Summary of SI2360 Analytical Mechanics and Classical Field Theory

Yashar Honarmandi
yasharh@kth.se

12 juni 2020

Sammanfattning

Innehåll

1 Variational Calculus

The objective of Variational Calculus In variational calculus, we study the optimization of functionals, typically integrals, which are maps from functions to scalars.

Formulating the Problem We will primarily be interested in the following problem, as well as its derivatives:

Consider a function (or set of functions) q which assumes fixed values at a and b and the functional

$$S\left(q, \frac{\mathrm{d}q}{\mathrm{d}\tau}\right) = \int_a^b \mathrm{d}\tau\, F\left(q, \frac{\mathrm{d}q}{\mathrm{d}\tau}\right).$$

Describe the function q such that S has an extremum.

What derivatives of this are interesting? For instance, functionals that depend on higher derivatives and functionals based on multiple integrals, which generalize from this problem. But problems with different boundary conditions in q, entirely without boundary conditions in q or even with a functional as a boundary condition are also of interest. However, to introduce the involved techniques, we will first be studying problems of the above form.

The Variation of a Function The strategy for solving these problems is to assume that there exists a function that extremizes the functional and that it belongs to a family of functionals where all members satisfy the boundary conditions. We also assume that the set of functions in the family can be parametrized with a parameter α such that $\alpha = 0$ corresponds to the extremum[1], allowing us to analyze the problem using the tools of single-variable calculus.

We will be interested in small deviations from the extremum, and linearizing about the extremum gives rise to the quantity

$$\left.\frac{\mathrm{d}q}{\mathrm{d}\alpha}\right|_0 \mathrm{d}\alpha\,.$$

We define this to be the variation of q and denote it as δq.

The Variation of a Functional We now in a similar way define the variation of a functional as

$$\delta S = \left.\frac{\mathrm{d}S}{\mathrm{d}\alpha}\right|_0 \mathrm{d}\alpha\,.$$

Differentiating under the integral sign, and now working under the assumption that multiple q^i are involved, yields

$$\begin{aligned}\delta S &= \mathrm{d}\alpha \int_a^b \mathrm{d}\tau\, \partial_{q^i} F \frac{\mathrm{d}q^i}{\mathrm{d}\alpha} + \partial_{\dot{q}^i} F \frac{\mathrm{d}\dot{q}^i}{\mathrm{d}\alpha} \\ &= \int_a^b \mathrm{d}\tau\, \partial_{q^i} F\, \delta q^i + \partial_{\dot{q}^i} F\, \delta \dot{q}^i\,,\end{aligned}$$

where the dot represents a derivative with respect to τ.

We see that the variation operation behaves exactly like a derivative, and according to the definition, it commutes with all other derivatives that may be involved, assuming sufficient smoothness. These results will therefore be used without further argument.

As a side note, what in god's name is the partial derivative of a function with respect to another function? And what about a derivative with respect to a derivative? And should the derivative with respect to a derivative not contain some information about the derivative of the function the derivative of which we are differentiating with respect to? These are all very good and somewhat complex questions.

[1] An example of such a parametrization is $q(\alpha) = q + \alpha\eta$, where q is the actual extremum and η is some function that fits the boundary condition. The following steps can thus be performed in terms of η, if that feels more reasonable.

Mathematically, it is somewhat beyond me to give a proper answer. But physically, we can think of it the following way: The functions q^i, as we will see later, will represent the path of the system, and τ will be replaced with the time t. As such, by varying q^i we vary the path of the system, and by varying $\dot{q}^i$ we vary the velocity with which the system traverses the path. What we are doing with variational calculus is testing out an infinite number of paths to find the extremum of the action, and surely we must both try out different paths and traverse them with different velocities in order to find the extremum. That is why the functions and their derivatives can be treated separately.

What is a derivative with respect to a function? To answer this, recall that at every point τ, the functions are just numbers. When studying the variation of a functional, we are in truth studying its dependence on α. In order to do this, the chain rule states that we must first compute an outer partial derivative, which is exactly what the derivative with respect to a function is. This can in turn be indirectly translated to say something about how the functional changes when the function is varied.

Solving the Variational Problem We have now tried varying the functional around the extremum. Single-variable calculus gives the condition that $\frac{\mathrm{d}S}{\mathrm{d}\alpha} = 0$ at the extremum, which is equivalent to $\delta S = 0$. In other words,

$$\int_a^b \mathrm{d}\tau\, \partial_{q^i} F\, \delta q^i + \partial_{\dot{q}^i} F\, \delta \dot{q}^i = 0.$$

We can integrate this by parts to obtain

$$\left[\partial_{\dot{q}^i} F\, \delta q^i\right]_a^b + \int_a^b \mathrm{d}\tau \left(\partial_{q^i} F - \frac{\mathrm{d}}{\mathrm{d}\tau}\partial_{\dot{q}^i} F\right) \delta q^i = 0.$$

The first term from the integration by parts can be handled in two ways. If the variational problem has fixed boundary conditions, the families q^i must have been chosen such that all functions in the family satisfy the boundary conditions. Thus the variations of these at the endpoints vanish. Otherwise, the two arising terms might be used as boundary conditions themselves (the need for this arises due to the problem in question being second-order, and thus requiring two conditions). As q^i may be varied in any way possible, it is thus clear that when using these boundary terms as conditions, they must be set equal to zero.

The remaining integral might of course happen to be zero for the given choice of families of q^i. But the extremum is an extremum no matter what choice I make. So by changing up the problem - for instance, by reparametrizing the q^i or study an entirely different family in a similar way - I still obtain the same results. This must imply that the integral is zero no matter what δq^i is. And for this to be true, the only possibility is for the integrand to always be zero. To be absolutely sure, we can try varying only one coordinate at a time. This implies that

$$\partial_{q^i} F - \frac{\mathrm{d}}{\mathrm{d}\tau}\partial_{\dot{q}^i} F = 0$$

for all i, always. Solutions to this set of equations are thus our extrema, and are called the Euler-Lagrange equations.

Variational Problems With Higher-Order Derivatives What if the integrand also involves higher-order derivatives of the q^i? We can retrace the above steps mostly, but we will have to perform an extra (series of) integration(s) by parts. For instance, by including the second derivative and adding one extra integration, you should be able to show (unless this is wrong) that the extremum solution satisfies

$$\partial_{q^i} F - \frac{\mathrm{d}}{\mathrm{d}\tau}\partial_{\dot{q}^i} F + \frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\partial_{\ddot{q}^i} F = 0.$$

The Functional Derivative For a variational problem with fixed boundary conditions we obtained

$$\delta S = \int_a^b \mathrm{d}\tau \left(\partial_{q^i} F - \frac{\mathrm{d}}{\mathrm{d}\tau}\partial_{\dot{q}^i} F\right) \delta q^i\,.$$

In particular, if only a single q^i is varied, the entire variation can be traced back to the effect of a single functional. This inspires us to define the functional derivative $\frac{\delta S}{\delta q^i}$ as the function such that

$$\delta S = \int_a^b \mathrm{d}\tau \, \frac{\delta S}{\delta q^i} \, \delta q^i \, .$$

Using this definition, the Euler-Lagrange equations may be written as

$$\frac{\delta S}{\delta q^i} = 0.$$

2 Group Theory

Definition of a Group A group is a set of objects G with an operation $G \times G \to G, \ (a, b) \to ab$ such that

- If $a, b \in G$ then $ab \in G$.
- $a(bc) = (ab)c$ for all $a, b, c \in G$.
- There exists an identity e such that $ae = ea = a$ for all $a \in G$.
- There exists for every element a an inverse $a^{-1} \in G$ such that $aa^{-1} = a^{-1}a = e$.

Groups can be

- cyclic, i.e. all elements in the group are powers of a single element.
- finitie, i.e. groups containing a finite number of elements, or infinte.
- discrete, i.e. all elements in the group can be labelled with some index, por continuous.
- commutative, i.e. $ab = ba$ for all elements in the group, or non-commutative.

Subgroups If $G = \{g_\alpha\}$ and the subset $H = \{h_\alpha\}$ is also a group, we call H a subgroup of G and write $H < G$.

Conjugacy Classes Two group elements a and b are conjugate if there exists an element g such that

$$a = gbg^{-1}.$$

We write $a \sim b$.

Equivalence Relations An equivalence relation is a relation (here denoted $=$) between two things such that

- $a = b \equiv b = a$.
- $a = b, \ b = c \implies a = c$.

Example: Conjugacy as an Equivalence Relation

Homomorphisms and Isomorphisms A homomorphisms is a map $f : G \to H$ such that $f(g_1)f(g_2) = f(g_1 g_2)$. If the map is bijective, f is called an isomorphism.

Direct Products GIven two groups F and G, we define $F \times G$ as the set of ordered pairs of elements of the two groups. The group action of $F \times G$ is the group actions of F and G separately on the elements in the ordered pair.

Generators For discrete groups, the generators of a group is the smallest set of elements in the group such that all other elements in the group can be composed by the elements in the set. For continuous groups, we will use the term generators to refer to elements such that any group element can be written as real powers of this element.

Point Groups Point groups are symmetries of, for instance, a crystal structure that leave at least one point in the structure invariant. Examples include

- rotations.
- reflections.
- spatial inversions.

Combining these with certain discrete translation, you obtain the space groups of the crystal. Space groups are the groups of all symmetries of a crystal.

Dihedral Groups The dihedral group D_n is the group of transformations that leave an n-sided polygon invariant.

Lie Groups A Lie group is a group containing a manifold with the group operation and inverse operation being smooth maps. My current understanding of the significance of this is that it allows us to differentiate and expand the group elements with respect to certain parameters.

The group elements are denoted $g(\boldsymbol{\theta})$, where $g(\mathbf{0}) = 1$. We write them as

$$g(\boldsymbol{\theta}) = e^{\theta_a T_a}$$

where we have introduced the generators T_a. This is reasonable, partially because it is reasonable for a smooth operation to have an addition operation (I think), and in this case the series definition of the exponential function yields exactly that any element is generated by powers of the generators. Note however that the generators themselves are typically not group elements, which might leave a hole in this reasoning. In addition, the question of whether the group is Abelian leaves it questionable whether the addition of generators makes sense, but I will return to this concern. There is probably a deeper understanding to this, and what I am saying may even be completely incorrect. Depending on the context, the exponent may also contain a factor $-i$, but this discussion will omit it.

The Lie Algebra Expanding an element around the identity yields

$$g(\boldsymbol{\theta}) \approx 1 + \theta_a T_a.$$

First of all, we note that performing this for exponentials of a single generator yields

$$e^{\theta_i T_i} e^{\theta_j T_j} \approx (1 + \theta_i T_i)(1 + \theta_j T_j),$$

which is equal to $e^{\theta_i T_i + \theta_j T_j}$ to first order even for a non-commutative group. Hence the addition of generators is reasonable. Furthermore, this implies that set of generators is a vector space, termed the Lie algebra.

The Lie Bracket Having seen that the exponential notation makes sense, we study it for a non-commutative group. The element $e^{-\theta_i T_i} e^{-\theta_j T_j} e^{\theta_i T_i} e^{\theta_j T_j}$ is equal to the identity for a commutative group, and we would like to study this in the general case. We have

$$\begin{aligned}
e^{\theta_i T_i} &\approx 1 + \theta_i T_i + \frac{1}{2}\theta_i^2 T_i^2, \\
e^{\theta_i T_i} e^{\theta_j T_j} &\approx \left(1 + \theta_i T_i + \frac{1}{2}\theta_i^2 T_i^2\right)\left(1 + \theta_j T_j + \frac{1}{2}\theta_j^2 T_j^2\right) \approx 1 + \theta_i T_i + \theta_j T_j + \frac{1}{2}\left(\theta_i^2 T_i^2 + \theta_j^2 T_j^2\right) + \theta_i \theta_j T_i T_j.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
e^{-\theta_i T_i} e^{-\theta_j T_j} e^{\theta_i T_i} e^{\theta_j T_j} \approx & 1 + \theta_i T_i + \theta_j T_j + \frac{1}{2}\left(\theta_i^2 T_i^2 + \theta_j^2 T_j^2\right) + \theta_i \theta_j T_i T_j - \theta_i T_i \left(1 + \theta_i T_i + \theta_j T_j\right) \\
& - \theta_j T_j \left(1 + \theta_i T_i + \theta_j T_j\right) + \frac{1}{2}\left(\theta_i^2 T_i^2 + \theta_j^2 T_j^2\right) + \theta_i \theta_j T_i T_j \\
= & 1 + \theta_i \theta_j T_i T_j - \theta_i \theta_j T_i T_j - \theta_i \theta_j T_j T_i + \theta_i \theta_j T_i T_j \\
= & 1 + \theta_i \theta_j \left[T_a, T_b\right]
\end{aligned}$$

to second order, where we have introduced the Lie bracket

$$[T_i, T_j] = T_i T_j - T_j T_i.$$

Hence, the non-commutativity of Lie groups close to the identity is described by the Lie brackets, which is why we study them. Furthermore, we have

$$e^{-\theta_i T_i} e^{-\theta_j T_j} e^{\theta_i T_i} e^{\theta_j T_j} \approx e^{\theta_i \theta_j [T_a, T_b]},$$

implying that the Lie bracket belongs to the vector space spanned by the generators and allowing us to write

$$[T_a, T_b] = f_{a,b,c} T_c.$$

The constants $f_{a,b,c}$ are called structure constants.

