we recognize the second term as the usual Newtonian kinetic energy. We therefore identify $E = p^0 c$. Since the first term is constant it plays no measurable role in classical mechanics but it suggests that there is intrinsic energy associated with the mass of an object. This conjecture is confirmed by observations of nuclear decay. In such decays, the mass of the initial particle is greater than the sum of the masses of the product particles, with the energy difference

$$\Delta E = m_{initial}c^2 - \sum m_{final}c^2$$

correctly showing up as kinetic energy.

Suppose a muon is produced in the upper atmosphere moving downward at $v = .99c$ relative to the surface of Earth. If it decays after a proper time $\tau = 2.2 \times 10^{-6}$ seconds, how far would it travel if there were no time dilation? Would it reach Earth's surface? How far does it actually travel relative to Earth? Note that many muons are seen reaching Earth's surface.

A free neutron typically decays into a proton, an electron, and an antineutrino. How much kinetic energy is shared by the final particles?

Suppose a proton at Fermilab travels at $.99c$. Compare Newtonian energy, $\frac{1}{2}mv^2$ to the relativistic energy $p^0 c$.

12.3 Acceleration

Next we consider acceleration. We define the acceleration 4-vector to be the proper-time rate of change of 4-velocity,

$$\begin{aligned} a^\alpha &= \frac{du^\alpha}{d\tau} \\ &= \frac{dt}{d\tau}\frac{d\left(\gamma\left(c, v^i\right)\right)}{dt} \\ &= \gamma\left(-\frac{1}{2}\gamma^3\left(-2\frac{v^m a_m}{c^2}\right)\left(c, v^i\right) + \gamma\left(0, a^i\right)\right) \\ &= \frac{v^m a_m}{c^2}\gamma^3 u^\alpha + \gamma^2\left(0, a^i\right) \end{aligned}$$

Is this consistent with our expectations? We know, for instance, that

$$u^\alpha u_\alpha = -c^2$$

which means that

$$\begin{aligned} 0 &= \frac{d}{d\tau}\left(-c^2\right) \\ &= 2\frac{du^\alpha}{d\tau}u_\alpha \end{aligned}$$

Therefore, the 4-velocity and 4-acceleration are orthogonal, which we easily verify directly,

$$\begin{aligned} u^\alpha a_\alpha &= \gamma\left(c, v^i\right)\cdot\left(\frac{v^m a_m}{c^2}\gamma^3 u^\alpha + \gamma^2\left(0, a^i\right)\right) \\ &= -v^m a_m\gamma^3 + \gamma^3 a^i v_i \\ &= 0 \end{aligned}$$

Now compute $a^\alpha a_\alpha$:

$$\begin{aligned} a^\alpha a_\alpha &= \left(\frac{v^m a_m}{c^2}\gamma^3 u^\alpha + \gamma^2\left(0, a^i\right)\right)a_\alpha \\ &= \gamma^2\left(0, a^i\right)a_\alpha \\ &= \gamma^2\left(0, a^i\right)\cdot\left(\frac{v^m a_m}{c^2}\gamma^4\left(c, v_i\right) + \gamma^2\left(0, a_i\right)\right) \\ &= \frac{v^m a_m}{c^2}\gamma^6 a^i v_i + \gamma^4 a^i a_i \\ &= \gamma^4\left(a^i a_i + \gamma^2\frac{\left(v^m a_m\right)^2}{c^2}\right) \end{aligned}$$

This expression gives the acceleration of a particle moving with relative velocity v^i when the acceleration in the instantaneous rest frame of the particle is given by the $v^i = 0$ expression

$$a^\alpha a_\alpha = a^i a_i$$

We consider two cases. First, suppose v^i is parallel to a^i. Then since $a^\alpha a_\alpha$ is invariant, the 3-acceleration is given by

$$a^\alpha a_\alpha = \gamma^6 a^i a_i$$

or

$$a^i a_i = a^\alpha a_\alpha \left(1 - \frac{v^2}{c^2}\right)^3$$

where $a^\alpha a_\alpha$ is independent of v^i. Therefore, as the particle nears the speed of light, its apparent 3-acceleration decreases dramatically. When the acceleration is orthogonal to the velocity, the exponent is reduced,

$$a^i a_i = a^\alpha a_\alpha \left(1 - \frac{v^2}{c^2}\right)^2$$

12.4 Equations of motion

The relativistic action for a particle in a potential is surprisingly simple. To derive a suitable equation of motion, we once again start with arc length. Suppose we have a timelike curve $x^\alpha(\lambda)$. Then distance along the curve is given by

$$\tau = -\frac{1}{c^2}\int \sqrt{(-v^\alpha v_\alpha)}d\lambda$$

where

$$v^\alpha = \frac{dx^\alpha}{d\lambda}$$

Since the integral is reparameterization invariant, there is no loss of generality if we use the 4-velocity in place of v^a and write

$$\tau_C = -\frac{1}{c^2}\int_C \sqrt{(-u^\alpha u_\alpha)}d\tau$$

Then the path of extremal proper time is given by the Euler-Lagrange equation

$$\frac{d}{d\tau}\frac{\partial}{\partial u^\beta}\left(-\frac{1}{c^2}u^\alpha u_\alpha\right) = 0$$

that is, vanishing 4-acceleration,

$$\frac{du^\alpha}{d\tau} = 0$$

Show that vanishing 4-velocity implies vanishing 3-velocity.

12.5 Relativistic action with a potential

We can easily generalize this expression to include a potential. For relativistic problems it is possible to keep the action reparameterization invariant. To do so, we must multiply the line element by a function instead of adding the function. This gives

$$\tau_C = \frac{1}{c}\int_C \phi\sqrt{(-u^\alpha u_\alpha)}d\tau$$

The Euler-Lagrange equation is

$$\begin{aligned}\frac{d}{d\tau}\left(-\phi\left(-u^\alpha u_\alpha\right)^{-1/2}u_\alpha\right) - \left(-u^\alpha u_\alpha\right)^{1/2}\frac{\partial\phi}{\partial x^\alpha} &= 0\\ \frac{1}{c^2}\frac{d}{d\tau}\left(\phi u_\alpha\right) + \frac{\partial\phi}{\partial x^\alpha} &= 0\end{aligned}$$

where we have simplified using the normalization condition $c = \left(-u^\alpha u_\alpha\right)^{1/2}$. Expanding the derivatives, and rearranging,

$$\begin{aligned}0 &= \frac{1}{c^2}\frac{du_\alpha}{d\tau}\phi + \frac{1}{c^2}u_\alpha\frac{d\phi}{d\tau} + \frac{\partial\phi}{\partial x^\alpha}\\ &= \frac{1}{c^2}\frac{du_\alpha}{d\tau}\phi + \left(\frac{1}{c^2}u_\alpha u^\beta + \delta^\beta_\alpha\right)\frac{\partial\phi}{\partial x^\beta}\end{aligned}$$

Notice that

$$P_\alpha{}^\beta = \delta_\alpha^\beta + \frac{1}{c^2} u_\alpha u^\beta$$

is a projection operator.

A projection operator is an operator which is idempotent, that is, it is its own square. Show that $P^\beta{}_\alpha$ is a projection operator by showing that

$$P_\alpha{}^\mu P_\mu{}^\beta = P_\alpha{}^\beta$$

and show that $u^\alpha P_\alpha{}^\beta = 0$ and $P_\beta{}^\alpha u_\alpha = 0$.

Now we have

$$0 = \frac{1}{c^2}\frac{du_\alpha}{d\tau}\phi + P_\alpha{}^\beta \frac{\partial \phi}{\partial x^\beta}$$

This projection operator is necessary because the acceleration term is orthogonal to u^α. Dividing by ϕ, we see that

$$\frac{1}{c^2}\frac{du_\alpha}{d\tau} = -P_\alpha{}^\beta \frac{\partial \ln \phi}{\partial x^\beta}$$

If we identify

$$\phi = \exp\left(\frac{V}{mc^2}\right)$$

then we arrive at the desired equation of motion

$$m\frac{du_\alpha}{d\tau} = -P_\alpha{}^\beta \frac{\partial V}{\partial x^\beta}$$

which now is seen to follow as the extremum of the functional

$$S\left[x^a\right] = \frac{1}{c}\int_C e^{\frac{V}{mc^2}} \left(-u^\alpha u_\alpha\right)^{1/2} d\tau \tag{49}$$

See the exercises for other ways of arriving at this result.

It is suggestive to notice that the integrand is simply the usual line element multiplied by a scale factor.

$$d\sigma^2 = \frac{1}{c^2} e^{\frac{2V}{mc^2}} \left(-u^\alpha u_\alpha\right) d\tau^2 = -\frac{1}{c^2} e^{\frac{2V}{mc^2}} \eta_{\alpha\beta} dx^\alpha dx^\beta$$

This is called a *conformal line element* because it is formed from a metric which is related to the flat space metric by a conformal factor, $e^{\frac{V}{mc^2}}$,

$$g_{\alpha\beta} = e^{\frac{2V}{mc^2}} \eta_{\alpha\beta}$$

Conformal transformations will appear again when we study Hamiltonian mechanics.

We can generalize the action further by observing that the potential is the integral of the force along a curve,

$$V = -\int_C F_\alpha dx^\alpha$$

The potential is defined only when this integral is single valued. By Stoke's theorem, this occurs if and only if the force is curl-free. But even for general forces we can write the action as

$$S\left[x^a\right] = \frac{1}{c}\int_C e^{-\frac{1}{mc^2}\int_C F_\alpha dx^\alpha} \left(-u^\alpha u_\alpha\right)^{1/2} d\tau$$

In this case, variation leads to

$$\begin{aligned}
0 &= \frac{1}{c}\int_C \left(e^{-\frac{1}{mc^2}\int_C F_\beta dx^\beta} \left(-u^\mu u_\mu\right)^{-1/2} \left(-u_\alpha \delta u^\alpha\right)\right) d\tau \\
&\quad -\frac{1}{c}\int_C \left(e^{-\frac{1}{mc^2}\int_C F_\nu dx^\nu} \frac{1}{mc^2} \left(-u^\mu u_\mu\right)^{1/2} \delta x^\alpha \frac{\partial}{\partial x^\alpha}\int_C F_\beta dx^\beta\right) d\tau \\
&= \frac{1}{c}\int_C \left(\frac{d}{d\tau}\left(\frac{u_\alpha}{c} e^{-\frac{1}{mc^2}\int_C F_\alpha dx^\alpha}\right) - \frac{1}{mc^2} e^{\frac{1}{mc^2}\int_C F_\beta dx^\beta} F_\alpha c\right) \delta x^\alpha d\tau
\end{aligned}$$

The equation of motion is

$$0 = \left(\frac{du_\alpha}{d\tau} - \frac{1}{mc^2} u_\alpha u^\beta \frac{\partial}{\partial x^\beta} \int_C F_\alpha dx^\alpha - \frac{1}{m} F_\alpha \right) e^{\frac{1}{mc^2} \int_C F_\beta dx^\beta}$$

and therefore,

$$m \frac{du_\alpha}{d\tau} = P_\alpha \ {}^\beta F_\beta$$

This time the equation holds for an arbitrary force.

Finally, consider the non-relativistic limit of the action. If $v << c$ and $V << mc^2$ then to lowest order,

$$\begin{aligned} S[x^a] &= \int_C e^{\frac{V}{mc^2}} d\tau \\ &= \int_C \left(1 + \frac{V}{mc^2} \right) \frac{1}{\gamma} dt \\ &= \frac{1}{mc^2} \int_C \left(mc^2 + V \right) \sqrt{1 - \frac{v^2}{c^2}} dt \\ &= -\frac{1}{mc^2} \int_C \left(mc^2 \left(-1 + \frac{v^2}{2c^2} \right) - V \right) dt \\ &= -\frac{1}{mc^2} \int_C \left(-mc^2 + \frac{1}{2} mv^2 - V \right) dt \end{aligned}$$

Discarding the multiplier and irrelevant constant mc^2 in the integral we recover

$$S_{Cl} = \int_C \left(\frac{1}{2} mv^2 - V \right) dt$$

Since the conformal line element is a more fundamental object than the classical action, this may be regarded as another derivation of the classical form of the Lagrangian, $L = T - V$.

Consider the action

$$S[x^a] = \int \left(mu^\alpha u_\alpha + \phi \right) d\tau$$

This is no longer reparameterization invariant, so we need an additional Lagrange multiplier term to enforce the constraint,

$$+\lambda \left(u^\alpha u_\alpha + c^2 \right)$$

so the action becomes

$$S_1[x^a] = \int \left(mu^\alpha u_\alpha + c^2 \phi + \lambda \left(u^\alpha u_\alpha + c^2 \right) \right) d\tau$$

1. Write the Euler-Lagrange equations (including the one arising from the variation of λ).

2. The constraint implies $u^\alpha \frac{du_\alpha}{d\tau} = 0$. Solve for λ by contracting the equation of motion with u^α, using $u^\alpha \frac{du_\alpha}{d\tau} = 0$, and integrating. You should find that

$$\lambda = -\frac{1}{2} \left(\phi + a \right)$$

3. Substitute λ back into the equation of motion and show that the choice

$$\ln \left(\frac{\phi - 2m + a}{\phi_0 - 2m + a} \right) = \frac{1}{mc^2} V$$

gives the correct equation of motion.

4. Show, using the constraint freely, that S_1 is a multiple of the action of eq.(49).

Consider the action

$$S_2\left[x^a\right] = \int \left(mc^2 \sqrt{1 - \frac{v^2}{c^2}} + V \right) dt$$

1. Show that S_2 has the correct low-velocity limit, $L = T - V$.
2. Show that the Euler-Lagrange equation following from the variation of S_2 is *not* covariant. S_2 is therefore unsatisfactory.

Consider the action

$$S_3\left[x^a\right] = \int \left(mu^\alpha u_\alpha - 2V \right) d\tau$$

1. Show that the Euler-Lagrange equation for S_3 is

$$\frac{d}{d\tau} m u_\alpha = -\frac{\partial V}{\partial x^\alpha}$$

2. Show that the constraint, $u^\alpha u_\alpha = -c^2$ is not satisfied for general potentials V.
3. Show that S_3 has the wrong low-velocity limit.

Find the condition for the existence of a function $U\left(x^\alpha\right)$ such that

$$F^\alpha = \eta^{\alpha\beta} \frac{\partial U}{\partial x^\beta}$$

Find the condition for existence of a function $V\left(x^\alpha\right)$ satisfing

$$F^\alpha = P^{\alpha\beta} \frac{\partial U}{\partial x^\beta}$$

where

$$P^{\alpha\beta} = \eta^{\alpha\beta} + \frac{1}{c^2} u^\alpha u^\beta$$

is a projection operator.

13 The symmetry of Newtonian mechanics

Recall the distinction between the Newtonian dynamical law

$$F^i = ma^i$$

and Newtonian measurement theory, which centers on ratios of inner products of vectors,

$$\langle \mathbf{u}, \mathbf{v} \rangle = g_{mn} u^m v^n$$

or equivalently, on ratios of infinitesimal line elements,

$$ds^2 = g_{mn} dx^m dx^n$$

While the second law is invariant under the global Galilean transformations of eq.(33)

$$x^i\left(\mathbf{q}, t\right) = M^i{}_j q^j + a^i + v^i t$$

the inner products of vectors are invariant under local rotations while line elements are invariant under diffeomorphisms. When we consider transformations preserving ratios of line elements, we are led to the conformal group.

To construct the conformal group, notice that the ratio of two line elements (evaluated at a given point in space) is unchanged if both elements are multiplied by a common factor. Thus, is does not matter whether we measure with the metric g_{mn} or a multiple of g_{mn},

$$\tilde{g}_{mn} = e^{2\phi} g_{mn}$$

We will consider only positive multiples, but ϕ may be a function of position. Clearly, any transformation that preserves g_{mn} also preserves g_{mn} up to an overall factor, so the conformal group includes all of the Euclidean transformations. In addition, multiplying all dynamical variables with units of $(length)^n$ by the n^{th} power of a given constant λ will produce only an overall factor. Such transformatoins are called *dilatations*. For example, if x^i is the position of a particle then x^i is replaced by λx^i. For a force F^i we first need to express the units of F^i as a power of length. To do this we need two standards of measurement. Thus, beginning with MKS units, we can use one constant, $v_0 = 1m/\sec$, with units of velocity to convert seconds to meters, and a second constant, $h_0 = 1kg\text{-}m^2/\sec$ so that v_0/h_0 converts kilograms to inverse meters. Any other choice is equally valid. Using these constants, F^i is re-expressed as $F^i/h_0 v_0^2$ and thus measured in

$$\frac{kgm}{\sec^2}\frac{\sec^2}{kgm^4} = \frac{1}{m^3}$$

We therefore replace F^i by F^i/λ^3. In this way we may express all physical variables with powers of length only. In classical mechanics the constants h_0 and v_0 drop out of all physical predictions.

In addtion to these seven transformations – three rotations, three translations, and three dilatations – there are three more transformations which are conformal. In the next section we find all conformal transformations systematically.

13.1 The conformal group of Euclidean 3-space

We may find all conformal transformations by starting with the Euclidean-Cartesian metric, δ_{ij}, and finding all infinitesimal transformations

$$y^i = x^i + \varepsilon^i(x)$$

that preserve the line element up to an overall factor,

$$\delta_{ij} dy^i dy^j = e^{\phi} \delta_{ij} dx^i dx^j$$

The small functions ε^i are then infinitesimal generators of the conformal transformations. Keeping only first order terms, they must satisfy

$$\begin{aligned}
\delta_{ij}\left(dx^i + \partial_k \varepsilon^i dx^k\right)\left(dx^j + \partial_m \varepsilon^j dx^m\right) &= (1+\phi)\,\delta_{ij} dx^i dx^j \\
\delta_{ij}\left(\partial_m \varepsilon^j dx^i dx^m + \partial_k \varepsilon^i dx^k dx^j\right) &= \phi \delta_{ij} dx^i dx^j \\
\partial_j \varepsilon_i + \partial_i \varepsilon_j &= \phi \delta_{ij}
\end{aligned}$$

We therefore need solutions to

$$\partial_j \varepsilon_i + \partial_i \varepsilon_j - \phi \delta_{ij} = 0$$

Notice that when $\phi = 0$, these are just the generators of the Euclidean group. To see their form, consider

$$\partial_j \varepsilon_i + \partial_i \varepsilon_j = 0$$

Take another derivative and cycle the indices:

$$\begin{aligned}
\partial_k \partial_j \varepsilon_i + \partial_k \partial_i \varepsilon_j &= 0 \\
\partial_j \partial_i \varepsilon_k + \partial_j \partial_k \varepsilon_i &= 0 \\
\partial_i \partial_k \varepsilon_j + \partial_i \partial_j \varepsilon_k &= 0
\end{aligned}$$

Add the first two and subtract the third, then use the commutation of second partials,

$$\begin{aligned}\partial_k\partial_j\varepsilon_i + \partial_k\partial_i\varepsilon_j + \partial_j\partial_i\varepsilon_k + \partial_j\partial_k\varepsilon_i - \partial_i\partial_k\varepsilon_j - \partial_i\partial_j\varepsilon_k &= 0\\ 2\partial_j\partial_k\varepsilon_i &= 0\end{aligned}$$

Therefore, ε_i has vanishing second derivative and must be linear in x^i :

$$\varepsilon_i = a_i + b_{ij}x^j$$

Substituting into the condition for homogeneous solutions,

$$0 = \partial_j\varepsilon_i + \partial_i\varepsilon_j = b_{ji} + b_{ij}$$

so that b_{ij} must be antisymmetric. Clearly a_i generates translations and b_{ij} generates rotations.

Now we return to the conformal case. First, take the trace:

$$\begin{aligned}\delta^{ji}\left(\partial_j\varepsilon_i + \partial_i\varepsilon_j\right) - 3\phi &= 0\\ \partial^i\varepsilon_i &= \frac{3}{2}\phi\end{aligned}$$

so we must solve

$$\partial_j\varepsilon_i + \partial_i\varepsilon_j - \frac{2}{3}\partial^m\varepsilon_m\delta_{ij} = 0$$

Cycle the same way as before to get

$$\begin{aligned}0 &= \partial_k\partial_j\varepsilon_i + \partial_k\partial_i\varepsilon_j - \frac{2}{3}\partial_k\partial^m\varepsilon_m\delta_{ij}\\ 0 &= \partial_j\partial_i\varepsilon_k + \partial_j\partial_k\varepsilon_i - \frac{2}{3}\partial_j\partial^m\varepsilon_m\delta_{ik}\\ 0 &= \partial_i\partial_k\varepsilon_j + \partial_i\partial_j\varepsilon_k - \frac{2}{3}\partial_i\partial^m\varepsilon_m\delta_{kj}\end{aligned}$$

so that

$$\begin{aligned}0 &= \partial_k\partial_j\varepsilon_i + \partial_k\partial_i\varepsilon_j - \frac{2}{3}\partial_k\partial^m\varepsilon_m\delta_{ij}\\ &\quad +\partial_j\partial_i\varepsilon_k + \partial_j\partial_k\varepsilon_i - \frac{2}{3}\partial_j\partial^m\varepsilon_m\\ &\quad \delta_{ik} - \partial_i\partial_k\varepsilon_j - \partial_i\partial_j\varepsilon_k + \frac{2}{3}\partial_i\partial^m\varepsilon_m\delta_{kj}\\ \partial_j\partial_k\varepsilon_i &= \frac{1}{3}\partial_k\partial^m\varepsilon_m\delta_{ij} + \frac{1}{3}\partial_j\partial^m\varepsilon_m\delta_{ik} - \frac{1}{3}\partial_i\partial^m\varepsilon_m\delta_{kj}\end{aligned}$$

Now take a trace on jk

$$\partial^k\partial_k\varepsilon_i = -\frac{1}{3}\partial_i\left(\partial^m\varepsilon_m\right)$$

Now look at the third derivative:

$$\partial_n\partial_j\partial_k\varepsilon_i = \frac{1}{3}\partial_n\partial_k\partial^m\varepsilon_m\delta_{ij} + \frac{1}{3}\partial_n\partial_j\partial^m\varepsilon_m\delta_{ik} - \frac{1}{3}\partial_n\partial_i\partial^m\varepsilon_m\delta_{kj}$$

and trace nj :

$$\begin{aligned}\partial^j\partial_j\partial_k\varepsilon_i &= \frac{1}{3}\partial_i\partial_k\partial^m\varepsilon_m + \frac{1}{3}\partial^j\partial_j\partial^m\varepsilon_m\delta_{ik} - \frac{1}{3}\partial_k\partial_i\partial^m\varepsilon_m\\ \partial_k\partial^j\partial_j\varepsilon_i &= \frac{1}{3}\partial^j\partial_j\partial^m\varepsilon_m\delta_{ik}\\ -\frac{1}{3}\partial_k\partial_i\left(\partial^m\varepsilon_m\right) &= \frac{1}{3}\partial^j\partial_j\partial^m\varepsilon_m\delta_{ik}\\ &= -\frac{1}{9}\partial^m\partial_m\partial^n\varepsilon_n\delta_{ik}\\ \partial_k\partial_i\left(\partial^m\varepsilon_m\right) &= \frac{1}{3}\partial^m\partial_m\partial^n\varepsilon_n\delta_{ik}\end{aligned}$$

Substitute into the third derivative equation,

$$\partial_n\partial_j\partial_k\varepsilon_i = \frac{1}{9}\partial^m\partial_m\partial^s\varepsilon_s\left(\delta_{nk}\delta_{ij} + \delta_{nj}\delta_{ik} - \delta_{ni}\delta_{kj}\right)$$

Now trace nk,

$$\partial^k\partial_j\partial_k\varepsilon_i = \frac{1}{3}\partial^m\partial_m\partial^s\varepsilon_s\delta_{ij}$$

Compare with the contraction on jk :

$$\begin{aligned}\partial_n\partial^k\partial_k\varepsilon_i &= \frac{1}{9}\partial^m\partial_m\partial^s\varepsilon_s\left(\delta_{ni} + \delta_{ni} - 3\delta_{ni}\right)\\ &= -\frac{1}{9}\partial^m\partial_m\partial^s\varepsilon_s\delta_{ni}\end{aligned}$$

Together these show that

$$\begin{aligned}\partial_j\partial^k\partial_k\varepsilon_i &= 0\\ \partial^m\partial_m\partial^s\varepsilon_s &= 0\end{aligned}$$

Substituting into the triple derivative,

$$\partial_n\partial_j\partial_k\varepsilon_i = 0$$

Now we can write ε_i as a general quadratic expression in x^i,

$$\varepsilon_i = a_i + b_{ij}x^j + c_{ijk}x^jx^k$$

where we require $c_{ijk} = c_{ikj}$. Substituting into the original equation we have:

$$\begin{aligned}0 &= \partial_j\varepsilon_i + \partial_i\varepsilon_j - \frac{2}{3}\partial^k\varepsilon_k\delta_{ij}\\ &= b_{ij} + 2c_{ijk}x^k + b_{ji} + 2c_{jik}x^k - \frac{2}{3}\left(b^m{}_m + 2c^m{}_{mk}x^k\right)\delta_{ij}\end{aligned}$$

Therefore, separating the x-dependent terms from the constant terms, we must have, for all x^i,

$$\begin{aligned}0 &= b_{ij} + b_{ji} - \frac{2}{3}b^m{}_m\delta_{ij}\\ 0 &= 2c_{ijk}x^k + 2c_{jik}x^k - \frac{2}{3}2c^m{}_{mk}x^k\delta_{ij}\end{aligned}$$

The first,

$$b_{ij} + b_{ji} = \frac{2}{3}b^m{}_m\delta_{ij}$$

shows that the symmetric part of b_{ij} is determined purely by the trace $b = b^m{}_m$, while the antisymmetric part is arbitrary. If we let $\varepsilon_{ij} = -\varepsilon_{ji}$ be an arbitrary, constant, antisymmetric matrix then

$$b_{ij} = \varepsilon_{ij} + \frac{1}{3}b\delta_{ij}$$

For the equation involving c_{ijk}, we can strip off the x^k to write

$$c_{ijk} + c_{jik} = \frac{2}{3}c^m{}_{mk}\delta_{ij}$$

Once again cycling the indices, adding and subtracting:

$$\begin{aligned} c_{ijk} + c_{jik} &= \frac{2}{3} c^m{}_{mk} \delta_{ij} \\ c_{jki} + c_{kji} &= \frac{2}{3} c^m{}_{mi} \delta_{jk} \\ c_{kij} + c_{ikj} &= \frac{2}{3} c^m{}_{mj} \delta_{ik} \end{aligned}$$

so adding the first two and subtracting the third,

$$2c_{jik} = \frac{2}{3} \left(c^m{}_{mk} \delta_{ij} + c^m{}_{mi} \delta_{jk} - c^m{}_{mj} \delta_{ik} \right)$$

Now letting $c_k = c^m{}_{mk}$ we see that the full infinitesimal transformation may be written as

$$\begin{aligned} \varepsilon_i &= a_i + b_{ij} x^j + c_{ijk} x^j x^k \\ &= a_i + \varepsilon_{ij} x^j + \frac{1}{3} b x_i + \frac{1}{3} \left(2c_k x^k x_i - c_i x^2 \right) \end{aligned} \tag{50}$$

We know that a_i is a translation and it is not hard to see that ε_{ij} generates a rotation. In addition we have four more independent transformations, with parameters b and c_i. We could find the form of the transformations these produce by compounding many infinitesimal transformations into a finite one, but if we can just find a 4-parameter family of conformal transformations that are neither translations nor rotations, then we must have all the transformations that exist. One of the necessary transformations, called a dilatation, is trivial – simply multiply x^i by a positive constant:

$$y^i = e^\lambda x^i$$

Keeping the parameter positive keeps us from simulataneously performing a reflection, so these transformations are continuously connected to the identity.

The remaining four transformations follow from a simple observation. Define an *inversion* of x^i to be

$$y^i = \frac{x^i}{x^2}$$

Computing the effect of an inversion on the line element, we have

$$\begin{aligned} dy^i &= \frac{dx^i}{x^2} - \frac{2x^i x_k dx^k}{\left(x^2\right)^2} \\ \left(ds'\right)^2 &= \delta_{ij} dy^i dy^j \\ &= \delta_{ij} \left(\frac{dx^i}{x^2} - \frac{2x^i x_k dx^k}{\left(x^2\right)^2} \right) \left(\frac{dx^j}{x^2} - \frac{2x^j x_m dx^m}{\left(x^2\right)^2} \right) \\ &= \left(\frac{1}{x^2} \right)^2 \delta_{ij} dx^i dx^j - \frac{2\left(x_m dx^m\right)^2}{\left(x^2\right)^3} - \frac{2\left(x_k dx^k\right)^2}{\left(x^2\right)^3} + \frac{4x^2 \left(x_m dx^m\right)^2}{\left(x^2\right)^2} \\ &= \left(\frac{1}{x^2} \right)^2 ds^2 \end{aligned}$$

Because the line element is changed by only a factor, inversions are conformal transformations.

Before using inversions to construct new transformations, we note that the inverse of the origin is not defined. However, it is easy to correct this by adding a single point to 3-space. Consider $y^i = \frac{x^i}{x^2}$ to be a new coordinate patch on an extended manifold. The overlap with the x^i coordinates consists of all values except $x^i = 0$ and $y^i = 0$. The new point that we want to add is the one with y^i coordinate $y^i = 0$. This

point is called the "point at infinity." The manifold becomes compact, and requires two coordinate patches to cover. Inversions are defined for all points of this extended manifold.

While inversion is simply a discrete transformation, we easily construct a 3-parameter family of transformations by sandwiching other transformations between two inversions. Since each transformation is separately conformal, sequences of them must also be conformal. The result gives nothing new for rotations or dilatations, but for translations we find that inverting, translating, then inverting again gives

$$y^i = \frac{\frac{x^i}{x^2} + b^i}{\left(\frac{x^i}{x^2} + b^i\right)^2} = \frac{x^i + x^2 b^i}{1 + b_i x^i + b^2 x^2}$$

These transformations, which are simply translations of the point at infinity, are called *special conformal transformations*.Since we now have the required total of 10 independent conformal transformations, we have found the full.

Show that an inversion, followed by a dilatation, followed by another inversion is just a dilatation by a different parameter. What is the parameter?

Show that an inversion followed by a rotation, followed by another inversion, is just the original rotation.

An infinitesimal dilatation is given by

$$y^i = x^i + bx^i$$

where b is infinitesimal. Take a limit of infinitely many infinitesimal dilatations, with $\lim_{n\to\infty}(nb) = \lambda$, to show that a finite dilatation is given by $y^i = e^\lambda x^i$.