Representations A representation is a homomorphism $D : G \to GL(V)$, where $GL(V)$ is the group of all invertible linear transformations on V. The group elements thus act on V according to

$$D(g_1)D(g_2)v = D(g_1 g_2)v, \ v \in V.$$

Reducible and irreducible representations Two representations are equivalent if they satisfy $S^{-1}DS = D'$, where S is a matrix representing a change of basis. Some representations can be written as direct sums in certain bases. For these, there is a basis where the representation is block diagonal. These are reducible. Those that cannot are irreducible.

Example: Rotations in Two Dimensions Consider a rotation of an infinitesimal displacement d$\mathbf{x}$ with a rotation R. The requirement for length to be preserved implies $R^T R = 1$.

Consider now a rotation by a small angle $\delta\theta$. Taylor expanding it in terms of the angle yields

$$R(\delta\theta) \approx 1 + A\,\delta\theta\,.$$

The requirement for R to be orthogonal yields $A^T = -A$. We choose the solution

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

We can now write the rotation matrix as

$$R(\delta\theta) = \begin{bmatrix} 1 & \delta\theta \\ -\delta\theta & 1 \end{bmatrix}.$$

We would now like to construct a large rotation in terms of smaller rotations as

$$R(\theta) = \lim_{N\to\infty} \left(1 + \frac{\theta}{N} J\right)^N = e^{\theta J}.$$

We can write this as an infinite series and use the fact that $J^2 = -1$ to obtain

$$R(\theta) = \cos\theta + J \sin\theta.$$

Example: Rotations in Three Dimensions The argument done for two dimensions does not use the dimensionality, so we conclude that even for higher dimensions, $R^T R = 1$. Expanding a small rotation around the identity yields that the first-order term must include an antisymmetric matrix. The space of antisymmetric 3×3 matrices is three-dimensional. We thus choose the basis

$$J_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}, J_y = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, J_z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Exponentiating yields

$$R(\theta) = e^{\sum \theta_i J_i} = e^{\theta \cdot \mathbf{J}}.$$

In physics we usually extract a factor i such that the basis matrices are Hermitian, and the rotation becomes

$$R(\theta) = e^{i\theta\cdot\mathbf{J}}.$$

The set of generators of these rotations constitutes the Lie algebra.

We know in general that rotations in three dimensions do not commute. In fact, we obtain in general that

$$R(\theta)R(\theta')R^{-1}(\theta) = \theta_a\theta'_b\left[J_a, J_b\right],$$

where $[J_a, J_b]$ is the commutator. This commutator satisfies

$$[J_a, J_b]^T = \left[J_b^T, J_a^T\right] = [-J_b, -J_a] = -\left[J_a, J_b\right],$$

which implies

$$[J_a, J_b] = f_{a,b,c}J_c.$$

It can be shown that

$$[J_i, J_j] = \varepsilon_{i,j,k}J_k,$$

or in a physics context (where a factor i is extracted):

$$[J_i, J_j] = i\varepsilon_{i,j,k}J_k.$$

3 Differential Geometry

Coordinates A general set of coordinates on $\mathbb{R}^n$ is n numbers $\chi^a, a = 1, \ldots, n$ that uniquely define a point in the space.

Example: Cartesian Coordinates In cartesian coordinates we introduce an orthonormal basis $\mathbf{e}_i$. We can then write $\mathbf{x} = \chi^i\mathbf{e}_i$. This example is, however, not very illustrative.

Basis Vectors When working with both Cartesian and non-Cartesian coordinate systems, there are two different choices of coordinate bases.

The first is the tangent basis of vectors

$$\mathbf{E}_a = \partial_{\chi^a}\mathbf{r} = \partial_a\mathbf{r}.$$

The second is the dual basis

$$\mathbf{E}^a = \vec{\boldsymbol{\nabla}}\chi^a.$$

Example: Cartesian Coordinates In Cartesian coordinates we have $\mathbf{r} = x^i\mathbf{e}_i$. The basis vectors are fixed and orthonormal, meaning $\mathbf{E}_a = \delta^i_a\mathbf{e}_i = \mathbf{e}_a$. Likewise, we have $\mathbf{E}^a = \vec{\boldsymbol{\nabla}}x^a = \mathbf{e}_a$. As we can see, the tangent and dual basis are equal in Cartesian coordinates.

Example: Polar Coordinates A slightly more non-trivial example is polar coordinates, where we have

$$\mathbf{r} = r(\cos\phi\mathbf{e}_x + \sin\phi\mathbf{e}_y).$$

The tangent basis vectors are thus

$$\mathbf{E}_r = \cos\phi\mathbf{e}_x + \sin\phi\mathbf{e}_y,\ \mathbf{E}_\phi = r(-\sin\phi\mathbf{e}_x + \cos\phi\mathbf{e}_y).$$

It is not quite obvious how to find the dual basis vectors - to compute the gradient, you would need to express the polar coordinates in terms of Cartesian coordinates, and this might not even be possible (in the case of polar coordinates, it isn't). To circumvent this problem, we compute the gradients of x and y, yielding

$$\mathbf{e}_x = \cos\phi\vec{\boldsymbol{\nabla}}r - r\sin\phi\vec{\boldsymbol{\nabla}}\phi,\ \mathbf{e}_y = \sin\phi\vec{\boldsymbol{\nabla}}r + r\cos\phi\vec{\boldsymbol{\nabla}}\phi.$$

The solutions to this are

$$\mathbf{E}^r = \cos\phi\mathbf{e}_x + \sin\phi\mathbf{e}_y,\ \mathbf{E}^\phi = \frac{1}{r}(-\sin\phi\mathbf{e}_x + \cos\phi\mathbf{e}_y).$$

Note that the use of Cartesian basis vectors was necessary in order to express the tangent and dual basis in terms of something sensible - otherwise, we would have no sense of space or direction.

Orthogonality We can now compute the scalar product

$$\mathbf{E}_a \cdot \mathbf{E}^b = \partial_a \mathbf{r} \cdot \vec{\boldsymbol{\nabla}} \chi^b = (\partial_a x^i \mathbf{e}_i) \cdot (\partial_{x^j} \chi^b \mathbf{e}_j) = \partial_a x^i \partial_{x^j} \chi^b \delta_{ij} = \partial_a x^i \partial_{x^i} \chi^b.$$

According to the chain rule, this is simply equal to $\partial_a \chi^b$, which again is equal to δ_a^b.

Note that the vectors in the tangent and dual bases are not necessarily orthogonal amongst themselves.

Vector Components Any vector can now be written as

$$\mathbf{v} = v^a \mathbf{E}_a = v_a \mathbf{E}^a.$$

The v^a are called contravariant components and the v_a are called covariant components.

Up until now we have not been careful about where we place the indices. This will now change. In addition, we add to the convention of Einstein summation the idea that the balance of raised and lowered indices msut be preserved by an equality.

Changes of Basis and Coordinate Transformations Suppose we perform the change of basis, expressed in the dual basis as

$$(\mathbf{E}_b)' = L_b^a \mathbf{E}_a.$$

This must be due to a change of coordinates. The chain rule dictates

$$(\mathbf{E}_b)' = \partial_b' \mathbf{r} = \partial_a \mathbf{r} \partial_b' \chi^a = \partial_b' \chi^a \mathbf{E}_a,$$

which identifies the transformation coefficients as

$$L_b^a = \partial_b' \chi^a.$$

Similarly, expressing the change of basis in the dual basis yields

$$(\mathbf{E}^b)' = K_a^b \mathbf{E}^a$$

yields

$$(\mathbf{E}^b)' = \vec{\boldsymbol{\nabla}} (\chi')^b = \partial_a (\chi')^b \vec{\boldsymbol{\nabla}} \chi^a = \partial_a (\chi')^b \mathbf{E}^a,$$

identifying the transformation coefficients as

$$K_a^b = \partial_a (\chi')^b.$$

These transformation coefficients satisfy

$$L_a^c K_c^b = \partial_a' \chi^c \partial_c (\chi')^b = \delta_a^b,$$

and hence the two transformations are inverses of each other.

Transformations of Vectors Under a coordinate transformation, a single contravariant vector component is given by

$$(v')^b = (\mathbf{E}^b)' \cdot v^a \mathbf{E}_a = v^a \vec{\boldsymbol{\nabla}} (\chi')^b \cdot \partial_a \mathbf{r} = v^a \partial_a (\chi')^b.$$

Likewise, the covariant components are given by

$$v_b' = \mathbf{E}_b' \cdot v_a \mathbf{E}^a = v_a \partial_b' \mathbf{r} \cdot \vec{\boldsymbol{\nabla}} \chi^a = v_a \partial_b' \chi^a.$$

The covariant components thus transform in the same way as the tangent basis, while the contravariant components change in the opposite way. This is the reason for the nomenclature.

Tangents to Curves The tangent to a curve is given by

$$\dot{\boldsymbol{\gamma}} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \partial_a \mathbf{x} \frac{\mathrm{d}\chi^a}{\mathrm{d}t} = \dot{\chi}^a \mathbf{E}_a.$$

We see that it is naturally expressed in terms of the tangent basis.

Gradients The gradient of a function is given by

$$\vec{\boldsymbol{\nabla}} f = \partial_a f \vec{\boldsymbol{\nabla}} \chi^a = \partial_a f \mathbf{E}^a.$$

We see that it is naturally expressed in terms of the dual basis.

Rates of change along a curve The rate of change of a quantity along a path is given by

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \partial_a f \frac{\mathrm{d}\chi^a}{\mathrm{d}t} = \vec{\boldsymbol{\nabla}} f \cdot \dot{\boldsymbol{\gamma}}.$$

4 Differentiation and Integration in Orthogonal Coordinates

To tie together what we have learned thus far with what we studied in Vector Calculus, we will study differentiation and integration in orthogonal coordinate systems. For this part of the summary we will take a break from the oh-so strict indexing rules established above.

Defining Relation Orthogonal coordinate systems are defined by the relation

$$\mathbf{E}_a \cdot \mathbf{E}_b = h_a^2 \delta_{ab} \text{ (no sum)}.$$

Orthonormal Basis Based on the orthogonality conditions, we define the orthogonal basis vectors

$$\mathbf{e}_a = \frac{1}{h_a} \mathbf{E}_a \text{ (no sum)}.$$

The normalization thus implies

$$h_a = \sqrt{\sum_i (\partial_a x^i)^2}.$$

Physical Components The physical components of a vector is its projection onto the orthonormal basis vectors, denoted with a tilde.

Relation to Dual Basis By expanding the dual basis vectors in terms of their physical components, we obtain

$$\tilde{E}_a^b = \mathbf{e}_b \cdot \mathbf{E}^a = \frac{1}{h_b} \delta_b^a \text{ (no sum)}.$$

This implies

$$\mathbf{E}^a = \tilde{E}_b^a \mathbf{e}_b = \frac{1}{h_a} \delta_a^b \mathbf{e}_b = \frac{1}{h_a} \mathbf{e}_a \text{ (no sum)},$$

and thus

$$\mathbf{e}_a = h_a \mathbf{E}^a.$$

We see that the dual basis would have been an equally good starting point for describing orthogonal systems.

Line Integrals Using our previous knowledge of rates of change along a curve, we have

$$\begin{aligned} \int_\Gamma \mathrm{d}\mathbf{r} \cdot \mathbf{v} &= \int_\Gamma \mathrm{d}\chi^a \, \mathbf{E}_a \cdot \mathbf{v} \\ &= \int_\Gamma \mathrm{d}\tau \, \dot{\chi}^a \mathbf{E}_a \cdot \mathbf{v} \\ &= \int_\Gamma \mathrm{d}\tau \sum_a \dot{\chi}^a h_a \tilde{v}_a. \end{aligned}$$

Specifically, when integrating along a χ^c coordinate line, we can use the coordinate as a parameter, yielding

$$\int_\Gamma \mathrm{d}\mathbf{r} \cdot \mathbf{v} = \int_\Gamma \mathrm{d}\tau \, \frac{\mathrm{d}\chi^a}{\mathrm{d}\chi^c} \mathbf{E}_a \cdot \mathbf{v} = \int_\Gamma \mathrm{d}\chi^c \, h_c v_c.$$

Surface Integrals Consider a coordinate level surface S_c. In three dimensions we have

$$\begin{aligned}\mathrm{d}\mathbf{S} &= \partial_a \mathbf{r} \times \partial_b \mathbf{r}\, \mathrm{d}\chi^a\, \mathrm{d}\chi^b \\ &= h_a h_b \mathbf{e}_a \mathbf{e}_b\, \mathrm{d}\chi^a\, \mathrm{d}\chi^b \\ &= \pm h_a h_b\, \mathrm{d}\chi^a\, \mathrm{d}\chi^b\, \mathbf{e}_c.\end{aligned}$$

We immediately identify the unit normal and area element.

The final results (hopefully) generalize to other dimensionalities, but I could not see any way of bypassing the need for the cross product in three dimensions.

Volume Integrals Consider an infinitesimal volume element separated by $2n$ coordinate surfaces corresponding to coordinate values χ^a and $\chi^a + \mathrm{d}\chi^a$. Using what we did with line integrals

$$\mathrm{d}V = \prod_i h_i\, \mathrm{d}\chi^i\,.$$

We identify the Jacobian as $\mathcal{J} = \prod_i h_i$.

5 Tensors

Definition A tensor of rank N is a multilinear map from N vectors to a scalar.

Components of a Tensor The components of a tensor are defined by

$$T(\mathbf{E}^{a_1}, \ldots, \mathbf{E}^{a_N}) = T^{a_1,\ldots,a_N}.$$

These are called the contravariant components of the tensor, and the covariant components are defined similarly. Mixed components can also be defined.

Basic Operations on Tensors Tensors obey the following rules:

$$\begin{aligned}(T_1 + T_2)(\mathbf{w}_1, \ldots, \mathbf{w}_{n_2}) &= T_1(\mathbf{w}_1, \ldots, \mathbf{w}_{n_2}) + T_2(\mathbf{w}_1, \ldots, \mathbf{w}_{n_2}), \\ (kT)(\mathbf{w}_1, \ldots, \mathbf{w}_{n_2}) &= kT(\mathbf{w}_1, \ldots, \mathbf{w}_{n_2}).\end{aligned}$$

In component form:

$$\begin{aligned}(T_1 + T_2)^{a_1\ldots a_n} &= T_1^{a_1\ldots a_n} + T_2^{a_1\ldots a_n}, \\ (kT)^{a_1\ldots a_n} &= kT^{a_1\ldots a_n}.\end{aligned}$$

Coordinate Transformations Under a change of coordinates, the components of the tensor then transform according to

$$\begin{aligned}T^{a_1\ldots a_n} &= T((\mathbf{E}_{a_1})', \ldots, (\mathbf{E}_{a_n})') \\ &= T(L^{b_1}_{a_1}\mathbf{E}_{b_1}, \ldots, L^{b_n}_{a_n}\mathbf{E}_{b_n}) \\ &= L^{b_1}_{a_1} \ldots L^{b_n}_{a_n} T(\mathbf{E}_{b_1}, \ldots, \mathbf{E}_{b_n}) \\ &= L^{b_1}_{a_1} \ldots L^{b_n}_{a_n} T^{b_1,\ldots,b_n}.\end{aligned}$$

Many introductions to tensors define tensors according to this relation. And now you know where it comes from.

The Tensor Product Given two tensors T_1 and T_2 of ranks n_1 and n_2, we kan define the rank $n_1 + n_2$ tensor $T_1 \otimes T_2$ as

$$(T_1 \otimes T_2)(\mathbf{v}_1, \ldots, \mathbf{v}_{n_1}, \mathbf{w}_1, \ldots, \mathbf{w}_{n_2}) = T_1(\mathbf{v}_1, \ldots, \mathbf{v}_{n_1}) T_2(\mathbf{w}_1, \ldots, \mathbf{w}_{n_2}).$$

In component form:

$$(T_1 \otimes T_2)^{a_1\ldots a_{n_1+n_2}} = T_1^{a_1\ldots a_{n_1}} T_2^{a_{n_1+1}\ldots a_{n_1+n_2}}.$$

A Base For The Space of Tensors Using the tensor product, all tensors can be written as linear combinations of certain basis elements due to their bilinearity. Define

$$e_{a_1 \dots a_n} = \mathbf{E}_{a_1} \otimes \dots \otimes \mathbf{E}_{a_n}$$

to be the tensor that satisfies

$$e_{a_1 \dots a_n}(\mathbf{E}^{b_1}, \dots, \mathbf{E}^{b_n}) = (\mathbf{E}_{a_1} \cdot \mathbf{E}^{b_1}) \dots (\mathbf{E}_{a_n} \cdot \mathbf{E}^{b_n}) = \delta^{b_1}_{a_1} \dots \delta^{b_n}_{a_n}.$$

Then any tensor can be written as

$$T = T^{a_1 \dots a_n} e_{a_1 \dots a_n}$$

where the $T^{a_1 \dots a_n}$ are exactly the contravariant components of T.

Tensors as Linear Transforms Between Tensors A rank n tensor can also be viewed as a linear map from rank m tensors to rank $n-m$ tensors. To do this, we first define, given T, the rank $n-m$ tensor $\tilde{T}(\mathbf{w}_1 \otimes \dots \otimes \mathbf{w}_m)$ such that

$$(\tilde{T}(\mathbf{w}_1 \otimes \dots \otimes \mathbf{w}_m))(\mathbf{v}_1, \dots, \mathbf{v}_{n-m}) = T(\mathbf{w}_1, \dots, \mathbf{w}_m, \mathbf{v}_1, \dots, \mathbf{v}_{n-m}).$$

This map is also linear in all the $\mathbf{w}_i$. Next, given a rank $n-m$ tensor $\tilde{T}$, one can define the rank $n-m$ tensor $T(\mathbf{w}_1, \dots, \mathbf{w}_m)$ such that

$$T(\mathbf{w}_1, \dots, \mathbf{w}_m, \mathbf{v}_1, \dots, \mathbf{v}_{n-m}) = (\tilde{T}(\mathbf{w}_1 \otimes \dots \otimes \mathbf{w}_m))(\mathbf{v}_1, \dots, \mathbf{v}_{n-m}).$$

This is a linear rank n tensor.

Tensor Contraction Given a complete set of vectors $\mathbf{v}_i$ and their dual $\mathbf{v}^i$ such that $\mathbf{v}_i \cdot \mathbf{v}^i = \delta^j_i$, the contraction $e_{12}T$ of two arguments of a rank n tensor is the tensor of rank $n-2$ satisfying

$$(e_{12}T)(\mathbf{w}_1, \dots, \mathbf{w}_{n-2}) = T(\mathbf{v}_i, \mathbf{v}^i, \mathbf{w}_1, \dots, \mathbf{w}_{n-2}).$$

In component form:

$$(e_{12}T)^{a_1 \dots a_{n-2}} = T^{c\ a_1 \dots a_{n-2}}_{c}.$$

The definition is similar (I assume) for the contraction of other arguments.

Tensor Fields A tensor field is a map from coordinate space to a tensor.

6 Advanced Differential Geometry

In this part we will expand on the previously discussed concepts of differential geometry, mainly by incorporating our knowledge of tensors into it.

The Metric Tensor The metric tensor g is a rank 2 tensor. We start by defining it as $g(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$, but more generally the metric tensor defines the inner product.

The metric tensor is symmetric. Its components satisfy

$$v_a = \mathbf{E}_a \cdot v^b \mathbf{E}_b = g(\mathbf{E}_a, \mathbf{E}_b) v^b = g_{ab} v^b,$$

and likewise

$$v^a = g^{ab} v_b$$

where $\mathbf{v}$ is a vector. This demonstrates the capabilities of the metric to raise and lower indices.

We note that

$$v_a = g_{ab} v^b = g_{ab} g^{bc} v_c,$$

which implies $g_{ab} g^{bc} = \delta^c_a$.

Example: The Metric in Polar Coordinates The contravariant components of the metric tensor, according to the definition, are

$$g_{rr} = \mathbf{E}_r \cdot \mathbf{E}_r = 1,\ g_{r\phi} = g_{\phi r} = \mathbf{E}_r \cdot \mathbf{E}_\phi = 0,\ g_{\phi\phi} = \mathbf{E}_\phi \cdot \mathbf{E}_\phi = r^2.$$

Likewise, the covariant components are

$$g^{rr} = \mathbf{E}^r \cdot \mathbf{E}^r = 1,\ g_{r\phi} = g_{\phi r} = \mathbf{E}^r \cdot \mathbf{E}^\phi = 0,\ g_{\phi\phi} = \mathbf{E}^\phi \cdot \mathbf{E}^\phi = \frac{1}{r^2}.$$

Christoffel Symbols When computing the derivative of a vector quantity, one must account both for the change in the quantity itself and the change in the basis vectors. We define the Christoffel symbols according to

$$\partial_b \mathbf{E}_a = \Gamma^c_{\ ba} \mathbf{E}_c.$$

These can be computed according to

$$\mathbf{E}^c \cdot \partial_b \mathbf{E}_a = \mathbf{E}^c \cdot \Gamma^d_{\ ba} \mathbf{E}_d = \delta^c_d \Gamma^d_{\ ba} = \Gamma^c_{\ ba}.$$

Note that

$$\partial_a \mathbf{E}_b = \partial_a \partial_b \mathbf{r} = \partial_b \partial_a \mathbf{r} = \partial_b \mathbf{E}_a,$$

which implies

$$\Gamma^c_{\ ba} = \Gamma^c_{\ ab}.$$

Do the Christoffel symbols define a tensor? Someone should probably study that.

Example: Christoffel Symbols in Polar Coordinates To compute these, we need partial derivative of the basis vectors. We have

$$\partial_r \mathbf{E}_r = \mathbf{0},\ \partial_\phi \mathbf{E}_r = \partial_r \mathbf{E}_\phi = \frac{1}{r}\mathbf{E}_\phi,\ \partial_\phi \mathbf{E}_\phi = -r\mathbf{E}_r.$$

We thus obtain

$$\Gamma^a_{\ rr} = 0,\ \Gamma^r_{\ r\phi} = 0,\ \Gamma^\phi_{\ r\phi} = \frac{1}{r},\ \Gamma^r_{\ \phi\phi} = -r,\ \Gamma^\phi_{\ \phi\phi} = 0.$$

Covariant Derivatives The partial derivate of $\mathbf{v} = v^a \mathbf{E}_a$ with respect to χ^a is given by

$$\partial_a \mathbf{v} = \mathbf{E}_b \partial_a v^b + v^b \partial_a \mathbf{E}_b = \mathbf{E}_b \partial_a v^b + v^b \Gamma^c_{\ ab} \mathbf{E}_c.$$

Renaming the summation indices yields

$$\partial_a \mathbf{v} = \mathbf{E}_b (\partial_a v^b + v^c \Gamma^b_{\ ac}),$$

which contains one term from the change in the coordinates and one term from the change in basis.

Realizing that derivatives of vector quantities must take both of these into account in order to transform like a tensor, we would like to define a differentiation operation that takes both of these to account when differentiating vector components. This is the covariant derivative. We define its action on contravariant vector components as

$$\vec{\boldsymbol{\nabla}}_a v^b = \partial_a v^b + v^c \Gamma^b_{\ ac},$$

such that

$$\partial_a \mathbf{v} = E_b \vec{\boldsymbol{\nabla}}_a v^a.$$

In a similar fashion we would like to define its action on covariant vector components. To do this, we use the fact that

$$\partial_a (\mathbf{E}_b \cdot \mathbf{E}^c) = \partial_a \delta^c_b = 0.$$

The product rule yields

$$\mathbf{E}_b \cdot \partial_a \mathbf{E}^c + \mathbf{E}^c \cdot \partial_a \mathbf{E}_b = \mathbf{E}_b \cdot \partial_a \mathbf{E}^c + \mathbf{E}^c \cdot \Gamma^d_{\ ab} \mathbf{E}_d = \mathbf{E}_b \cdot \partial_a \mathbf{E}^c + \delta^c_d \cdot \Gamma^d_{\ ab} = \mathbf{E}_b \cdot \partial_a \mathbf{E}^c + \Gamma^c_{\ ab},$$

which implies

$$\partial_a \mathbf{E}^c = -\Gamma^c_{\ ab} \mathbf{E}^b.$$

Repeating the steps above now yields

$$\vec{\boldsymbol{\nabla}}_a v_b = \partial_a v_b - \Gamma^c_{\ ab} v_c.$$

Covariant Derivatives of Tensor Fields

Christoffel Symbols and the Metric The derivatives of the metric tensor are given by

$$\partial_c g_{ab} = \mathbf{E}_a \cdot \partial_c \mathbf{E}_b + \mathbf{E}_b \cdot \partial_c \mathbf{E}_a = \mathbf{E}_a \cdot \Gamma^d_{\ cb} \mathbf{E}_d + \mathbf{E}_b \cdot \Gamma^d_{\ ca} \mathbf{E}_d = \Gamma^d_{\ cb} g_{ad} + \Gamma^d_{\ ca} g_{bd}.$$

Multiplying by g^{ea} and summing over a yields

$$g^{ea} \partial_c g_{ab} = \Gamma^d_{\ cb} g_{ad} g^{ea} + \Gamma^d_{\ ca} g_{bd} g^{ea} = \Gamma^d_{\ cb} g_{da} g^{ae} + \Gamma^d_{\ ca} g_{bd} g^{ea} = \Gamma^e_{\ cb} + \Gamma^d_{\ ca} g_{bd} g^{ea}.$$

The hope is that this can be used to obtain an expression for the Christoffel symbols. To try to do that, we will compare this to the expression obtained by switching c and b. This expression is

$$g^{ea} \partial_b g_{ac} = \Gamma^e_{\ bc} + \Gamma^d_{\ ba} g_{cd} g^{ea} = \Gamma^e_{\ cb} + \Gamma^d_{\ ba} g_{cd} g^{ea},$$

yielding

$$\begin{aligned}
\Gamma^e_{\ cb} &= \frac{1}{2} \left(g^{ea} \partial_c g_{ab} + g^{ea} \partial_b g_{ac} - \Gamma^d_{\ ca} g_{bd} g^{ea} - \Gamma^d_{\ ba} g_{cd} g^{ea} \right) \\
&= \frac{1}{2} g^{ea} \left(\partial_c g_{ab} + \partial_b g_{ac} - \Gamma^d_{\ ac} g_{bd} - \Gamma^d_{\ ac} g_{cd} \right) \\
&= \frac{1}{2} g^{ea} \left(\partial_c g_{ab} + \partial_b g_{ac} - \partial_a g_{bc} \right).
\end{aligned}$$