Let y^i be a special conformal transformation of x^i,

$$y^i = \frac{x^i + x^2 b^i}{1 + b_i x^i + b^2 x^2}$$

Prove directly that this transformation is conformal by showing that

$$dy^i dy_i = \delta_{ij} dy^i dy^j = \frac{dx^i dx_i}{\left(1 + 2x^n b_n + b^2 x^2\right)^2}$$

Hint: Consider how the norm changes under each inversion, translation, and inversion separately, then combine the results.

Show that for an infinitesimal special conformal transformation, that is,

$$y^i = \frac{x^i + x^2 c^i}{1 + c_i x^i + c^2 x^2}$$

with c_i infinitesimal,

$$\delta x^i = y^i - x^i$$

agrees with the c_i-dependent part of eq.(50).

13.2 The relativisic conformal group

As noted in the introduction, we will begin with the relativistic conformal group, then restrict to the non-relativistic case. The derivation of the relativistic conformal group is entirely analogous to the Euclidean case. The only difference is that we start in spacetime rather than Euclidean space, demanding transformations such that

$$\eta_{\alpha\beta} dy^\alpha dy^\beta = e^\phi \eta_{\alpha\beta} dx^\alpha dx^\beta$$

where $\eta_{\alpha\beta}$ is the Minkowski metric and the Greek indices run from 0 to 3. Linearizing by setting $y^\alpha = x^\alpha + \varepsilon^\alpha$, we now require the condition

$$\partial_\alpha \varepsilon_\beta + \partial_\beta \varepsilon_\alpha = \frac{1}{2} \partial^\mu \varepsilon_\mu \eta_{\alpha\beta} \tag{51}$$

The argument is the same as above, with the result that

$$\varepsilon_\alpha = a_\alpha + \varepsilon_{\alpha\beta}x^\beta + \frac{1}{3}bx_\alpha + \frac{1}{3}\left(2c_\beta x^\beta x_\alpha - c_\alpha x^2\right) \tag{52}$$

There are now four spacetime translations generated by a_α, six Lorentz transformations generated by $\varepsilon_{\alpha\beta} = -\varepsilon_{\beta\alpha}$, one dilatation generated by b, and four translations of the point at infinity with parameters c_α, giving a total of fifteen conformal transformations of spacetime.

Starting from eq.(51), repeat the steps of the preceeding section to derive the form of the relativistic conformal generators given in eq.(52).

13.3 A linear representation for conformal transformations

It is usually easier to calculate the effect of transformations if those transformations act linearly. But the form of the special conformal transformations is very nonlinear. We now introduce define a new space, related to Euclidean 3-space, but constructed so that the conformal transformations all act linearly.

To find a linear representation of the conformal group we let

$$y^i = \frac{x^i}{x^2}$$

be the inverse of x^i. Now introduce a pair of new parameters, α and β and define

$$\begin{aligned} x^i &= \frac{w^i}{\alpha} \\ y^i &= \frac{w^i}{\beta} \end{aligned}$$

Then since $x^i y_i = 1$, we have

$$\begin{aligned} \alpha\beta - w^i w_i &= 0 \\ \beta &= \frac{w^i w_i}{\alpha} \end{aligned}$$

Now consider the effect of conformal transformations on x^i and y^i. Applying, respectively, rotations, dilatations, translatations and special conformal transformations to x^i and y^i, we find:

$$\begin{aligned} \bar{x}^i &= R^i{}_j x^j \\ \bar{x}^i &= e^\phi x^i \\ \bar{x}^i &= x^i + a^i \\ \bar{x}^i &= \frac{x^i + x^2 b^i}{1 + b_i x^i + b^2 x^2} \end{aligned}$$

for x^i and

$$\begin{aligned} \bar{y}^i &= R^i{}_j x^j \\ \bar{y}^i &= e^\phi x^i \\ \bar{y}^i &= \frac{y^i + y^2 a^i}{1 + a_i y^i + a^2 y^2} \\ \bar{y}^i &= y^i + b^i \end{aligned}$$

for y^i.

Next, suppose these transformations are also allowed to change α (and therefore β) in a way yet to be specified. For each type of conformal transformation, we will choose the way α and β transform so that $w^i w_i - \alpha\beta = 0$. First, under rotations, let

$$\begin{aligned} \bar{w}^i &= R^i{}_j w^j \\ \bar{\alpha} &= \alpha \\ \bar{\beta} &= \beta \end{aligned}$$

Next, for dilatations, we set

$$\begin{aligned} \bar{w}^i &= w^i \\ \bar{\alpha} &= e^{-\phi}\alpha \\ \bar{\beta} &= e^{\phi}\beta \end{aligned}$$

For translations set

$$\begin{aligned} \bar{w}^i &= w^i + \alpha a^i \\ \bar{\alpha} &= \alpha \\ \bar{\beta} &= \frac{1}{\alpha}\left(w^i + \alpha a^i\right)\left(w_i + \alpha a_i\right) \\ &= \frac{1}{\alpha}\left(w^i w_i + 2\alpha a^i w_i + \alpha^2 a^2\right) \\ &= \beta + 2a^i w_i + a^2\alpha \end{aligned}$$

and for special conformal transformations let

$$\begin{aligned} \bar{w}^i &= w^i + \beta b^i \\ \bar{\alpha} &= \frac{1}{\beta}\left(w^i w_i + 2\beta b^i w_i + \beta^2 b^2\right) \\ &= \alpha + 2b^i w_i + b^2\beta \\ \bar{\beta} &= \beta \end{aligned}$$

In each case we have chosen β so that the quadratic form

$$\alpha\beta - w^i w_i$$

vanishes. But also, each of the transformations is linear – for example, the special conformal transformations are now given by

$$\begin{pmatrix} \bar{w}^i \\ \bar{\alpha} \\ \bar{\beta} \end{pmatrix} = \begin{pmatrix} \delta^i_j & 0 & b^i \\ 2b_j & 1 & b^2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w^j \\ \alpha \\ \beta \end{pmatrix}$$

We can think of the quadratic form as an inner product on a vector space,

$$\begin{aligned} s^2 &= \eta_{AB} w^A w^B = w^i w_i - \alpha\beta \\ w^A &= \left(w^i, \alpha, \beta\right) \end{aligned}$$

If we let

$$\begin{aligned} \alpha &= v - u \\ \beta &= v + u \end{aligned}$$

then the inner product takes the form

$$s^2 = w^i w_i + u^2 - v^2$$

and the metric is

$$\eta_{AB} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix}$$

The linear transformations that preserve this metric form the pseudo-orthogonal group, $O(4,1)$. Since the dimension of any orthogonal group in n-dimensions is $\frac{n(n-1)}{2}$, $O(4,1)$ must have dimension $\frac{5\cdot 4}{2} = 10$, exactly the dimension of the conformal transformations. The conformal group of Euclidean 3-space is therefore $O(4,1)$.

We collect the matrices representing conformal transformations:

Rotations:

$$\begin{pmatrix} \bar{w}^i \\ \bar{\alpha} \\ \bar{\beta} \end{pmatrix} = \begin{pmatrix} R^i{}_j & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w^j \\ \alpha \\ \beta \end{pmatrix}$$

Dilatations:

$$\begin{pmatrix} \bar{w}^i \\ \bar{\alpha} \\ \bar{\beta} \end{pmatrix} = \begin{pmatrix} \delta^i_j & 0 & 0 \\ 0 & e^{-\phi} & 0 \\ 0 & 0 & e^{\phi} \end{pmatrix} \begin{pmatrix} w^j \\ \alpha \\ \beta \end{pmatrix}$$

Translations:

$$\begin{pmatrix} \bar{w}^i \\ \bar{\alpha} \\ \bar{\beta} \end{pmatrix} = \begin{pmatrix} \delta^i_j & a^i & 0 \\ 0 & 1 & 0 \\ 2a_j & a^2 & 1 \end{pmatrix} \begin{pmatrix} w^j \\ \alpha \\ \beta \end{pmatrix}$$

Special conformal transformations:

$$\begin{pmatrix} \bar{w}^i \\ \bar{\alpha} \\ \bar{\beta} \end{pmatrix} = \begin{pmatrix} \delta^i_j & 0 & b^i \\ 2b_j & 1 & b^2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w^j \\ \alpha \\ \beta \end{pmatrix}$$

This is a linear representation for the conformal group of compactified Euclidean 3-space.

14 A new arena for mechanics

Recall the procedure we followed in building Euclidean 3-space from the Euclidean group. We constructed the Euclidean spatial manifold by regarding as equivalent all points related by rotations. We showed that within the 6-dimensional Euclidean group there is a 3-dimensional set of such equivalent points – one for each point in R^3. We concluded by taking the space R^3 as the arena for mechanics, observing that it was a space which by construction had Euclidean symmetry. The rotational symmetry of R^3 guarantees isotropy of space while the translation invariance guarantees homogeneity.

However, if we take account of the need for a standard of measurement, we find that it is the conformal group and not the Euclidean group that best describes Newtonian measurement theory. We therefore repeat the construction starting from the conformal group. The first question is which transformations to regard as equivalent. Certainly we still want to equate points of the group manifold related by rotations (or Lorentz transformations in the relativistic case), but we also want dilatational symmetry of the final space. This means that we should regard points in the group space as equivalent if they are related by dilatations as well. Special conformal transformations, on the other hand, are a type of translation and should be treated the same way as the usual translations.

Consider the Euclidean case first. There are three rotations and one dilatation, so in the 10-dimensional space of conformal transformations we treat certain 4-dimensional subspaces as equivalent. We therefore find a 6-dimensional family of such 4-dimensional subspaces. This 6-dimensional family of points is the arena

for our generalization of mechanics. In the relativistic case we have six Lorentz transformations and one dilatation, accounting for seven of the fifteen conformal transformations. This will leave an 8-dimensional set of equivalent points, so the arena is 8-dimensional.

14.1 Dilatation covariant derivative

Recall the Christoffel connectionn we introduced in order to make Newton's law covariant under the diffeomorphism group:

$$\begin{aligned} g_{ij} &= \sum_i \delta_{mn} \frac{\partial x^m}{\partial y^i} \frac{\partial x^n}{\partial y^j} \\ g^{ij} &= \left[g^{-1}\right]_{ij} \\ \Gamma^i{}_{mn} &= \frac{1}{2} g^{ik} \left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right) \end{aligned}$$

The connection allows us to write a generalization of the derivative which has the property that

$$D_k v^i = \partial_k v^i + v^m \Gamma^i_{mk}$$

transforms as a $\binom{1}{1}$ tensor whenever v^i transforms as a $\binom{1}{0}$ tensor. The way it works is this: when v^j is transformed by a Jacobian matrix,

$$v^i \to J^i{}_m v^m$$

the connection transforms with two parts. The first part is appropriate to a $\binom{1}{2}$ tensor, while the second, inhomogenous term contains the derivative of the Jacobian matrix,

$$\Gamma^i_{jk} \to J^i{}_m \Gamma^m_{rs} \bar{J}^r{}_j \bar{J}^s{}_k - \left(\partial_s J^i{}_j\right) \bar{J}^s{}_k$$

This transformation property happens automatically because of the way Γ^i_{jk} is built from the metric. As a result, the covariant derivative transforms as

$$\begin{aligned} D_k v^i \quad \to \quad & \partial_k v^i + v^m \Gamma^i_{mk} = \bar{J}^s{}_k \partial_s \left(J^i{}_m v^m\right) \\ & + J^m{}_n v^n \left(J^i{}_l \Gamma^l_{rs} \bar{J}^r{}_m \bar{J}^s{}_k - \left(\partial_s J^i{}_r\right) \bar{J}^s{}_k \bar{J}^r{}_m\right) \\ = \quad & \bar{J}^s{}_k J^i{}_m \partial_s v^m + v^m \bar{J}^s{}_k \partial_s J^i{}_m \\ & + J^i{}_l \Gamma^l_{rs} v^r \bar{J}^s{}_k - \bar{J}^s{}_k \left(\partial_s J^i{}_r\right) v^r \\ = \quad & J^i{}_m \partial_k v^m + J^i{}_l \Gamma^l_{rs} v^r \bar{J}^s{}_k \\ = \quad & J^i{}_m \left(\partial_s v^m + \Gamma^m_{rs} v^r\right) \bar{J}^s{}_k \\ = \quad & \left(D_s v^m\right) J^i{}_m \bar{J}^s{}_k \end{aligned}$$

The extra derivative terms cancel out, leaving just the homogeneous transformation law we need.

With the conformal group, the Christoffel connection still works to make derivatives covariant with respect to rotations, but we must also make derivatives covariant with respect to scale changes. This is not difficult to do, since scale changes are simpler than diffeomorphisms. Since a vector v^i which has units of (length)^n will transform as

$$v^i \to e^{n\phi} v^i$$

its derivative will change by

$$\partial_m \left(e^{n\phi} v^i\right) = n e^{n\phi} v^i \partial_m \phi + e^{n\phi} \partial_m v^i$$

We can remove the extra term involving the derivative of the dilatation by adding a vector W_m to the derivative and demanding that W_m change by the troublesome term. Thus, setting

$$D_m v^i = \partial_m v^i + n W_m v^i$$

and demanding that W_m transform to $W_m - \partial_m \phi$ we have

$$\begin{aligned} D_m v^i &\rightarrow \partial_m \left(e^{n\phi} v^i\right) + n \left(W_m - \partial_m \phi\right) \left(e^{n\phi} v^i\right) \\ &= n e^{n\phi} v^i \partial_m \phi + e^{n\phi} \partial_m v^i + n W_m e^{n\phi} v^i - n \partial_m \phi e^{n\phi} v^i \\ &= e^{n\phi} \left(\partial_m v^i + n W_m v^i\right) \\ &= e^{n\phi} D_m v^i \end{aligned}$$

This is fine for v^i, but we also have to correct changes in the Christoffel connection. Start from

$$\Gamma^i{}_{mn} = \frac{1}{2} g^{ik} \left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right)$$

and performing a dilatation. Since the metric changes according to

$$g_{mn} \rightarrow e^{2\phi} g_{mn}$$

the connection changes by

$$\begin{aligned} \Gamma^i{}_{mn} &\rightarrow \frac{1}{2} e^{-2\phi} g^{ik} \left(\partial_m \left(e^{2\phi} g_{kn}\right) + \partial_n \left(e^{2\phi} g_{km}\right) - \partial_k \left(e^{2\phi} g_{mn}\right)\right) \\ &= \frac{1}{2} g^{ik} \left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right) \\ &\quad + g^{ik} \left(g_{kn} \partial_m \phi + g_{km} \partial_n \phi - g_{mn} \partial_k \phi\right) \end{aligned}$$

We need terms involving W_m to cancel the last three terms. It is easy to see that the needed form of Γ^i_{mn} is

$$\begin{aligned} \Gamma^i{}_{mn} &= \frac{1}{2} g^{ik} \left(\partial_m g_{kn} + \partial_n g_{km} - \partial_k g_{mn}\right) \\ &\quad + g^{ik} \left(g_{kn} W_m + g_{km} W_n - g_{mn} W_k\right) \end{aligned} \tag{53}$$

The new vector, W_m, is called the *Weyl vector* after Hermann Weyl, who introduced it in 1918. Weyl suggested at the time that in a relativistic model W_α might correspond to the electromagnetic vector potential, $A_\alpha = (\varphi, A_i)$. However, this theory is unphysical because it leads to measureable changes in the sizes of atoms when they pass through electromagnetic fields. This flaw was pointed out by Einstein in Weyl's original paper. A decade later, however, Weyl showed that electromagnetism could be written as a closely related theory involving phase invariance rather than scale invariance. This was the first modern gauge theory.

Check that the form of the connection given in eq.(53) is invariant under dilatation.

Combining the two covariances – rotations and dilatations – we write the covariant derivative of v^i as a combination of the two forms,

$$D_m v^i = \partial_m v^i + v^n \Gamma^i{}_{nm} + n W_m v^i$$

where $\Gamma^i{}_{mn}$ is given by the dilatation-invariant form in eq.(53).

14.2 Consequences of the covariant derivative

The dilatationally and rotationally covariant derivative has a surprising property – it applies not only to vectors and higher rank tensors, but even scalars and constants as well. The reason for this is that even constants may have dimensions, and if we rescale the units, the "constant" value will change accordingly. Mass, for example, when expressed in units of length

$$\left[\frac{m v_0}{h_0}\right] = [m]^{-1}$$

has units of inverse length. Therefore, its covariant derivative is

$$D_i m = \partial_i m - (-1) W_i m$$

If the vector W_i is nonzero then the condition for constant mass is

$$D_i m = 0$$

which integrates to give the position-dependent result

$$m = m_0 \exp\left(-\int W_i dx^i\right)$$

This is not surprising. When we change from meters to centimeters, the value of all masses increases by a factor of 100. The presence of W_i now allows us to choose position-dependent units of length, giving a continuous variation of the *value* of m. Notice, however, that the ratio of two masses transported together remains constant because the exponential factor cancels out.

There is a well-known difficulty with this factor though. Suppose we move two masses along *different* paths, C_1 and C_2, with common endpoints. Then their ratio is

$$\begin{aligned}\frac{m_1}{m_2} &= \frac{m_{01}\exp\left(-\int_{C_1} W_i dx^i\right)}{m_{02}\exp\left(-\int_{C_2} W_i dx^i\right)} \\ &= \frac{m_{01}}{m_{02}}\exp\left(-\oint_{C_1-C_2} W_i dx^i\right)\end{aligned}$$

This is a measurable consequence of the geometry unless

$$\oint_{C_1-C_2} W_i dx^i = 0$$

for all closed paths $C_1 - C_2$. By Stokes' theorem, this is the case if and only if W_i is the gradient of some function f. Any consistent interpretation of a dilatation-covariant geometry must make a consistent interpretation of this factor. We will re-examine the consequences of the present formulation of mechanics for this factor below.

14.3 Biconformal geometry

We now bring these features together. Consider the relativistic conformal group. The space we get by regarding points as classes of rotationally and dilatationally equivalent group elements, we get an 8-dimensional space corresponding to the parameters a^α and b^α in the translation and special conformal transformations,

$$\begin{aligned}\tilde{x}^\alpha &= x^\alpha + a^\alpha \\ \tilde{x}^\alpha &= \frac{x^\alpha + x^2 b^\alpha}{1 + b_\beta x^\beta + b^2 x^2}\end{aligned}$$

These parameters do not have the same units. Clearly, a^α has units of length like x^α, but b_α must have units of inverse length. As a reminder of their different roles it is convenient to write the length-type coordinate with its index up and the inverse-length type coordinate with its index down, that is, (a^α, b_α) or, as we write from now on, $\xi^A = (x^\alpha, y_\alpha)$. Since these become the coordinates of our new arena for mechanics, called biconformal space, biconformal space contains different types of geometric object in one space. Like the coordinates, we will write vectors in biconformal spaces as pairs of 4-vectors in which the first is a vector and the second a form:

$$w^A = (u^\alpha, v_\alpha)$$

Forms are written with the indices in the opposite positions. For example, the Weyl vector is usually written as a form. We write its eight components as

$$W_A = (W_\alpha, S^\alpha)$$

or in a basis (dx^α, dy_α)as

$$W_A d\xi^A = W_\alpha dx^\alpha + S^\alpha dy_\alpha$$

The Weyl vector plays a central role in the dynamics of biconformal spaces. Using the structure of the conformal group, one can write a set of differential equations which determine the form of the metric and connection, including the Weyl vector. All known solutions for biconformal spaces have the same form for the Weyl vector,

$$\begin{aligned} W_\alpha &= -\kappa y_\alpha \\ S^\alpha &= 0 \end{aligned}$$

where κ is constant. As a form, this means that we may write

$$W_A d\xi^A = -\kappa y_\alpha dx^\alpha$$

The Weyl vector transforms inhomogeneously according to

$$W_A = W_A - \partial_A \phi$$

Show that, by a suitable choice of ϕ, the Weyl vector may be written with components

$$W_A d\xi^A = -\frac{\kappa}{2} y_\alpha dx^\alpha + \frac{\kappa}{2} x^\alpha dy_\alpha$$

Recalling the non-integrability of lengths when W_A is not a gradient, we compute the components of the curl. Using the symmetric form of W_A,

$$\begin{aligned} \frac{\partial W_A}{\partial \xi^B} - \frac{\partial W_B}{\partial \xi^A} &= \begin{pmatrix} \frac{\partial W_\alpha}{\partial x^\beta} - \frac{\partial W_\beta}{\partial x^\alpha} & \frac{\partial W_\alpha}{\partial y_\beta} - \frac{\partial S^\beta}{\partial x^\alpha} \\ \frac{\partial S^\alpha}{\partial x^\beta} - \frac{\partial W_\beta}{\partial y_\alpha} & \frac{\partial S^\alpha}{\partial y_\beta} - \frac{\partial S^\beta}{\partial y_\alpha} \end{pmatrix} \\ &= \kappa \begin{pmatrix} 0 & -\delta^\beta_\alpha \\ \delta^\alpha_\beta & 0 \end{pmatrix} \end{aligned}$$

This object is antisymmetric and non-degenerate; any such object is called a *symplectic form.* It is the central feature of Hamiltonian mechanics, because its inverse leads to an extremely useful operator called the *Poisson bracket.* The Poisson bracket is a special case of a more general object called the *biconformal bracket.*

A function, $f(x^\alpha, y_\alpha) = f\left(\xi^A\right)$, on biconformal space is called a *dynamical variable.*

Let Ω^{AB} be κ times the inverse to the curl of the Weyl vector,

$$\Omega^{AB} = \begin{pmatrix} 0 & \delta^\alpha_\beta \\ -\delta^\beta_\alpha & 0 \end{pmatrix}$$

Then the biconformal bracket of two dynamical variables, f, g on biconformal space is defined by

$$\begin{aligned} \{f, g\} &= \Omega^{AB} \frac{\partial f}{\partial \xi^A} \frac{\partial g}{\partial \xi^B} \\ &= \delta^\alpha_\beta \frac{\partial f}{\partial x^\alpha} \frac{\partial g}{\partial y_\beta} - \delta^\beta_\alpha \frac{\partial f}{\partial y_\alpha} \frac{\partial g}{\partial x^\beta} \\ &= \frac{\partial f}{\partial x^\alpha} \frac{\partial g}{\partial y_\alpha} - \frac{\partial f}{\partial y_\alpha} \frac{\partial g}{\partial x^\alpha} \end{aligned}$$

A pair of dynamical variables f and g are said to be *canonically conjugate* if their Poisson brackets satisfy

$$\begin{aligned} \{f, f\} &= 0 \\ \{f, g\} &= 1 \\ \{g, g\} &= 0 \end{aligned}$$

We will find considerable use for canonical variables.
Show that

$$\begin{aligned}\left\{x^\alpha, x^\beta\right\} &= 0\\ \{x^\alpha, y_\beta\} &= \delta^\alpha_\beta\\ \{y_\alpha, y_\beta\} &= 0\end{aligned}$$

so that the coordinates x^α are canonically conjugate to the coordinates y_β. Show that these biconformal brackets may be written compactly as

$$\left\{\xi^A, \xi^B\right\} = \Omega^{AB}$$

14.4 Motion in biconformal space

As shown above, generic spaces with local dilatational symmetry allow measurable changes of lengths along different paths. The situation is no different in biconformal spaces. If we define lengths l_1 and l_2 using the Minkowski metric $\eta_{\alpha\beta}$ then follow their motions along paths C_1 and C_2, respectively, then the final ratio of their lengths is given by

$$\frac{l_1}{l_2} = \frac{l_{01}}{l_{02}} \exp\left(\oint_{C_1 - C_2} W_A d\xi^A\right)$$

We note in passing that biconformal spaces possess a natural, dilatation invariant, 8-dimensional metric, but using the Minkowski metric lets us make connection with our usual way of measuring in spacetime.

We now postulate that classical motion in biconformal space is along extremals of dilatation, defined by the variation of the functional

$$\begin{aligned}S &= -\int_C W_A d\xi^A\\ &= \int_C y_\alpha dx^\alpha.\end{aligned}$$

Variation of S leads to

$$\begin{aligned}0 &= \int_C \left(\delta y_\alpha \frac{dx^\alpha}{d\lambda} + y_\alpha \frac{d\delta x^\alpha}{d\lambda}\right) d\lambda\\ &= \int_C \left(\delta y_\alpha \frac{dx^\alpha}{d\lambda} - \frac{dy_\alpha}{d\lambda}\delta x^\alpha\right) d\lambda\end{aligned}$$

so, since the x^α and y_α variations are independent, we have the Euler-Lagrange equations for a straight line:

$$\begin{aligned}\frac{dx^\alpha}{d\lambda} &= 0\\ \frac{dy_\alpha}{d\lambda} &= 0\end{aligned}$$

The result is independent of parameterization. At this point we could introduce a potential and examine the consequences, but there is a more direct way to arrive at classical mechanics. We simply impose the invariance of the time coordinate.

14.5 Hamiltonian dynamics and phase space

One of the distinguishing features of classical mechanics is the universality of time. In keeping with this, we now require $x^0 = ct$ to be an invariant parameter so that $\delta t = 0$. Then varying the corresponding canonical

bracket we find

$$\begin{aligned} 0 &= \delta\left\{t, \xi^A\right\} \\ &= \left\{\delta t, \xi^A\right\} + \left\{t, \delta\xi^A\right\} \\ &= \frac{\partial t}{\partial x^0}\frac{\partial\left(\delta\xi^A\right)}{\partial y_0} \end{aligned}$$

so that variations of all of the biconformal coordinates must be independent of y_0 :

$$\frac{\partial\left(\delta\xi^A\right)}{\partial y_0} = 0$$

In particular, this applies to y_0 itself,

$$\frac{\partial\left(\delta y_0\right)}{\partial y_0} = 0$$

so that y_0 plays no role as an independent variable. Instead, δy_0, and hence y_0 itself, depend only on the remaining coordinates,

$$y_0 = -\frac{1}{h_0 c}H\left(y_i, x^j, t\right)$$

The function H is called the *Hamiltonian*. Our inclusion of h_0 and c in its definition give it units of energy. Its existence is seen to be a direct consequence of the universality of classical time.

Fixing the time for all observers effectively reduces the 8-dimensional biconformal space to a 6-dimensional biconformal space in which t parameterizes the motion. We call this 6-dimensional space, as well as its $6N$-dimensional generalization below, *phase space*. The restriction of the biconformal bracket to phase space is called the *Poisson bracket*.

Returning to the postulate of the previous section, the action functional becomes

$$S = \int_C y_\alpha dx^\alpha = \int_C \left(y_i\frac{dx^i}{dt} - \frac{1}{h_0}H\right)dt.$$

In this form we may identify y_i as proportional to the momentum canonically conjugate to x^i, since

$$\begin{aligned} \frac{\partial L}{\partial\dot{x}^i} &= y_i - \frac{1}{h_0}\frac{\partial H\left(x^i, y_i, t\right)}{\partial\dot{x}^i} \\ &= y_i \end{aligned}$$

Henceforth we identify

$$y_i = \frac{p_i}{h_0}$$

where the inclusion of the constant h_0 gives the momentum p_i its usual units.

Variation of S now leads to

$$\begin{aligned} 0 &= \delta S \\ &= \frac{1}{h_0}\int\left(\delta p_i\frac{dx^i}{dt} - \frac{dp_i}{dt}\delta x^i - \delta H dt)\right)dt \\ &= \frac{1}{h_0}\int\left(\delta p_i\frac{dx^i}{dt} - \frac{dp_i}{dt}\delta x^i - \frac{\partial H}{\partial x^i}\delta x^i - \frac{\partial H}{\partial p_i}\delta p_i\right)dt \end{aligned}$$

which immediately gives:

$$\frac{dx^i}{dt} = \frac{\partial H}{\partial p_i} \tag{54}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial x^i}\,. \tag{55}$$

Eqs.(54) and (55) are *Hamilton's equations of motion*. They form the basis of Hamiltonian dynamics.

The Lagrangian and the Hamiltonian are related by

$$L = p_i \frac{dx^i}{dt} - H$$

It follows from this relationship that all of the x^i dependences of L arises from H and vice-versa, so that

$$\frac{\partial L}{\partial x^i} = -\frac{\partial H}{\partial x^i} \tag{56}$$

Furthermore, since the Hamiltonian depends on x^i and p_i while the Lagrangian depends on x^i and $\dot{x}^i$, eq.(54) is simply a consequence of the relationship between L and H. Combining eq.(56) with the definition of the conjugate momentum, $p_i = \frac{\partial L}{\partial \dot{x}^i}$ we can rewrite eq.(55) as

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} = \frac{\partial L}{\partial x^i}$$

Thus, Hamilton's equations imply the Euler-Lagrange equation, even though they are first order equations and the Euler-Lagrange equation is second order.

We also have

$$\begin{aligned} E &= \dot{x}^i \frac{\partial L}{\partial \dot{x}^i} - L \\ &= \dot{x}^i p_i - p_i \frac{dx^i}{dt} + H \\ &= H \end{aligned}$$

so when H has no explicit dependence on time it is equal to the energy. Finally, notice that the arbitrary constant h_0 drops out of the equations of motion.

14.5.1 Multiparticle mechanics

In the case of N particles, the action becomes a functional of N distinct curves, $C_i, i = 1, \ldots, N$

$$S = -\sum_{i=1}^{N} \int_{C_i} W_A d\xi^A \tag{57}$$

As for the single particle case, the invariance of time constrains p_0. However, since $W_A d\xi^A = -y_\alpha \mathbf{d}x^\alpha$ is to be evaluated on N different curves, there will be N distinct coordinates x_n^α and momenta, p_α^n. We therefore have

$$\begin{aligned} 0 &= \delta\left\{x_m^0, p_0^n\right\} \\ &= \left\{\delta x_m^0, p_0^n\right\} + \left\{x_m^0, \delta p_0^n\right\} \\ &= \frac{\partial\left(\delta p_0^n\right)}{\partial p_0^k}\frac{\partial x_m^0}{\partial x_k^0} \end{aligned} \tag{58}$$

Now, since time is universal in non-relativistic physics, we may set $x_m^0 = t$ for all m. Therefore, $\frac{\partial x_m^0}{\partial x_k^0} = 1$ and we have

$$\frac{\partial\left(\delta p_0^n\right)}{\partial p_0^k} = 0 \tag{59}$$

which implies that each p_0^n is a function of spatial components only,

$$p_0^n = p_0^n\left(x_k^i, p_i^k\right)$$

This means that each p_0^n is sufficiently general to provide a generic Hamiltonian. Conversely, any single N-particle Hamiltonian may be written as a sum of N identical Hamiltonians,

$$H = \frac{1}{N}\sum_{n=1}^{N} H$$

so that eq.(57) becomes

$$\begin{aligned} S &= \frac{1}{h_0}\sum_{i=1}^{N}\int_{C_i}\left(-p_0^n dt + p_i^n dx_n^i\right) \\ &= \frac{1}{h_0}\int_{C_i}\left(\sum_{i=1}^{N} p_i^n dx_n^i - H\left(x_k^i, p_i^k\right) dt\right) \end{aligned}$$

The introduction of multiple biconformal coordinates has consequences for the conformal equations that describe biconformal space, for once we write the Weyl vector as

$$W_A d\xi^A = -\frac{1}{h_0}\sum_{i=1}^{N} p_\alpha^n \mathbf{d}x_n^\alpha = -i\sum_{i=1}^{N} y_\alpha^n \mathbf{d}x_n^\alpha$$

then we have, in effect, introduced an $8N$-dimensional biconformal space and a corresponding $6N$-dimensional phase space. This is the usual procedure in Hamiltonian mechanics, and it is required by the equations that describe biconformal space.