Curve Length Consider some curve parametrized by t, and let $\dot{\gamma}$ denote its tangent. The curve length is given by

$$\mathrm{d}s^2 = \mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{x} = g(\dot{\gamma}, \dot{\gamma}) \, \mathrm{d}t^2 = g_{ab} \dot{\chi^a} \dot{\chi^b} \, \mathrm{d}t^2 \, .$$

The curve length is now given by

$$L = \int \mathrm{d}t \, \sqrt{g_{ab} \dot{\chi^a} \dot{\chi^b}}.$$

Geodesics A geodesic is a curve that extremises the curve length between two points. From variational calculus, it is known that such curves satisfy the Euler-Lagrange equations, and we would like a differential equation that describes such a curve. By defining $\mathcal{L} = \sqrt{g_{ab} \dot{\chi}^a \dot{\chi}^b}$, the Euler-Lagrange equations for the curve length becomes

$$\partial_{\chi^a} \mathcal{L} - \frac{\mathrm{d}}{\mathrm{d}t} \partial_{\dot{\chi}^a} \mathcal{L} = 0.$$

The Euler-Lagrange equation thus becomes

$$\begin{aligned}
\frac{1}{2\mathcal{L}} \dot{\chi}^b \dot{\chi}^c \partial_a g_{bc} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{1}{2\mathcal{L}} g_{bc} (\dot{\chi}^b \delta^c_a + \dot{\chi}^c \delta^b_a) \right) &= 0, \\
\frac{1}{2\mathcal{L}} \dot{\chi}^b \dot{\chi}^c \partial_a g_{bc} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{1}{2\mathcal{L}} (g_{ba} \dot{\chi}^b + g_{ac} \dot{\chi}^c \right) &= 0, \\
\frac{1}{2\mathcal{L}} \dot{\chi}^b \dot{\chi}^c \partial_a g_{bc} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{1}{\mathcal{L}} g_{ac} \dot{\chi}^c \right) &= 0.
\end{aligned}$$

Expanding the time derivative yields

$$\frac{1}{2\mathcal{L}}\dot{\chi}^b\dot{\chi}^c\partial_a g_{bc} - \frac{1}{\mathcal{L}}\frac{\mathrm{d}}{\mathrm{d}t}(g_{ac}\dot{\chi}^c) + g_{ac}\dot{\chi}^c\frac{1}{\mathcal{L}^2}\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}t} = \frac{1}{2\mathcal{L}}\dot{\chi}^b\dot{\chi}^c\partial_a g_{bc} - \frac{1}{\mathcal{L}}\frac{\mathrm{d}}{\mathrm{d}t}(g_{ac}\dot{\chi}^c) + \frac{1}{\mathcal{L}}g_{ac}\dot{\chi}^c\frac{\mathrm{d}\ln\mathcal{L}}{\mathrm{d}t} = 0.$$

The curve may be reparametrized such that $\mathcal{L}$ is equal to 1 everywhere, yielding

$$\frac{1}{2\mathcal{L}}\left(\dot{\chi}^a\dot{\chi}^b\partial_c g_{ab} - \frac{\mathrm{d}}{\mathrm{d}t}(2\dot{\chi}^a g_{ac})\right) = 0.$$

We note that the expression in the paranthesis is the Euler-Lagrange equation for the integral of $\mathcal{L}^2$, a nice fact for the future. Expanding the derivative yields

$$\frac{1}{\mathcal{L}}\left(\frac{1}{2}\dot{\chi}^a\dot{\chi}^b\partial_c g_{ab} - g_{ac}\ddot{\chi}^a - \dot{\chi}^a\dot{\chi}^b\partial_b g_{ac}\right) = 0.$$

To remove the metric from the second derivative, we multiply by $-g^{cd}\mathcal{L}$ to obtain

$$\begin{aligned}
g_{ac}g^{cd}\ddot{\chi}^a + \frac{1}{2}\dot{\chi}^a\dot{\chi}^b g^{cd}(2\partial_b g_{ac} - \partial_c g_{ab}) &= 0,\\
g_{ac}g^{cd}\ddot{\chi}^a + \frac{1}{2}\dot{\chi}^a\dot{\chi}^b g^{cd}(\partial_b g_{ac} + \partial_a g_{bc} - \partial_c g_{ab}) &= 0,\\
\ddot{\chi}^d + \frac{1}{2}\dot{\chi}^a\dot{\chi}^b g^{cd}(\partial_b g_{ac} + \partial_a g_{bc} - \partial_c g_{ab}) &= 0.
\end{aligned}$$

This is the geodesic equation. It may alternatively be written in terms of the Christoffel symbols as

$$\ddot{\chi}^d + \Gamma^d_{\ ab}\dot{\chi}^a\dot{\chi}^b = 0.$$

Christoffel Symbols and the Geodesic Equation Consider a straight line with a tangent vector of constant magnitude. In euclidean space, this is a geodesic. This curve satisfies

$$\mathbf{0} = \frac{\mathrm{d}\dot{\boldsymbol{\gamma}}}{\mathrm{d}t} = (\dot{\boldsymbol{\gamma}}\cdot\vec{\boldsymbol{\nabla}})\dot{\boldsymbol{\gamma}} = \dot{\chi}^a\partial_a\dot{\boldsymbol{\gamma}} = \dot{\chi}^a(\vec{\boldsymbol{\nabla}}_a\dot{\chi}^d)\mathbf{E}_d = (\dot{\chi}^a\partial_a\dot{\chi}^d + \dot{\chi}^a\dot{\chi}^c\Gamma^d_{\ ac})\mathbf{E}_d.$$

Comparing this to the geodesic equation yields

$$\Gamma^d_{\ ab} = \frac{1}{2}g^{dc}(\partial_b g_{ac} + \partial_a g_{cb} - \partial_c g_{ab}).$$

A better approach would have been to go through the derivation of the geodesic equation again, identifying the Christoffel symbols as you go, but I am not sure if that is what I did in the previous paragraph. In any case we have already obtained this result.

Manifolds A manifold is a set which is locally isomorphic to $\mathbb{R}^n$. We will take this to mean that we can locally impose coordinates χ^a on the manifold.

Along with the definition of a manifold come certain other definitions. A local description of the manifold is called a chart. A collection of charts such that the combination of the charts describe the entire manifold is called an atlas. n is called the dimension of the manifold.

Manifolds and Vectors Even though manifolds are locally isomorphic to Euclidean space, the vectors that were previously developed do not make sense when applied to this Euclidean space.

Example: Tangent Vectors on S_2 Consider S_2, the unit sphere in $\mathbb{R}^3$, and suppose you cover it with a layer of water like an ocean, introduce north and south poles, place two sailors on opposite sides of the equator and tell both of them to sail south at some given speed. In practice, this means that they should both travel in their local $-y$ direction. Assuming vectors in the two spaces to make sense, you would conclude that the sailors are sailing in the same direction at the same speed and could not possibly hit each other. The accident which would occur at the south pole would of course prove you wrong. This example is one, very verbose, way of expressing why the vectors in the local Euclidean spaces do not make sense.

This argument seems to have one hole in it, namely that S_2 is implicitly embedded in $\mathbb{R}^3$. Using this fact, the collision between the sailors could be deduced using the previously developed concepts of vectors. The reason why this would work is that you could impose a position vector in $\mathbb{R}^3$ onto every point on S_2. This is not a feature of more general manifold, meaning that this hole does not exist for more general manifolds.

Tangent Vectors Tangent vectors describe how scalar fields change with displacement along a curve. In Euclidean space the tangent basis was composed of derivatives with respect to the set of coordinates. In general curved spaces, we define

$$\mathbf{E}_a = \partial_a.$$

Derivatives are linear operators, so at least the set of tangent bases span some vector space and it makes sense to call a derivative a vector. A general tangent vector is now

$$X = X^a \mathbf{E}_a = X^a \partial_a.$$

To get more of a sense of how this can be related to vectors, consider the directional derivative

$$\vec{\boldsymbol{\nabla}}_{\mathbf{n}} = \mathbf{n} \cdot \vec{\boldsymbol{\nabla}} = n^a \partial_a.$$

When applied to Euclidean space, there is a direct correspondence between $\mathbf{n}$ and the directional derivative, as $\vec{\boldsymbol{\nabla}}_{\mathbf{n}} \mathbf{x} = \mathbf{n}$. For more general manifolds, tangent vectors are defined to be directional derivatives. Note that this definition carries with it the same dependence on position as was previously warned about.

Tangent vectors transform according to

$$X^a \partial_a = X^a \partial_a (\chi')^b \partial'_b,$$

implying the transformation rule

$$(X')^a = \partial_b (\chi')^a X^b,$$

which is the same as the transformation rule for contravariant vector components in Euclidean space.

Dual Vectors To define dual vectors, we first introduce the dual space as the set of all linear operations from the tangent space to real numbers. This is also a vector space. The basis for the space is defined such that

$$\mathbf{E}^a(\partial_b) = \delta^a_b.$$

In Euclidean space the dual basis was constructed from the gradient. The only concept here that carries over to manifolds is a definition based on small changes in the coordinates. More specifically, for any smooth scalar field f we define a dual vector field according to

$$df(X) = Xf = X^a \partial_a f$$

and call it the differential (it is still not clear to me whether this has anything to do with the actual differential of a function). This has a similar structure to an inner product if the dual vector field has components $df_a = \partial_a f$. These components correspond to those of the gradient in Euclidean space. The basis we desire is $d\chi^a$.

The components transform according to

$$\partial_a f = \partial'_b f \partial_a (\chi')^b,$$

which is the transformation rule for covariant vector components.

The Geometry of Curved Space We can also impose a metric tensor such that $\mathbf{v} \cdot \mathbf{w} = g_{ab} v^a w^b$, where the metric tensor is symmetric and positive definite.

Dual vectors can be defined as linear maps from tangent vectors to scalars, i. e. on the form

$$V(\mathbf{w}) = V_a w^a.$$

In particular, the dual vector $\mathrm{d}f$ can be defined as

$$\mathrm{d}f\,(\mathbf{v}) = v^a \partial_a f = \frac{\mathrm{d}f}{\mathrm{d}t}$$

along a curve with $\mathbf{v}$ as a tangent. A basis for the space of dual vectors is $e^a = \mathrm{d}\chi^a$. The tangent and dual spaces, if a metric exists, are related by $v_a = g_{ab} v^b$.

Curve lengths are defined and computed as before. By defining geodesics as curves that extremize path length, this gives a set of Christoffel symbols and therefore a covariant derivative and a sense of what it means for a vector to change along a curve.

7 Classical Mechanics

Lagrangian Mechanics and Differential Geometry In Lagrangian mechanics, configuration space is the space of all possible configurations of a system. We can impose coordinates q^a on this space in order to use what we know from the previous parts on differential geometry. Note that the term configuration does not exclude the discussion of systems other than the purely mechanical with Lagrangian mechanics. This is a clear advantage of Lagrangian mechanics as opposed to Newtonian mechanics.

Hamilton's Principle Hamilton's principle replaces Newton's laws as the fundamental law of nature that is postulated in order to start off the theory. To formulate it, we first define the action of a system over time is defined as

$$S = \int \mathrm{d}t\, \mathcal{L},$$

where $\mathcal{L}$ is the Lagrangian. Hamilton's principle states that for the motion of the system in configuration space, $\delta S = 0$.

The Lagrangian The Lagrangian is essential in Lagrangian mechanics. For a system of particles in a conservative force field, it can be constructed as $\mathcal{L} = E_\mathrm{k} - V$. V is the potential energy and is taken to be a function only of the coordinates. Note that this is not the only way to construct a Lagrangian - for instance, adding the total time derivative of some function adds a constant term to the action, and so has no effect on the variational analysis. In addition, Lagrangian mechanics can be used to describe entirely different systems in which terms such as kinetic energy do not make sense. In fact, this is an important feature of analytical mechanics.

Lagrange's Equations of Motion Hamilton's principle can be expressed as

$$\delta S = \int \mathrm{d}t\, \delta\mathcal{L} = \int \mathrm{d}t \left(\partial_{q^a}\mathcal{L} - \frac{\mathrm{d}}{\mathrm{d}t}\partial_{\dot{q}^a}\mathcal{L}\right)\delta q^a = 0.$$

The equations of motion of the system are thus of the form

$$\partial_{q^a}\mathcal{L} - \frac{\mathrm{d}}{\mathrm{d}t}\partial_{\dot{q}^a}\mathcal{L} = 0.$$

How to Construct a Theory In Lagrangian mechanics, the recipe for formulating a theory to describe a system is simple: Introduce its Lagrangian and extremize the action to find the equations describing the system.

Example: An Inductor Consider a lone inductor with inductance L over which some (possibly time-dependent) potential V is applied. According to classical electrodynamics, we have that

$$V = L\frac{\mathrm{d}I}{\mathrm{d}t}.$$

Let us see if we can reconstruct this using Lagrangian mechanics.