Phase space includes $3N$ dimensions corresponding to positions of N particles in our usual Newtonian 3-space. This space contains complete information about the spatial configuration of the system. Sometimes constraints reduce the effective dimension of these spatial dimensions. Therefore, by *configuration space*, we mean the space of all possible postions of the particles comprising the system, or the complete set of possible values of the degrees of freedom of the problem. Thus, configuration space for N particles is a subspace of the $3N$-dimensional space of all values of x_n^i. By *momentum space*, we mean the $3N$-dimensional space of all possible values of all of the conjugate momenta.

Finally, *phase space* is the $6N$-dimensional space of all possible values of both position and momentum. Let

$$\xi^A = \left(q_1^i, p_j^1, q_2^i, p_j^2, \ldots, q_N^i, p_j^N\right)$$

where $A = 1, \ldots, 6N$. Then the $6N$ variables ξ^A provide a set of coordinates for phase space. In subsequent sections we will write Hamilton's equations in terms of these, thereby treating all $2N$ directions on an equal footing. However, before developing the powerful tools of phase space, we revisit the issue of measurable size change that occurs in locally dilatational spaces.

14.6 Measurement and Hamilton's principal function

In biconformal spaces, the Weyl vector is not the gradient of a function so that its curl does not vanish. Therefore, we can consider experiments in which we could expect to measure relative size change. Why is no such size change ever detected?

Suppose a system characterized by an initial length, l_0, moves dynamically from the point x_0 to the point x_1, along an allowed (i.e., classical) path C_1. In order to measure a relative size change, we must employ a standard of length, e.g., a ruler with length, λ_0. Suppose the ruler moves dynamically from x_0 to x_1 along any classical path C_2. If the integral of the dilatational curvature over surfaces bounded by C_1 and C_2 does not vanish, the relative sizes of the two objects will be different.

As shown above, the ratio of lengths changes by the exponential of

$$\oint_{C_1 - C_2} W_A d\xi^A$$

By Stokes' theorem, this integral is equal to the surface integral of the curl of W_A, which in turn is given by

$$\begin{aligned}
\frac{\partial W_A}{\partial \xi^B} - \frac{\partial W_B}{\partial \xi^A} &= \begin{pmatrix} \frac{\partial W_\alpha}{\partial x^\beta} - \frac{\partial W_\beta}{\partial x^\alpha} & \frac{\partial W_\alpha}{\partial y_\beta} - \frac{\partial S^\beta}{\partial x^\alpha} \\ \frac{\partial S^\alpha}{\partial x^\beta} - \frac{\partial W_\beta}{\partial y_\alpha} & \frac{\partial S^\alpha}{\partial y_\beta} - \frac{\partial S^\beta}{\partial y_\alpha} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial W_\alpha}{\partial x^\beta} - \frac{\partial W_\beta}{\partial x^\alpha} & \frac{\partial W_\alpha}{\partial y_\beta} \\ -\frac{\partial W_\beta}{\partial y_\alpha} & \end{pmatrix} \\
&= \begin{pmatrix} 0 & \frac{\partial H}{\partial x^j} + \frac{\partial p_j}{\partial t} & \frac{\partial H}{\partial p_j} \\ -\frac{\partial p_i}{\partial t} - \frac{\partial H}{\partial x^i} & -\frac{\partial y_i}{\partial x^j} + \frac{\partial y_j}{\partial x^i} & -\delta_i^j \\ -\frac{\partial H}{\partial p_i} & \delta_i^j & \end{pmatrix} \\
&= \begin{pmatrix} 0 & \frac{\partial H}{\partial x^j} & \frac{\partial H}{\partial p_j} \\ -\frac{\partial H}{\partial x^i} & 0 & -\delta_i^j \\ -\frac{\partial H}{\partial p_i} & \delta_i^j & \end{pmatrix}
\end{aligned}$$

since $S^\alpha = 0$. This matrix can be written as the sum of antisymmetric product of two triples of 8-vectors. Let

$$\begin{aligned}
[U_i]_A &= \left(\frac{\partial H}{\partial x_i}, 0, 0, 0; 0, \delta_i^m\right) \\
\left[V^i\right]_A &= \left(-\frac{\partial H}{\partial p_i}, \delta_i^n; 0, 0, 0, 0\right)
\end{aligned}$$

$$\begin{aligned}
[U_i]_A \left[V^i\right]_B - [U_i]_B \left[V^i\right]_A &= \left(\frac{\partial H}{\partial x^i}, 0, 0, 0; 0, \delta_i^m\right)\left(-\frac{\partial H}{\partial p_i}, \delta_n^i; 0, 0, 0, 0\right) \\
&\quad - \left(-\frac{\partial H}{\partial p_i}, \delta_m^i; 0, 0, 0, 0\right)\left(\frac{\partial H}{\partial x^i}, 0, 0, 0; 0, \delta_i^n\right) \\
&= \begin{pmatrix} -\frac{\partial H}{\partial x^i}\frac{\partial H}{\partial p_i} & \frac{\partial H}{\partial x^n} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{\partial H}{\partial p_m} & \delta_n^m & 0 & 0 \end{pmatrix} \\
&\quad - \begin{pmatrix} -\frac{\partial H}{\partial p_i}\frac{\partial H}{\partial x^i} & 0 & 0 & -\frac{\partial H}{\partial p_n} \\ \frac{\partial H}{\partial x^m} & 0 & 0 & \delta_m^n \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & \frac{\partial H}{\partial x^n} & 0 & \frac{\partial H}{\partial p_n} \\ -\frac{\partial H}{\partial x^m} & 0 & 0 & -\delta_m^n \\ 0 & 0 & 0 & 0 \\ -\frac{\partial H}{\partial p_m} & \delta_n^m & 0 & 0 \end{pmatrix}
\end{aligned}$$

This agrees with the curl of the Weyl vector, so we may write

$$\frac{\partial W_A}{\partial \xi^B} - \frac{\partial W_B}{\partial \xi^A} = [U_i]_A \left[V^i\right]_B - [U_i]_B \left[V^i\right]_A$$

Now we write U_i and V_i as 1-forms:

$$[U_i]_A \, \mathbf{d}\xi^A = \frac{\partial H}{\partial x_i} dt + dp_i$$

$$\left[V^i\right]_A \mathbf{d}\xi^A = -\frac{\partial H}{\partial p_i}dt + dx^i$$

Evaluated along the classical paths where Hamilton's equations hold, these forms both vanish! This means that the curl of the Weyl vector is zero when evaluated along classical paths of motion.

Another way to see the same result is by constructing

$$\int_C W_A d\xi^A$$

where C is a solution to Hamilton's equations. Starting at a fixed initial point $\left(x_0^i, p_{0i}\right)$ there is a unique path in phase space satisfying Hamilton's equations. This is one of the great advantages of working in phase space – unlike configuration space, there is only one allowed classical path through any point. Integrating along this path to a point $\left(x^i, p_i\right)$ gives a single value,

$$\mathcal{S}\left(x^i, p_i\right) = \int_{x_0^i, p_{0i}}^{x^i, p_i} W_A d\xi^A$$

Therefore, $\mathcal{S}\left(x^i, p_i\right)$ is a function rather than a functional.

This argument immediately raises a question. To compare the evolving system to a ruler carried along a different path we need a different allowed path. Yet we claimed that the initial conditions uniquely specify the path! How can we reconcile the need for an independently moving standard with the uniqueness of phase space paths?

The answer lies in the fact that the ruler and system can be compared if the initial *positions* agree, even if the *momenta* are different. But this means that we need a stronger result than the single-valuedness of the integral for $\mathcal{S}\left(x^i, p_i\right)$. We need the values to be independent of the initial and final momenta, p_{0i} and p_i. This independence is guaranteed by the collapse of $W_A d\xi^A$ to $W_\alpha dx^\alpha$, which is a derived property of biconformal spaces. Integrating the Weyl vector along any configuration space path gives the same result as integrating it along any path that projects to that configuration space path because W is horizontal – it depends only on dx^α, not on dp_α. This is an example of a fibre bundle – the possible values of momentum form a "fibre" over each point of configuration space, and the integral of the Weyl vector is independent of which point on the fibre a path goes through.

Since the integral of the Weyl vector is a function of position, independent of path, then it is immediately obvious that no physical size change is measurable. Consider our original comparison of a system with a ruler. If both of them move from x_0 to x_1 and the dilatation they experience depends only on x_1, then at that point they will have experienced identical scale factors. Such a change is impossible to observe from relative comparison – the lengths are still the same.

This observation can also be formulated in terms of the gauge freedom. Since we may write the integral of the Weyl vector as a function of position, the integral of the Weyl vector along every classical path may be removed by the gauge transformation [104]

$$e^{-\mathcal{S}(x)}. \tag{60}$$

Then in the new gauge,

$$\begin{aligned} \int W'_\alpha dx^\alpha &= \int \left(W_\alpha - \partial_\alpha \mathcal{S}(x)\right) dx^\alpha \\ &= \int W_\alpha dx^\alpha - \mathcal{S}(x) \\ &= 0 \end{aligned} \tag{61}$$

as long as the integration is restricted to classical paths of motion. Note that we have removed all possible

integrals with one gauge choice. This reflects the difference between a function and a functional. It again follows that no classical objects ever display measurable length change.

In the next section, we provide a second proof of the existence of Hamilton's principal function, this time using differential forms. Then we derive some consequences of the existence of Hamilton's principal function.

14.7 A second proof of the existence of Hamilton's principal function

Proving the existence of Hamilton's principal function $\mathcal{S}$ provides an excellent example of the usefulness of differential forms. The Weyl vector written as a 1-form is

$$\mathbf{W} = p_i \mathbf{d}x_i - H\mathbf{d}t$$

We can do more than this, however. One of the most powerful uses of differential forms is for finding integrability conditions. To use them in this way, we must define the exterior derivative $\mathbf{d}$ of a 1-form. Given a 1-form,

$$\omega = \omega_i \mathbf{d}x^i$$

we define the exterior derivative of ω to be the antisymmetric $\binom{0}{2}$ tensor

$$\frac{1}{2}\left(\partial_j \omega_i - \partial_i \omega_j\right)$$

Normally this is written as a *2-form,*

$$\mathbf{d}\omega = \partial_j \omega_i \mathbf{d}x^j \wedge \mathbf{d}x^i$$

Here the wedge product, $\wedge$, means that we automatically antisymmetrize the product of basis forms,

$$\mathbf{d}x^j \wedge \mathbf{d}x^i = -\mathbf{d}x^i \wedge \mathbf{d}x^j$$

This antisymmetry is the only property of the wedge product we need.

Now we can use the following theorem (the converse to the Poincare lemma). Suppose (in a star-shaped region) that

$$\mathbf{d}\theta = 0 \tag{62}$$

for some p-form θ. The p-form is then said to be *closed*. It follows that

$$\theta = \mathbf{d}\sigma \tag{63}$$

for some (p-1)-form σ. If θ can be written in this form, it is said to be *exact*. Thus, closed forms are exact. Conversely, if $\theta = \mathbf{d}\sigma$, then $\mathbf{d}\theta = 0$ because $\mathbf{d}\theta = \mathbf{d}^2\sigma = 0$ (since $\mathbf{d}^2 = 0$ always), so exact forms are closed.

You are already familiar with certain applications of the converse to the Poincaré lemma. The simplest is for the differential of a one-form. Suppose

$$\mathbf{d}\theta = 0 \tag{64}$$

Then there exists a function f (a function is a zero-form) such that

$$\theta = \mathbf{d}f \tag{65}$$

This is the usual condition for the integrability of a conservative force. Think of f as the negative of the potential and θ as the force, $\theta = F_i \mathbf{d}x^i$. Then the integrability condition

$$\mathbf{d}\theta = 0 \tag{66}$$

is just

$$\begin{aligned}
0 &= \mathbf{d}\theta \\
&= \mathbf{d}F_i \wedge \mathbf{d}x^i \\
&= \frac{\partial F_i}{\partial x^j}\mathbf{d}x^j \wedge \mathbf{d}x^i \\
&= \frac{1}{2}\frac{\partial F_i}{\partial x^j}\left(\mathbf{d}x^j \wedge \mathbf{d}x^i - \mathbf{d}x^i \wedge \mathbf{d}x^j\right) \\
&= \frac{1}{2}\left(\frac{\partial F_i}{\partial x^j}\mathbf{d}x^j \wedge \mathbf{d}x^i - \frac{\partial F_i}{\partial x^j}\mathbf{d}x^i \wedge \mathbf{d}x^j\right) \\
&= \frac{1}{2}\left(\frac{\partial F_i}{\partial x^j}\mathbf{d}x^j \wedge \mathbf{d}x^i - \frac{\partial F_j}{\partial x^i}\mathbf{d}x^j \wedge \mathbf{d}x^i\right) \\
&= \frac{1}{2}\left(\frac{\partial F_i}{\partial x^j} - \frac{\partial F_j}{\partial x^i}\right)\mathbf{d}x^j \wedge \mathbf{d}x^i
\end{aligned}$$

This is equivalent to the vanishing of the curl of $\mathbf{F}$, the usual condition for the existence of a potential.

The Poincaré lemma and its converse give us a way to tell when a 1-form is the differential of a function. Given the Weyl vector

$$\omega = p_i\mathbf{d}x_i - H\mathbf{d}t \tag{67}$$

we have $\omega = \mathbf{d}\mathcal{S}$ if and only if $\mathbf{d}\omega = 0$. Therefore, compute

$$\begin{aligned}
\mathbf{d}\omega &= \mathbf{d}p_i \wedge \mathbf{d}x_i - \frac{\partial H}{\partial x_i}\mathbf{d}x_i \wedge \mathbf{d}t - \frac{\partial H}{\partial p_i}\mathbf{d}p_i \wedge \mathbf{d}t - \frac{\partial H}{\partial t}\mathbf{d}t \wedge \mathbf{d}t \\
&= \mathbf{d}p_i \wedge \mathbf{d}x_i - \frac{\partial H}{\partial x_i}\mathbf{d}x_i \wedge \mathbf{d}t - \frac{\partial H}{\partial p_i}\mathbf{d}p_i \wedge \mathbf{d}t
\end{aligned} \tag{68}$$

where the last term vanishes because $\mathbf{d}t \wedge \mathbf{d}t = 0$. We can rearrange this result as

$$\begin{aligned}
\mathbf{d}\omega &= \mathbf{d}p_i \wedge \mathbf{d}x_i - \frac{\partial H}{\partial x_i}\mathbf{d}x_i \wedge \mathbf{d}t - \frac{\partial H}{\partial p_i}\mathbf{d}p_i \wedge \mathbf{d}t \\
&= \left(\mathbf{d}p_i + \frac{\partial H}{\partial x_i}\mathbf{d}t\right) \wedge \left(\mathbf{d}x_i - \frac{\partial H}{\partial p_i}\mathbf{d}t\right)
\end{aligned} \tag{69}$$

where we again use $\mathbf{d}t \wedge \mathbf{d}t = 0$ and also use $\mathbf{d}x_i \wedge \mathbf{d}t = -\mathbf{d}t \wedge \mathbf{d}x_i$. Since each factor is one of Hamilton's equations, this condition is clearly satisfied when the equations of motion are satisfied. Therefore, the Weyl vector must be the exterior derivative of a function,

$$\omega = \mathbf{d}\mathcal{S}$$

and we have proved the existence of Hamilton's principal function $\mathcal{S}$. To find the function we just integrate,

$$\mathcal{S} = \int \omega$$

Since the integrability condition is satisfied, it doesn't matter which solution to Hamilton's equations we integrate along – all give the same value.

We end this section with some properties of Hamilton's principal function. We have

$$\omega = p_i\mathbf{d}x_i - H\mathbf{d}t$$

On the other hand, since $\omega = \mathbf{d}\mathcal{S}$, we can expanding the differential as

$$\mathbf{d}\mathcal{S} = \frac{\partial \mathcal{S}}{\partial x_i}\mathbf{d}x_i + \frac{\partial \mathcal{S}}{\partial p_i}\mathbf{d}p_i + \frac{\partial \mathcal{S}}{\partial t}\mathbf{d}t$$

Equating the two expressions for ω term by term shows that

$$\begin{aligned}\frac{\partial \mathcal{S}}{\partial p_i} &= 0 \\ \frac{\partial \mathcal{S}}{\partial x_i} &= p_i \\ \frac{\partial \mathcal{S}}{\partial t} &= -H\end{aligned}$$

We will require these relationships when we develop Hamilton-Jacobi theory.

14.8 Phase space and the symplectic form

We now explore some of the properties of phase space and Hamilton's equations.

One advantage of the Hamiltonian formulation is that there is now one equation for each initial condition. This gives the space of all qs and ps a uniqueness property that configuration space (the space spanned by the qs only) doesn't have. For example, a projectile which is launched from the origin. Knowing only this fact, we still don't know the path of the object – we need the initial velocity as well. As a result, many possible trajectories pass through each point of configuration space. By contrast, the initial point of a trajectory in phase space gives us *both* the initial position and the initial momentum. There can be only one path of the system that passes through that point.

Systems with any number of degrees of freedom may be handled in this way. If a system has N degrees of freedom then its phase space is the $2N$-dimensional space of all possible values of both position and momentum. We define *configuration space* to be the space of all possible postions of the particles comprising the system, or the complete set of possible values of the degrees of freedom of the problem. Thus, configuration space is the N-dimensional space of all values of q_i. By *momentum space*, we mean the N-dimensional space of all possible values of all of the conjugate momenta. Hamilton's equations then consist of $2N$ first order differential equations for the motion in phase space.

We illustrate these points with the simple example of a one dimensional harmonic oscillator.

Let a mass, m, free to move in one direction, experience a Hooke's law restoring force, $F = -kx$. Solve Hamilton's equations and study the motion of system in phase space. The Lagrangian for this system is

$$\begin{aligned}L &= T - V \\ &= \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2\end{aligned}$$

The conjugate momentum is just

$$p = \frac{\partial L}{\partial \dot{x}} = m\dot{x}$$

so the Hamiltonian is

$$\begin{aligned}H &= p\dot{x} - L \\ &= \frac{p^2}{m} - \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2 \\ &= \frac{p^2}{2m} + \frac{1}{2}kx^2\end{aligned}$$

Hamilton's equations are

$$\begin{aligned}\dot{x} &= \frac{\partial H}{\partial p} = \frac{p}{m} \\ \dot{p} &= -\frac{\partial H}{\partial x} = -kx \\ \frac{\partial H}{\partial t} &= -\frac{\partial L}{\partial t} = 0\end{aligned}$$

Note that Hamilton's equations are two first-order equations. From this point on the coupled linear equations

$$\begin{aligned} \dot{p} &= -kx \\ \dot{x} &= \frac{p}{m} \end{aligned}$$

may be solved in any of a variety of ways. Let's treat it as a matrix system,

$$\frac{d}{dt}\begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} & \frac{1}{m} \\ -k & \end{pmatrix}\begin{pmatrix} x \\ p \end{pmatrix} \tag{70}$$

The matrix $M = \begin{pmatrix} & -k \\ \frac{1}{m} & \end{pmatrix}$ has eigenvalues $\omega = \left(\sqrt{\frac{k}{m}}, -\sqrt{\frac{k}{m}}\right)$ and diagonalizes to

$$\begin{pmatrix} -i\omega & 0 \\ 0 & i\omega \end{pmatrix} = AMA^{-1}$$

where

$$\begin{aligned} A &= \frac{1}{2i\sqrt{km}}\begin{pmatrix} i\sqrt{km} & 1 \\ i\sqrt{km} & -1 \end{pmatrix} \\ A^{-1} &= \begin{pmatrix} -1 & -1 \\ -i\sqrt{km} & i\sqrt{km} \end{pmatrix} \\ \omega &= \sqrt{\frac{k}{m}} \end{aligned}$$

Therefore, multiplying eq.(70) on the left by A and inserting $1 = A^{-1}A$,

$$\frac{d}{dt}A\begin{pmatrix} x \\ p \end{pmatrix} = A\begin{pmatrix} & \frac{1}{m} \\ -k & \end{pmatrix}A^{-1}A\begin{pmatrix} x \\ p \end{pmatrix} \tag{71}$$

we get decoupled equations in the new variables:

$$\begin{pmatrix} a \\ a^\dagger \end{pmatrix} = A\begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left(x - \frac{ip}{\sqrt{km}}\right) \\ \frac{1}{2}\left(x + \frac{ip}{\sqrt{km}}\right) \end{pmatrix} \tag{72}$$

The decoupled equations are

$$\frac{d}{dt}\begin{pmatrix} a \\ a^\dagger \end{pmatrix} = \begin{pmatrix} -i\omega & 0 \\ 0 & i\omega \end{pmatrix}\begin{pmatrix} a \\ a^\dagger \end{pmatrix} \tag{73}$$

or simply

$$\begin{aligned} \dot{a} &= -i\omega a \\ \dot{a}^\dagger &= -i\omega a^\dagger \end{aligned}$$

with solutions

$$\begin{aligned} a &= a_0 e^{-i\omega t} \\ a^\dagger &= a_0^\dagger e^{i\omega t} \end{aligned}$$

The solutions for x and p may be written as

$$\begin{aligned} x &= x_0 \cos\omega t + \frac{p_0}{m\omega}\sin\omega t \\ p &= -m\omega x_0 \sin\omega t + p_0 \cos\omega t \end{aligned}$$

Notice that once we specify the initial point in phase space, (x_0, p_0), the entire solution is determined. This solution gives a parameterized curve in phase space. To see what curve it is, note that

$$\begin{aligned}
\frac{m^2\omega^2x^2}{2mE} + \frac{p^2}{2mE} &= \frac{m^2\omega^2x^2}{p_0^2 + m^2\omega^2x_0^2} + \frac{p^2}{p_0^2 + m^2\omega^2x_0^2} \\
&= \frac{m^2\omega^2}{p_0^2 + m^2\omega^2x_0^2}\left(x_0\cos\omega t + \frac{p_0}{m\omega}\sin\omega t\right)^2 \\
&\quad + \frac{1}{p_0^2 + m^2\omega^2x_0^2}\left(-m\omega x_0\sin\omega t + p_0\cos\omega t\right)^2 \\
&= \frac{m^2\omega^2x_0^2}{p_0^2 + m^2\omega^2x_0^2} + \frac{p_0^2}{p_0^2 + m^2\omega^2x_0^2} \\
&= 1
\end{aligned}$$

or

$$m^2\omega^2x^2 + p^2 = 2mE$$

This describes an ellipse in the xp plane. The larger the energy, the larger the ellipse, so the possible motions of the system give a set of nested, non-intersecting ellipses. Clearly, every point of the xp plane lies on exactly one ellipse.

The phase space description of classical systems are equivalent to the configuration space solutions and are often easier to interpret because more information is displayed at once. The price we pay for this is the doubled dimension – paths rapidly become difficult to plot. To ofset this problem, we can use *Poincaré sections* – projections of the phase space plot onto subspaces that cut across the trajectories. Sometimes the patterns that occur on Poincaré sections show that the motion is confined to specific regions of phase space, even when the motion never repeats itself. These techniques allow us to study systems that are *chaotic*, meaning that the phase space paths through nearby points diverge rapidly.

Now consider the general case of N degrees of freedom. Let

$$\xi^A = \left(q^i, p_j\right)$$

where $A = 1, \ldots, 2N$. Then the $2N$ variables ξ^A provide a set of coordinates for phase space. We would like to write Hamilton's equations in terms of these, thereby treating all $2N$ directions on an equal footing.

In terms of ξ^A, we have

$$\begin{aligned}
\frac{d\xi^A}{dt} &= \begin{pmatrix} \dot{q}^i \\ \dot{p}_j \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial H}{\partial p_i} \\ -\frac{\partial H}{\partial q^j} \end{pmatrix} \\
&= \Omega^{AB}\frac{\partial H}{\partial \xi^B}
\end{aligned}$$

where the presence of Ω^{AB} in the last step takes care of the difference in signs on the right. Here Ω^{AB} is just the inverse of the symplectic form found from the curl of the dilatation, given by

$$\Omega^{AB} = \begin{pmatrix} 0 & \delta^i_j \\ -\delta^j_i & 0 \end{pmatrix}$$

Its occurrence in Hamilton's equations is an indication of its central importance in Hamiltonian mechanics. We may now write Hamilton's equations as

$$\frac{d\xi^A}{dt} = \Omega^{AB}\frac{\partial H}{\partial \xi^B} \tag{74}$$

Consider what happens to Hamilton's equations if we want to change to a new set of phase space coordinates, $\chi^A = \chi^A(\xi)$. Let the inverse transformation be $\xi^A(\chi)$. The time derivatives become

$$\frac{d\xi^A}{dt} = \frac{\partial \xi^A}{\partial \chi^B}\frac{d\chi^B}{dt}$$

while the right side becomes

$$\Omega^{AB}\frac{\partial H}{\partial \xi^B} = \Omega^{AB}\frac{\partial \chi^C}{\partial \xi^B}\frac{\partial H}{\partial \chi^C}$$

Equating these expressions,

$$\frac{\partial \xi^A}{\partial \chi^B}\frac{d\chi^B}{dt} = \Omega^{AB}\frac{\partial \chi^D}{\partial \xi^B}\frac{\partial H}{\partial \chi^D}$$

we multiply by the Jacobian matrix, $\frac{\partial \chi^C}{\partial \xi^A}$ to get

$$\begin{aligned}\frac{\partial \chi^C}{\partial \xi^A}\frac{\partial \xi^A}{\partial \chi^B}\frac{d\chi^B}{dt} &= \frac{\partial \chi^C}{\partial \xi^A}\Omega^{AB}\frac{\partial \chi^D}{\partial \xi^B}\frac{\partial H}{\partial \chi^D} \\ \delta^C_B\frac{d\chi^B}{dt} &= \frac{\partial \chi^C}{\partial \xi^A}\Omega^{AB}\frac{\partial \chi^D}{\partial \xi^B}\frac{\partial H}{\partial \chi^D}\end{aligned}$$

and finally

$$\frac{d\chi^C}{dt} = \frac{\partial \chi^C}{\partial \xi^A}\Omega^{AB}\frac{\partial \chi^D}{\partial \xi^B}\frac{\partial H}{\partial \chi^D}$$

Defining the symplectic form in the new coordinate system,

$$\tilde{\Omega}^{CD} \equiv \frac{\partial \chi^C}{\partial \xi^A}\Omega^{AB}\frac{\partial \chi^D}{\partial \xi^B}$$

we see that Hamilton's equations are entirely the same if the transformation leaves the symplectic form invariant,

$$\tilde{\Omega}^{CD} = \Omega^{CD}$$

Any linear transformation $M^A{}_B$ leaving the symplectic form invariant,

$$\Omega^{AB} \equiv M^A{}_C M^B{}_D \Omega^{CD}$$

is called a symplectic transformation. Coordinate transformations which are symplectic transformations at each point are called *canonical*. Therefore those functions $\chi^A(\xi)$ satisfying

$$\Omega^{CD} \equiv \frac{\partial \chi^C}{\partial \xi^A}\Omega^{AB}\frac{\partial \chi^D}{\partial \xi^B}$$

are canonical transformations. Canonical transformations preserve Hamilton's equations.

14.9 Poisson brackets

We may also write Hamilton's equations in terms of the Poisson brackets. Recall that the Poisson bracket of any two dynamical variables f and g is given by

$$\{f, g\} = \Omega^{AB}\frac{\partial f}{\partial \xi^A}\frac{\partial g}{\partial \xi^B}$$

The importance of this product is that it too is preserved by canonical transformations. We see this as follows.

Let ξ^A be any set of phase space coordinates in which Hamilton's equations take the form of eq.(74), and let f and g be any two dynamical variables, that is, functions of these phase space coordinates, ξ^A. The Poisson bracket of f and g is given above. In a different set of coordinates, $\chi^A(\xi)$, we have

$$\begin{aligned}
\{f,g\}' &= \Omega^{AB}\frac{\partial f}{\partial \chi^A}\frac{\partial g}{\partial \chi^B} \\
&= \Omega^{AB}\left(\frac{\partial \xi^C}{\partial \chi^A}\frac{\partial f}{\partial \xi^C}\right)\left(\frac{\partial \xi^D}{\partial \chi^B}\frac{\partial g}{\partial \xi^D}\right) \\
&= \left(\frac{\partial \xi^C}{\partial \chi^A}\Omega^{AB}\frac{\partial \xi^D}{\partial \chi^B}\right)\frac{\partial f}{\partial \xi^C}\frac{\partial g}{\partial \xi^D}
\end{aligned}$$

Therefore, if the coordinate transformation is canonical so that

$$\frac{\partial \xi^C}{\partial \chi^A}\Omega^{AB}\frac{\partial \xi^D}{\partial \chi^B} = \Omega^{CD}$$

then we have

$$\{f,g\}' = \Omega^{AB}\frac{\partial f}{\partial \xi^C}\frac{\partial g}{\partial \xi^D} = \{f,g\}$$

and the Poisson bracket is unchanged. We conclude that canonical transformations preserve all Poisson brackets.

An important special case of the Poisson bracket occurs when one of the functions is the Hamiltonian. In that case, we have

$$\begin{aligned}
\{f,H\} &= \Omega^{AB}\frac{\partial f}{\partial \xi^A}\frac{\partial H}{\partial \xi^B} \\
&= \frac{\partial f}{\partial x^i}\frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p^i}\frac{\partial H}{\partial x_i} \\
&= \frac{\partial f}{\partial x^i}\frac{dx^i}{dt} - \frac{\partial f}{\partial p^i}\left(-\frac{dp_i}{dt}\right) \\
&= \frac{df}{\partial t} - \frac{\partial f}{\partial t}
\end{aligned}$$

or simply,

$$\frac{df}{\partial t} = \{f,H\} + \frac{\partial f}{\partial t}$$

This shows that as the system evolves classically, the total time rate of change of any dynamical variable is the sum of the Poisson bracket with the Hamiltonian and the partial time derivative. If a dynamical variable has no explicit time dependence, then $\frac{\partial f}{\partial t} = 0$ and the total time derivative is just the Poisson bracket with the Hamiltonian.

The coordinates now provide a special case. Since neither x^i nor p_i has any explicit time dependence, with have

$$\begin{aligned}
\frac{dx^i}{dt} &= \left\{H, x^i\right\} \\
\frac{dp_i}{dt} &= \{H, p_i\}
\end{aligned} \tag{75}$$

and we can check this directly:

$$\frac{dq_i}{dt} = \left\{H, x^i\right\}$$

$$
\begin{aligned}
&= \sum_{j=1}^{N}\left(\frac{\partial x^i}{\partial x^j}\frac{\partial H}{\partial p_j}-\frac{\partial x^i}{\partial p_j}\frac{\partial H}{\partial x^j}\right)\\
&= \sum_{j=1}^{N}\delta_{ij}\frac{\partial H}{\partial p_j}\\
&= \frac{\partial H}{\partial p_i}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{dp_i}{dt} &= \{H,p_i\}\\
&= \sum_{j=1}^{N}\left(\frac{\partial p_i}{\partial q_j}\frac{\partial H}{\partial p_j}-\frac{\partial p_i}{\partial p_j}\frac{\partial H}{\partial q_j}\right)\\
&= -\frac{\partial H}{\partial q_i}
\end{aligned}
$$

Notice that since q_i, p_i and are all independent, and do not depend explicitly on time, $\frac{\partial q_i}{\partial p_j} = \frac{\partial p_i}{\partial q_j} = 0 = \frac{\partial q_i}{\partial t} = \frac{\partial p_i}{\partial t}$.

Finally, we define the *fundamental Poisson brackets.* Suppose x^i and p_j are a set of coordinates on phase space such that Hamilton's equations hold in the either the form of eqs.(75) or of eqs.(74). Since they themselves are functions of (x^m, p_n) they are dynamical variables and we may compute their Poisson brackets with one another. With $\xi^A = (x^m, p_n)$ we have

$$
\begin{aligned}
\left\{x^i,x^j\right\}_\xi &= \Omega^{AB}\frac{\partial x^i}{\partial \xi^A}\frac{\partial x^j}{\partial \xi^B}\\
&= \sum_{m=1}^{N}\left(\frac{\partial x^i}{\partial x^m}\frac{\partial x^j}{\partial p_m}-\frac{\partial x^i}{\partial p_m}\frac{\partial x^j}{\partial x^m}\right)\\
&= 0
\end{aligned}
$$

for x^i with x^j,

$$
\begin{aligned}
\left\{x^i,p_j\right\}_\xi &= \Omega^{AB}\frac{\partial x^i}{\partial \xi^A}\frac{\partial p_j}{\partial \xi^B}\\
&= \sum_{m=1}^{N}\left(\frac{\partial x^i}{\partial x^m}\frac{\partial p_j}{\partial p_m}-\frac{\partial x^i}{\partial p_m}\frac{\partial p_j}{\partial x^m}\right)\\
&= \sum_{m=1}^{N}\delta^i_m\delta^m_j\\
&= \delta^i_j
\end{aligned}
$$

for x^i with p_j and finally

$$
\begin{aligned}
\{p_i,p_j\}_\xi &= \Omega^{AB}\frac{\partial p_i}{\partial \xi^A}\frac{\partial p_j}{\partial \xi^B}\\
&= \sum_{m=1}^{N}\left(\frac{\partial p_i}{\partial x^m}\frac{\partial p_j}{\partial p_m}-\frac{\partial p_i}{\partial p_m}\frac{\partial p_j}{\partial x^m}\right)\\
&= 0
\end{aligned}
$$

for p_i with p_j. The subscript ξ on the bracket indicates that the partial derivatives are taken with respect to the coordinates $\xi^A = \left(x^i, p_j\right)$. We summarize these relations as

$$\left\{\xi^A, \xi^B\right\}_\xi = \Omega^{AB}$$

We summarize the results of this subsection with a theorem: Let the coordinates ξ^A be canonical. Then a transformation $\chi^A(\xi)$ is canonical if and only if it satisfies the fundamental bracket relation

$$\left\{\chi^A, \chi^B\right\}_\xi = \Omega^{AB}$$

For proof, note that the bracket on the left is defined by

$$\left\{\chi^A, \chi^B\right\}_\xi = \Omega^{CD} \frac{\partial \chi^A}{\partial \xi^C} \frac{\partial \chi^B}{\partial \xi^D}$$

so in order for χ^A to satisfy the canonical bracket we must have

$$\Omega^{CD} \frac{\partial \chi^A}{\partial \xi^C} \frac{\partial \chi^B}{\partial \xi^D} = \Omega^{AB} \tag{76}$$

which is just the condition shown above for a coordinate transformation to be canonical. Conversely, suppose the transformation $\chi^A(\xi)$ is canonical and $\left\{\xi^A, \xi^B\right\}_\xi = \Omega^{AB}$. Then eq.(76) holds and we have

$$\left\{\chi^A, \chi^B\right\}_\xi = \Omega^{CD} \frac{\partial \chi^A}{\partial \xi^C} \frac{\partial \chi^B}{\partial \xi^D} = \Omega^{AB}$$

so χ^A satisfies the fundamental bracked relation.

In summary, each of the following statements is equivalent:

1. $\chi^A(\xi)$ is a canonical transformation.
2. $\chi^A(\xi)$ is a coordinate transformation of phase space that preserves Hamilton's equations.
3. $\chi^A(\xi)$ preserves the symplectic form, according to

$$\Omega^{AB} \frac{\partial \xi^C}{\partial \chi^A} \frac{\partial \xi^D}{\partial \chi^B} = \Omega^{CD}$$

4. $\chi^A(\xi)$ satisfies the fundamental bracket relations

$$\left\{\chi^A, \chi^B\right\}_\xi = \Omega^{CD} \frac{\partial \chi^A}{\partial \xi^C} \frac{\partial \chi^B}{\partial \xi^D}$$

These bracket relations represent a set of integrability conditions that must be satisfied by any new set of canonical coordinates. When we formulate the problem of canonical transformations in these terms, it is not obvious what functions $q^i\left(x^j, p_j\right)$ and $\pi_i\left(x^j, p_j\right)$ will be allowed. Fortunately there is a simple procedure for generating canonical transformations, which we develop in the next section.

We end this section with three examples of canonical transformations.

14.9.1 Example 1: Coordinate transformations

Let the new configuration space variable, q^i, be and an arbitrary function of the spatial coordinates:

$$q^i = q^i\left(x^j\right)$$

and let π_j be the momentum variables corresponding to q^i. Then $\left(q^i, \pi_j\right)$ satisfy the fundamental Poisson bracket relations iff:

$$\begin{aligned}\left\{q^i, q^j\right\}_{x,p} &= 0 \\ \left\{q^i, \pi_j\right\}_{x,p} &= \delta^i_j \\ \left\{\pi_i, \pi_j\right\}_{x,p} &= 0\end{aligned}$$

Check each:

$$\begin{aligned}\left\{q^i, q^j\right\}_{x,p} &= \sum_{m=1}^{N}\left(\frac{\partial q^i}{\partial x^m}\frac{\partial q^j}{\partial p_m} - \frac{\partial q^i}{\partial p_m}\frac{\partial q^j}{\partial x^m}\right) \\ &= 0\end{aligned}$$

since $\frac{\partial q^j}{\partial p_m} = 0$. For the second bracket,

$$\begin{aligned}\delta^i_j &= \left\{q^i, \pi_j\right\}_{x,p} \\ &= \sum_{m=1}^{N}\left(\frac{\partial q^i}{\partial x^m}\frac{\partial \pi_j}{\partial p_m} - \frac{\partial q^i}{\partial p_m}\frac{\partial \pi_j}{\partial x^m}\right) \\ &= \sum_{m=1}^{N}\frac{\partial q^i}{\partial x^m}\frac{\partial \pi_j}{\partial p_m}\end{aligned}$$

Since q^i is independent of p_m, we can satisfy this only if

$$\frac{\partial \pi_j}{\partial p_m} = \frac{\partial x^m}{\partial q^j}$$

Integrating gives

$$\pi_j = \frac{\partial x^n}{\partial q^j}p_n + c_j$$

with c_j an arbitrary constant. The presence of c_j does not affect the value of the Poisson bracket. Choosing $c_j = 0$, we compute the final bracket:

$$\begin{aligned}\left\{\pi_i, \pi_j\right\}_{x,p} &= \sum_{m=1}^{N}\left(\frac{\partial \pi_i}{\partial x^m}\frac{\partial \pi_j}{\partial p_m} - \frac{\partial \pi_i}{\partial p_m}\frac{\partial \pi_j}{\partial x^m}\right) \\ &= \sum_{m=1}^{N}\left(\frac{\partial^2 x^n}{\partial x^m \partial q^i}p_n\frac{\partial x^m}{\partial q^j} - \frac{\partial x^m}{\partial q^i}\frac{\partial^2 x^n}{\partial x^m \partial q^j}p_n\right) \\ &= \sum_{m=1}^{N}\left(\frac{\partial x^m}{\partial q^j}\frac{\partial}{\partial x^m}\frac{\partial x^n}{\partial q^i} - \frac{\partial x^m}{\partial q^i}\frac{\partial}{\partial x^m}\frac{\partial x^n}{\partial q^j}\right)p_n \\ &= \sum_{m=1}^{N}\left(\frac{\partial}{\partial q^j}\frac{\partial x^n}{\partial q^i} - \frac{\partial}{\partial q^i}\frac{\partial x^n}{\partial q^j}\right)p_n \\ &= 0\end{aligned}$$

Therefore, the transformations

$$\begin{aligned}q^j &= q^j(x^i) \\ \pi_j &= \frac{\partial x^n}{\partial q^j}p_n + c_j\end{aligned}$$

is a canonical transformation for any functions $q^i(x)$. This means that the symmetry group of Hamilton's equations is at least as big as the symmetry group of the Euler-Lagrange equations.

14.9.2 Example 2: Interchange of x and p.

The transformation

$$\begin{aligned} q^i &= p_i \\ \pi_i &= -x^i \end{aligned}$$

is canonical. We easily check the fundamental brackets:

$$\begin{aligned} \left\{q^i, q^j\right\}_{x,p} &= \left\{p_i, p_j\right\}_{x,p} = 0 \\ \left\{q^i, \pi_j\right\}_{x,p} &= \left\{p_i, -x^j\right\}_{x,p} \\ &= -\left\{p_i, x^j\right\}_{x,p} \\ &= +\left\{x^j, p_i\right\}_{x,p} \\ &= \delta_i^j \\ \left\{\pi_i, \pi_j\right\}_{x,p} &= \left\{-x^i, -x^j\right\}_{x,p} = 0 \end{aligned}$$

Interchange of x^i and p_j, with a sign, is therefore canonical. The use of generalized coordinates does not include such a possibility, so Hamiltonian dynamics has a larger symmetry group than Lagrangian dynamics.

For our last example, we first show that the composition of two canonical transformations is also canonical. Let $\psi(\chi)$ and $\chi(\xi)$ both be canonical. Defining the composition transformation, $\psi(\xi) = \psi(\chi(\xi))$, we compute

$$\begin{aligned} \Omega^{CD}\frac{\partial\psi^A}{\partial\xi^C}\frac{\partial\psi^B}{\partial\xi^D} &= \Omega^{CD}\left(\frac{\partial\psi^A}{\partial\chi^E}\frac{\partial\chi^E}{\partial\xi^C}\right)\left(\frac{\partial\psi^B}{\partial\chi^F}\frac{\partial\chi^F}{\partial\xi^D}\right) \\ &= \frac{\partial\chi^E}{\partial\xi^C}\frac{\partial\chi^F}{\partial\xi^D}\Omega^{CD}\left(\frac{\partial\psi^A}{\partial\chi^E}\right)\left(\frac{\partial\psi^B}{\partial\chi^F}\right) \\ &= \Omega^{EF}\left(\frac{\partial\psi^A}{\partial\chi^E}\right)\left(\frac{\partial\psi^B}{\partial\chi^F}\right) \\ &= \Omega^{AB} \end{aligned}$$

so that $\psi(\chi(\xi))$ is canonical.

14.9.3 Example 3: Momentum transformations

By the previous results, the composition of an arbitratry coordinate change with x, p interchanges is canonical. Consider the effect of composing (a) an interchange, (b) a coordinate transformation, and (c) an interchange.

For (a), let

$$\begin{aligned} q_1^i &= p_i \\ \pi_i^1 &= -x^i \end{aligned}$$

Then for (b) we choose an arbitrary function of q_1^i :

$$\begin{aligned} Q^i &= Q^i\left(q_1^j\right) = Q^i(p_j) \\ P_i &= \frac{\partial q_1^n}{\partial Q^i}\pi_n = -\frac{\partial p_n}{\partial Q^i}x^n \end{aligned}$$

Finally, for (c), another interchange:

$$\begin{aligned} q^i &= P_i = -\frac{\partial p_n}{\partial Q^i}x^n \\ \pi_i &= -Q^i = -Q^i(p_j) \end{aligned}$$

This establishes that replacing the momenta by any three independent functions of the momenta, preserves Hamilton's equations.

14.10 Generating functions

There is a systematic approach to canonical transformations using generating functions. We will give a simple example of the technique. Given a system described by a Hamiltonian $H(x^i, p_j)$, we seek another Hamiltonian $H'(q^i, \pi_j)$ such that the equations of motion have the same form, namely

$$\begin{aligned} \frac{dx^i}{dt} &= \frac{\partial H}{\partial p_i} \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial x^i} \end{aligned}$$

in the original system and

$$\begin{aligned} \frac{dq^i}{dt} &= \frac{\partial H'}{\partial \pi_i} \\ \frac{d\pi_i}{dt} &= -\frac{\partial H'}{\partial q^i} \end{aligned}$$

in the transformed variables. The principle of least action must hold for each pair:

$$\begin{aligned} S &= \int \left(p_i dx^i - H dt\right) \\ S' &= \int \left(\pi_i dq^i - H' dt\right) \end{aligned}$$

where S and S' differ by at most a constant. Correspondingly, the integrands may differ by the addition of a total differential, $df = \frac{df}{dt} dt$, since this will integrate to a surface term and therefore will not contribute to the variation. Notice that this corresponds exactly to a local dilatation, which produces a change

$$\begin{aligned} W'_\alpha dx^\alpha &= W_\alpha dx^\alpha - df \\ &= W_\alpha dx^\alpha - \frac{df}{dt} dt \end{aligned}$$

In general we may therefore write

$$p_i dx^i - H dt = \pi_i dq^i - H' dt + df$$

A convenient way to analyze the condition is to solve it for the differential df

$$df = p_i dx^i - \pi_i dq^i + (H' - H)\, dt$$

For the differential of f to take this form, it must be a function of x^i, q^i and t, that is, $f = f(x^i, q^i, t)$. Therefore, the differential of f is

$$df = \frac{\partial f}{\partial x^i} dx^i + \frac{\partial f}{\partial q^i} dq^i + \frac{\partial f}{\partial t} dt$$

Equating the expressions for df we match up terms to require

$$p_i = \frac{\partial f}{\partial x^i} \tag{77}$$

$$\pi_i = -\frac{\partial f}{\partial q^i} \tag{78}$$

$$H' = H + \frac{\partial f}{\partial t} \tag{79}$$

The first equation

$$p_i = \frac{\partial f(x^j, q^j, t)}{\partial x^i} \tag{80}$$

gives q^i implicitly in terms of the original variables, while the second determines π_i. Notice that we may pick any function $q^i = q^i(p_j, x^j, t)$. This choice fixes the form of π_i by the eq.(78), while the eq.(79) gives the new Hamiltonian in terms of the old one. The function f is the generating function of the transformation.

15 General solution in Hamiltonian dynamics

We conclude with the crowning theorem of Hamiltonian dynamics: a proof that for any Hamiltonian dynamical system there exists a canonical transformation to a set of variables on phase space such that the paths of motion reduce to single points. Clearly, this theorem shows the power of canonical transformations! The theorem relies on describing solutions to the Hamilton-Jacobi equation, which we introduce first.

15.1 The Hamilton-Jacobi Equation

We have the following equations governing Hamilton's principal function.

$$\begin{aligned} \frac{\partial \mathcal{S}}{\partial p_i} &= 0 \\ \frac{\partial \mathcal{S}}{\partial x_i} &= p_i \\ \frac{\partial \mathcal{S}}{\partial t} &= -H \end{aligned}$$

Since the Hamiltonian is a given function of the phase space coordinates and time, $H = H(x_i, p_i, t)$, we combine the last two equations:

$$\frac{\partial \mathcal{S}}{\partial t} = -H(x_i, p_i, t) = -H(x_i, \frac{\partial \mathcal{S}}{\partial x_i}, t)$$

This first order differential equation in $s+1$ variables $(t, x_i; i = 1, \ldots s)$ for the principal function $\mathcal{S}$ is the Hamilton-Jacobi equation. Notice that the Hamilton-Jacobi equation has the same general form as the Schrödinger equation and is equally difficult to solve for all but special potentials. Nonetheless, we are guaranteed that a complete solution exists, and we will assume below that we can find it. Before proving our central theorem, we digress to examine the exact relationship between the Hamilton-Jacobi equation and the Schrödinger equation.

15.2 Quantum Mechanics and the Hamilton-Jacobi equation

The Hamiltonian-Jacobi equation provides the most direct link between classical and quantum mechanics. There is considerable similarity between the Hamilton-Jacobi equation and the Schrödinger equation:

$$\begin{aligned} \frac{\partial \mathcal{S}}{\partial t} &= -H(x_i, \frac{\partial \mathcal{S}}{\partial x_i}, t) \\ i\hbar \frac{\partial \psi}{\partial t} &= H(\hat{x}_i, \hat{p}_i, t) \end{aligned}$$

We make the relationship precise as follows.

Suppose the Hamiltonian in each case is that of a single particle in a potential:

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{x})$$

Write the quantum wave function as

$$\psi = Ae^{\frac{i}{\hbar}\varphi}$$

The Schrödinger equation becomes

$$\begin{aligned}
i\hbar\frac{\partial\left(Ae^{\frac{i}{\hbar}\varphi}\right)}{\partial t} &= -\frac{\hbar^2}{2m}\bigtriangledown^2\left(Ae^{\frac{i}{\hbar}\varphi}\right) + V\left(Ae^{\frac{i}{\hbar}\varphi}\right) \\
i\hbar\frac{\partial A}{\partial t}e^{\frac{i}{\hbar}\varphi} - Ae^{\frac{i}{\hbar}\varphi}\frac{\partial\varphi}{\partial t} &= -\frac{\hbar^2}{2m}\bigtriangledown\cdot\left(e^{\frac{i}{\hbar}\varphi}\bigtriangledown A + \frac{i}{\hbar}Ae^{\frac{i}{\hbar}\varphi}\bigtriangledown\varphi\right) + VAe^{\frac{i}{\hbar}\varphi} \\
&= -\frac{\hbar^2}{2m}e^{\frac{i}{\hbar}\varphi}\left(\frac{i}{\hbar}\bigtriangledown\varphi\bigtriangledown A + \bigtriangledown^2 A\right) \\
&\quad -\frac{\hbar^2}{2m}e^{\frac{i}{\hbar}\varphi}\left(\frac{i}{\hbar}\bigtriangledown A\cdot\bigtriangledown\varphi + \frac{i}{\hbar}A\bigtriangledown^2\varphi\right) \\
&\quad -\frac{\hbar^2}{2m}\left(\frac{i}{\hbar}\right)^2 e^{\frac{i}{\hbar}\varphi}\left(A\bigtriangledown\varphi\cdot\bigtriangledown\varphi\right) \\
&\quad +VAe^{\frac{i}{\hbar}\varphi}
\end{aligned}$$

Then cancelling the exponential,

$$\begin{aligned}
i\hbar\frac{\partial A}{\partial t} - A\frac{\partial\varphi}{\partial t} &= -\frac{i\hbar}{2m}\bigtriangledown\varphi\bigtriangledown A - \frac{\hbar^2}{2m}\bigtriangledown^2 A \\
&\quad -\frac{i\hbar}{2m}\bigtriangledown A\cdot\bigtriangledown\varphi - \frac{i\hbar}{2m}A\bigtriangledown^2\varphi \\
&\quad +\frac{1}{2m}\left(A\bigtriangledown\varphi\cdot\bigtriangledown\varphi\right) + VA
\end{aligned}$$

Collecting by powers of $\hbar$,

$$\begin{aligned}
O\left(\hbar^0\right) &: -\frac{\partial\varphi}{\partial t} = \frac{1}{2m}\bigtriangledown\varphi\cdot\bigtriangledown\varphi + V \\
O\left(\hbar^1\right) &: \frac{1}{A}\frac{\partial A}{\partial t} = -\frac{1}{2m}\left(\frac{2}{A}\bigtriangledown A\cdot\bigtriangledown\varphi + \bigtriangledown^2\varphi\right) \\
O\left(\hbar^2\right) &: 0 = -\frac{\hbar^2}{2m}\bigtriangledown^2 A
\end{aligned}$$

The zeroth order terms is the Hamilton-Jacobi equation, with $\varphi = \mathcal{S}$:

$$\begin{aligned}
-\frac{\partial\mathcal{S}}{\partial t} &= \frac{1}{2m}\bigtriangledown\mathcal{S}\cdot\bigtriangledown\mathcal{S} + V \\
&= \frac{1}{2m}\mathbf{p}^2 + V(x)
\end{aligned}$$

where $p = \bigtriangledown\mathcal{S}$. Therefore, the Hamilton-Jacobi equation is the $\hbar \to 0$ limit of the Schrödinger equation.

15.3 Trivialization of the motion

We now seek a solution, in principle, to the complete mechanical problem. The solution is to find a canonical transformation that makes the motion trivial. Hamilton's principal function, the solution to the Hamilton-Jacobi equation, is the generating function of this canonical transformation.

To begin, suppose we have a solution to the Hamilton-Jacobi equation of the form

$$\mathcal{S} = g(t, x_1, \ldots, x_s, \alpha_1, \ldots, \alpha_s) + A$$

where the α_i and A provide $s+1$ constants describing the solution. Such a solution is called a *complete* integral of the equation, as opposed to a *general* integral which depends on arbitrary functions. We will show below that a complete solution leads to a general solution. We use $\mathcal{S}$ as a generating function.

Our canonical transformation will take the variables (x_i, p_i) to a new set of variables $\left(\beta^i, \alpha_i\right)$. Since $\mathcal{S}$ depends on the old coordinates x_i and the new momenta α_i, we have the relations

$$\begin{aligned} p_i &= \frac{\partial \mathcal{S}}{\partial x_i} \\ \beta_i &= \frac{\partial \mathcal{S}}{\partial \alpha_i} \\ H' &= H + \frac{\partial \mathcal{S}}{\partial t} \end{aligned}$$

Notice that the new Hamiltonian, H', vanishes because the Hamiltonian-Jacobi equation is satisfied by $\mathcal{S}$!. With $H' = 0$, Hamilton's equations in the new canonical coordinates are simply

$$\begin{aligned} \frac{d\alpha_i}{dt} &= \frac{\partial H'}{\partial \beta_i} = 0 \\ \frac{d\beta_i}{dt} &= -\frac{\partial H'}{\partial \alpha_i} = 0 \end{aligned}$$

with solutions

$$\begin{aligned} \alpha_i &= const. \\ \beta_i &= const. \end{aligned}$$

The system remains at the phase space point (α_i, β_i). To find the motion in the original coordinates as functions of time and the $2s$ constants of motion,

$$x_i = x_i(t; \alpha_i, \beta_i)$$

we can algebraically invert the s equations

$$\beta_i = \frac{\partial g(x_i, t, \alpha_i)}{\partial \alpha_i}$$

The momenta may be found by differentiating the principal function,

$$p_i = \frac{\partial \mathcal{S}(x_i, t, \alpha_i)}{\partial x_i}$$

Therefore, solving the Hamilton-Jacobi equation is the key to solving the full mechanical problem. Furthermore, we know that a solution exists because Hamilton's equations satisfy the integrability equation for $\mathcal{S}$.

We note one further result. While we have made use of a *complete* integral to solve the mechanical problem, we may want a *general* integral of the Hamilton-Jacobi equation. The difference is that a complete integral of an equation in $s+1$ variables depends on $s+1$ constants, while a general integral depends on s *functions*. Fortunately, a complete integral of the equation can be used to construct a general integral, and there is no loss of generality in considering a complete integral. We see this as follows. A complete solution takes the form

$$\mathcal{S} = g(t, x_1, \ldots, x_s, \alpha_1, \ldots, \alpha_s) + A$$

To find a general solution, think of the constant A as a function of the other s constants, $A(\alpha_1, \ldots, \alpha_s)$. Now replace each of the α_i by a function of the coordinates and time, $\alpha_i \to h_i(t, x_i)$. This makes $\mathcal{S}$ depend on

arbitrary functions, but we need to make sure it still solves the Hamilton-Jacobi equation. It will provided the partials of $\mathcal{S}$ with respect to the coordinates remain unchanged. In general, these partials are given by

$$\frac{\partial \mathcal{S}}{\partial x_i} = \left(\frac{\partial \mathcal{S}}{\partial x_i}\right)_{h_i=const.} + \left(\frac{\partial \mathcal{S}}{\partial h_k}\right)_{x=const.} \frac{\partial h_k}{\partial x_i}$$

We therefore still have solutions provided

$$\left(\frac{\partial \mathcal{S}}{\partial h_k}\right)_{x=const.} \frac{\partial h_k}{\partial x_i} = 0$$

and since we want h_k to be an arbitrary function of the coordinates, we demand

$$\left(\frac{\partial \mathcal{S}}{\partial h_k}\right)_{x=const.} = 0$$

Then

$$\frac{\partial \mathcal{S}}{\partial h_k} = \frac{\partial}{\partial h_k}\left(g(t, x_i, \alpha_i) + A(\alpha_i)\right) = 0$$

and we have

$$A(\alpha_1, \ldots, \alpha_s) = const. - g$$

This just makes A into some specific function of x^i and t.

Since the partials with respect to the coordinates are the same, and we haven't changed the time dependence,

$$\mathcal{S} = g(t, x_1, \ldots, x_s, h_1, \ldots, h_s) + A\left(h_i\right)$$

is a general solution to the Hamilton-Jacobi equation.

15.3.1 Example 1: Free particle

The simplest example is the case of a free particle, for which the Hamiltonian is

$$H = \frac{p^2}{2m}$$

and the Hamilton-Jacobi equation is

$$\frac{\partial S}{\partial t} = -\frac{1}{2m}\left(S'\right)^2$$

Let

$$S = f(x) - Et$$

Then $f(x)$ must satisfy

$$\frac{df}{dx} = \sqrt{2mE}$$

and therefore

$$\begin{aligned} f(x) &= \sqrt{2mE}x - c \\ &= \pi x - c \end{aligned}$$

where c is constant and we write the integration constant E in terms of the new (constant) momentum. Hamilton's principal function is therefore

$$S\left(x, \pi, t\right) = \pi x - \frac{\pi^2}{2m}t - c$$

Then, for a generating function of this type we have

$$\begin{aligned} p &= \frac{\partial S}{\partial x} = \pi \\ q &= \frac{\partial S}{\partial \pi} = x - \frac{\pi}{m}t \\ H' &= H + \frac{\partial S}{\partial t} = H - E \end{aligned}$$

Because $E = H$, the new Hamiltonian, H', is zero. This means that both q and π are constant. The solution for x and p follows immediately:

$$\begin{aligned} x &= q + \frac{\pi}{m}t \\ p &= \pi \end{aligned}$$

We see that the new canonical variables (q, π) are just the initial position and momentum of the motion, and therefore do determine the motion. The fact that knowing q and π is equivalent to knowing the full motion rests here on the fact that S generates motion along the classical path. In fact, given initial conditions (q, π), we can use Hamilton's principal function as a generating function but treat π as the *old* momentum and x as the *new* coordinate to reverse the process above and generate $x(t)$ and p.

15.3.2 Example 2: Simple harmonic oscillator

For the simple harmonic oscillator, the Hamiltonian becomes

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2$$

and the Hamilton-Jacobi equation is

$$\frac{\partial S}{\partial t} = -\frac{1}{2m}\left(S'\right)^2 + \frac{1}{2}kx^2$$

Letting

$$S = f(x) - Et$$

as before, $f(x)$ must satisfy

$$\frac{df}{dx} = \sqrt{2m\left(E - \frac{1}{2}kx^2\right)}$$

and therefore

$$\begin{aligned} f(x) &= \int \sqrt{2m\left(E - \frac{1}{2}kx^2\right)}dx \\ &= \int \sqrt{\pi^2 - mkx^2}dx \end{aligned}$$

where we have set $E = \frac{\pi^2}{2m}$. Now let $\sqrt{mk}x = \pi \sin y$. The integral is immediate:

$$\begin{aligned} f(x) &= \int \sqrt{\pi^2 - mkx^2}dx \\ &= \frac{\pi^2}{\sqrt{mk}} \int \cos^2 y dy \\ &= \frac{\pi^2}{2\sqrt{mk}}\left(y + \sin y \cos y\right) \end{aligned}$$

Hamilton's principal function is therefore

$$\begin{aligned} S(x,\pi,t) &= \frac{\pi^2}{2\sqrt{mk}}\left(\sin^{-1}\left(\sqrt{mk}\frac{x}{\pi}\right)+\sqrt{mk}\frac{x}{\pi}\sqrt{1-mk\frac{x^2}{\pi^2}}\right) \\ &\quad -\frac{\pi^2}{2m}t-c \\ &= \frac{\pi^2}{2\sqrt{mk}}\sin^{-1}\left(\sqrt{mk}\frac{x}{\pi}\right)+\frac{x}{2}\sqrt{\pi^2-mkx^2}-\frac{\pi^2}{2m}t-c \end{aligned}$$

and we may use it to generate the canonical change of variable.

This time we have

$$\begin{aligned} p &= \frac{\partial S}{\partial x} \\ &= \frac{\pi}{2}\frac{1}{\sqrt{1-mk\frac{x^2}{\pi^2}}}+\frac{1}{2}\sqrt{\pi^2-mkx^2}+\frac{x}{2}\frac{-mkx}{\sqrt{\pi^2-mkx^2}} \\ &= \frac{1}{\sqrt{\pi^2-mkx^2}}\left(\frac{\pi^2}{2}+\frac{1}{2}\left(\pi^2-mkx^2\right)-\frac{mkx^2}{2}\right) \\ &= \sqrt{\pi^2-mkx^2} \\ q &= \frac{\partial S}{\partial \pi} \\ &= \frac{\pi}{\sqrt{mk}}\sin^{-1}\left(\sqrt{mk}\frac{x}{\pi}\right)+\frac{\pi^2}{2\sqrt{mk}}\frac{1}{\sqrt{1-mk\frac{x^2}{\pi^2}}}\left(-\sqrt{mk}\frac{x}{\pi^2}\right) \\ &\quad +\frac{x}{2}\frac{\pi}{\sqrt{\pi^2-mkx^2}}-\frac{\pi}{m}t \\ &= \frac{\pi}{\sqrt{mk}}\sin^{-1}\left(\sqrt{mk}\frac{x}{\pi}\right)-\frac{\pi}{m}t \\ H' &= H+\frac{\partial S}{\partial t}=H-E=0 \end{aligned}$$

The first equation relates p to the energy and position, the second gives the new position coordinate q, and third equation shows that the new Hamiltonian is zero. Hamilton's equations are trivial, so that π and q are constant, and we can invert the expression for q to give the solution. Setting $\omega=\sqrt{\frac{k}{m}}$, the solution is

$$\begin{aligned} x(t) &= \frac{\pi}{m\omega}\sin\left(\frac{m\omega}{\pi}q+\omega t\right) \\ &= A\sin(\omega t+\phi) \end{aligned}$$

where

$$\begin{aligned} q &= A\phi \\ \pi &= Am\omega \end{aligned}$$

The new canonical coordinates therefore characterize the initial amplitude and phase of the oscillator.