We first need to impose coordinates on the system. I choose the lone coordinate q as the amount of charge carried across the inductor to ground. Suppose that the system had a Lagrangian $\mathcal{L} = \frac{1}{2}L\dot{q}^2 + qV$. The equation of motion is

$$V - \frac{\mathrm{d}}{\mathrm{d}t}(L\dot{q}) = 0,\ L\ddot{q} = V.$$

Recognizing that $\dot{q} = I$, we arrive at the desired result

$$V = L\frac{\mathrm{d}I}{\mathrm{d}t}.$$

At this point I feel it only reasonable to discuss how I (or rather, Goldstein) arrived at this particular Lagrangian. To the trained eye it is very clear that the given equation of motion would result from that

Lagrangian, but surely the core idea cannot be to just guess Lagrangians based on what we already know? Well, yes it can. The goal of physics is to describe reality, so if someone told you that their measurements implied the given equations of motion, is it not your job as a theorist to construct a theory which describes that experiment? And how do you that, if not by constructing an appropriate Lagrangian? It might seem somewhat ad hoc to construct theories based on desired results in this fashion, but the most important check for a theory to satisfy is its compliance with the reality around us. If that is ad hoc to you, then perhaps mathematics will feel more satisfactory to you than physics.

There is also an argument from the physics we already know. Having introduced the coordinate q, we know that the loss of potential energy is qV. In addition, the instantaneous power absorbed by the inductor is $P = IV = \dot{q}V$. Using the already known equation of motion, this can be written as $P = L\dot{q}\ddot{q}$. Integrating this over time gives that the inductor has energy $\frac{1}{2}L\dot{q}^2$. This does not have pure coordinate dependance, so we can use it as a kinetic term in the Lagrangian. Thus we arrive at the Lagrangian we used.

Again I feel my own skepticism, as it seems that the Lagrangian we constructed already contained some information about the system which it describes. It seems that the Lagrangian method couldn't possibly arrive at a different result, so now we are just using what we already know to rederive what we knew to begin with. And in a certain sense, this is correct. That doesn't mean that Lagrangian mechanics is useless or superfluous. The logical structure of physical theory is Babylonian, meaning that it is made to contain certain core results (namely, experimental results) and is constructed from certain starting points (namely, postulates). Beyond this it is non-directional, meaning that there is no need for certain results to build on others in a specific order. Translated and isolated to our example, there is no difference between going from the Lagrangian to the equations of motion and going from the equations of motion to the Lagrangian. Of course, physics as a whole value the Lagrangian formalism higher, as it is a more consistent way of doing a wide range of physics than simply starting from a wide variety of individual relations between quantities.

Example: The Lagrangian Mechanics of a Charged Particle Consider a particle with mass m and charge q in an external electromagnetic field. Newton's equations for this particle is

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}).$$

In terms of the potentials we have

$$\mathbf{E} = -\vec{\boldsymbol{\nabla}}\phi - \partial_t \mathbf{A},\ \mathbf{B} = \vec{\boldsymbol{\nabla}} \times \mathbf{A}$$

yielding

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q\left(-\vec{\boldsymbol{\nabla}}\phi - \partial_t \mathbf{A} + \mathbf{v} \times (\vec{\boldsymbol{\nabla}} \times \mathbf{A})\right).$$

The components of the latter term are

$$\begin{aligned} V_i &= \varepsilon_{ijk} v_j \varepsilon_{kmn} \partial_m A_n && (1)\\ &= \varepsilon_{kij}\varepsilon_{kmn} v_j \partial_m A_n && (2)\\ &= (\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm}) v_j \partial_m A_n && (3)\\ &= v_j \partial_i A_j - v_j \partial_j A_i, && (4) \end{aligned}$$

hence the magnetic term can be expanded to yield

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = q\left(-\vec{\boldsymbol{\nabla}}\phi - \partial_t \mathbf{A} + \vec{\boldsymbol{\nabla}}(\mathbf{v}\cdot\mathbf{A}) - (\mathbf{v}\cdot\vec{\boldsymbol{\nabla}})\mathbf{A}\right) = q\left(-\vec{\boldsymbol{\nabla}}\phi - \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} + \vec{\boldsymbol{\nabla}}(\mathbf{v}\cdot\mathbf{A})\right).$$

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + q\left(\vec{\boldsymbol{\nabla}}\phi + \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} - \vec{\boldsymbol{\nabla}}(\mathbf{v}\cdot\mathbf{A})\right) = \mathbf{0}.$$

To construct a Lagrangian with the same equations of motion, note that all terms but the ones containing the magnetic potential are easily created as $\mathcal{L} = \frac{1}{2}mv^2 - q\phi$. To create the remaining terms, add a term $q\mathbf{v}\cdot\mathbf{A}$ for a total Lagrangian of

$$\mathcal{L} = \frac{1}{2}mv^2 - q\phi + q\mathbf{v}\cdot\mathbf{A}.$$

The corresponding equations of motion in (Cartesian) vector notation are

$$\vec{\boldsymbol{\nabla}}_{\mathbf{r}}\mathcal{L} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\vec{\boldsymbol{\nabla}}_{\mathbf{v}}\mathcal{L}\right) = \mathbf{0}.$$

With this Lagrangian we obtain

$$\begin{aligned} -q\vec{\boldsymbol{\nabla}}\phi + q\vec{\boldsymbol{\nabla}}(\mathbf{v}\cdot\mathbf{A}) - \frac{\mathrm{d}}{\mathrm{d}t}\left(m\mathbf{v} + q\mathbf{A}\right) &= \mathbf{0}, \\ m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + q\left(\vec{\boldsymbol{\nabla}}\phi + \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} - \vec{\boldsymbol{\nabla}}(\mathbf{v}\cdot\mathbf{A})\right) &= \mathbf{0}. \end{aligned}$$

Great - we did it!

Kinetic Energy Kinetic energy is defined in terms of the rank 2 kinetic metric tensor as

$$E_{\mathrm{k}} = \frac{1}{2}T_{ab}\dot{q}^a\dot{q}^b.$$

The Kinetic Metric Consider a system with no potential energy. The Lagrangian simply becomes $\mathcal{L} = \frac{1}{2}T_{ab}\dot{\chi}^a\dot{\chi}^b$. The action computed from this Lagrangian is very similar to the integral of curve length (or, rather its square, the extremum of which was noted to be the same), except g_{ab} has been replaced by T_{ab}. This inspires us to name T_{ab} the kinetic metric. From this we may repeat the process performed in the section on differential geometry, for instance to obtain the geodesic equations as the equations of motion or to introduce Christoffel symbols.

Motion of a Classical System By defining $a^b = \dot{\chi}^a\vec{\boldsymbol{\nabla}}_a\dot{\chi}^b$, the previous work leads us to a system with no potential satisfying $a^b = \ddot{\chi}^b + \Gamma^b{}_{ac}\dot{\chi}^a\dot{\chi}^c = 0$. In other words, a system with no potential moves along the geodesics of the kinetic metric.

For a system with a potential, only the $\partial_{\chi^a}\mathcal{L}$ term is affected, and

$$a^b = -T^{ba}\partial_a V = T^{ba}F,$$

which is a generalization of Newton's second law.

Legendre Transforms To illustrate the Legendre transform, consider a function $f(x,y)$ and $g(x,y,u) = ux - f(x,y)$. Its total derivative is given by

$$\mathrm{d}g = u\,\mathrm{d}x + x\,\mathrm{d}u - \partial_x f\,\mathrm{d}x - \partial_y f\,\mathrm{d}y\,.$$

By choosing $u = \partial_x f$, we obtain

$$\mathrm{d}g = x\,\mathrm{d}u - \partial_y f\,\mathrm{d}y\,,$$

implying that g is only a function of u and y. To obtain g, invert the definition of u to obtain $x(u,y)$.

The Legendre transform is thus, as shown, a way to transform a function from one set of variables to another. If the function to be transformed describes the behaviour of some system, the Legendre transform allows us to describe the behaviour of the system in terms of different variables, which might be practical.

Hamiltonian Mechanics The Lagrangian equations of motion are n coupled second-order differential equations. Hamiltonian mechanics starts with trying to decouple these into $2n$ first-order differential equations. To illustrate how this is done, consider an equation of motion

$$\partial_{q^a}\mathcal{L} - \frac{\mathrm{d}}{\mathrm{d}t}\partial_{\dot{q}^a}\mathcal{L} = 0.$$

The second-order term is the place to hack away at. We thus define the generalized momenta

$$p_a = \partial_{\dot{q}^a}\mathcal{L}.$$

Up until now, we have mathematically treated the coordinates and velocities as variables, making no difference between them in a technical manner. The next step is now to replace the velocities with momenta. As the Lagrangian describes the system, we do this by Legendre transforming the Lagrangian.

We now define the Hamiltonian

$$\mathcal{H} = p_i \dot{q}^i - \mathcal{L}.$$

From this definition we would like to obtain new equations of motion. This is done by computing the differential of the Hamiltonian. We have

$$\mathrm{d}\mathcal{L} = \partial_{q^i}\mathcal{L}\,\mathrm{d}q^i + \partial_{\dot{q}^i}\mathcal{L}\,\mathrm{d}\dot{q}^i + \partial_t\mathcal{L}\,\mathrm{d}t\,.$$

The definition of the general momenta and the equations of motion allows us to write this as

$$\mathrm{d}\mathcal{L} = \dot{p}_i\,\mathrm{d}q^i + p_i\,\mathrm{d}\dot{q}^i + \partial_t\mathcal{L}\,\mathrm{d}t\,.$$

The differential of the Hamiltonian is

$$\mathrm{d}\mathcal{H} = p_i\,\mathrm{d}\dot{q}^i + \dot{q}^i\,\mathrm{d}p_i - \dot{p}_i\,\mathrm{d}q^i - p_i\,\mathrm{d}\dot{q}^i - \partial_t\mathcal{L}\,\mathrm{d}t = \dot{q}^i\,\mathrm{d}p_i - \dot{p}_i\,\mathrm{d}q^i - \partial_t\mathcal{L}\,\mathrm{d}t\,,$$

serving as an example of how the Legendre transform works. This implies that the equations of motion are

$$\dot{p}_i = -\partial_{q_i}\mathcal{H},\ \dot{q}_i = \partial_{p_i}\mathcal{H}.$$

Alternatively, we could obtain this from Hamilton's principle. The action is given by

$$S = \int \mathrm{d}t\, p_i\dot{q}^i - \mathcal{H},$$

the extremization of which is a variational problem in the coordinates, momenta and their time derivatives. The equations of motion yield

$$-\partial_{q_i}\mathcal{H} - \frac{\mathrm{d}}{\mathrm{d}t}(p_i) = 0,\ \dot{q}^i - \partial_{p_i}\mathcal{H} = 0,$$

which are the same equations of motion.

In Lagrangian mechanics, we considered paths in configuration space. In Hamiltonian mechanics, we instead consider paths in phase space, i.e. a space where the points are (q, p). In this space, paths do not intersect as the system is deterministic. Paths in phase space are periodic for integrable systems and fill out the accessible parts of phase space for chaotic systems.

We note that

$$\begin{aligned}\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} &= \partial_{q_i}\mathcal{H}\dot{q}_i + \partial_{p_i}\mathcal{H}\dot{p}_i + delt\mathcal{H} \\ &= -\dot{p}_i\dot{q}_i + \dot{q}_i\mathcal{H}\dot{p}_i + delt\mathcal{H} \\ &= \partial_t\mathcal{H},\end{aligned}$$

and so the Hamiltonian is conserved if it has no explicit time dependence.

Example: The Hamiltonian Mechanics of a Charged Particle Starting with the Lagrangian

$$\mathcal{L} = \frac{1}{2}mv^2 - q\phi + q\mathbf{v}\cdot\mathbf{A}.$$

we have

$$\mathbf{p} = \vec{\boldsymbol{\nabla}}_{\mathbf{v}}\mathcal{L} = m\mathbf{v} + q\mathbf{A},$$

and the Hamiltonian is thus

$$\begin{aligned}\mathcal{H} &= \mathbf{p}\cdot\mathbf{v} - \mathcal{L} \\ &= \mathbf{p}\cdot\frac{1}{m}(\mathbf{p} - q\mathbf{A}) - \frac{1}{2m}(\mathbf{p} - q\mathbf{A})^2 + q\phi - \frac{q}{m}(\mathbf{p} - q\mathbf{A})\cdot\mathbf{A} \\ &= \frac{1}{m}\left(\mathbf{p}\cdot(\mathbf{p} - q\mathbf{A}) - \frac{1}{2}(\mathbf{p} - q\mathbf{A})^2 - q\mathbf{A}\cdot(\mathbf{p} - q\mathbf{A})\right) + q\phi \\ &= \frac{1}{2m}\left(\mathbf{p} - q\mathbf{A}\right)^2 + q\phi.\end{aligned}$$

Poisson Brackets Consider a function $f(\mathbf{q}, \mathbf{p}, t)$. Its time derivative is given by

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}t} &= \partial_{q_i} f \dot{q}_i + \partial_{p_i} f \dot{p}_i + \partial_t f \\ &= \partial_{q_i} f \partial_{p_i} \mathcal{H} - \partial_{p_i} f \partial_{q_i} \mathcal{H} + \partial_t f \\ &= \{f, \mathcal{H}\} + \partial_t f,\end{aligned}$$

where we now have defined the Poisson bracket. It is bilinear and satisfies

$$\begin{aligned}\{f, g\} &= -\{g, f\}, \\ \{fg, h\} &= f\{g, h\} + \{f, h\} g, \\ \{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} &= 0.\end{aligned}$$

The expression above implies that if $\{f, \mathcal{H}\} = 0$ and f does not depend explicitly on time, then it is a constant of motion.