15.3.3 Example 3: One dimensional particle motion

Now suppose a particle with one degree of freedom moves in a potential $U(x)$. Little is changed. The the Hamiltonian becomes

$$H=\frac{p^2}{2m}+U$$

and the Hamilton-Jacobi equation is

$$\frac{\partial S}{\partial t} = -\frac{1}{2m}\left(S'\right)^2 + U(x)$$

Letting

$$S = f(x) - Et$$

as before, $f(x)$ must satisfy

$$\frac{df}{dx} = \sqrt{2m\left(E - U\left(x\right)\right)}$$

and therefore

$$\begin{aligned} f(x) &= \int \sqrt{2m\left(E - U\left(x\right)\right)}dx \\ &= \int \sqrt{\pi^2 - 2mU\left(x\right)}dx \end{aligned}$$

where we have set $E = \frac{\pi^2}{2m}$. Hamilton's principal function is therefore

$$S\left(x, \pi, t\right) = \int \sqrt{\pi^2 - 2mU\left(x\right)}dx - \frac{\pi^2}{2m}t - c$$

and we may use it to generate the canonical change of variable.

This time we have

$$\begin{aligned} p &= \frac{\partial S}{\partial x} = \sqrt{\pi^2 - 2mU\left(x\right)} \\ q &= \frac{\partial S}{\partial \pi} = \frac{\partial}{\partial \pi}\left(\int_{x_0}^{x} \sqrt{\pi^2 - 2mU\left(x\right)}dx\right) - \frac{\pi}{m}t \\ H' &= H + \frac{\partial S}{\partial t} = H - E = 0 \end{aligned}$$

The first and third equations are as expected, while for q we may interchange the order of differentiation and integration:

$$\begin{aligned} q &= \frac{\partial}{\partial \pi}\left(\int \sqrt{\pi^2 - 2mU\left(x\right)}dx\right) - \frac{\pi}{m}t \\ &= \int \frac{\partial}{\partial \pi}\left(\sqrt{\pi^2 - 2mU\left(x\right)}\right)dx - \frac{\pi}{m}t \\ &= \int \frac{\pi dx}{\sqrt{\pi^2 - 2mU\left(x\right)}} - \frac{\pi}{m}t \end{aligned}$$

To complete the problem, we need to know the potential. However, even without knowing $U(x)$ we can make sense of this result by combining the expression for q above to our previous solution to the same problem. There, conservation of energy gives a first integral to Newton's second law,

$$\begin{aligned} E &= \frac{p^2}{2m} + U \\ &= \frac{1}{2}m\left(\frac{dx}{dt}\right)^2 + U \end{aligned}$$

so we arrive at the familiar quadrature

$$t - t_0 = \int dt = \int_{x_0}^{x} \frac{mdx}{\sqrt{2m\left(E - U\right)}}$$

Substituting into the expression for q,

$$\begin{aligned} q &= \int^{x} \frac{\pi dx}{\sqrt{\pi^2 - 2mU(x)}} - \frac{\pi}{m}\int_{x_0}^{x} \frac{m dx}{\sqrt{2m(E-U)}} - \frac{\pi}{m}t_0 \\ &= \int^{x} \frac{\pi dx}{\sqrt{\pi^2 - 2mU(x)}} - \int_{x_0}^{x} \frac{\pi dx}{\sqrt{\pi^2 - 2mU(x)}} - \frac{\pi}{m}t_0 \\ &= \int^{x_0} \frac{\pi dx}{\sqrt{\pi^2 - 2mU(x)}} - \frac{\pi}{m}t_0 \end{aligned}$$

We once again find that q is a constant characterizing the initial configuration. Since t_0 is the time at which the position is x_0 and the momentum is p_0, we have the following relations:

$$\frac{p^2}{2m} + U(x) = \frac{p_0^2}{2m} + U(x_0) = E = const.$$

and

$$t - t_0 = \int_{x_0}^{x} \frac{dx}{\sqrt{\frac{2}{m}(E-U)}}$$

which we may rewrite as

$$t - \int^{x} \frac{dx}{\sqrt{\frac{2}{m}(E-U)}} = t_0 - \int^{x_0} \frac{dx}{\sqrt{\frac{2}{m}(E-U)}} = \frac{m}{\pi}q = const.$$

Part IV
Bonus sections

We did not have time for the following topics or their applications, but you may find them interesting or useful. The section on gauge theory is not complete, but has a full treatment of differential forms.

16 Classical spin, statistics and pseudomechanics

16.1 Spin

Now that we have a gauge theory of mechanics, we can ask further about the representation of the gauge symmetry. A representation of a group is the vector space on which the group acts. The largest class of objects on which our symmetry acts will be the class determining the *covering group*. This achieves the fullest realization of our symmetry. For example, while the Euclidean group $ISO(3)$ leads us to the usual formulation of Lagrangian mechanics, we can ask if we might not achieve something new by gauging the covering group, $ISpin(3) \cong ISU(2)$. This extension, which places spinors in the context of classical physics, depends only on symmetry, and therefore is completely independent of quantization.

There are numerous advantages to the spinorial extension of classical physics. After Cartan's discovery of spinors as linear representations of orthogonal groups in 1913 ([61],[62]) and Dirac's use of spinors in the Dirac equation ([63],[64]), the use of spinors for other areas of relativistic physics was pioneered by Penrose ([65],[66]). Penrose developed spinor notation for general relativity that is one of the most powerful tools of the field. For example, the use of spinors greatly simplifies Petrov's classification of spacetimes (compare Petrov [67] and Penrose ([65],[68]), and tremendously shortens the proof of the positive mass theorem (compare Schoen and Yau ([69],[70],[71]) and Witten [72]). Penrose also introduced the idea and techniques of *twistor spaces*. While Dirac spinors are representations of the Lorentz symmetry of Minkowski space, twistors are the spinors associated with larger conformal symmetry of compactified Minkowski space.

Their overlap with string theory as twistor strings is an extremely active area of current research in quantum field theory (see [73] and references thereto). In nonrelativistic classical physics, the use of Clifford algebras (which, though they do not provide a spinor representation in themselves, underlie the definition of the spin groups) has been advocated by Hestenes in the "geometric algebra" program [9].

It is straightforward to include spinors in a classical theory. We provide a simple example.

For the rotation subgroup of the Euclidean group, we can let the group act on complex 2-vectors, χ^a, $a = 1, 2$. The resulting form of the group is $SU(2)$. In this representation, an ordinary 3-vector such as the position vector x^i is written as a traceless Hermitian matrix,

$$\begin{aligned} X &= x^i \sigma_i \\ [X]^{ab} &= x^i \left[\sigma_i\right]^{ab} \end{aligned}$$

where σ_i are the Pauli matrices. It is easy to write the usual Lagrangian in terms of X :

$$L = \frac{m}{4} tr\left(\dot{X}\dot{X}\right) - V\left(X\right)$$

where V is any scalar-valued function of X. However, we now have the additional complex 2-vectors, χ^a, available. Consider a Dirac-type kinetic term

$$\lambda \chi_a \left(i\dot{\chi}^a - \mu\chi^a\right)$$

and potential

$$V\left(\chi^a\right) = \lambda \bar{\chi}^a B^i \sigma_{iab} \chi^b + \ldots$$

Notice there is no necessity to introduce fermions and the concomitant anticommutation relations – we regard these spinors as commuting variables. A simple action therefore takes the form

$$S = \int dt \left(\frac{m}{4} tr\left(\dot{X}\dot{X}\right) + \bar{\chi}_a \left(i\dot{\chi}^a - \mu\chi^a\right) - V\left(X\right) - \lambda \bar{\chi}^a B^i \sigma_{iab} \chi^b\right)$$

The equations of motion are then

$$\begin{aligned} m\ddot{x}^i &= -\sigma^{iab} \frac{\partial V}{\partial X^{ab}} \\ \dot{\chi}^a &= -i\mu\chi^a - i\lambda B^i \sigma_{iab} \chi^b \end{aligned}$$

together with the complex conjugate of the second. The first reproduces the usual equation of motion for the position vector. Assuming a constant vector B^i, we can easily solve the second. Setting $\chi = \psi e^{-i\mu t}$, ψ must satisfy,

$$\dot{\psi} = -i\lambda B^i \sigma_i{}^{ab} \psi_b$$

This describes steady rotation of the spinor,

$$\psi = e^{-i\lambda B} \psi_0$$

The important thing to note here is that, while the spinors ψ rotate with a single factor of $e^{i\mathbf{w}\cdot\sigma}$, a vector such as X rotates as a matrix and therefore requires two factors of the rotation

$$X' = e^{-i\mathbf{w}\cdot\sigma} X e^{i\mathbf{w}\cdot\sigma}$$

This illustrates the 2 : 1 ratio of rotation angle characteristic of spin 1/2. The new degrees of freedom therefore describe classical spin and we see that spin is best thought of as a result of the symmetries of classical physics, rather than as a necessarily quantum phenomenon. Similar results using the covering group of the Lorentz group introduce Dirac spinors naturally into relativity theory. Indeed, as noted above, 2-component spinor notation is a powerful tool in general relativity, where it makes such results as the Petrov classification or the positivity of mass transparent.

16.2 Statistics and pseudomechanics

The use of spinors brings immediately to mind the exclusion principle and the spin-statistics theorem. We stressed that spin and statistics are independent. Moreover, spin, as described above, follows from the use of the covering group of any given orthogonal group and is therefore classical. For statistics, on the other hand, the situation is not so simple. In quantum mechanics, the difference between Bose-Einstein and Fermi-Dirac statistics is a consequence of the combination of anticommuting variables with the use of discrete states. In classical physics we do not have discrete states. However, nothing prevents us from introducing anticommuting variables. In its newest form, the resulting area of study is called *pesudomechanics*.

The use of anticommuting, or Grassmann variables in classical physics actually has an even longer history than spin. The oldest and most ready example is the use of the wedge product for differential forms

$$\mathbf{d}x \wedge \mathbf{d}y = -\mathbf{d}y \wedge \mathbf{d}x$$

This gives a grading of $(-)^p$ to all p-forms. Thus, if ω is a p-form and η a q-form,

$$\begin{aligned} \omega &= \omega_{i_1 \cdots i_p} \mathbf{d}x^{i_i} \wedge \cdots \wedge \mathbf{d}x^{i_p} \\ \eta &= \omega_{i_1 \cdots i_q} \mathbf{d}x^{i_1} \wedge \cdots \wedge \mathbf{d}x^{i_q} \end{aligned}$$

Then their (wedge) product is even or odd depending on whether pq is even or odd:

$$\omega \wedge \eta = (-)^{pq}\, \eta \wedge \omega$$

Nonetheless, p-forms rotate as rank-$\binom{0}{p}$ tensors under $SO(3)$ (or $SO(n)$), in violation of the familiar spin-statistics theorem. Under $SU(2)$ they rotate as rank-$\binom{0}{2p}$ tensors, not as spinors.

Another appearance of anticommuting variables in classical mechanics stems from the insights of supersymmetric field theory. Before supersymmetry, continuous symmetries in classical systems were characterized by Lie algebras, with each element of the Lie algebra generating a symmetry transformation. The Lie algebra is a vector space characterized by a closed commutator product and the Jacobi identity. Supersymmetries are extensions of the normal Lie symmetries of physical systems to include symmetry generators (Grassmann variables) that anticommute. Like the grading of differential forms, all transformations of the graded Lie algebra are assigned a grading, 0 or 1, that determines whether a commutator or commutator is appropriate, according to

$$[T_p, T_q] \equiv T_p T_q - (-)^{pq}\, T_q T_p$$

where $p, q \in \{0, 1\}$. Thus, two transformations which both have grading 1 have anticommutation relations with one another, while all other combinations satisfy commutation relations.

Again, there is nothing intrinsically "quantum" about such generalized symmetries, so we can consider classical supersymmetric field theories and even supersymmetrized classical mechanics. Since anticommuting fields correspond to fermions in quantum mechanics, we may continue to call variables fermionic when used classically, even though their statistical properties may not be Fermi-Dirac. Perhaps more importantly, we arrive at a class of classical action functionals whose quantization leads directly to Pauli or Dirac spinor equations.

The development of pseudomechanics was pioneered by Casalbuoni ([74], [75], see also Freund [76]), who showed that it was possible to formuate an $\hbar \to 0$ limit of a quantum system in such a way that the spinors remain but their magnitude is no longer quantized. Conversely, the resulting classical action leads to the Pauli-Schrödinger equation when quantized. Similarly, Berezin and Marinov [77], and Brink, Deser, Zumino, di Vecchia and Howe [78] introduced four anticommuting variables, θ^α to write the pre-Dirac action. We display these actions below, after giving a simplified example. Since these approaches moved from quantum fields to classical equations, they already involved spinor representations. However, vector versions (having anticommuting variables without spinors) are possible as well. Our example below is of the latter type. Our development is a slight modification of that given by Freund [76].

To construct a simple pseudomechanical model, we introduce a superspace formulation, extending the usual "bosonic" 3-space coordinates x_i by three additional anticommuting coordinates, θ^a,

$$\left\{\theta^a, \theta^b\right\} = 0$$

Consider the motion of a particle described by $\left(x_i\left(t\right), \theta^a\left(t\right)\right)$, and the action functional

$$S = \int dt \left(\frac{1}{2} m \dot{x}^i \dot{x}^i + \frac{i}{2} \theta^a \dot{\theta}^a - V\left(x^i, \theta^b\right)\right)$$

Notice that $\theta^2 = 0$ for any anticommuting variable, so the linear velocity term is the best we can do. For the same reason, the Taylor series in θ^a of the potential $V\left(x^i, \theta^b\right)$ terminates:

$$V\left(x^i, \theta^b\right) = V_0\left(x^i\right) + \psi_a\left(x^i\right)\theta^a + \frac{1}{2}\varepsilon_{abc} B^a\left(x^i\right)\theta^b\theta^c + \frac{1}{3!}\kappa\left(x^i\right)\varepsilon_{abc}\theta^a\theta^b\theta^c$$

Since the coefficients remain functions of x^i, we have introduced four new fields into the problem. However, they are not all independent. If we change coordinates from θ^a to some new anticommuting variables, setting

$$\begin{aligned} \theta^a &= \chi^a + \xi B^a_{bc}\chi^b\chi^c + C^a\varepsilon_{bcd}\chi^b\chi^c\chi^d \\ B^a_{bc} &= B^a_{[bc]} \end{aligned}$$

where ζ is an anticommuting constant, the component functions in $H\left(\theta^b\right)$ change according to

$$\begin{aligned} V &= V_0 + \psi_a\chi^a + \left(\psi_a\xi B^a_{bc} + \frac{1}{2}\varepsilon_{abc}B^a\right)\chi^b\chi^c \\ &\quad + \left(\varepsilon_{afb}B^a\xi B^f_{cd} + \frac{1}{3!}\kappa\varepsilon_{bcd} + \psi_a C^a\varepsilon_{bcd}\right)\chi^b\chi^c\chi^d \end{aligned}$$

The final term vanishes if we choose

$$\xi B^a_{bc} = \frac{\kappa + 6\psi_a C^a}{4B^2}\left(\delta^a_b B_c - \delta^a_c B_b\right)$$

while no choice of B^a_{bc} can make the second term vanish because $\psi_a\xi B^a_{bc}$ is nilpotent while $\frac{1}{2}\varepsilon_{abc}B^a$ is not. Renaming the coefficient functions, V takes the form

$$V\left(\theta^b\right) = V_0 + \psi^a\theta^a + \frac{1}{2}\varepsilon_{abc}B^a\theta^b\theta^c$$

Now, without loss of generality, the action takes the form

$$S = \int dt \left(\frac{1}{2} m \dot{x}^i \dot{x}^i + \frac{i}{2}\theta^a\dot{\theta}^a - V_0 - \psi_a\theta^a - \frac{1}{2}\varepsilon_{abc}B^a\theta^b\theta^c\right)$$

Varying, we get two sets of equations of motion:

$$\begin{aligned} m\ddot{x}^i &= -\frac{\partial V}{\partial x^i} \\ &= -\frac{\partial V_0}{\partial x^i} + \frac{\partial\psi^a}{\partial x^i}\theta^a + \frac{1}{2}\varepsilon_{abc}\frac{\partial B^a}{\partial x^i}\theta^b\theta^c \\ \dot{\theta}^a &= i\psi^a + i\varepsilon^a{}_{bc}B^b\theta^c \end{aligned}$$

Clearly this generalizes Newton's second law. The coefficients in the first equation depend only on x^i, so terms with different powers of θ^a must vanish separately. Therefore, B^a and ψ^a are constant and we can integrate the θ^a equation immediately. Since $\left[J_b\right]^a{}_c = \varepsilon_{cb}{}^a$ satisfies

$$\left[J_a, J_b\right]^c{}_d = \varepsilon^e{}_{ba}\left[J_e\right]^c{}_d$$

we see that $B^b\varepsilon^a{}_{bc}$ is an element of the Lie algebra of $SO(3)$. Exponentiating to get an element of the rotation group, the solution for θ^a is

$$\theta^a = i\psi^a t + e^{iB^b t\varepsilon^a{}_{bc}}\theta_0^c$$

The solution for x^i depends on the force, $-\frac{\partial V}{\partial x^i}$, in the usual way.

It is tempting to interpret the θ^a variables as spin degrees of freedom and B^a as the magnetic field. Then the solution shows that the spin precesses in the magnetic field. However, notice that $B^b\varepsilon^a{}_{bc}$ is in $SO(3)$, not the spin group $SU(2)$. The coordinates θ^a therefore provide an example of fermionic, spin-1 objects.

One of the goals of early explorations of pseudomechanics was to ask what classical equations lead to the Pauli and Dirac equations when quantized. Casalbuoni ([74],[75], see also [76]) showed how to introduce classical, anticommuting spinors using an $\hbar \to 0$ limit of a quantum system. Conversely, the action

$$S = \int dt \left(\frac{1}{2} m\dot{\mathbf{x}}^2 + \frac{i}{2}\theta^a\dot{\theta}^a - V_0(\mathbf{x}) - (\mathbf{L}\cdot\mathbf{S})\, V_{LS} - \kappa\frac{1}{2}(\mathbf{S}\cdot\mathbf{B}) \right)$$

where $\mathbf{L}$ is the orbital angular momentum, $\mathbf{S} = -\frac{i}{2}\varepsilon^a{}_{bc}\theta^b\theta^c$, and V_{LS} is a spin-orbit potential, leads to the Pauli-Schrödinger equation when quantized. Similarly, Berezin and Marinov [77], Brink, Deser, Zumino, and di Vecchia and Howe [78] introduced four anticommuting variables, θ^α to write the pre-Dirac action,

$$S_{Dirac} = \int d\lambda \left(-m\sqrt{-v^\alpha v_\alpha} + \frac{i}{2}\left(\theta_\beta \frac{d\theta^\beta}{d\lambda} + u_\alpha\theta^\alpha u_\beta \frac{d\theta^\beta}{d\lambda} - \alpha\left(u_\alpha\theta^\alpha + \theta_5\right)\right)\right)$$

where

$$\begin{aligned} v^\alpha &= \frac{dx_\alpha}{d\lambda} \\ u^\alpha &= \frac{v^\alpha}{\sqrt{-v^2}} \end{aligned}$$

and α is a Lagrange multiplier. The action, S_{Dirac}, is both reparameterization invariant and Lorentz invariant. Its variation leads to the usual relativistic mass-energy-momentum relation together with a constraint. When the system is quantized, imposing the constraint on the physical states gives the Dirac equation.

Evidently, the quantization of these actions is also taken to include the entension to the relevant covering group.

16.3 Spin-statistics theorem

Despite the evident classical independence of spin and statistics, there exists a limited spin-statistics theorem due to Morgan [79]. The theorem is proved from Poincaré invariance, using extensive transcription of quantum methods into the language of Poisson brackets – an interesting accomplishment in itself. A brief statement of the theorem is the following:

Let L be a pseudoclassical, Poincaré-invariant Lagrangian, built quadratically from the dynamical variables. If L is invariant under the combined action of charge conjugation (C) and time reversal (T) then integer spin variables are even Grassmann quantities while odd-half-integer spin variables are odd Grassmann quantities.

The proof relies on extending the quantum notions of charge conjugation and time reversal. As in quantum mechanics, charge conjugation is required to include complex conjugation. For fermionic variables, Morgan requires reversal of the order of Grassmann variables under conjugation

$$(\eta\xi)^* = \xi^*\eta^*$$

This insures the reality property $(\eta\xi^*)^* = \eta\xi^*$, but this is not a necessary condition for complex Grassmann numbers. For example, the conjugate of the complex 2-form

$$\mathbf{d}z \wedge \mathbf{d}z^*$$

is clearly just

$$\mathbf{d}z^* \wedge \mathbf{d}z$$

and is therefore pure imaginary. We must therefore regard the TC symmetry required by the proof as somewhat arbitrary.

Similarly, for time reversal, [79] requires both

$$\begin{aligned} t &\rightarrow -t \\ \tau &\rightarrow -\tau \end{aligned}$$

Whether this is an allowed Poincaré transformation depends on the precise definition of the symmetry. If we define Poincaré transformations as those preserving the infintesimal line element, $d\tau$, then reversing proper time is not allowed. Of course, we could define Poincaré transformations as preserving the quadratic form, $d\tau^2 = g_{\alpha\beta}dx^\alpha dx^\beta$, in which case the transformation is allowed.

Despite its shortcomings, the proof is interesting because it identifies a set of conditions under which a classical pseudomechanics action obeys the spin statistics theorem. This is an interesting class of theories and it would be worth investigating further. Surely there is some set of properties which can be associated with the classical version of the theorem. Perhaps a fruitful approach would be to assume the theorem and derive the maximal class of actions satisfying it.

There are other questions we might ask of spinorial and graded classical mechanics. A primary question is whether there are any actual physical systems which are well modeled by either spinors or graded variables. If such systems exist, are any of them supersymmetric? What symmetries are associated with spinorial and fermionic variables? Is there a generalization of the Noether theorem to these variables? What are the resulting conserved quantities? What is the supersymmetric extension of familiar problems such as the Kepler or harmonic oscillator?

The statistical behavior of fermionic classical systems is not clear. Quantum mechanically, of course, Fermi-Dirac statistics follow from the limitation of discrete states to single occupancy. This, in turn, follows from the action of an anticommuting raising operator on the vacuum:

$$\begin{aligned} a^\dagger |0\rangle &= |1\rangle \\ a^\dagger a^\dagger &= 0 \end{aligned}$$

Since classical states are not discrete, there may be no such limitation. Do anticommuting classical variables therefore satisfy Bose-Einstein statistics? If so, how do Fermi-Dirac quantum states become Bose-Einstein in the classical limit?

The introduction of pseudomechanics has led to substantial formal work on supermanifolds and symplectic supermanifolds. See [80], [81] and references therein.

17 Gauge theory

Recall our progress so far. Starting from our direct experience of the world, we have provisionally decided to work in a 3-dimensional space, together with time. Because we want to associate physical properties with objects which move in this space rather than with the space itself, we demand that the space be homogeneous and isotropic. This led us to the construction of Euclidean 3-space. Next, in order to make measurements in this space, we introduced a metric. In order to describe uniform motion, we digressed to study the functional derivative. Vanishing functional derivative gives us a criterion for a straight line, or geodesic.

The next question we addressed concerned the description of matter. Since we do not want our description to depend on the way we choose to label points, we sought quantities which are invariant under the relevant transformations. We found that, in order to construct invariant, measurable quantities, it is useful to introduce tensors. Equations relating tensors do not change form when we change coordinates, and scalars formed from tensors are invariant.

We are now ready to describe motion, by which we mean the evolution in time of various measurable quantities. Foremost among these physical observables, of course, is the position of the particle or other object of study. But the measurable quantities also include invariant formed from the velocity, momentum and so on.

In order to describe the evolution of vectors and other tensors, we need to associate a vector space with each point of the physical arena – that is, we choose a set of reference frames. However, because the Newtonian, Lagrangian, and Hamiltonian formulations of mechanics are based on different symmetries, the arenas are different. Thus, while the arena for Newtonian mechanics is Euclidean 3-space, the arena for Lagrangian mechanics may be larger or smaller depending on the number of degrees of freedom and the number of constraints, while the arena for Hamiltonian dynamics differs not only in dimension but also in the underlying symmetry. As a result, when we assign a basis to the relevant manifold, we will require different symmetries of the basis in different formulations.

Once we know the symmetry we wish to work with and have selected the relevant class of reference frames, we need to know how a frame at one point and time is related to a frame at another. Our formulation will be general enough to allow arbitrary changes of frame from place to place and time to time. The tool that allows this is the connection. Just as the metric gives the distance between two nearby points of a manifold, the connection tells us how two infinitesimally separated reference frames are related.

To introduce connections for arbitrary symmetries, we need to develop two tools: Lie algebras to describe the symmetry, and differential forms to cast the problem in a coordinate invariant form.

We can then turn to our discussion of motion.

17.1 Group theory

Recall our definition of a group and a Lie group:

A *group* is a pair $G = \{S, \circ\}$ where S is a set and $\circ$ is an operation mapping pairs of elements in S to elements in S (i.e., $\circ : S \times S \to S$. This implies closure) and satisfying the following conditions:

1. Existence of an identity: $\exists\ e \in S$ such that $e \circ a = a \circ e = a$, $\forall a \in S$.
2. Existence of inverses: $\forall\ a \in S$, $\exists\ a^{-1} \in S$ such that $a \circ a^{-1} = a^{-1} \circ a = e$.
3. Associativity: $\forall\ a, b, c \in S$, $a \circ (b \circ c) = (a \circ b) \circ c = a \circ b \circ c$

A Lie group is a group $G = \{S, \circ\}$ for which the set S is a manifold.

We have considered several familiar examples of groups. Here we describe a few of the most important classes of Lie group.

The set of non-degenerate linear transformations of a real, n-dimensional vector space forms a Lie group. This class of Lie groups is important enough to have its own name: $GL(n; R)$, or more simply, $GL(n)$ when the field (usually R or C) is unambiguous. The GL stands for General Linear. The transformations may be represented by $n \times n$ matrices with nonzero determinant. For any $A \in GL(n; R)$ we demand $\det A \neq 0$,

since the matrix A must be invertible. The identity is the identity matrix. To see that matrix multiplication is associative we simply note the equality of the double sums

$$M_{ij}\left(M_{jk}M_{kl}\right) = \left(M_{ij}M_{jk}\right)M_{kl}$$

Since each A can be written in terms of n^2 real numbers, $GL\left(n\right)$ has dimension n^2 (note that the nonvanishing determinant does not reduce the number of real numbers required to specify the transformations). $GL\left(n\right)$ provides an example of a Lie group with more than one *connected component*. We can imagine starting with the identity element and smoothly varying the parameters that define the group elements, thereby sweeping out curves in the space of all group elements. If such continuous variation can take us to every group element, we say the group is *connected*. If there remain elements that cannot be connected to the identity by such a continuous variation (actually a curve in the group manifold), then the group has more than one *component*. $GL\left(n\right)$ is of this form because as we vary the parameters to move from element to element of the group, the determinant of those elements also varies smoothly. But since the determinant of the identity is 1 and no element can have determinant zero, we can never get to an element that has negative determinant. The elements of $GL\left(n\right)$ with negative determinant are related to those of positive determinant by a discrete transformation: if we pick any element of $GL\left(n\right)$ with negative determinant, and multiply it by each element of $GL\left(n\right)$ with positive determinant, we get a new element of negative determinant. This shows that the two components of $GL\left(n\right)$ are in 1 to 1 correspondence. In odd dimensions, a suitable 1 to 1 mapping is given by $-\mathbf{1}$, which is called the *parity* transformation.

We will be concerned with *linear representations* of Lie groups. As described previously, this means that each group element may be written as a matrix and the group multiplication is correctly given by the usual form of matrix multiplication. Since $GL\left(n\right)$ is the set of *all* linear, invertible transformations in n-dimensions, all Lie groups with linear representations must be subgroups of $GL\left(n\right)$ for some n and some field. We now look at three classes of such subgroups.

The simplest subgroup of $GL\left(n\right)$ removes the second component to give a connected Lie group. In fact, it is useful to factor out the determinant entirely, because the operation of multiplying by a constant commutes with every other transformation of the group. In this way, we arrive at a *simple group*, one in which each transformation has nontrivial effect on some other transformations. For a general matrix $A \in GL\left(n\right)$ with positive determinant, let

$$A = \left(\det A\right)^{\frac{1}{n}} \ \hat{A}$$

Then $\det \hat{A} = 1$. Since

$$\det\left(\hat{A}\hat{B}\right) = \det\hat{A}\det\hat{B} = 1$$

the set of all $\hat{A}$ closes under matrix multiplication. We also have $\det\hat{A}^{-1} = 1$, and $\det 1 = 1$, so the set of all $\hat{A}$ forms a Lie group. This group is called the Special Linear group, $SL(n)$, where special refers to the unit determinant.

Frequently, the most useful way to characterize a group is by a set of objects that group transformations leave invariant. In this way, we produce the orthogonal, unitary and symplectic groups. Let us examine the general case of an invariant matrix, M. We start with a theorem.

Consider the subset of $GL\left(n;R\right)$ that leaves a fixed matrix M invariant under a similarity transformation:

$$H = \left\{A | A \in GL(n), AMA^{-1} = M\right\}$$

Then H is also a Lie group.

Proof: First, H is closed, since if

$$\begin{aligned} AMA^{-1} &= M \\ BMB^{-1} &= M \end{aligned}$$

then the product AB is also in H because

$$\left(AB\right)M\left(AB\right)^{-1} = ABMB^{-1}A^{-1}$$

$$\begin{aligned} &= A\left(BMB^{-1}\right)A^{-1} \\ &= AMA^{-1} \\ &= M \end{aligned}$$

The identity is present because

$$IMI^{-1} = M$$

To see that H includes inverses of all its elements, notice that $\left(A^{-1}\right)^{-1} = A$. Using this, we start with

$$AMA^{-1} = M$$

and multiply on the left by A^{-1} and on the right by A to get

$$\begin{aligned} A^{-1}AMA^{-1}A &= A^{-1}MA \\ M &= A^{-1}MA \\ &= \left(A^{-1}\right)M\left(A^{-1}\right)^{-1} \end{aligned}$$

The last line is the statement that A^{-1} leaves M invariant, and is therefore in H. Finally, matrix multiplication is associative, so H is a group, concluding our proof.

Consider the subset of $GL(n;R)$ that leaves a fixed matrix M invariant under a transpose-similarity transformation:

$$H' = \left\{A | A \in GL(n), AMA^t = M\right\}$$

Show that H' is a Lie group.

Now, fix a (nondegenerate) matrix M and consider the group,

$$H' = \left\{A | A \in GL(n), AMA^t = M\right\}$$

that leaves M invariant. Suppose M is generic, so it has nondegenerate symmetric and antisymmetric parts:

$$\begin{aligned} M &= \frac{1}{2}\left(M + M^t\right) + \frac{1}{2}\left(M - M^t\right) \\ &\equiv M_s + M_a \end{aligned}$$

Then, for any A in H,

$$AMA^t = M$$

implies

$$A\left(M_s + M_a\right)A^t = \left(M_s + M_a\right) \tag{81}$$

The transpose of this equation must also hold,

$$\begin{aligned} A\left(M_s^t + M_a^t\right)A^t &= \left(M_s^t + M_a^t\right) \\ A\left(M_s - M_a\right)A^t &= \left(M_s - M_a\right) \end{aligned} \tag{82}$$

so adding and subtracting eqs.(81) and (82) gives two independent constraints on A :

$$\begin{aligned} AM_sA^t &= M_s \\ AM_aA^t &= M_a \end{aligned}$$

That is, A must separately leave the M_s and M_a invariant. Therefore, the largest subgroups of G that we can form in this way are found by demanding that M be either symmetric or antisymmetric. These conditions define the *orthogonal* and *symplectic* groups, respectively. We look at each case.

If M is symmetric, then we can always choose a basis for the vector space on which the transformations act in such a way that M is diagonal; indeed we can go further, for rescaling the basis we can make every diagonal element into $+1$ or -1. Therefore, any symmetric M may be put in the form

$$M \to \eta_{ij}^{(p,q)} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & \ddots & \\ & & & & & -1 \end{pmatrix} \tag{83}$$

where there are p terms $+1$ and q terms -1. Then η provides a pseudo-metric, since for any vector v^i, we may define

$$\langle v, v \rangle = \eta_{ij} v^i v^j = \sum_{i=1}^{p} \left(v^i\right)^2 - \sum_{i=p+1}^{p+q} \left(v^i\right)^2$$

(We use summation signs here because we are violating the convention of always summing one raised and one lowered index.). The *signature* of η is $s = p - q$ and the dimension is $n = p + q$. Groups preserving a (p, q) pseudo-metric are the (pseudo-)orthogonal groups, denoted $O(p, q)$. If we demand in addition that all transformations have unit determinant, we have the special orthogonal groups, $SO(p, q)$. This class of Lie groups includes the $O(3, 1)$ Lorentz metric, as well as the $O(3)$ case of Euclidean 3-space.

Now suppose M is antisymmetric. This case arises in classical Hamiltonian dynamics, where we have canonically conjugate variables satisfying fundamental Poisson bracket relations.

$$\begin{aligned} \left\{q^i, q^j\right\}_{x\pi} &= 0 = \{p_i, p_j\}_{x\pi} \\ \left\{p_j, q^i\right\}_{x\pi} &= -\left\{q^i, p_j\right\}_{x\pi} = \delta_j^i \end{aligned}$$

If we define a single set of coordinates including both p_i and q^i,

$$\xi^a = (q^i, p_j)$$

where if $i, j = 1, 2, \ldots, n$ then $a = 1, 2, \ldots, 2n$, then the fundamental brackets may be written in terms of an antisymmetric matrix Ω^{ab} as

$$\{\xi^a, \xi^b\} = \Omega^{ab}$$

where

$$\Omega^{ab} = \begin{pmatrix} 0 & -\delta_j^i \\ \delta_j^i & 0 \end{pmatrix} = -\Omega^{ba} \tag{84}$$

Since canonical transformations are precisely diffeomorphisms that preserve the fundamental brackets, the canonical transformations at any fixed point is the Lie group of which preserves Ω^{ab}. These transformations, and in general the subgroup of $GL(n)$ preserving any nondegenerate antisymmetric matrix, is called the *symplectic group*. We have a similar result here as for the (pseudo-) orthogonal groups – we can always choose a basis for the vector space that puts the invariant matrix Ω^{ab} in the form given in eq.(84). From the form of eq.(84) we suspect, correctly, that the symplectic group is always even dimensional. Indeed, the determinant of an antisymmetric matrix in odd dimensions is always zero, so such an invariant cannot be non-degenerate. The notation for the symplectic groups is therefore $Sp(2n)$.

17.2 Lie algebras

If we want to work with large Lie groups, working directly with the transformation matrices becomes prohibitively difficult. Instead, most of the information we need to know about the group is already present

in the infinitesimal transformations. Unlike the group multiplication, the combination of the infinitesimal transformations is usually fairly simple.

The relationship between the Lie algebra and the Lie group is easily seen even in the simple case of $SO(2)$. Let the general element be written as

$$A(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

and consider those transformations that are close to the identity. Since the identity is $A(0)$, these will be the transformations $A(\varepsilon)$ with $\varepsilon << 1$. Expanding in a Taylor series, we keep only terms to first order:

$$\begin{aligned} A(\varepsilon) &= \begin{pmatrix} \cos\varepsilon & -\sin\varepsilon \\ \sin\varepsilon & \cos\varepsilon \end{pmatrix} \\ &= \begin{pmatrix} 1 & -\varepsilon \\ \varepsilon & 1 \end{pmatrix} + O\left(\varepsilon^2\right) \\ &= \mathbf{1} + \varepsilon \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \end{aligned}$$

The only information here besides the identity is the matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

but remarkably, this is enough to recover the whole group! For general Lie groups, we get one generator for each continuous parameter labeling the group elements. The set of all linear combinations of these generators is a vector space called the *Lie algebra* of the group, denoted by the same abbreviation in lower case letters. The Lie algebra of $SO(2)$ is therefore $so(2)$. We will give the full defining set of properties of a Lie algebra below.

To recover a general element of the group, we perform infinitely many infinitesimal transformations. Applying $A(\varepsilon)$ n times rotates the plane by an angle $n\varepsilon$:

$$A(n\varepsilon) = (A(\varepsilon))^n = \left(\mathbf{1} + \varepsilon \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right)^n$$

Expanding the power on the right using the binomial expansion,

$$A(n\varepsilon) = \sum_{k=0}^{n} \binom{n}{k} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^k \varepsilon^k \mathbf{1}^{n-k}$$

We take the limit as $\varepsilon \to 0$ and $n \to \infty$, holding the product $n\varepsilon = \theta$ finite. Then:

$$A(\theta) = \lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \sum_{k=0}^{n} \binom{n}{k} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^k \varepsilon^k$$

Consider the combinatoric factor:

$$\begin{aligned} \lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \binom{n}{k} \varepsilon^k &= \lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \frac{n!}{k!\,(n-k)!} \varepsilon^k \\ &= \lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \frac{n(n-1)\cdots(n-k+1)}{k!} \varepsilon^k \\ &= \lim_{\substack{\varepsilon \to 0 \\ n \to \infty}} \frac{1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{k!} n^k \varepsilon^k \\ &= \frac{1}{k!}\theta^k \end{aligned}$$

Therefore

$$\begin{aligned} A(\theta) &= \lim_{\substack{\varepsilon\to 0\\ n\to\infty}} \sum_{k=0}^{n} \frac{1}{k!} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^k \theta^k \\ &\equiv \exp\left(\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \theta \right) \end{aligned}$$

where in the last step we *define* the exponential of a matrix to be the power series in the second line. Quite generally, since we know how to take powers of matrices, we can define the exponential of any matrix, M, by its power series:

$$\exp M \equiv \sum_{k=0}^{\infty} \frac{1}{k!} M^k$$

Next, we check that the exponential form of $A(\theta)$ actually is the original class of transformations. To do this we first examine powers of $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$:

$$\begin{aligned} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 &= \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -\mathbf{1} \\ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^3 &= -\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^4 &= \mathbf{1} \end{aligned}$$

The even terms are plus or minus the identity, while the odd terms are always proportional to the generator, $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Therefore, we divide the power series into even and odd parts, and remove the matrices from the sums:

$$\begin{aligned} A(\theta) &= \sum_{k=0}^{\infty} \frac{1}{k!} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^k \theta^k \\ &= \sum_{m=0}^{\infty} \frac{1}{(2m)!} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^{2m} \theta^{2m} + \sum_{m=0}^{\infty} \frac{1}{(2m+1)!} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^{2m+1} \theta^{2m+1} \\ &= \mathbf{1} \left(\sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} \theta^{2m} \right) + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} \theta^{2m+1} \\ &= \mathbf{1} \cos\theta + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \sin\theta \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \end{aligned}$$

The generator has given us the whole group back.

17.2.1 The Lie algebra $so(3)$

To begin to see the power of this technique, let's look at $SO(3)$, the subgroup of elements of $O(3)$ with unit determinant.The defining property of $O(3)$ is the invariance of the Euclidean metric

$$\eta_{ij} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$$

by inserting identities to write

$$\left[A^t A\right]_{mn} = \left[A^t\right]_m{}^i \eta_{ij} \left[A\right]^j{}_n = \eta_{mn}$$

To see that this property also means that $O(3)$ transformations preserve the Pythagorean length of vectors, contract $x^m x^n$ on both sides of the invariance equation to get

$$\left[A^t\right]_m{}^i \eta_{ij} \left[A\right]^j{}_n x^m x^n = \eta_{mn} x^m x^n$$

Then, defining

$$y^i = A^i{}_m x^m$$

we have

$$y^i \eta_{ij} y^j = \eta_{mn} x^m x^n$$

But this just says that

$$\left(y^1\right)^2 + \left(y^3\right)^2 + \left(y^3\right)^2 = \left(x^1\right)^2 + \left(x^3\right)^2 + \left(x^3\right)^2$$

so $O(3)$ preserves the Pythagorean norm.

Next we restrict to $SO(3)$. Since every element of $O(3)$ satisfies

$$A^t A = 1$$

we have

$$\begin{aligned} 1 &= \det(1) \\ &= \det\left(A^t\right) \det(A) \\ &= (\det(A))^2 \end{aligned}$$

so either $\det A = 1$ or $\det A = -1$. Defining the parity transformation to be

$$P = \begin{pmatrix} -1 & & \\ & -1 & \\ & & -1 \end{pmatrix}$$

then every element of $O(3)$ is of the form A or PA, where A is in $SO(3)$. Because P is a discrete transformation and not a continuous set of transformations, $O(3)$ and $SO(3)$ have the same Lie algebra.

As in the 2-dimensional case, we look at transformations infinitesimally close to the identity. Let

$$A^i{}_j = \delta^i_j + \varepsilon^i{}_j$$

where all components of $\varepsilon^i{}_j$ are small. Then

$$\begin{aligned} \eta_{mn} &= \eta_{ij} \left(\delta^i_m + \varepsilon^i{}_m\right) \left(\delta^j_n + \varepsilon^j{}_n\right) \\ &= \eta_{ij}\delta^i_m\delta^j_n + \eta_{ij}\delta^i_m\varepsilon^j{}_n + \eta_{ij}\varepsilon^i{}_m\delta^j_n + \eta_{ij}\varepsilon^i{}_m\varepsilon^j{}_n \\ &= \eta_{mn} + \varepsilon_{nm} + \varepsilon_{mn} + O(\varepsilon^2) \end{aligned} \tag{85}$$

where $\varepsilon_{mn} = \eta_{mj}\varepsilon^j{}_n$. Dropping the second order term and cancelling η_{mn} on the left and right, we see that the generators ε_{mn} must be antisymmetric:

$$\varepsilon_{nm} = -\varepsilon_{mn} \tag{86}$$

We are dealing with 3×3 matrices here, but note the power of index notation! There is actually nothing in the preceeding calculation that is specific to $n = 3$, and we could draw all the same conclusions up to this

point for $O(p,q)$! For the 3×3 case, every antisymmetric matrix is of the form

$$\begin{aligned} A(a,b,c) &= \begin{pmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{pmatrix} \\ &= a\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \end{aligned}$$

and therefore a linear combination of the three *generators*

$$\begin{aligned} J_1 &= \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ J_2 &= \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ J_3 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \end{aligned} \tag{87}$$

Notice that any three independent, antisymmetric matrices could serve as the generators. The Lie algebra is defined as the entire vector space

$$v = v^1 J_1 + v^2 J_2 + v^3 J_3$$

and the generators form a basis.

The generators, J_i, close under commutation. For example

$$\begin{aligned} [J_1, J_2] &= \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\ &= J_3 \end{aligned}$$

Similarly, we find that $[J_2, J_3] = J_1$ and $[J_3, J_1] = J_2$. Since the commutator is antisymmetric, we can summarize these results in one equation,

$$[J_i, J_j] = \varepsilon_{ij}{}^k J_k \tag{88}$$

This closure of commutators of basis vectors induces a well-defined commutator product on the entire Lie algebra.

Let two generic elements of $so(3)$ be written as

$$\begin{aligned} u &= u^i J_i \\ v &= v^k J_k \end{aligned}$$

Use the commutators of the basis, eq.(88) to compute $[u, v]$.

Compute the double commutator,

$$[J_1, [J_2, J_3]]$$

Prove the *Jacobi identity*

$$[J_1, [J_2, J_3]] + [J_2, [J_3, J_1]] + [J_3, [J_1, J_2]] = 0$$

These properties – a vector space with a closed commutator algebra satisfying the Jacobi identity, hold for any Lie algebra, and provide the definition,

A *Lie algebra* is a finite dimensional vector space V together with a bilinear, antisymmetric (commutator) product satisfying

1. For all $u, v \in V$, the product $[u, v] = -[v, u] = w$ is in V.
2. All $u, v, w \in V$ satisfy the Jacobi identity

$$[u, [v, w]] + [v, [w, u]] + [w, [u, v]] = 0$$

As we showed with $so(3)$, these properties may be expressed in terms of a basis. Let $\{J_a \mid a = 1, \ldots, n\}$ be a vector basis for V. Then we may compute the commutators of the basis,

$$[J_a, J_b] = w_{ab}$$

where for each a and each b, w_{ab} is some vector in V. We may expand each w_{ab} in the basis as well,

$$w_{ab} = c_{ab}{}^{c} J_c$$

for some constants $c_{ab}{}^{c}$. The $c_{ab}{}^{c} = -c_{ba}{}^{c}$ are called the *Lie structure constants*. These constants are always real, regardless of the representation. The basis therefore satisfies,

$$[J_a, J_b] = c_{ab}{}^{c} J_c$$

which is sufficient, using linearity, to determine the commutators of all elements of the algebra:

$$\begin{aligned} [u, v] &= \left[u^a J_a, v^b J_b\right] \\ &= u^a v^b \left[J_a, J_b\right] \\ &= u^a v^b c_{ab}{}^{c} J_c \\ &= w^c J_c \\ &= w \end{aligned} \tag{89}$$

Show that if the Jacobi identity holds among all generators,

$$[J_a, [J_b, J_c]] + [J_b, [J_c, J_a]] + [J_c, [J_a, J_b]] = 0$$

then it holds for all triples of vectors u, v, w in the algebra.

17.2.2 The Lie algebras $so(p, q)$

Notice that most of the calculations above for $O(3)$ actually apply to any of the pseudo-orthogonal groups $O(p, q)$. In the general case, the form of the generators is still given by eq.(86), with η_{mn} replaced by $\eta^{(p,q)}_{mn}$ of eq.(83). That is, ε_{mn} is still antisymmetric but now it is defined by $\varepsilon_{mn} = \eta^{(p,q)}_{mj} \varepsilon^{j}{}_{n}$.

The trickiest part of the general case is finding a methodical way to write the generators. We choose the following set of antisymmetric matrices as generators:

$$\left[\varepsilon^{(rs)}\right]_{mn} = (\delta^r_m \delta^s_n - \delta^r_n \delta^s_m)$$

On the left, the (rs) labels tell us *which* generator we are talking about, while the m and n indices are the matrix components. The net effect is that the labels tell us which components of the matrix are nonzero. For example,

$$\left[\varepsilon^{(12)}\right]_{mn} = \left(\delta^1_m \delta^2_n - \delta^1_n \delta^2_m\right) = \begin{pmatrix} 0 & 1 & 0 & \cdots \\ -1 & 0 & 0 & \\ 0 & 0 & 0 & \\ \vdots & & & \ddots \end{pmatrix}$$

and so on. To compute the Lie algebra, we need the mixed form of the generators,

$$\begin{aligned}\left[\varepsilon^{(rs)}\right]^m{}_n &= \eta^{mk}\left[\varepsilon^{(rs)}\right]_{kn} \\ &= \eta^{mk}\delta^r_k\delta^s_n - \eta^{mk}\delta^r_n\delta^s_k \\ &= \eta^{mr}\delta^s_n - \eta^{ms}\delta^r_n\end{aligned}$$

We can now calculate

$$\begin{aligned}\left[\left[\varepsilon^{(uv)}\right],\left[\varepsilon^{(rs)}\right]\right]^m{}_n &= \left[\varepsilon^{(uv)}\right]^m{}_k\left[\varepsilon^{(rs)}\right]^k{}_n - \left[\varepsilon^{(rs)}\right]^m{}_k\left[\varepsilon^{(uv)}\right]^k{}_n \\ &= \left(\eta^{mu}\delta^v_k - \eta^{mv}\delta^u_k\right)\left(\eta^{kr}\delta^s_n - \eta^{ks}\delta^r_n\right) \\ &\quad - \left(\eta^{mr}\delta^s_k - \eta^{ms}\delta^r_k\right)\left(\eta^{ku}\delta^v_n - \eta^{kv}\delta^u_n\right) \\ &= \eta^{mu}\eta^{vr}\delta^s_n - \eta^{mu}\eta^{vs}\delta^r_n \\ &\quad -\eta^{mv}\eta^{ur}\delta^s_n + \eta^{mv}\eta^{us}\delta^r_n \\ &\quad -\eta^{mr}\eta^{su}\delta^v_n + \eta^{ms}\eta^{ru}\delta^v_n \\ &\quad +\eta^{mr}\eta^{sv}\delta^u_n - \eta^{ms}\eta^{rv}\delta^u_n\end{aligned}$$

Rearranging to collect the terms as generators, and noting that each must have free m and n indices, we get

$$\begin{aligned}\left[\left[\varepsilon^{(uv)}\right],\left[\varepsilon^{(rs)}\right]\right]^m{}_n &= \eta^{vr}\left(\eta^{mu}\delta^s_n - \eta^{ms}\delta^u_n\right) \\ &\quad -\eta^{vs}\left(\eta^{mu}\delta^r_n - \eta^{mr}\delta^u_n\right) \\ &\quad -\eta^{ur}\left(\eta^{mv}\delta^s_n - \eta^{ms}\delta^v_n\right) \\ &\quad +\eta^{us}\left(\eta^{mv}\delta^r_n - \eta^{mr}\delta^v_n\right) \\ &= \eta^{vr}\left[\varepsilon^{(us)}\right]^m{}_n - \eta^{vs}\left[\varepsilon^{(ur)}\right]^m{}_n \\ &\quad -\eta^{ur}\left[\varepsilon^{(vs)}\right]^m{}_n + \eta^{us}\left[\varepsilon^{(vr)}\right]^m{}_n\end{aligned} \tag{90}$$

Finally, we can drop the matrix indices. It is important that we can do this, because it demonstrates that the Lie algebra is a relationship among the different generators that doesn't depend on whether the operators are written as matrices or not. The result, valid for any $o(p,q)$, is

$$\left[\varepsilon^{(uv)},\varepsilon^{(rs)}\right] = \eta^{vr}\varepsilon^{(us)} - \eta^{vs}\varepsilon^{(ur)} - \eta^{ur}\varepsilon^{(vs)} + \eta^{us}\varepsilon^{(vr)} \tag{91}$$

By the arguements of this section, we could have written the generators of $so(3)$ as

$$\left[\varepsilon^{(jk)}\right]^m{}_n = \eta^{mj}\delta^k_n - \eta^{mk}\delta^j_k$$

and therefore the generators, J_i, that we chose must be linear combinations of these. Show that

$$[J_k]^m{}_n = \frac{1}{2}\varepsilon_{ijk}\left[\varepsilon^{(jk)}\right]^m{}_n$$

Show that the $O(p,q)$ Lie algebra in eq.(91) reduces to the $O(3)$ Lie algebra in eq.(88) when $(p,q) = (3,0)$. (Hint: go back to eq.(91) and multiply the whole equation by $\varepsilon_{wuv}\varepsilon_{trs}$ and use $J_i = \frac{1}{2}\varepsilon_{ijk}\varepsilon^{(jk)}$).

17.2.3 Lie algebras: a general approach

The defining properties of a Lie algebra follow from the properties of a Lie group. To see this, let's work from the group side. We have a linear representation of the group, so each elements may be written as a

matrix that depend on, say, N continuous parameters. Choosing the parameters so that $g(0)$ is the identity, we can therefore expand $g\left(x^1, x^2, \ldots, x^N\right) = g\left(x^a\right)$ near the identity in a Taylor series of the form

$$\begin{aligned} g\left(x^1, \ldots, x^N\right) &= 1 + \left.\frac{\partial g}{\partial x^a}\right|_{\mathbf{x}=0} x^a + \frac{1}{2}\left.\frac{\partial^2 g}{\partial x^a x^b}\right|_{\mathbf{x}=0} x^a x^b + \ldots \\ &\equiv 1 + J_a x^a + \frac{1}{2} K_{ab} x^a x^b + \ldots \end{aligned}$$

Now let's look at the consequences of the properties of the group on the infinitesimal generators, J_a. First, there exists a group product, which must close:

$$\begin{aligned} g(x_1^a) g\left(x_2^b\right) &= g(x_3^a) \\ \left(1 + J_a x_1^a + \ldots\right)\left(1 + J_a x_2^a + \ldots\right) &= 1 + J_a x_3^a + \ldots \\ 1 + J_a x_1^a + J_a x_2^a + \ldots &= 1 + J_a x_3^a + \ldots \end{aligned}$$

so that to linear order,

$$J_a x_1^a + J_a x_2^a = J_a x_3^a$$

This requires the generators to combine linearly under addition and scalar multiplication. Next, we require an identity operator. This just means that the zero vector lies in the space of generators, since $g(0, \ldots, 0) = 1 = 1 + J_a 0^a$. For inverses, we have

$$\begin{aligned} g(x_1^a) g^{-1}\left(x_2^b\right) &= 1 \\ \left(1 + J_a x_1^a + \ldots\right)\left(1 + J_a x_2^a + \ldots\right) &= 1 \\ 1 + J_a x_1^a + J_a x_2^a &= 1 \end{aligned}$$

so that $x_2^a = -x_1^a$, guaranteeing an additive inverse in the space of generators. These properties together make the set $\{x^a J_a\}$ a vector space.

Now we need the commutator product. For this, consider the (closed!) product of group elements

$$g_1 g_2 g_1^{-1} g_2^{-1} = g_3$$

We need to compute this in a Taylor series to second order, so we need the inverse to second order.

Show to second order that the inverse of

$$g \equiv 1 + J_a x^a + \frac{1}{2} K_{ab} x^a x^b + \ldots$$

is

$$g^{-1} \equiv 1 - J_b x^b + \frac{1}{2}\left(J_a J_b + J_b J_a - K_{ab}\right) x^a x^b + \ldots$$

Now, expanding to second order in the Taylor series,

$$\begin{aligned} g_3 &= 1 + J_a z^a(x,y) + \frac{1}{2} K_{ab} z^a(x,y) z^b(x,y) \\ &= \left(1 + J_a x^a + \frac{1}{2} K_{ab} x^a x^b\right)\left(1 + J_b\ y^b + \frac{1}{2} K_{bc} y^b y^c\right) \\ &\quad \times \left(1 - J_c x^c + \left(J_c J_d - \frac{1}{2} K_{cd}\right) x^c x^d\right) \\ &\quad \times \left(1 - J_d y^d + \left(J_d J_e - \frac{1}{2} K_{de}\right) y^d y^e\right) \\ &= \left(1 + J_b x^b + J_b y^b + J_a J_b x^a y^b + \frac{1}{2} K_{bc} y^b y^c + \frac{1}{2} K_{ab} x^a x^b\right) \end{aligned}$$

$$
\begin{aligned}
&\times \left(1 - J_d x^d - J_d y^d + J_d J_e y^d y^e + J_c J_d x^c y^d \right. \\
&\left. + J_c J_d x^c x^d - \frac{1}{2} K_{de} y^d y^e - \frac{1}{2} K_{cd} x^c x^d \right) \\
= \quad & 1 - J_d x^d - J_d y^d + J_d J_e y^d y^e + J_c J_d x^c y^d + J_c J_d x^c x^d \\
& - \frac{1}{2} K_{de} y^d y^e - \frac{1}{2} K_{cd} x^c x^d + \left(J_b x^b + J_b y^b\right)\left(1 - J_d x^d - J_d y^d\right) \\
& + J_a J_b x^a y^b + \frac{1}{2} K_{bc} y^b y^c + \frac{1}{2} K_{ab} x^a x^b
\end{aligned}
$$

Collecting terms,

$$
\begin{aligned}
g_3 &= 1 + J_a z^a(x,y) + \cdots \\
&= 1 - J_d x^d - J_d y^d + J_b x^b + J_b y^b \\
&\quad + J_d J_e y^d y^e + J_c J_d x^c y^d + J_c J_d x^c x^d - J_b J_d x^b x^d \\
&\quad - J_b J_d y^b x^d - J_b J_d x^b y^d - J_b J_d y^b y^d + J_a J_b x^a y^b \\
&\quad + \frac{1}{2} K_{bc} y^b y^c + \frac{1}{2} K_{ab} x^a x^b - \frac{1}{2} K_{de} y^d y^e - \frac{1}{2} K_{cd} x^c x^d \\
&= 1 + J_c J_d x^c y^d - J_b J_d y^b x^d \\
&= 1 + J_c J_d x^c y^d - J_d J_c x^c y^d \\
&= 1 + [J_c, J_d] x^c y^d
\end{aligned}
$$

Equating the expansion of g_3 to the collected terms we see that for the lowest order terrms to match we must have

$$
[J_c, J_d]\, x^c y^d = J_a z^a\,(x,y)
$$

for some z^a. Since x^c and y^d are already small, we may expand z^a in a Taylor series in them

$$
z^a = a^a + b^a_b x^b + c^a_b y^b + c_{bc}{}^a x^b y^c + \ldots
$$

The first three terms must vanish, because $z^a = 0$ if either $x^a = 0$ or $y^a = 0$. Therefore, at lowest order, $z^a = c_{bc}{}^a x^b y^c$ and therefore the commutator must close

$$
[J_c, J_d] = c_{bc}{}^a J_a
$$

Finally, the Lie group is associative: if we have three group elements, g_1, g_2 and g_3, then

$$
g_1\,(g_2 g_3) = (g_1 g_2)\, g_3
$$

To first order, this simply implies associativity for the generators

$$
J_a\,(J_b J_c) = (J_a J_b)\, J_c
$$

Now consider the Jacobi identity:

$$
\begin{aligned}
0 &= [J_a, [J_b, J_c]] + [J_b, [J_c, J_a]] + [J_c, [J_a, J_b]] \\
&= [J_a, (J_b J_c - J_c J_b)] + [J_b, (J_c J_a - J_a J_c)] \\
&\quad + [J_c, (J_a J_b - J_b J_a)] \\
&= J_a\,(J_b J_c) - J_a\,(J_c J_b) - (J_b J_c)\, J_a + (J_c J_b)\, J_a \\
&\quad + J_b\,(J_c J_a) - J_b\,(J_a J_c) - (J_c J_a)\, J_b + (J_a J_c)\, J_b \\
&\quad + J_c\,(J_a J_b) - J_c\,(J_b J_a) - (J_a J_b)\, J_c + (J_b J_a)\, J_c \\
&= J_a\,(J_b J_c) - (J_a J_b)\, J_c
\end{aligned}
$$

$$
\begin{aligned}
&-J_a\left(J_cJ_b\right)+\left(J_aJ_c\right)J_b\\
&-\left(J_bJ_c\right)J_a+J_b\left(J_cJ_a\right)\\
&+\left(J_cJ_b\right)J_a-J_c\left(J_bJ_a\right)\\
&-J_b\left(J_aJ_c\right)+\left(J_bJ_a\right)J_c\\
&+J_c\left(J_aJ_b\right)-\left(J_cJ_a\right)J_b
\end{aligned}
$$

From the final arrangement of the terms, we see that it is satisfied identically provided the multiplication is associative.

Therefore, the definition of a Lie algebra is a necessary consequence of being built from the infinitesimal generators of a Lie group. The conditions are also sufficient, though we won't give the proof here.

The correspondence between Lie groups and Lie algebras is not one to one, because in general several Lie groups may share the same Lie algebra. However, groups with the same Lie algebra are related in a simple way. Our example above of the relationship between $O(3)$ and $SO(3)$ is typical – these two groups are related by a discrete symmetry. Since discrete symmetries do not participate in the computation of infinitesimal generators, they do not change the Lie algebra. The central result is this: for every Lie algebra there is a unique maximal Lie group called the *covering group* such that every Lie group sharing the same Lie algebra is the quotient of the covering group by a discrete symmetry group. This result suggests that when examining a group symmetry of nature, we should always look at the covering group in order to extract the greatest possible symmetry. Following this suggestion for Euclidean 3-space and for Minkowski space leads us directly to the use of *spinors*. Spinors form the vector space on which the linear representation of the covering group acts. Thus, we see that making the maximal use of of symmetry makes the appearance of spinors in quantum mechanics, general relativity and quantum field theory seem natural.

17.3 Differential forms

In section 4.2.2 we defined forms as the vector space of linear mappings from curves to the reals. This suggests a generalization, since we know how to integrate over surfaces and volumes as well as curves. In higher dimensions we also have higher order multiple integrals. We now consider the integrands of arbitrary multiple integrals

$$\int f(\mathbf{x})dl,\ \int\int f(\mathbf{x})dS,\ \int\int\int f(\mathbf{x})dV \tag{92}$$

Much of their importance lies in the coordinate invariance of the resulting integrals.