Canonical Transformations A canonical transformation is a transformation $(\mathbf{q}, \mathbf{p}) \to (\mathbf{Q}, \mathbf{P})$ such that the Hamiltonian H, which is the old Hamiltonian expressed in terms of the new coordinates, also satisfies Hamilton's equations, i.e.

$$\dot{Q}_i = \partial_{P_i} H, \ \dot{P}_i = -\partial_{Q_i} H.$$

Symplectic Description of Canonical Transforms It is not clear that all transformations are canonical, so we would like to find a way to check whether a given transformation is canonical. We will do this by introducing a different way of handling mechanics, termed the symplectic approach. This approach considers the canonical transformation as a map from a point $\mathbf{x}$ to a point $\mathbf{y}$ between old and new phase spaces, where we have

$$x_i = \begin{cases} q^i, & i < n, \\ p_{i-n}, & i > n \end{cases}$$

and similarly for $\mathbf{y}$. The equations of motion can thus be written

$$\dot{\mathbf{x}} = J \vec{\boldsymbol{\nabla}}_{\mathbf{x}} H, \ J_{ij} = \delta_{i+n,j} - \delta_{i,j+n}.$$

To begin our investigation, we consider a canonical transformation which is time-independent. Under such a transformation, the Hamiltonian is unchanged. In order to find a criterion, we will study the equations of motion under a canonical transformation. We have

$$\dot{y}^i = M_{ij} \dot{x}^j, \ M_{ij} = \partial_{x^j} y^i,$$

and we thus recognize M as the Jacobian of the coordinate change. Inserting the equations of motion yields

$$\dot{\mathbf{y}} = M J \vec{\boldsymbol{\nabla}}_{\mathbf{x}} \mathcal{H}.$$

On the other hand, according to the chain rule, we have

$$\partial_{y^i} \mathcal{H} = \partial_{x^j} H \partial_{y^i} x^j,$$

in matrix form

$$\vec{\boldsymbol{\nabla}}_{\mathbf{y}} \mathcal{H} = (M^T)^{-1} \vec{\boldsymbol{\nabla}}_{\mathbf{x}} \mathcal{H}.$$

In order for Hamilton's equations to be satisfied, a sufficient condition is thus

$$MJ = J(M^T)^{-1}, \ J = MJM^T.$$

To show that it is necessary, you could reverse this proof. It can also be shown that this is true for time-dependent transforms. Someone should probably do that.

Using this formalism, we can also rewrite Poisson brackets. We have

$$\{u, v\}_{\mathbf{x}} = \partial_{x^i} u J_{ij} \partial_{x^j} v = (\vec{\boldsymbol{\nabla}}_{\mathbf{x}} u)^T J \vec{\boldsymbol{\nabla}}_{\mathbf{x}} v.$$

Canonical Transformations and Poisson Brackets The requirement for a canonical transformation can be restated in terms of Poisson brackets. A transformation is canonical if and only if it preserves

$$\{q_i, q_j\} = \{p_i, p_j\} = 0,\ \{q_i, p_j\} = \delta_{ij}.$$

To show this, we apply the symplectic approach, which yields

$$\left\{x^i, x^j\right\}_{\mathbf{x}} = \delta_{ik} J_{km} \delta_{mj} = J_{ij},$$

the exact same relations that should be conserved by a canonical transformation.

Under the assumption that the equations of motion are satisfied in both the old and new phase space, the above relation holds under the change $\mathbf{x} \to \mathbf{y}$. The question remains, however, whether the new phase space variables satisfy this relation in terms of the old ones. The truth of this statement would imply that the transformation is canonical, so we want to show this. We have

$$\left\{y^i, y^j\right\}_{\mathbf{x}} = \partial_{x^k} y^i J_{km} \partial_{x^m} y^j = M_{ik} J_{km} M_{jm} = (MJM^T)_{ij}.$$

Under the assumption that the transformation is canonical, we have

$$\left\{y^i, y^j\right\}_{\mathbf{x}} = J_{ij}.$$

Thus the stated criterion is sufficient. Assuming this yields that it is necessary, in perfect analogy with the previous proof.

Generators of Canonical Transforms As both the old and new phase space variables must obey Hamilton's equations, they both satisfy the principle of least action. The action is given by

$$S = \int \mathrm{d}t\, p_i \dot{q}_i - \mathcal{H}$$

in the old variables and

$$\int \mathrm{d}t\, P_i \dot{Q}_i - H,$$

in the new variables. One way for the two to be extremal at the same time is if

$$\lambda(p_i \dot{q}_i - \mathcal{H}) = P_i \dot{Q}_i - H + \frac{\mathrm{d}F}{\mathrm{d}t}.$$

We note that λ is essentially a matter of units - for instance, if you scale the Hamiltonian and all coordinates by a factor λ, the scaled coordinates satisfy Hamilton's equations. Thus the general case can be obtained by combining any given transformation with a scaling, and we only need to study the case $\lambda = 1$.

The time derivative of F is

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \partial_{q_i} F \dot{q}_i + \partial_{p_i} F \dot{p}_i + \partial_{Q_i} F \dot{Q}_i + \partial_{P_i} F \dot{P}_i + \partial_t F.$$

Inserting this into the requirement above gives

$$p_i \dot{q}_i - \mathcal{H} = P_i \dot{Q}_i - H + \partial_{q_i} F \dot{q}_i + \partial_{p_i} F \dot{p}_i + \partial_{Q_i} F \dot{Q}_i + \partial_{P_i} F \dot{P}_i + \partial_t F.$$

Comparing similar terms yields

$$p_i \dot{q}_i = \partial_{q_i} F \dot{q}_i,\ 0 = \partial_{p_i} F \dot{p}_i,\ 0 = P_i \dot{Q}_i + \partial_{Q_i} F \dot{Q}_i,\ 0 = \partial_{P_i} F \dot{P}_i,\ -\mathcal{H} = -H + \partial_t F,$$

and finally

$$\partial_{q_i} F = p_i,\ \partial_{p_i} F = 0,\ \partial_{Q_i} F = -P_i,\ \partial_{P_i} F = 0,\ \mathcal{H} = H - \partial_t F.$$

The solution to this is $F = F_1(\mathbf{q}, \mathbf{Q}, t)$, and is called a generator of the canonical transform. We dub it a generator because, given the expression for the old and new coordinates and a given F_1, the momenta can be obtained according to

$$\partial_{q_i} F_1 = p_i,\ \partial_{Q_i} F_1 = -P_i.$$

In other words, the generator contains information about the nature of the canonical transformation. Naturally, given the complete nature of the transformation, one can solve the equation above to obtain its generator.

The form of the generatore is not unique. Another choice is a function $F = F_2 - P_i Q_i$. Inserting this into the above criterion yields

$$\partial_{q_i} F_2 = p_i,\ \partial_{p_i} F_2 = 0,\ \partial_{Q_i} F_2 = 0,\ \partial_{P_i} F = Q_i,\ \mathcal{H} = H - \partial_t F_2.$$

The solution to this is $F_2 = F_2(\mathbf{q}, \mathbf{P}, t)$.

A third choice is a function $F = F_3 + p_i q_i$. Inserting this into the above criterion yields

$$\partial_{q_i} F_3 = 0,\ \partial_{p_i} F = -q_i,\ \partial_{Q_i} F = -P_i,\ \partial_{P_i} F = 0,\ \mathcal{H} = H - \partial_t F_3.$$

The solution to this is $F_3 = F_3(\mathbf{p}, \mathbf{Q}, t)$.

A fourth choice is a function $F = F_4 + p_i q_i - P_i Q_i$. Inserting this into the above criterion yields

$$\partial_{q_i} F_4 = 0,\ \partial_{p_i} F_4 = -q_i,\ \partial_{Q_i} F_4 = 0,\ \partial_{P_i} F_4 = Q_i,\ \mathcal{H} = H - \partial_t F_4.$$

The solution to this is $F_4 = F_4(\mathbf{p}, \mathbf{P}, t)$.

Going from one type of generator to another looks very similar to a Legendre transform, and computationally is a (somewhat) clear demonstration of what the Legendre transform does. However, the one thing separating it from a Legendre transform is the fact that performing this transformation to obtain a generator is not always possible. In addition, it might not be possible to find a generator of a certain kind, and you might have to use a mixed type of generator.

Example: A Failed Choice of Generator Suppose that we want to perform a canonical transform that preserves the first coordinate. Looking for a generator of the first kind, we find that it must satisfy

$$\partial_{q_1} F = p_1,\ \partial_{Q_1} F = -P_1.$$

However, as the two coordinates are equal, the partial derivatives represent equivalent operations. The only way to resolve this is for the transformation to satisfy $p_1 = -P_1$ - otherwise you cannot find a generator of the first kind.

Symmetries and Quasi-Symmetries Consider a one-parameter transformation $t \to \tau(t, s),\ q^a \to Q^a(q, s)$, where s is the parameter with respect to which the system is transformed, such that $\tau(t, 0) = t,\ Q^a(q, 0) = q^a$. A quasi-symmetry of a system with Lagrangian $\mathcal{L}$ is a transformation such that

$$\varepsilon\, \delta\mathcal{L} = \mathcal{L}(Q, \dot{Q}, \tau) - \mathcal{L}(q, \dot{q}, t) = \varepsilon \frac{\mathrm{d}F}{\mathrm{d}t}\, \mathrm{d}s$$

for some F. The case $F = 0$ is referred to as symmetries of the system.

Noether's Theorem Noether's theorem relates symmetries - or, more specifically, quasi-symmetries, of physical systems to conservation laws.

The variation of the Lagrangian under variation of coordinates and time can be written as

$$\delta\mathcal{L} = \partial_{q^a}\mathcal{L}\, \delta q^a + \partial_{\dot{q}^a}\mathcal{L}\, \delta\dot{q}^a + \partial_t \mathcal{L}\, \delta t\,.$$

The total time derivative of the Lagrangian is given by

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}t} = \partial_t \mathcal{L} + \partial_{q^a}\mathcal{L}\dot{q}^a + \partial_{\dot{q}^a}\mathcal{L}\ddot{q}^a,$$

which yields

$$\delta\mathcal{L} = \partial_{q^a}\mathcal{L}(\delta q^a - \dot{q}^a\, \delta t) + \partial_{\dot{q}^a}\mathcal{L}(\delta\dot{q}^a - \ddot{q}^a\, \delta t) + \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}t}\, \delta t\,.$$

The equations of motion are $\partial_{q^a}\mathcal{L} = \frac{\mathrm{d}}{\mathrm{d}t}\partial_{\dot{q}^a}\mathcal{L}$. For a set of coordinates that satisfy this - a so-called on-shell solution - we have

$$\begin{aligned}\delta\mathcal{L} &= \frac{\mathrm{d}}{\mathrm{d}t}\partial_{\dot{q}^a}\mathcal{L}(\delta q^a - \dot{q}^a\, \delta t) + \partial_{\dot{q}^a}\mathcal{L}(\delta\dot{q}^a - \ddot{q}^a\, \delta t) + \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}t}\, \delta t \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\partial_{\dot{q}^a}\mathcal{L}(\delta q^a - \dot{q}^a\, \delta t) + \mathcal{L}\, \delta t\right).\end{aligned}$$

If the variation corresponds to a quasi-symmetry of the system, then this is equal to a total time derivative of F, and the quantity

$$J = F - \partial_{\dot{q}^a}\mathcal{L}\,\delta q^a + (\dot{q}^a\partial_{\dot{q}^a}\mathcal{L} - \mathcal{L})\,\delta t$$

thus satisfies $\frac{\mathrm{d}J}{\mathrm{d}t} = 0$. We can introduce the general momenta and the Hamiltonian to rewrite this as

$$J = F - p_a\,\delta q^a + \mathcal{H}\,\delta t\,.$$

We arrive at the conclusion that J is a conserved quantity under a quasi-symmetry of the system. Identifying the conservation laws of a system is thus a matter of identifying the quasi-symmetries of a system and computing J under that transformation.

Example: A Free Particle in Space Consider a free particle in space. Its Lagrangian is given by $\mathcal{L} = \frac{1}{2}m\dot{\mathbf{x}}^2$, and the variation of this is

$$\delta\mathcal{L} = m\dot{\mathbf{x}}\cdot\delta\dot{\mathbf{x}}\,.$$

Its general momentum is

$$\mathbf{p} = \partial_{\dot{\mathbf{x}}}\mathcal{L} = m\dot{\mathbf{x}}.$$

The Hamiltonian is

$$\mathcal{H} = \mathbf{p}\cdot\dot{\mathbf{x}} - \mathcal{L} = \frac{1}{2}m\dot{\mathbf{x}}^2.$$

We now want to identify quasi-symmetries of the system that make the variation of the Lagrangian either zero or the time derivative of some quantity. A key idea here is that we are only allowed to change the variations (or so I think).