One of the important properties of integrands is that they can all be regarded as oriented. If we integrate a line integral along a curve from A to B we get a number, while if we integrate from B to A we get minus the same number,

$$\int_A^B f(\mathbf{x})dl=-\int_B^A f(\mathbf{x})dl \tag{93}$$

We can also demand oriented surface integrals, so the surface integral

$$\int\int \mathbf{A}\cdot\mathbf{n}\ dS \tag{94}$$

changes sign if we reverse the direction of the normal to the surface. This normal can be thought of as the cross product of two basis vectors within the surface. If these basis vectors' cross product is taken in one order, $\mathbf{n}$ has one sign. If the opposite order is taken then $-\mathbf{n}$ results. Similarly, volume integrals change sign if we change from a right- or left-handed coordinate system.

We can build this alternating sign into our convention for writing differential forms by introducing a formal antisymmetric product, called the *wedge* product, symbolized by $\wedge$, which is defined to give these differential elements the proper signs. Thus, surface integrals will be written as integrals over the products

$$\mathbf{d}x\wedge\mathbf{d}y,\mathbf{d}y\wedge\mathbf{d}z,\mathbf{d}z\wedge\mathbf{d}x$$

with the convention that $\wedge$ is antisymmetric:

$$\mathbf{d}x \wedge \mathbf{d}y = -\mathbf{d}y \wedge \mathbf{d}x$$

under the interchange of any two basis forms. This automatically gives the right orientation of the surface. Similarly, the volume element becomes

$$\mathbf{V} = \mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z$$

which changes sign if any pair of the basis elements are switched.

We can go further than this by formalizing the full integrand. For a line integral, the general form of the integrand is a linear combination of the basis differentials,

$$A_x \mathbf{d}x + A_y \mathbf{d}y + A_z \mathbf{d}z$$

Notice that we simply add the different parts. Similarly, a general surface integrand is

$$A_z \mathbf{d}x \wedge \mathbf{d}y + A_y \mathbf{d}z \wedge \mathbf{d}x + A_x \mathbf{d}y \wedge \mathbf{d}z$$

while the volume integrand is

$$f(x)\, \mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z$$

These objects are called *differential forms*.

Clearly, differential forms come in several types. Functions are called 0 -forms, line elements 1-forms, surface elements 2-forms, and volume forms are called 3-forms. These are all the types that exist in 3 -dimensions, but in more than three dimensions we can have p-forms with p ranging from zero to the dimension, d, of the space. Since we can take arbitrary linear combinations of p-forms, they form a vector space, Λ_p.

We can always wedge together any two forms. We assume this wedge product is associative, and obeys the usual distributive laws. The wedge product of a p-form with a q-form is a $(p+q)$-form.

Notice that the antisymmetry is all we need to rearrange any combination of forms. In general, wedge products of even order forms with any other forms commute while wedge products of pairs of odd-order forms anticommute. In particular, functions (0-forms) commute with all p-forms. Using this, we may interchange the order of a line element and a surface area, for if

$$\begin{aligned} \mathbf{l} &= A\mathbf{d}x \\ \mathbf{S} &= B\mathbf{d}y \wedge \mathbf{d}z \end{aligned}$$

then

$$\begin{aligned} \mathbf{l} \wedge \mathbf{S} &= (A\ \mathbf{d}x) \wedge (B\ \mathbf{d}y \wedge \mathbf{d}z) \\ &= A\ \mathbf{d}x \wedge B\ \mathbf{d}y \wedge \mathbf{d}z \\ &= AB\ \mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z \\ &= -AB\ \mathbf{d}y \wedge \mathbf{d}x \wedge \mathbf{d}z \\ &= AB\ \mathbf{d}y \wedge \mathbf{d}z \wedge \mathbf{d}x \\ &= \mathbf{S} \wedge \mathbf{l} \end{aligned}$$

but the wedge product of two line elements changes sign, for it

$$\begin{aligned} \mathbf{l}_1 &= A\mathbf{d}x \\ \mathbf{l}_2 &= B\mathbf{d}y + C\mathbf{d}z \end{aligned}$$

then

$$\begin{aligned}
\mathbf{l}_1 \wedge \mathbf{l}_2 &= (A\ \mathbf{d}x) \wedge (B\mathbf{d}y + C\mathbf{d}z) \\
&= A\ \mathbf{d}x \wedge B\mathbf{d}y + A\ \mathbf{d}x \wedge C\mathbf{d}z \\
&= AB\ \mathbf{d}x \wedge \mathbf{d}y + AC\ \mathbf{d}x \wedge \mathbf{d}z \\
&= -AB\ \mathbf{d}y \wedge \mathbf{d}x - AC\ \mathbf{d}z \wedge \mathbf{d}x \\
&= -B\mathbf{d}y \wedge A\mathbf{d}x - C\mathbf{d}z \wedge A\mathbf{d}x \\
&= -\mathbf{l}_2 \wedge \mathbf{l}_1
\end{aligned} \tag{95}$$

For any odd-order form, ω, we immediately have

$$\omega \wedge \omega = -\omega \wedge \omega = 0$$

In 3-dimensions there are no 4-forms because anything we try to construct must contain a repeated basis form. For example

$$\begin{aligned}
\mathbf{l} \wedge \mathbf{V} &= (A\ \mathbf{d}x) \wedge (B\ \mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z) \\
&= AB\ \mathbf{d}x \wedge \mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z \\
&= 0
\end{aligned}$$

since $\mathbf{d}x \wedge \mathbf{d}x = 0$. The same occurs for anything we try. Of course, if we have more dimensions then there are more independent directions and we can find nonzero 4-forms. In general, in d-dimensions we can find d-forms, but no $(d+1)$-forms.

Now suppose we want to change coordinates. How does an integrand change? Suppose Cartesian coordinates (x, y) in the plane are given as some functions of new coordinates (u, v). Then we already know that differentials change according to

$$\mathbf{d}x = \mathbf{d}x\,(u, v) = \frac{\partial x}{\partial u}\mathbf{d}u + \frac{\partial x}{\partial v}\mathbf{d}v$$

and similarly for $\mathbf{d}y$, applying the usual rules for partial differentiation. Notice what happens when we use the wedge product to calculate the new area element:

$$\begin{aligned}
\mathbf{d}x \wedge \mathbf{d}y &= \left(\frac{\partial x}{\partial u}\mathbf{d}u + \frac{\partial x}{\partial v}\mathbf{d}v\right) \wedge \left(\frac{\partial y}{\partial u}\mathbf{d}u + \frac{\partial y}{\partial v}\mathbf{d}v\right) \\
&= \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\mathbf{d}v \wedge \mathbf{d}u + \frac{\partial x}{\partial u}\frac{\partial y}{\partial v}\mathbf{d}u \wedge \mathbf{d}v \\
&= \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right)\mathbf{d}u \wedge \mathbf{d}v \\
&= \mathcal{J}\ \mathbf{d}u \wedge \mathbf{d}v
\end{aligned}$$

where

$$\mathcal{J} = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

is the Jacobian of the coordinate transformation. This is exactly the way that an area element changes when we change coordinates! Notice the Jacobian coming out automatically. We couldn't ask for more – the wedge product not only gives us the right signs for oriented areas and volumes, but gives us the right transformation to new coordinates. Of course the volume change works, too.

In eq.(95), showing the anticommutation of two 1-forms, identify the property of form multiplication used in each step (associativity, anticommutation of basis forms, commutation of 0-forms, etc.).

Show that under a coordinate transformation

$$\begin{aligned} x &\rightarrow x(u,v,w) \\ y &\rightarrow y(u,v,w) \\ z &\rightarrow z(u,v,w) \end{aligned}$$

the new volume element is just get the full Jacobian times the new volume form,

$$\mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z = \mathcal{J}(xyz;uvw)\ \mathbf{d}u \wedge \mathbf{d}v \wedge \mathbf{d}w$$

So the wedge product successfully keeps track of p-dim volumes and their orientations in a coordinate invariant way. Now any time we have an integral, we can regard the integrand as being a differential form. But all of this can go much further. Recall our proof that 1-forms form a vector space. Thus, the differential, $\mathbf{d}x$, of $x(u,v)$ given above is just a gradient. It vanishes along surfaces where x is constant, and the components of the vector

$$\left(\frac{\partial x}{\partial u}, \frac{\partial x}{\partial v}\right)$$

point in a direction normal to those surfaces. So symbols like $\mathbf{d}x$ or $\mathbf{d}u$ contain directional information. Writing them with a boldface $\mathbf{d}$ indicates this vector character. Thus, we write

$$\mathbf{A} = A_i \mathbf{d}x^i$$

Let

$$f(x,y) = axy$$

Show that the vector with components

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

is perpendicular to the surfaces of constant f.

Let's sum up. We have defined forms, have written down their formal properties, and have use those properties to write them in components. Then, we defined the wedge product, which enables us to write p-dimensional integrands as p-forms in such a way that the orientation and coordinate transformation properties of the integrals emerges automatically.

Though it is 1-forms, $A_i \mathbf{d}x^i$ that correspond to vectors, we have defined a product of basis forms that we can generalize to more complicated objects. Many of these objects are already familiar. Consider the product of two 1-forms.

$$\begin{aligned} \mathbf{A} \wedge \mathbf{B} &= A_i \mathbf{d}x^i \wedge B_j \mathbf{d}x^j \\ &= A_i B_j \mathbf{d}x^i \wedge \mathbf{d}x^j \\ &= \frac{1}{2} A_i B_j \left(\mathbf{d}x^i \wedge \mathbf{d}x^j - \mathbf{d}x^j \wedge \mathbf{d}x^i\right) \\ &= \frac{1}{2}\left(A_i B_j \mathbf{d}x^i \wedge \mathbf{d}x^j - A_i B_j \mathbf{d}x^j \wedge \mathbf{d}x^i\right) \\ &= \frac{1}{2}\left(A_i B_j \mathbf{d}x^i \wedge \mathbf{d}x^j - A_j B_i \mathbf{d}x^i \wedge \mathbf{d}x^j\right) \\ &= \frac{1}{2}\left(A_i B_j - A_j B_i\right) \mathbf{d}x^i \wedge \mathbf{d}x^j \end{aligned}$$

The coefficients

$$A_i B_j - A_j B_i$$

are essentially the components of the cross product. We will see this in more detail below when we discuss the curl.

17.4 The exterior derivative

We may regard the differential of any function, say $f(x, y, z)$, as the 1-form:

$$\begin{aligned} \mathbf{d}f &= \frac{\partial f}{\partial x}\mathbf{d}x + \frac{\partial f}{\partial y}\mathbf{d}y + \frac{\partial f}{\partial z}\mathbf{d}z \\ &= \frac{\partial f}{\partial x^i}\mathbf{d}x^i \end{aligned}$$

Since a function is a 0-form then we can imagine an operator $\mathbf{d}$ that differentiates any 0-form to give a 1-form. In Cartesian coordinates, the coefficients of this 1-form are just the Cartesian components of the gradient.

The operator $\mathbf{d}$ is called the *exterior derivative*, and we may apply it to any p-form to get a $(p+1)$-form. The extension is defined as follows. First consider a 1-form

$$\mathbf{A} = A_i \mathbf{d}x^i$$

We define

$$\mathbf{d}\mathbf{A} = \mathbf{d}A_i \wedge \mathbf{d}x^i$$

Similarly, since an arbitrary p-form in n-dimensions may be written as

$$\omega = A_{i_1 i_2 \cdots i_p} \mathbf{d}x^{i_1} \wedge \quad \mathbf{d}x^{i_2} \cdots \wedge \mathbf{d}x^{i_p}$$

we define the exterior derivative of ω to be the $(p+1)$-form

$$\mathbf{d}\omega = \mathbf{d}A_{i_1 i_2 \cdots i_p} \wedge \mathbf{d}x^{i_1} \wedge \mathbf{d}x^{i_2} \cdots \wedge \mathbf{d}x^{i_p}$$

Let's see what happens if we apply $\mathbf{d}$ twice to the Cartesian coordinate, x, regarded as a function of x, y and z:

$$\begin{aligned} \mathbf{d}^2 x &= \mathbf{d}(\mathbf{d}x) \\ &= \mathbf{d}(1\mathbf{d}x) \\ &= \mathbf{d}(1) \wedge \mathbf{d}x \\ &= 0 \end{aligned}$$

since all derivatives of the constant function $f = 1$ are zero. The same applies if we apply $\mathbf{d}$ twice to *any* function:

$$\begin{aligned} \mathbf{d}^2 f &= \mathbf{d}(\mathbf{d}f) \\ &= \mathbf{d}\left(\frac{\partial f}{\partial x^i}\mathbf{d}x^i\right) \\ &= \mathbf{d}\left(\frac{\partial f}{\partial x^i}\right) \wedge \mathbf{d}x^i \\ &= \left(\frac{\partial^2 f}{\partial x^j \partial x^i}\mathbf{d}x^j\right) \wedge \mathbf{d}x^i \\ &= \frac{\partial^2 f}{\partial x^j \partial x^i}\mathbf{d}x^j \wedge \mathbf{d}x^i \end{aligned}$$

By the same argument we used to get the components of the curl, we may write this as

$$\begin{aligned}\mathbf{d}^2 f &= \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^j \partial x^i} - \frac{\partial^2 f}{\partial x^i \partial x^j}\right)\mathbf{d}x^j \wedge \mathbf{d}x^i \\ &= 0\end{aligned}$$

since partial derivatives commute.

Prove the Poincaré Lemma: $\mathbf{d}^2\omega = 0$ where ω is an arbitrary p-form.

Next, consider the effect of $\mathbf{d}$ on an arbitrary 1-form. We have

$$\begin{aligned}\mathbf{dA} &= \mathbf{d}\left(A_i \mathbf{d}x^i\right) \\ &= \left(\frac{\partial A_i}{\partial x^j}\mathbf{d}x^j\right) \wedge \mathbf{d}x^i \\ &= \frac{1}{2}\left(\frac{\partial A_i}{\partial x^j} - \frac{\partial A_j}{\partial x^i}\right)\mathbf{d}x^j \wedge \mathbf{d}x^i \end{aligned} \tag{96}$$

We have the components of the curl of the vector $\mathbf{A}$. We must be careful here, however, because these are the components of the curl only in Cartesian coordinates. Later we will see how these components relate to those in a general coordinate system. Also, recall from Section (4.2.2) that the components A_i are distinct from the usual vector components A^i. These differences will be resolved when we give a detailed discussion of the metric in Section (5.6). Ultimately, the action of $\mathbf{d}$ on a 1-form gives us a coordinate invariant way to calculate the curl.

Finally, suppose we have a 2-form expressed as

$$\mathbf{S} = A_z \mathbf{d}x \wedge \mathbf{d}y + A_y \mathbf{d}z \wedge \ \mathbf{d}x + A_x \mathbf{d}y \wedge \mathbf{d}z$$

Then applying the exterior derivative gives

$$\begin{aligned}\mathbf{dS} &= \mathbf{d}A_z \wedge \mathbf{d}x \wedge \mathbf{d}y + \mathbf{d}A_y \wedge \mathbf{d}z \wedge \mathbf{d}x + \mathbf{d}A_x \wedge \mathbf{d}y \wedge \mathbf{d}z \\ &= \frac{\partial A_z}{\partial z}\mathbf{d}z \wedge \mathbf{d}x \wedge \mathbf{d}y + \frac{\partial A_y}{\partial y}\mathbf{d}y \wedge \mathbf{d}z \wedge \mathbf{d}x + \frac{\partial A_x}{\partial x}\mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z \\ &= \left(\frac{\partial A_z}{\partial z} + \frac{\partial A_y}{\partial y} + \frac{\partial A_x}{\partial x}\right)\mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z \end{aligned} \tag{97}$$

so that the exterior derivative can also reproduce the divergence.

Fill in the missing steps in the derivation of eq.(97).

Compute the exterior derivative of the arbitrary 3-form, $\mathbf{A} = f(x, y, z)\,\mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z$.

17.5 The Hodge dual

To truly have the curl in eq.(97) or the curl in eq.(96), we need a way to turn a 2-form into a vector, i.e., a 1-form and a way to turn a 3-form into a 0-form. This leads us to introduce the Hodge dual, or star, operator, *.

Notice that in 3-dim, both 1-forms and 2-forms have three independent components, while both 0- and 3-forms have one component. This suggests that we can define an invertible mapping between these pairs. In Cartesian coordinates, suppose we set

$$\begin{aligned}{}^*(\mathbf{d}x \wedge \mathbf{d}y) &= \mathbf{d}z \\ {}^*(\mathbf{d}y \wedge \mathbf{d}z) &= \mathbf{d}x \\ {}^*(\mathbf{d}z \wedge \mathbf{d}x) &= \mathbf{d}y \\ {}^*(\mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z) &= 1\end{aligned}$$

and further require the star to be its own inverse,

$$^{**} = 1$$

With these rules we can find the Hodge dual of any form in 3-dim.

Show that the dual of a general 1-form,

$$\mathbf{A} = A_i \mathbf{d}x^i$$

is the 2-form

$$\mathbf{S} = A_z \mathbf{d}x \wedge \mathbf{d}y + A_y \mathbf{d}z \wedge \mathbf{d}x + A_x \mathbf{d}y \wedge \mathbf{d}z$$

Show that for an arbitrary (Cartesian) 1-form

$$\mathbf{A} = A_i \mathbf{d}x^i$$

that

$$^*\mathbf{d}^*\mathbf{A} = divA$$

Write the curl of $\mathbf{A}$

$$curl\,(A) = \left(\frac{\partial A_y}{\partial z} - \frac{\partial A_z}{\partial y}\right)\mathbf{d}x + \left(\frac{\partial A_z}{\partial x} - \frac{\partial A_x}{\partial z}\right)\mathbf{d}y + \left(\frac{\partial A_x}{\partial y} - \frac{\partial A_y}{\partial x}\right)\mathbf{d}z$$

in terms of the exterior derivative and the Hodge dual.

Write the Cartesian dot product of two 1-forms in terms of wedge products and duals.

We have now shown how three operations – the wedge product $\wedge$, the exterior derivative $\mathbf{d}$, and the Hodge dual * – together encompass the usual dot and cross products as well as the divergence, curl and gradient. In fact, they do much more – they extend all of these operations to arbitrary coordinates and arbitrary numbers of dimensions. To explore these generalizations, we must first explore properties of the metric and look at coordinate transformations. This will allow us to define the Hodge dual in arbitrary coordinates.

17.6 Transformations

Since the use of orthonormal frames is simply a convenient choice of basis, no information is lost in restricting our attention to them. We can always return to general frames if we wish. But as long as we maintain the restriction, we can work with a reduced form of the symmetry group. Arbitrary coordinate transformations – diffeomorphisms – preserve the class of frames, but only orthogonal transformations preserve orthonormal frames. Nonetheless, the class of tensors is remains unchanged – there is a 1-1, onto correspondence between tensors with diffeomorphism covariance and those with orthogonal covariance.

The correspondence between general frame and orthonormal frame tensors is provided by the orthonormal frame itself. Given an orthonormal frame

$$\mathbf{e}^a = e_m{}^a \mathbf{d}x^m$$

we can use the coefficient matrix $e_m{}^a$ and its inverse to transform back and forth between orthonormal and coordinate indices. Thus, given any vector in an arbitrary coordinate basis,

$$\mathbf{v} = v^m \frac{\partial}{\partial x^m}$$

we may insert the identity in the form

$$\delta^m_n = e_n{}^a e_a{}^m$$

to write

$$\begin{aligned}\mathbf{v} &= v^n \delta_n^m \frac{\partial}{\partial x^m} \\ &= v^n e_n{}^a e_a{}^m \frac{\partial}{\partial x^m} \\ &= \left(v^n e_n{}^a\right) \mathbf{e}_a \\ &= v^a \mathbf{e}_a\end{aligned}$$

The mapping

$$v^a = v^n e_n{}^a$$

is invertible because $e_n{}^a$ is invertible. Similarly, any tensor, for example

$$T^{m_1 \ldots m_r}{}_{n_1 \ldots n_s}$$

may be written in an orthonormal basis by using one factor of $e_m{}^a$ or $e_a{}^n$ for each linear slot:

$$T^{a_1 \ldots a_r}{}_{b_1 \ldots b_s} = T^{m_1 \ldots m_r}{}_{n_1 \ldots n_s} e_{m_1}{}^{a_1} \ldots e_{m_n}{}^{a_r} e_{b_1}{}^{n_1} \ldots e_{b_s}{}^{n_s}$$

Similar expressions may be written for tensors with their contravariant and covariant indices in other orders.

We showed in Section (3) that the components of the metric are related to the Cartesian components by

$$g_{jk} = \frac{\partial x^m}{\partial y^j} \frac{\partial x^n}{\partial y^k} \eta_{mn}$$

where we have corrected the index positions and inserted the Cartesian form of the metric explicitly as $\eta_{mn} = diag(1, 1, 1)$. Derive the form of the metric in cylindrical coordinates directly from the coordinate transformation,

$$\begin{aligned}x &= x(\rho, \varphi, z) = \rho \cos \varphi \\ y &= y(\rho, \varphi, z) = \rho \sin \varphi \\ z &= z(\rho, \varphi, z) = z\end{aligned}$$

Notice that the identity matrix should exist in any coordinate system, since multiplying any vector by the identity should be independent of coordinate system. Show that the matrix $\delta^i{}_j$, defined to be the unit matrix in one coordinate system, has the same form in every other coordinate system. Notice that the upper index will transform like a contravariant vector and the lower index like a covariant vector. Also note that $\delta^i{}_j = \delta_j{}^i$.

Show that the inverse to the metric transforms as a contravariant second rank tensor. The easiest way to do this is to use the equation

$$g_{ij} g^{jk} = \delta_i^k$$

and the result of exercise 2, together with the transformation law for g_{ij}.

17.7 The Levi-Civita tensor in arbitrary coordinates

So far, we have only defined the Levi-Civita tensor in Cartesian coordinates, where it is given by the totally antisymmetric symbol

$$\varepsilon_{i_1 i_2 \ldots i_n}$$

in n dimensions. This symbol, however, is not quite a tensor because under a diffeomorphism it becomes

$$\varepsilon_{i_1 i_2 \ldots i_n} \frac{\partial x^{i_1}}{\partial y^{j_1}} \frac{\partial x^{i_2}}{\partial y^{j_2}} \ldots \frac{\partial x^{i_n}}{\partial y^{j_n}} = J \varepsilon_{j_1 j_2 \ldots j_n}$$

where

$$J = \det\left(\frac{\partial x^m}{\partial y^n}\right)$$

is the Jacobian of the coordinate transformation. The transformation is linear and homogeneous, but J is a density not a scalar. We can correct for this to form a tensor by dividing by another density. The most convenient choice is the determinant of the metric. Since the metric transforms as

$$g'_{mn} = \frac{\partial x^i}{\partial y^m}\frac{\partial x^j}{\partial y^n} g_{ij}$$

the determinants are related by

$$\begin{aligned} g' &= \det g'_{mn} \\ &= \det\left(\frac{\partial x^i}{\partial y^m} g_{ij} \frac{\partial x^j}{\partial y^n}\right) \\ &= \det\frac{\partial x^i}{\partial y^m} \deg t_{ij} \det\frac{\partial x^j}{\partial y^n} \\ &= J^2 g \end{aligned}$$

Therefore, in the combination

$$e_{i\ldots j} = \sqrt{g}\varepsilon_{i\ldots j}$$

the factors of J cancel, leaving

$$e'_{i\ldots j} = \sqrt{g'}\varepsilon_{i\ldots j}$$

so that $e_{i\ldots j}$ is a tensor. If we raise all indices on $e_{i_1 i_2 \ldots i_n}$, using n copies of the inverse metric, we have

$$\begin{aligned} e^{j_1 j_2 \cdots j_n} &= \sqrt{g} g^{j_1 i_1} g^{j_2 i_2} \ldots g^{j_n i_n} \varepsilon_{i_1 i_2 \ldots i_n} \\ &= \sqrt{g} g^{-1} \varepsilon^{j_1 j_2 \cdots j_n} \\ &= \frac{1}{\sqrt{g}} \varepsilon^{j_1 j_2 \cdots j_n} \end{aligned}$$

This is also a tensor.

17.8 Differential calculus

Define a p-form as a linear map from oriented p-dimensional volumes to the reals:

$$\Lambda_p : V_p \to R$$

Linearity refers to both the forms and the volumes. Thus, for any two p -forms, Λ_p^1 and Λ_p^2, and any constants a and b,

$$a\Lambda_p^1 + b\Lambda_p^2$$

is also a p-form, while for any two disjoint p-volumes, V_p^1 and V_p^2,

$$\Lambda_p\left(V_p^1 + V_p^2\right) = \Lambda_p\left(V_p^1\right) + \Lambda_p\left(V_p^2\right)$$

In Section 3, we showed for 1-forms that these conditions specify the differential of functions. For p-forms, they are equivalent to linear combinations of wedge products of p 1-forms.

Let $\mathbf{A}$ be a p-form in d-dimensions. Then, inserting a convenient normalization,

$$\mathbf{A} = \frac{1}{p!} A_{i_1 \ldots i_p} \mathbf{d}x^{i_1} \wedge \ldots \wedge \mathbf{d}x^{i_p}$$

The action of the exterior derivative, $\mathbf{d}$, on such a p-form is

$$\mathbf{dA} = \frac{1}{p!}\left(\frac{\partial}{\partial x^k} A_{i_1 \ldots i_p}\right) \mathbf{d}x^k \wedge \mathbf{d}x^{i_1} \wedge \ldots \wedge \mathbf{d}x^{i_p}$$

We also defined the wedge product as a distributive, associative, antisymmetric product on 1-forms:

$$\begin{aligned} \left(a\mathbf{d}x^i + b\mathbf{d}x^i\right) \wedge \mathbf{d}x^j &= a\mathbf{d}x^i \wedge \mathbf{d}x^j + b\mathbf{d}x^i \wedge \mathbf{d}x^j \\ \mathbf{d}x^i \wedge \left(\mathbf{d}x^j \wedge \mathbf{d}x^k\right) &= \left(\mathbf{d}x^i \wedge \mathbf{d}x^j\right) \wedge \mathbf{d}x^k \\ \mathbf{d}x^i \wedge \mathbf{d}x^j &= -\mathbf{d}x^j \wedge \mathbf{d}x^i \end{aligned}$$

A third operation, the Hodge dual, was provisionally defined in Cartesian coordinates, but now we can write its full definition. The dual of $\mathbf{A}$ is defined to be the $(d-p)$-form

$${}^*\mathbf{A} = \frac{1}{(d-p)!p!} A_{i_1 \ldots i_p} e^{i_1 \ldots i_p}{}_{i_{p+1} \ldots i_d} \mathbf{d}x^{i_{p+1}} \wedge \ldots \wedge \mathbf{d}x^{i_d}$$

Notice that we have written the first p indices of the Levi-Civita tensor in the superscript position to keep with our convention of always summing an up index with a down index. In Cartesian coordinates, these two forms represent the same array of numbers, but it makes a difference when we look at other coordinate systems.

Differential calculus is defined in terms of these three operations, $(\wedge, {}^*, \mathbf{d})$. Together, they allow us to perform all standard calculus operations in any number of dimensions and in a way independent of any coordinate choice.

17.8.1 Grad, Div, Curl and Laplacian

It is straightforward to write down the familiar operations of gradient and curl and divergence. We specify each, and apply each in polar coordinates, (ρ, θ, z). Recall that the metric in polar coordinates is

$$g_{mn} = \begin{pmatrix} 1 & & \\ & \rho^2 & \\ & & 1 \end{pmatrix}$$

its inverse is

$$g^{mn} = \begin{pmatrix} 1 & & \\ & \frac{1}{\rho^2} & \\ & & 1 \end{pmatrix}$$

and its determinant is

$$g = \det g_{mn} = \rho^2$$

Gradient The gradient of a function is given by the exterior derivative of a 0 -form,

$$\mathbf{d}f = \frac{\partial f}{\partial x^i} \mathbf{d}x^i$$

Notice that the coefficients are components of a type-$\binom{0}{1}$ tensor, so that if we want the gradient to be a vector, we require the metric:

$$[\nabla f]^i = g^{ij} \frac{\partial f}{\partial x^j}$$

For example, the gradient in polar coordinates has components

$$[\nabla f]^i = \begin{pmatrix} 1 & & \\ & \frac{1}{\rho^2} & \\ & & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial \rho} \\ \frac{\partial f}{\partial \varphi} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial \rho} \\ \frac{1}{\rho}\frac{\partial f}{\partial \varphi} \\ \frac{\partial f}{\partial z} \end{pmatrix}$$

so

$$\nabla f = \frac{\partial f}{\partial \rho}\hat{\rho} + \frac{1}{\rho}\frac{\partial f}{\partial \varphi}\hat{\varphi} + \frac{\partial f}{\partial z}\hat{\mathbf{k}}$$

Divergence The use of differential forms leads to an extremely useful expression for the divergence – important enough that it goes by the name of the divergence theorem. Starting with a 1-form, $\omega = \omega_i \mathbf{d}x^i$, we compute

$$\begin{aligned}
{}^*\mathbf{d}{}^*\omega &= {}^*\mathbf{d}{}^*\omega_i \mathbf{d}x^i \\
&= {}^*\mathbf{d}\left(\frac{1}{2}\omega_i e^i{}_{jk}\right)\mathbf{d}x^j \wedge \mathbf{d}x^k \\
&= \frac{1}{2}{}^*\,\mathbf{d}\left(\omega_i\sqrt{g}g^{in}\right)\varepsilon_{njk}\mathbf{d}x^j \wedge \mathbf{d}x^k \\
&= \frac{1}{2}{}^*\,\frac{\partial}{\partial x^m}\left(\omega_i\sqrt{g}g^{in}\right)\varepsilon_{njk}\mathbf{d}x^m \wedge \mathbf{d}x^j \wedge \mathbf{d}x^k \\
&= \frac{1}{2}\frac{\partial}{\partial x^m}\left(\omega_i\sqrt{g}g^{in}\right)\varepsilon_{njk}e^{mjk} \\
&= \frac{1}{2}\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^m}\left(\omega_i\sqrt{g}g^{in}\right)\varepsilon_{njk}\varepsilon^{mjk} \\
&= \frac{1}{2}\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^m}\left(\omega_i\sqrt{g}g^{in}\right)2\delta^m_n \\
&= \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^m}\left(\omega_i\sqrt{g}g^{im}\right)
\end{aligned}$$

In terms of the vector, rather than form, components of the original form, we may replace $\omega^i = g^{ij}\omega_j$ so that

$${}^*\mathbf{d}{}^*\omega = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^m}\left(\sqrt{g}\omega^m\right) = \nabla\cdot\omega$$

Since the operations on the left are all coordinate invariant, the in the middle is also. Notice that in Cartesian coordinates the metric is just δ_{ij}, with determinant 3, so the expression reduces to the familiar form of the divergence and

$$\nabla\cdot\omega = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^m}\left(\sqrt{g}\omega^m\right)$$

In polar coordinates we have

$$\begin{aligned}
\nabla\cdot\omega &= \frac{1}{\sqrt{\rho^2}}\frac{\partial}{\partial x^m}\left(\sqrt{\rho^2}\omega^m\right) \\
&= \frac{1}{\sqrt{\rho^2}}\left(\frac{\partial}{\partial\rho}\left(\sqrt{\rho^2}\omega^\rho\right) + \frac{\partial}{\partial\varphi}\left(\sqrt{\rho^2}\omega^\varphi\right) + \frac{\partial}{\partial z}\left(\sqrt{\rho^2}\omega^z\right)\right) \\
&= \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\omega^\rho\right) + \frac{\partial\omega^\varphi}{\partial\varphi} + \frac{\partial\omega^z}{\partial z}
\end{aligned}$$

Curl The curl is the dual of the exterior derivative of a 1-form. Thus, if $\omega = \omega_i\mathbf{d}x^i$ then

$$\begin{aligned}
{}^*\mathbf{d}\omega &= {}^*\frac{\partial}{\partial x_j}\omega_i\mathbf{d}x^j\mathbf{d}x^i \\
&= \left(e^{ji}{}_k\frac{\partial}{\partial x_j}\omega_i\right)\mathbf{d}x^k \\
&= e^{ji}{}_k g_{im}g^{mn}\frac{\partial}{\partial x_j}\omega_n\mathbf{d}x^k
\end{aligned}$$

$$
\begin{aligned}
&= e^{ji}{}_k g_{im} \left(\frac{\partial}{\partial x_j} \left(g^{mn} \omega_n \right) - \omega_n \frac{\partial}{\partial x_j} g^{mn} \right) \mathbf{d}x^k \\
&= e_{lmk} g^{lj} \left(\frac{\partial}{\partial x_j} \omega^m - \omega^s g_{sn} \frac{\partial}{\partial x_j} g^{mn} \right) \mathbf{d}x^k
\end{aligned}
$$

Now observe that

$$
\begin{aligned}
g_{sn} \frac{\partial}{\partial x_j} g^{mn} &= \frac{\partial}{\partial x_j} \left(g_{sn} g^{mn} \right) - g^{mn} \frac{\partial}{\partial x_j} \left(g_{sn} \right) \\
&= \frac{\partial}{\partial x_j} \delta_s^m - g^{mn} \frac{\partial}{\partial x_j} \left(g_{sn} \right) \\
&= -g^{mn} \frac{\partial}{\partial x_j} g_{sn}
\end{aligned}
$$

so that

$$
\begin{aligned}
{}^*\mathbf{d}\omega &= e_{lmk} g^{lj} \left(\frac{\partial}{\partial x_j} \omega^m + \omega^s g^{mn} \frac{\partial}{\partial x_j} g_{sn} \right) \mathbf{d}x^k \\
&= \left(e_{lmk} g^{lj} \frac{\partial}{\partial x_j} \omega^m + \omega^s e^{jn}{}_k \frac{\partial}{\partial x_j} g_{sn} \right) \mathbf{d}x^k
\end{aligned}
$$

Next consider

$$
\begin{aligned}
e^{jn}{}_k \frac{\partial}{\partial x_j} g_{sn} &= e^{jn}{}_k \partial_j g_{sn} \\
&= \frac{1}{2} e^{jn}{}_k \left(\partial_j g_{sn} - \partial_n g_{sj} \right) \\
&= \frac{1}{2} e^{jn}{}_k \left(\partial_j g_{sn} - \partial_n g_{sj} + \partial_s g_{jn} \right) \\
&= e^{jn}{}_k \Gamma_{nsj}
\end{aligned}
$$

This combines to

$$
\begin{aligned}
{}^*\mathbf{d}\omega &= \left(e_{lmk} g^{lj} \frac{\partial}{\partial x_j} \omega^m + \omega^s e^{jn}{}_k \frac{\partial}{\partial x_j} g_{sn} \right) \mathbf{d}x^k \\
&= \left(e_{lmk} g^{lj} \frac{\partial}{\partial x_j} \omega^m + \omega^s e^{jn}{}_k \Gamma_{nsj} \right) \mathbf{d}x^k \\
&= e^j_{mk} \left(\frac{\partial}{\partial x_j} \omega^m + g^{nm} \omega^s \Gamma_{nsj} \right) \mathbf{d}x^k \\
&= e^j_{mk} \left(\partial_j \omega^m + \omega^s \Gamma^m{}_{sj} \right) \mathbf{d}x^k \\
&= e^j_{mk} D_j \omega^m \mathbf{d}x^k \\
&= \left(e_{jmk} D^j \omega^m \right) \mathbf{d}x^k
\end{aligned}
$$

Therefore, if we raise the free index, the curl is

$$
\begin{aligned}
\left[\nabla \times \omega \right]^i &= g^{ik} \left(e_{jmk} D^j \omega^m \right) \\
&= \frac{1}{\sqrt{g}} \varepsilon^{ijk} D_j \omega_k
\end{aligned}
$$

Also consider

$$
\mathbf{d}^* \omega = \mathbf{d} \left(e^i{}_{jk} \omega_i \mathbf{d}x^j \mathbf{d}x^k \right)
$$

$$
\begin{aligned}
&= \mathbf{d}\left(e_{ijk}\omega^i \mathbf{d}x^j \mathbf{d}x^k\right) \\
&= \mathbf{d}\left(\sqrt{g}\varepsilon_{ijk}\omega^i \mathbf{d}x^j \mathbf{d}x^k\right) \\
&= \frac{\partial}{\partial x^m}\left(\sqrt{g}\omega^i \varepsilon_{ijk}\mathbf{d}x^m \mathbf{d}x^j \mathbf{d}x^k\right) \\
&= \left(e^{ji}{}_k \frac{\partial}{\partial x^j}\omega_i\right)\mathbf{d}x^k
\end{aligned}
$$

The simplest form computationally uses this to write

$$
{}^*\mathbf{d}\omega = \left[\nabla \times \omega\right]^i g_{ik}\mathbf{d}x^k
$$

To apply the formula, start with the components of the vector. In our familiar example in polar coordinates, let

$$
w^i = (w^\rho, w^\varphi, w^z)
$$

The corresponding form has components $\omega_i = g_{ij}w^j = \left(w^\rho, \rho^2 w^\varphi, w^z\right)$. Therefore, the exterior derivative is

$$
\begin{aligned}
\mathbf{d}\omega &= \mathbf{d}\left(w^\rho \mathbf{d}\rho + \rho^2 w^\varphi \mathbf{d}\varphi + w^z \mathbf{d}z\right) \\
&= \frac{\partial w^\rho}{\partial \varphi}\mathbf{d}\varphi \wedge \mathbf{d}\rho + \frac{\partial w^\rho}{\partial z}\mathbf{d}z \wedge \mathbf{d}\rho \\
&\quad + \frac{\partial}{\partial \rho}\left(\rho^2 w^\varphi\right)\mathbf{d}\rho \wedge \mathbf{d}\varphi + \frac{\partial}{\partial z}\left(\rho^2 w^\varphi\right)\mathbf{d}z \wedge \mathbf{d}\varphi \\
&\quad + \frac{\partial w^z}{\partial \rho}\mathbf{d}\rho \wedge \mathbf{d}z + \frac{\partial w^z}{\partial \varphi}\mathbf{d}\varphi \wedge \mathbf{d}z \\
&= \left(\frac{\partial}{\partial \rho}\left(\rho^2 w^\varphi\right) - \frac{\partial w^\rho}{\partial \varphi}\right)\mathbf{d}\rho \wedge \mathbf{d}\varphi + \left(\frac{\partial w^z}{\partial \varphi} - \frac{\partial}{\partial z}\left(\rho^2 w^\varphi\right)\right)\mathbf{d}\varphi \wedge \mathbf{d}z \\
&\quad + \left(\frac{\partial w^\rho}{\partial z} - \frac{\partial w^z}{\partial \rho}\right)\mathbf{d}z \wedge \mathbf{d}\rho
\end{aligned}
$$

Now the dual maps the basis as

$$
\begin{aligned}
{}^*\mathbf{d}\rho \wedge \mathbf{d}\varphi &= e^{123}g_{33}\mathbf{d}z = \frac{1}{\rho}\mathbf{d}z \\
{}^*\mathbf{d}\varphi \wedge \mathbf{d}z &= e^{231}g_{11}\mathbf{d}\rho = \frac{1}{\rho}\mathbf{d}\rho \\
{}^*\mathbf{d}z \wedge \mathbf{d}\rho &= e^{312}g_{22}\mathbf{d}\varphi = \rho\mathbf{d}\varphi
\end{aligned}
$$

so that

$$
\begin{aligned}
{}^*\mathbf{d}\omega &= \frac{1}{\rho}\left(\frac{\partial}{\partial \rho}\left(\rho^2 w^\varphi\right) - \frac{\partial w^\rho}{\partial \varphi}\right)\mathbf{d}z + \left(\frac{1}{\rho}\frac{\partial w^z}{\partial \varphi} - \rho\frac{\partial}{\partial z}\left(w^\varphi\right)\right)\mathbf{d}\rho \\
&\quad + \rho\left(\frac{\partial w^\rho}{\partial z} - \frac{\partial w^z}{\partial \rho}\right)\mathbf{d}\varphi
\end{aligned}
$$

Now, since

$$
{}^*\mathbf{d}\omega = \left[\nabla \times \omega\right]^i g_{ik}\mathbf{d}x^k
$$

we use the inverse metric on the components of ${}^*\mathbf{d}\omega$ to find

so that with $\omega_i = g_{ij}\omega^j$ we have

$$
\begin{aligned}
\left[\nabla \times \omega\right]^1 &= \frac{1}{\rho}\frac{\partial w^z}{\partial \varphi} - \rho\frac{\partial}{\partial z}\left(w^\varphi\right) \\
\left[\nabla \times \omega\right]^2 &= \frac{1}{\rho}\left(\frac{\partial \omega^\rho}{\partial z} - \frac{\partial \omega^z}{\partial \rho}\right) \\
\left[\nabla \times \omega\right]^3 &= \frac{1}{\rho}\left(\frac{\partial}{\partial \rho}\left(\rho^2 w^\varphi\right) - \frac{\partial w^\rho}{\partial \varphi}\right)
\end{aligned}
$$

Work out the form of the gradient, curl, divergence and laplacian in spherical coordinates. Express your results using a basis of unit vectors.

In an orthonormal vector basis the electric and magnetic fields and the current are

$$\begin{aligned}\mathbf{E} &= E^i \mathbf{e}_i \\ \mathbf{B} &= B^i \mathbf{e}_i \\ \mathbf{J} &= J^i \mathbf{e}_i\end{aligned}$$

Define equivalent forms in arbitrary coordinates by

$$\begin{aligned}\epsilon &= E^i g_{ij} \mathbf{d}x^j = E_j \mathbf{d}x^j \\ \beta &= \frac{1}{2} B^i e_{ijk} \mathbf{d}x^j \wedge \mathbf{d}x^k \\ \kappa &= J^i g_{ij} \mathbf{d}x^j\end{aligned}$$

Show that Maxwell's equations,

$$\begin{aligned}\nabla \cdot \mathbf{E} &= \frac{4\pi}{c}\rho \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \nabla \times \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} &= \frac{4\pi}{c}\mathbf{J}\end{aligned}$$

may be written in terms of ϵ, β, κ and ρ as

$$\begin{aligned}{}^*\mathbf{d}^*\epsilon &= \frac{4\pi}{c}\rho \\ \mathbf{d}\beta &= 0 \\ \mathbf{d}\varepsilon + \frac{1}{c}\frac{\partial \beta}{\partial t} &= 0 \\ {}^*\mathbf{d}^*\beta - \frac{1}{c}\frac{\partial \epsilon}{\partial t} &= \frac{4\pi}{c}\kappa\end{aligned}$$

The third equation may be proved as follows:

$$\begin{aligned}\mathbf{d}\varepsilon + \frac{1}{c}\frac{\partial \beta}{\partial t} &= \mathbf{d}\left(E^i g_{ij}\right)\mathbf{d}x^j + \frac{1}{c}\frac{\partial}{\partial t}\frac{1}{2}B^i e_{ijk}\mathbf{d}x^j \wedge \mathbf{d}x^k \\ &= \frac{\partial E_j}{\partial x^m}\mathbf{d}x^m \wedge \mathbf{d}x^j + \frac{1}{c}\frac{\partial}{\partial t}\frac{1}{2}B^i e_{ijk}\mathbf{d}x^j \wedge \mathbf{d}x^k \\ &= \frac{1}{2}\left(\frac{\partial}{\partial x^m}E_j - \frac{\partial}{\partial x^j}E_m\right)\mathbf{d}x^m \wedge \mathbf{d}x^j + \frac{1}{c}\frac{\partial}{\partial t}\frac{1}{2}B^i e_{ijk}\mathbf{d}x^j \wedge \mathbf{d}x^k \\ &= \frac{1}{4}\left(\frac{\partial}{\partial x^n}E_l - \frac{\partial}{\partial x^l}E_n\right)e^{inl}e_{ijk}\mathbf{d}x^j \wedge \mathbf{d}x^k + \frac{1}{c}\frac{\partial}{\partial t}\frac{1}{2}B^i e_{ijk}\mathbf{d}x^j \wedge \mathbf{d}x^k \\ &= \frac{1}{2}\left(\frac{1}{2}\left(\frac{\partial}{\partial x^n}E_l - \frac{\partial}{\partial x^l}E_n\right)e^{inl} + \frac{1}{c}\frac{\partial}{\partial t}B^i\right)e_{ijk}\mathbf{d}x^j \wedge \mathbf{d}x^k \\ &= \frac{1}{2}\left(e^{inl}\partial_n E_l + \frac{1}{c}\frac{\partial}{\partial t}B^i\right)e_{ijk}\mathbf{d}x^j \wedge \mathbf{d}x^k \\ &= \frac{1}{2}\left(\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{B}\right)^i e_{ijk}\mathbf{d}x^j \wedge \mathbf{d}x^k \\ &= 0\end{aligned}$$

From Maxwell's equations,

$$\begin{aligned} {}^*\mathbf{d}^*\epsilon &= \frac{4\pi}{c}\rho \\ \mathbf{d}\beta &= 0 \\ \mathbf{d}\varepsilon + \frac{1}{c}\frac{\partial\beta}{\partial t} &= 0 \\ {}^*\mathbf{d}^*\beta - \frac{1}{c}\frac{\partial\ \epsilon}{\partial t} &= \frac{4\pi}{c}\kappa \end{aligned}$$

show that

$$\frac{1}{c}\frac{\partial}{\partial t}\rho + {}^*\,\mathbf{d}^*\ \kappa = 0$$

Show that this equation is the continuity equation by writing it in the usual vector notation.

Using the homogeneous Maxwell equations

$$\begin{aligned} \mathbf{d}\beta &= 0 \\ \mathbf{d}\varepsilon + \frac{1}{c}\frac{\partial\beta}{\partial t} &= 0 \end{aligned}$$

show that the electric and magnetic fields arise from a potential.

Start with the magnetic equation

$$\mathbf{d}\beta = 0$$

Then the converse to the Poincaré lemma shows immediately that

$$\beta = \mathbf{dA}$$

for some 1-form $\mathbf{A}$. Substitute this result into the remaining homogeneous equation,

$$\begin{aligned} \mathbf{d}\varepsilon + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{dA} &= 0 \\ \mathbf{d}\left(\varepsilon + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{A}\right) &= 0 \end{aligned}$$

A second use of the converse to the Poincaré lemma shows that there exist a 0-form $-\varphi$ such that

$$\varepsilon + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{A} = -\mathbf{d}\varphi$$

and therefore

$$\varepsilon = -\mathbf{d}\varphi - \frac{1}{c}\frac{\partial}{\partial t}\mathbf{A}$$

References

[1] Here's a test Bibliography item.

[2] Tjiang, P. C. and Sutanto, S. H., On the Derivation of Conserved Quantities in Classical Mechanics, arXiv:physics/0302056v3.

[3] Muñoz, G., Vector constants of the motion and orbits in the Coulomb/Kepler problem, ArXiv:physics/0303106.

[4] Euler, L., Novi Comm. Acad. Sci. Petrop. 11 (1765) 144.

[5] Kustaanheimo, P. Annales Universitatis Turkuensis. Ser. AI. 73 (1964)

[6] Kustaanheimo, P. and Stiefel, E., J. Reine Angew. Mathematik 218 (1965) 204.

[7] Stiefel, E. and Scheifele, G., Linear and Regular Celestial Mechanics (Springer, 1971).

[8] Bartsch, T., The Kustaanheimo-Steifel transformation in geometric algebra, ArXiv:physics/0301017 v1

[9] Hestenes, D. and Sobcyzk, G., *Clifford Algebra to Geometric Calculus*, (Reidel Publishing Co., 1984).

[10] Ray, S. and Shamanna, J., Orbits in a central force field: bounded orbits, ArXiv:physics/0410149 v1.

[11] A. P. Balachandran, T. R. Govindaragan and B. Vijayalakshmi, Phys. Rev. D18 (1978) 1950.

[12] J. A. Kobussen, Acta Phys. Austr. **51** (1979) 293.

[13] C. C. Yan, Am. J. Phys. 46 (1978) 671.

[14] C. C. Yan, Am. J. Phys. 49 (1981) 269.

[15] Leubner, C. and Zoller, P., Gauge invariant interpretation of multiphoton transition probabilities, J. Phys. B: Atom. Molec. Phys. **13** (1980) 3613-3617.

[16] Leubner, C., Inequivalent Lagrangians from constants of the motion, Phys. Lett. 86A, no.2 (1981) 68-70.

[17] Lovelock, D., J. Math. Phys. **12**, (1971) 498.

[18] Wheeler, J. T., Nucl. Phys. **B273**, (1986) 732.

[19] Lie, S., Arch. Math. Naturvidenskab (1887) 129.

[20] Dirac, P. A. M., Proc. Camb. Phil. Soc. 29 (1933) 389.

[21] Dirac, P. A. M., Generalized Hamiltonian dynamics, Canad. J. Math. 2 (1950) 129.

[22] Whittaker, *Analytical Dynamics* (New York, 1944) 305.

[23] Wolsky, A. M., Amer. Journ. Phys., **39** (1971) 529.

[24] D. F. Currie and E. J. Saletan, J. Math. Phys. **7** (1966) 967.

[25] Currie, D. G., Jordan, T. F., and Sudarshan, E. G., Rev. Mod. Phys. 35 (1963) p 350.

[26] Y. Gelman and E. J. Saletan, q-Equivalent Particle Hamiltonians, Nuovo Cimento 18B (1973) 53.

[27] Bolza, O., *Lectures on the Calculus of Variations*, (Dover, New York, 1961) 30.

[28] S. Okubo, Phys. Rev. D22 (1980) 919.

[29] S. Okubo, Phys. Rev A23 (1981) 2776.

[30] López, G., About ambiguities appearing in the study of classical and quantum harmonic oscillator, Vevista Mexicana de Física **48** (1) 10-15.

[31] Ghosh, Shamanna, and Talukdar, Inequivalent Lagrangians for the damped harmonic oscillator, Can. Jour. of Physics **82** (2004) 561-567.

[32] Vujanovic, B., Group-variational procedure for finding first integrals of dynamical systems, Int. J. Non-linear Mechanics, Vol. 5, (1970) 269-278.

[33] Gorringe, V. M. and Leach, P. G. L., The first integrals and their Lie algebra of the most general autonomous Hamiltonian of the form $H = T + V$.

[34] Laplace, P. S., *Celestial Mechanics*, Vol. 1(Chelsea, NY, 1969) 344.

[35] Runge, C., *Vectoranalysis*, Vol. 1 (Hirzel, Leipzig, 1919) 70.

[36] Lenz, W., On the Course of the Motion and the Quantum States of the Disturbed Kepler Motion, Z. Phys. **24** (1924) 197-207.

[37] Goldstein, H., More on the prehistory of the Laplace of Runge-Lenz vector, Am. J. Phys. **44** (1976) 1123-1124.

[38] Goldstein, H., Poole, C. P., Safko, J. L., *Classical Mechanics (Third Edition)*, (Pearson Education, Inc., Addison-Wesley, San Francisco, 2002).

[39] Abelson, H. di Sessa, A., and Rudolph, L., Velocity space and the geometry of planetary orbits, Am. J. Phys. **43** , (1975) 579-589.

[40] Patera, R. P., Momentum-space derivation of the Runge-Lenz vector, Am. J. Phys. **49** (1981) 593-594.

[41] Derbes, D., Reinventing the wheel" Hodographic solutions to the Kepler problems, Am. J. Phys. **69** (2001) 481-489.

[42] Hamilton, W. R., On the applications of the method of quaternions to some dynamical questions, in *The Mathematical Papers of Sir William Rowan Hamilton*, edited by H. Halberstam and R. E. Ingram (Cambridge University Press, 1967), Vol. III, pp. 441-448.

[43] Bouda, A., From a mechanical Lagrangian to the Schrödinger equation. A modified version of the quantum Newton law, International Journal of Modern Physics A, Vol. 18, No. 19 (2003) 3347-3368.

[44] Bohm, D., Phys. Rev. **85** (1952) 166, 180.

[45] Bohm, D. and Vigier, J. P., Phys. Rev. **96** (1954) 208.

[46] Messiah, A., *Quantum Mechanics*, Dover Publications, Toronto (1999) pp. 214-242, reprint of *Quantum Mechanics* (John Wiley & Sons, USA 1958) Translated by G. H. Temmer from *Mécanique Quantique*.

[47] Floyd, E. R., Phys Rev **D25** (1982) 1547.

[48] Floyd, E. R., Phys Rev **D26** (1982) 1339.

[49] Landau, L. D., and Lifshitz, E. M., *Mechanics (Third Edition) Course of theoretical mechanics, Volume 1*, (Butterworth Heinmann, Oxford, 1976).

[50] Abraham, R. and Marsden, J. E., *Foundations of Mechanics, Second Edition*, (Addison-Wesley, New York, 1985) pp.462-472.

[51] Talukdar, B., Ghosh, S., and Shamanna, J., Canonical structure of the coupled Korteweg-de Vries equations, Can. J. Phys **82** (2004) 459-466.

[52] May, R., Simple Mathematical Models with Very Complicated Dynamics, Nature 261 (1976) 459-467.

[53] May, R. and Oster, G. F., Bifurcations and Dynamic Complexity in Simple Ecological Models, The American Naturalist 110 (1976) 573-599.

[54] May, R., Biological Populations with Nonoverlapping Generations:Stable Points, Stable Cycles, and Chaos, Science 186 (1974) 645-647.

[55] Wheeler, J. T., Gauging Newton's Law, ArXiv: hep-th\0305017.

[56] Anderson, L. B. and Wheeler, J. T., Quantum mechanics as a measurement theory on biconformal space, Arxiv: hep-th\0406159.

[57] Utiyama, R., Phys. Rev. **101** (1956) 1597.

[58] Kibble, T. W. B., J. Math. Phys. **2** (1961) 212.

[59] Wheeler, J. T., New conformal gauging and the electromagnetic theory of Weyl, J. Math. Phys. **39** (1998) 299, Arxiv: hep-th\9706214.

[60] Wehner, A. and Wheeler, J. T., Conformal actions in any dimension, Nuc. Phys. **B557** (1999) 380-406, Arxiv: hep-th backslash

[61] Cartan, E., Les groupes projectifs qui ne laissent invariante aucune multiplicité plane, Bull. Soc. Math. France, **41** (1913) 53-96.

[62] Cartan, E., *The theory of spinors*, English translation, Hermann, Paris, 1966, reprinted by Dover, New York, 1981. Originally printed in French in 1937 from Élie Cartan's lectures, collected and arranged by André Mercier.

[63] Dirac, P. A. M., Proc. Roy. Soc. Lon. **A117** (1928) 610.

[64] Dirac, P. A. M., Proc. Roy. Soc. Lon. **A126** (1930) 360.

[65] Penrose, R., A Spinor Approach to General Relativity, Ann. Phys. 10 (1960) 171-201.

[66] Penrose, R. and Rindler, W., *Spinors and space-time*, Vols. I and II (Cambridge University Press, Cambridge, 1984).

[67] Petrov, A. Z., *New methods in General Relativity* (in Russian), (Nauka, Moscow, 1966). English edition: Einstein spaces (Pergamon Press, 1969).

[68] Wald, R. M., *General Relativity*, (University of Chicago Press, Chicago, 1984), 373-375.

[69] Schoen, R. and Yau, S.-T., Proof of the Positive-Action Conjecture in Quantum Relativity, Phys. Rev. Lett, **42** (1979) 547-548.

[70] Schoen, R. and Yau, S.-T., Proof of the Positive Mass Theorem. II, Commun. Math. Phys.**79** (1981) 231-260.

[71] Schoen, R. and Yau, S.-T., Proof That the Bondi Mass is Positive, Phys. Rev. Lett, **48** (1982) 369-371.

[72] Witten, E., A New Proof of the Positive Energy Theorem, Commun. Math. Phys. **80** (1981) 381-402.

[73] Witten, E., Commun. Math. Phys. **252** (2004) 189-258.

[74] Casalbuoni, R., On the quantization of systems with anticomuting variables, Nuovo Cim. **33A** (1976) 115-125.

[75] Casalbuoni, R., The classical mechanics, fo Bose-Fermi systems, Nuovo Com. **33A** (1976a) 389-431.

[76] Freund, P. G. O., Supersymmetry, Cambridge University Press (1986), chapters 10 and 11.

[77] Berezin, F. A. and Marinov, M. S., Particle spin dynamics as the grassmann variant of classical mechanics, Ann. Phys. (NY) **104** (1977) 336-62.

[78] Brink, L., Deser, S., Zumino, B., di Vecchia, P. and Howe, P., Local sypersymmetry for spinning particles, Phys. Lett. **64B** (1976) 435-438.

[79] Morgan, J. A., Spin and statistics in classical mechanics, ArXiv: physics/0401070. This article provides a very accessible introduction to pseudomechanics, including 87 references on pseudomechanics and the history of the spin statistics theorem.

[80] Pimentel, B. M., Teixeira, R. G. and Tomazelli, J. L., Hamilton-Jacobi approach to Berezinian singular systems, ArXiv physics/9712016.

[81] Cariñena, J. F. and Figueroa, H., Hamiltonian vs. Lagrangian formulations of supermechanics, ArXiv dg-ga/9703016.

[82] J. Klauder, Phase space geometry in classical and quantum mechanics, quant-ph/0112010.

[83] P.A.M. Dirac, *The Principles of Quantum Mechanics, Fourth Edition,* (Oxford University Press, New York, 1958).

[84] C. Zachos, T. Curtright, Phase space quantization of field theory, Prog. Theor. Phys. Suppl. 135 (1999) 244-258.

[85] N.M.J. Woodhouse, *Geometrical Quantization*, (Clarendon, New York, 1992).

[86] R. P. Feynman, A. R. Hibbs, *Quantum Mechanics and Path integrals* (McGraw-Hill, New York, 1965).

[87] F. Berezin, Sov. Math. Izv. 38 (1974) 1116; Sov. Math. Izv. 39 (1975) 363.

[88] J. Isidro, Darboux theorem and quantization, hep-th/0112032.

[89] J. Isidro, Duality, Quantum mechanics and (almost) complex manifolds, hep-th/0304143.

[90] F. London, Z. Phys. 42, 375 (1927).

[91] J.T. Wheeler, Quantum measurement and geometry, Phys. Rev. D 41 (2) (1990) 431-441.

[92] E. Nelson, *Quantum Fluctuations* (Princeton University, Princeton, NJ, 1987).

[93] Wheeler, J. T., SU(3) x SU(2) x U(1) as the residual gauge group of the spacetime metric choice, *The Vancouver Meeting – Particles and Fields '91*, edited by D. Axen, D. Bryman and M. Comyn, World Scientific (1992) 854-856; J. T. Wheeler, SU(3) x SU(2) x U(1)SU(3): The residual symmetry of conformal gravity, *Mathematical Aspects of Classical Field Theory*, Contemporary Mathematics, Am. Math. Soc., Edited by M. J. Gotay, J. E. Marsden, & V. Moncrief (1992) 635-644.

[94] D. Hochberg, J. T. Wheeler, Scale invariant metric from conformal gauge theory, Bartol Research Institute Preprint, BA-89-24, (1989).

[95] A. Wehner, A new conformal gauge theory, Dissertation, Utah State University, 2001.

[96] J.T. Wheeler, New conformal gauging and the electromagnetic theory of Weyl, J. Math. Phys. 39 (1998) 299.

[97] A. Wehner, J.T. Wheeler, Conformal actions in any dimension, Nuc. Phys. B 577 (1999) 380-406.

[98] L.B. Anderson, J.T. Wheeler, Biconformal supergravity and the AdS/CFT conjecture. Nuc. Phys. B 686 (2004) 285-309. hep-th/030911

[99] E. A. Ivanov, J. Niederle, Gauge formulation of gravitation theories. I. The Poincaré, de Sitter, and conformal cases, Phys. Rev. D25, No 1 (1982) 976.

[100] E. A. Ivanov, J. Niederle, Gauge formulation of gravitation theories. II. The special conformal case, Phys. Rev. D25, No 4 (1982) 988.

[101] E. Bergshoeff, M. de Roo, B. de Wit, Conformal supergravity in ten dimensions, Nuc. Phys. B217 (1983) 489-530.

[102] A. Einstein, Sitz. Ber. Preuss. Akad. Wiss. 26, 478 (1918), including Weyl's reply.

[103] M. Kaku, P.K. Townsend, P. van Nieuwenhuizen, Properties of conformal supergravity, Phys. Rev. D 19 (10) (1979) 3166.

[104] J.T. Wheeler, Gauging Newton's Law, hep-th/0305017, submitted for publication.

[105] R. Abraham, J. Marsden, *Foundations of Mechanics* (Addison Wesley, New York, 1985).

[106] J. Jose, E. Saletan, *Classical Dynamics, A Contemporary Approach* (Cambridge University Press, Cambridge, 1998).

[107] L.D. Landau, E.M. Lifshitz, *Mechanics (Third Edition), Course of Theoretical Physics, Volume 1*, (Butterworth Heinmann, Oxford, 1976).

[108] E. Abers, *Quantum Mechanics*, (Pearson Education, New Jersey, 2004).

[109] H. Flanders, *Differential forms with applications to the physical sciences,* (Dover Publications, New York, 1989).

[110] J.S. Bell, Physics 1, (1965) 195.

[111] L. O'Raifeartaigh, *The Dawning of Gauge Theory*, (Princeton University Press, Princeton, 1997).

[112] J.P. Crawford, Cifford algebra: notes on the spinor metric and Lorentz, Poincaré, and conformal groups, J. Math. Phys. 32 (3) (1991) 576.