A first attempt is keeping $\delta\mathbf{x}$ constant and not varying time(a spatial translation), which implies $\delta\dot{\mathbf{x}} = \mathbf{0}$ and $\delta\mathcal{L} = 0$. This implies that F is constant. The conserved quantity is thus

$$J = F - \mathbf{p}\cdot\delta\mathbf{x} = F - \mathbf{p}\cdot\mathbf{c},$$

i.e. the momentum of the system is conserved. We also note that the constant F in this case is arbitrary, and we might as well have set it to 0. This will be the case at least sometimes.

A second attempt is varying time, i.e. $\delta t = 1$, but keeping the coordinates fixed, i.e. $\delta\mathbf{x} = 0$ (a time translation). This yields $\delta\dot{\mathbf{x}} = \mathbf{0}$ and $\delta\mathcal{L} = 0$. Once again F is constant and taken to be zero, and the conserved quantity is thus $J = H$, i.e. the Hamiltonian of the system is conserved.

A third attempt is to somehow make the scalar product in the variation of the Lagrangian zero, without varying time. An option is $\delta\mathbf{x} = \omega\times\mathbf{x}$, where ω is a constant vector. This yields $\delta\dot{\mathbf{x}} = \omega\times\dot{\mathbf{x}}$ and $\delta\mathcal{L} = 0$. The conserved quantity is thus

$$\begin{aligned} J &= -\mathbf{p}\cdot(\omega\times\mathbf{x}) \\ &= -\omega\cdot(\mathbf{x}\times\mathbf{p}). \end{aligned}$$

Since ω is constant, that means that $\mathbf{x}\times\mathbf{p}$, i.e. the angular momentum, is conserved.

Symmetry Groups in Classical Mechanics

Example: Newton's Second Law Newton's second law $m\ddot{\mathbf{x}} = -\vec{\boldsymbol{\nabla}}V$, assuming the potential to be fixed, has certain symmetry properties:

- The transformation $t \to t' = t + t_0$ is a symmetry, as $\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t'}$ and V is not changed under the transformation.
- The transformation $t \to \tau = -t$ is a symmetry as $\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\mathrm{d}\tau}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}\tau} = -\frac{\mathrm{d}}{\mathrm{d}\tau}$, which implies $\frac{\mathrm{d}^2}{\mathrm{d}t^2} = \frac{\mathrm{d}^2}{\mathrm{d}\tau^2}$ and V is not changed under the transformation.
- Considering a system of particles, if the forces between these only depend on differences between the position vectors, the translation $\mathbf{x}_i \to \mathbf{y}_i = \mathbf{x}_i + \mathbf{x}_0$ is a symmetry as it does not change any differences.

Example: Constraining Solutions Using Symmetries If a system is invariant under some transformation $\mathbf{x} \to \mathbf{R}(\mathbf{x})$, then any property u dependant on those coordinates satisfies $u(\mathbf{x}) = u(\mathbf{R}(\mathbf{x}))$.

Connection to Noether's Theorem We defined symmetries of the action as transformations that satisy $\delta\mathcal{L} = 0$. In particular, we can construct a set of transformations such that $\partial_s t = \delta t\,,\ \partial_s q^a = \delta q^a$, where s is the symmetry parameter. This is a one-parameter family of symmetries. By defining $T_s q(t,0) = q(t,s)$, these symmetries satisfy

$$T_{s_2} T_{s_1} q(t,0) = T_{s_1+s_2} q(t,0).$$

We see that these symmetries define a group.

Example: A Particle in a Moving Potential

Symmetry Breaking Suppose that a system is approximately symmetric under a a group G of transformations. When studying this system, a first approach is thus to treat it as symmetric under G and use the mathematics of G to compute corrections to the symmetric approximation. This is known as symmetry breaking.

Generators and Quasi-Symmetries Consider an infinitesimal canonical transformation of the form

$$q_i \to q_i + \alpha F_i,\ p_i \to p_i + \alpha E_i.$$

Computing the Poisson brackets of the new coordinates and momenta yields

$$\begin{aligned}
\{Q_i, P_j\} &= \{q_i, p_j\} + \{q_i, \alpha E_j\} + \{\alpha F_i, p_j\} + \{\alpha F_i, \alpha E_j\} \\
&= \delta_{ij} + \alpha(\{q_i, E_j\} + \{F_i, p_j\}) + \alpha^2 \{F_i, E_j\} \\
&= \delta_{ij} + \alpha(\delta_{ik}\partial_{p_k} E_j + \delta_{jk}\partial_{q_k} F_i) + \alpha^2 \{F_i, E_j\} \\
&= \delta_{ij} + \alpha(\partial_{p_i} E_j + \partial_{q_j} F_i) + \dots
\end{aligned}$$

and the requirement

$$\partial_{p_i} E_j = -\partial_{q_j} F_i.$$

A simple choice of solution is

$$E_j = -\partial_{q_j} G,\ F_i = \partial_{p_i} G$$

for some (smooth) function G, which may be said to generate the transformation.

Suppose now that we perform an infinitesimal transformation generated by G. We then obtain

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}\alpha} = \partial_{q^i}\mathcal{H}\frac{\mathrm{d}q^i}{\mathrm{d}\alpha} + \partial_{p_i}\mathcal{H}\frac{\mathrm{d}p_i}{\mathrm{d}\alpha} = \partial_{q^i}\mathcal{H}\partial_{p_i} G - \partial_{p_i}\mathcal{H}\partial_{q_i} G = \{\mathcal{H}, G\}\,.$$

Supposing G to be a quasi-symmetry, this implies $\dot{G} = \{\mathcal{H}, G\} = 0$, and G is conserved. In other words, if G generates a symmetry, then it is conserved.

Example: The Central Force Problem

Liouville's Theorem As paths in phase space do not cross, we can think of the time evolution of a system as a flow in phase space. The volume element is $\mathrm{d}V = \prod_i \mathrm{d}q_i\,\mathrm{d}p_i$. Liouville's theorem states that flow in phase space is incompressible.

To show this, consider the state at some point in time and after some infinitesimal time $\mathrm{d}t$. Denote the point in phase space at the start as $(\mathbf{q}, \mathbf{p})$ and after $\mathrm{d}t$ as $(\mathbf{q}', \mathbf{p}')$. To first order in time we have

$$q_i' = q_i + \dot{q}_i\,\mathrm{d}t = q_i + \partial_{p_i}\mathcal{H}\,\mathrm{d}t\,,\ p_i' = p_i + \dot{p}_i\,\mathrm{d}t = p_i - \partial_{q_i}\mathcal{H}\,\mathrm{d}t\,.$$

Considering a specific coordinate-momentum pair, we have

$$\begin{aligned}\mathrm{d}q_i'\,\mathrm{d}p_i' &= \left(\mathrm{d}q_i + \left(\partial_{q_i}\partial_{p_i}\mathcal{H}\,\mathrm{d}q_i + \partial_{p_i}^2\mathcal{H}\,\mathrm{d}p_i\right)\mathrm{d}t\right)\left(\mathrm{d}p_i - \left(\partial_{q_i}^2\mathcal{H}\,\mathrm{d}q_i + \partial_{p_i}\partial_{q_i}\mathcal{H}\,\mathrm{d}p_i\right)\mathrm{d}t\right)\\ &= \mathrm{d}q_i\,\mathrm{d}p_i + \left(-\,\mathrm{d}q_i\left(\partial_{q_i}^2\mathcal{H}\,\mathrm{d}q_i + \partial_{p_i}\partial_{q_i}\mathcal{H}\,\mathrm{d}p_i\right) + \mathrm{d}p_i\left(\partial_{q_i}\partial_{p_i}\mathcal{H}\,\mathrm{d}q_i + \partial_{p_i}^2\mathcal{H}\,\mathrm{d}p_i\right)\right)\mathrm{d}t\\ &= \mathrm{d}q_i\,\mathrm{d}p_i + \left(-\partial_{q_i}^2\mathcal{H}(\mathrm{d}q_i)^2 + (\partial_{q_i}\partial_{p_i}\mathcal{H} - \partial_{p_i}\partial_{q_i}\mathcal{H})\,\mathrm{d}q_i\,\mathrm{d}p_i + \partial_{p_i}^2\mathcal{H}(\mathrm{d}p_i)^2\right)\mathrm{d}t\,.\end{aligned}$$

The equations of motion imply that the terms containing two consecutive derivatives with respect to the same variable are equal to zero. Assuming the Hamiltonian to be sufficiently smooth, the cross-derivatives are equal. This means that this specific product is preserved. Multiplying them up for all coordinates and momenta yields

$$\mathrm{d}V' = \mathrm{d}V\,.$$

Liouville's Theorem in Statistical Mechanics We define $\rho(q,p,t)$ as the probability that a particle is close to (q,p). Alternatively, for a large number of particles, we can define it as the number of particles close to (q,p).

We have

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = 0,$$

implying

$$\partial_t\rho = -\{\rho,\mathcal{H}\}\,.$$

This is an equivalent statement of Liouville's theorem.

Hamilton-Jacobi theory Suppose that we could perform a canonical transformation on a system generated by a generator of the second kind such that $H = 0$. As the transformed Hamiltonian is constant, this implies that all Q_i and P_i are constant, and so the equations of motion are trivial. The equation

$$\mathcal{H} + \partial_t F_2 = 0$$

thus defines a differential equation for the generator of the transformation we seek. Identifying the generator as the action with no readily apparent argument yields the Hamilton-Jacobi equation

$$\mathcal{H} + \partial_t S = 0.$$

How does this define a differential equation for the action? The Hamiltonian can be written in terms of the generalized coordinates and momenta, and the momenta are, according to the argument above, derivatives of the generator (i.e. the action) with respect to the coordinates. Thus the different terms include some combination of the coordinates and the derivatives of the action, as well as the final time derivative.

Is the generator actually equal to the action? The total time derivative of the action is

$$\frac{\mathrm{d}S}{\mathrm{d}t} = p_i\dot{q}^i - \mathcal{H}.$$

On the other hand, considering the supposed action we have been studying, we have

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \partial_{q^i}S\dot{q}^i + \partial_{P^i}S\dot{P}^i + \partial_t S.$$

As the transformed system is trivial, the second (collection of) term(s) vanishes, and combining this with the equations defining the generator we are left with

$$\frac{\mathrm{d}S}{\mathrm{d}t} = p_i\dot{q}^i + \partial_t S.$$

Finally, using the Hamilton-Jacobi equation, we obtain

$$\frac{\mathrm{d}S}{\mathrm{d}t} = p_i\dot{q}^i - \mathcal{H}.$$

By solving the Hamilton-Jacobi equation, we see that we obtain the action. According to the theory of generators, it is a function of the old coordinates, the new momenta and time. As the new momenta are constant, these appear as integration constants in the action.

Example: The Harmonic Oscillator We first try studying the harmonic oscillator. By doing this we will demonstrate the general procedure of Hamilton-Jacobi theory.

We start with the familiar Hamiltonian

$$\mathcal{H} = \frac{1}{2m}p^2 + \frac{1}{2}m\omega^2 x^2.$$

We know that $p = \partial_x S$, and so the Hamilton-Jacobi equation becomes

$$\frac{1}{2m}\left(\partial_x S\right)^2 + \frac{1}{2}m\omega^2 x^2 + \partial_t S = 0.$$

Looking at this PDE, we see that it is non-homogenous and non-linear. Those are two words you typically do not want to hear in such a context. We have no boundary conditions, so it cannot be reduced to a Sturm-Liouville problem. In desperation, we turn to variable separation. The fact that different terms contain different powers of the action and its derivative is a hindrance to using a product anzats, so instead we will try an anzats of the form

$$S(x,t) = X(x) + T(t).$$

Inserting this into the Hamilton-Jacobi equation yields

$$\frac{1}{2m}\left(\frac{\mathrm{d}X}{\mathrm{d}x}\right)^2 + \frac{1}{2}m\omega^2 x^2 + \frac{\mathrm{d}T}{\mathrm{d}t} = 0.$$

Subtracting the time term, we see that each side must be equal to a constant which I dub α.

Solving the time part first, we have

$$T = -\alpha(t - t_0) + T(t_0).$$

Returning to the space part, we are left with

$$\frac{1}{2m}\left(\frac{\mathrm{d}X}{\mathrm{d}x}\right)^2 + \frac{1}{2}m\omega^2 x^2 = \alpha.$$

Solving this for the derivative yields

$$\frac{\mathrm{d}X}{\mathrm{d}x} = \sqrt{2m\alpha - m^2\omega^2 x^2}.$$

X is now given by integrating. The action is thus

$$S = \int_{x_0}^{x} \mathrm{d}y\, \sqrt{2m\alpha - m^2\omega^2 y^2} - \alpha(t - t_0) + T(t_0) + X(x_0).$$

Constant terms in the action are of course of no physical significance, and so we are free to set these to zero. We thus write

$$S = \int_{x_0}^{x} \mathrm{d}y\, \sqrt{2m\alpha - m^2\omega^2 y^2} - \alpha t.$$

We note that the action only depends on one integration constant α. This is therefore the canonical momentum P.

The next step is to compute the generalized (old) coordinates and momenta. To do this, we use the properties of the generator. We have

$$Q = \beta = \partial_P S = \partial_\alpha S = \int_{x_0}^{x} \mathrm{d}y\, \frac{m}{\sqrt{2m\alpha - m^2\omega^2 y^2}} - t,$$

$$p = \partial_x S = \sqrt{2m\alpha - m^2\omega^2 x^2},$$

where we have redubbed Q to the constant of motion β. In theory we now have a sufficient amount of equations to solve the system, as the relation for Q can be inverted to give q over time, and this will in turn give p. In this particular case, this can be done analytically, as

$$\begin{aligned}
\beta &= \int_{x_0}^{x} \mathrm{d}y \, \frac{m}{\sqrt{2m\alpha - m^2\omega^2 y^2}} - t \\
&= \sqrt{\frac{m}{2\alpha}} \int_{x_0}^{x} \mathrm{d}y \, \frac{1}{\sqrt{1 - \frac{m\omega^2 y^2}{2\alpha}}} - t \\
&= \frac{1}{\omega} \int_{s_0}^{s(x)} \mathrm{d}s \, \frac{1}{\sqrt{1 - s^2}} - t \\
&= \frac{1}{\omega} (\arcsin s(x) - \arcsin s_0) - t \\
&= \frac{1}{\omega} \left(\arcsin\left(\sqrt{\frac{m}{2\alpha}} \omega x \right) - \arcsin\left(\sqrt{\frac{m}{2\alpha}} \omega x_0 \right) \right) - t.
\end{aligned}$$

This yields

$$x = \sqrt{\frac{2\alpha}{m\omega^2}} \sin(\phi + \omega t),$$

where we have defined

$$\phi = \omega\beta + \arcsin\left(\sqrt{\frac{m}{2\alpha}} \omega x_0 \right).$$

We then obtain

$$p = \sqrt{2m\alpha - 2m\alpha \sin^2(\phi + \omega t)} = \sqrt{2m\alpha} \cos(\phi + \omega t),$$

which is the familiar solution of the harmonic oscillator.

Are the constants of motion also familiar? The answer is yes. We see that β corresponds to the phase in the position and momentum, which are known to be constant. We also see that the Hamiltonian is equal to α, and we also know that the Hamiltonian is constant. So that is nice.

Example: Hamilton-Jacobi theory in spherical coordinates

The Schrödinger equation from Hamilton-Jacobi theory

Quantum mechanics and the action

Integrable systems Consider a system with some Hamiltonian. This system is integrable if there is a canonical transformation $(q, p) \to (\theta, I)$ such that the transformed Hamiltonian only depends on the momenta. For such a system, the equations of motion become

$$\dot{\theta}_i = \partial_{I_i} H = \omega_i, \;\; \dot{I}_i = 0.$$

Structures of theory A theory in physics contain

- some notion of states.
- observables.
- a description of the dynamics of the system.
- predictions of experiments.

Example: Hamiltonian mechanics

8 Relativity

The Galilean group The Galilean group is the group of transformations between frames of reference under which the laws of physics are invariant. It consists of:

- Translations by a constant vector.
- Rotations of the coordinate axes.
- Boosts, i.e. translating the coordinates along a line with a constants speed.

It is based on a concept of absolute time. It turns out that the arc element $\mathrm{d}^2 s = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$ at a given time is preserved under all of these transformations.

The invariance of the laws of physics under these transformations corresponds to there being no special position or direction in the universe, and no special velocity. At least two of these claims have thus far not been disproved.

The emergence of special relativity It turned out that Maxwell's equations were not invariant under Galilean transformations.

DISCLAIMER: THIS IS SERIOUS HEAD CANON! The issue with Maxwell's equations is that they predict that electromagnetic waves travel at speed c. This should of course be the same in all frames of reference, according to Galilean relativity. No other wave phenomena had previously raised an issue as they travel through a medium. This medium naturally defines a certain frame of reference in which the physics are special, namely the rest frame of the medium. Only by transporting the medium with you when doing the boost will you reobtain the same physics. A natural idea to follow from this is that electromagnetic waves travel in a medium, so physicists started searching for it. After having found no evidence of its existence, most notably through Michelson and Morley's experiment, the conclusion was that there was no medium in which electromagnetic waves travelled, and thus the speed of light had to be one of the invariant properties under transformation between inertial frames of reference.

The constancy of the speed of light implies that the infinitesimal quantity

$$\mathrm{d}^2 s = c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2$$

is constant. We will soon replace the elements of the Galilean group with elements that keep this quantity.

Four-vectors and the Minkowski metric We now define the four-vector x^μ, where $\mu = 0, 1, 2, 3$ and $\mu = 0$ corresponds to ct and an inner product with the metric η. The metric is diagonal with $\eta_{11} = 1$ and $\eta_{ii} = -1$ otherwise. This is called the Minkowski metric.

Lorentz transformations We are now interested in transformations that preserve the new arc length. If the transformation is on the form $\mathbf{x}' = \Lambda\mathbf{x}$. Computing the arc length yields

$$\Lambda^T \nu \Lambda = \nu.$$

The transformations satisfying this constitute the Lorentz group, or $\mathrm{O}(1,3)$. Computing the determinant on either side yields $\det(\Lambda)^2 = 1$. The subgroup with determinant 1 (which preserve the direction of time) is the special Lorentz group $\mathrm{SO}(1,3)$.

The equation defining the group elements is symmetric, which imposes constraints on the elements of the matrix. The matrix has 16 elements, so the defining equation places 10 constraints on the coefficients of Λ. With 10 equations and 16 unknowns, we expect 6 linearly independent solutions.

The first three are rotations of space, written in block diagonal form as

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}.$$

Inserted into the defining equation, we obtain

$$\begin{bmatrix} 1 & 0 \\ 0 & -R^T R \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

This yields the familiar requirement $R^T R = 1$.

The remaining three transforms are Lorentz boosts corresponding to each axis. This can be shown explicitly for x, and a permutation of coordinates will yield the same result for a boost along any other axis. We believe it to be reasonable that such a transformation should not affect any other coordinates than the boosted coordinate and time. This means that the matrix will be on the form

$$\Lambda = \begin{bmatrix} \Lambda_x & 0 \\ 0 & 1 \end{bmatrix}.$$

The defining equation now yields

$$\Lambda_x^T \sigma_z \Lambda_x = \sigma_z, \ \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

We expand Λ_x around the identity as $\Lambda_x = 1 - \phi K$, where ϕ is independent of both coordinates and time. Inserting this into the above equation yields

$$(1 - \phi K^T)\sigma_z(1 - \phi K) = \sigma_z.$$

Expanding the bracket yields

$$\begin{aligned} (1 - \phi K^T)(\sigma_z - \phi\sigma_z K) &= \sigma_z, \\ \sigma_z - \phi\sigma_z K - \phi K^T \sigma_z + \phi^2 K^T \sigma_z K &= \sigma_z. \end{aligned}$$

Ignoring higher-order terms yields

$$\begin{aligned} \sigma_z K + K^T \sigma_z &= 0, \\ (\sigma_z K)^T &= -\sigma_z K. \end{aligned}$$

The generator K must therefore be

$$K = \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Now a transformation correpsonding to an arbitrary ϕ can be written as

$$\Lambda_x = e^{-\phi K} = \cosh\phi - \sinh\phi\sigma_x,$$

where the last equality comes from writing the exponential as an infinite series and using the fact that $K^2 = 1$. Perhaps someone should do this explicitly.

To identify the transformation more exactly, we consider two frames of reference in which the origins coincide at $t = 0$. Under such a transformation, we require that (ct, vt) map to $(ct', 0)$. This yields

$$-\sinh\phi ct + \cosh\phi vt = 0.$$

Defining $\gamma = \cosh\phi$ and applying hyperbolic identities yields

$$\begin{aligned} -\sqrt{\gamma^2 - 1}ct + \gamma vt = ct\left(\frac{v}{c}\gamma - \sqrt{\gamma^2 - 1}\right) = 0, \\ \gamma = \sqrt{\frac{1}{1 - \frac{v^2}{c^2}}}. \end{aligned}$$

The transformation can now be written as

$$\Lambda_x = \begin{bmatrix} \gamma & -\frac{v}{c}\gamma \\ -\frac{v}{c}\gamma & \gamma \end{bmatrix}.$$

The total matrix for a boost along any other coordinate axis can be found by permuting the elements in the transformation matrix for the x boost. This yields a basis of matrices, and a boost along an arbitrary direction can be found by taking linear combinations of these.

Adding velocities The product of two boosts is another boost. For two boosts along the same direction, we obtain $\Lambda(\phi_1)\Lambda(\phi_2) = \Lambda(\phi_1 + \phi_2)$. This can be used to show that the total boosted velocity is

$$v_3 = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}}.$$

Proper time Consider a particle at the origin in its rest frame. The arc length becomes $\mathrm{d}s^2 = c^2\,\mathrm{d}t^2$. As the left-hand side is invariant, so must the right-hand side be. This makes it natural to define the proper time

$$\mathrm{d}\tau = \frac{1}{c}\,\mathrm{d}s\,.$$

Relativistic kinematics Suppose that you wanted to define $\mathbf{u} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$ as the spatial part of velocity. Well, too bad, cause time transforms under a Lorentz transformation, so this thing will not behave linearly under Lorentz transformation. We need a better alternative.

Consider instead the rest frame S' of the particle, where it is resting at the origin. Its trajectory in the original inertial frame can be parametrized in terms of the proper time. Along a small trajectory we have

$$\mathrm{d}\tau = \frac{1}{c}\,\mathrm{d}s = \frac{1}{c}\sqrt{c^2\,\mathrm{d}t^2 - \mathrm{d}\mathbf{x}^2} = \mathrm{d}t\sqrt{1 - \left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\right)^2} \implies \frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma.$$

We can now define the four velocity

$$U = \frac{\mathrm{d}x}{\mathrm{d}\tau} = \frac{\mathrm{d}}{\mathrm{d}\tau}\begin{bmatrix} ct \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} c\frac{\mathrm{d}t}{\mathrm{d}\tau} \\ \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} \end{bmatrix} = \gamma\begin{bmatrix} c \\ \mathbf{u} \end{bmatrix}$$

where $\mathbf{u} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$. This quantity transforms like a four-vector, and is therefore the four-velocity.

9 Classical Field theory

Classical field theory can be considered a limit of classical dynamics when the number of particles is infinite. The system obtains new "coordinates" ϕ^a, which are functions of position and time. Summations over coordinates now become integrals over space.

Lagrangian Formulation of Field Theory The Lagrangian in a field theory now becomes

$$L = \int \mathrm{d}^d\mathbf{r}\,\mathcal{L}$$

where $\mathcal{L}$ is the Lagrangian density, which in general depends on all involved fields, their derivatives with respect to both space and time and space and time themselves. From this we can obtain the action, and extremize it to obtain the equations for the time evolution of the system. The equations of motion are of the form

$$\frac{\partial\mathcal{L}}{\partial\phi^a} - \frac{\partial}{\partial t}\frac{\partial\mathcal{L}}{\partial(\partial_t\phi^a)} - \frac{\partial}{\partial x^i}\frac{\partial\mathcal{L}}{\partial(\partial_i\phi^a)} = 0.$$

Alternatively, by defining $x^0 = ct$ for some speed c and extending the summation, we can write

$$\frac{\partial\mathcal{L}}{\partial\phi^a} - \frac{\partial}{\partial x^\mu}\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi^a)} = 0.$$

Example: A String

Example: The Electromagnetic Field

Example: The Schrödinger Equation

Hamiltonian Formulation In the Hamiltonian formalism, we define the momentum density

$$\pi_a = \partial_{\phi^a} \mathcal{L}.$$

The Hamiltonian is now given by

$$H = \int \mathrm{d}^D \mathbf{r}\, \mathcal{H},$$

where $\mathcal{H} = \pi_a \partial_t \phi^a - \mathcal{L}$. The Hamiltonian equations of motion become

$$\dot{\phi} = \frac{\delta H}{\delta \pi},\ \dot{\pi} = -\frac{\delta H}{\delta \phi}.$$

While the Hamiltonian formalism carries no issues with it in classical contexts, it does not generalize well to relativity due to the fact that it treats the time derivative differently to the spatial derivatives, which is a big no-no.

Example: A String

Reduction to Discrete Problems For a problem on a compact domain, one can Fourier expand the fields (and the momentum densities) to obtain a discrete set of Fourier coefficients, the dynamics of which can be studied. It is this approach which will be the basis for quantum mechanics, where the coefficients will be replaced by occupation operators.

For problems on non-compact domains, we instead employ the Fourier transform as a tool. However, we have not really helped ourselves in this case.

Symmetries of Field Theories Consider a field theory (on Euclidean space) described by the Lagrangian density $\mathcal{L}$. A symmetry of the system is a transformation of all involved coordinates and fields such that:

1. $\mathcal{L}$retains its functional form under the transformation - in other words, the expression for the Lagrangian density is unchanged.
2. The action is unchanged by the transformation.

Before proceeding, it would also be useful to clarify what kinds of transformation we are considering. Transformations in field theory concern both transformations of the coordinates according to

$$(x')^\mu = x^\mu + \delta x^\mu$$

and of the fields according to

$$(\phi')^a((x')^\mu) = \phi^a(x^\mu) + \delta\phi^a\,.$$

We will distinguish between the transformed fields and the change in the field at a particular point, given by

$$(\phi')^a(x^\mu) = \phi^a(x^\mu) + \bar{\delta}\phi^a$$

Nöether's Theorem Field theory also carries with it a version of Nöether's theorem, which will be covered here. A version will be presented here which is somewhat more restricted than the version presented for systems with discrete degrees of freedom - if you wanted to compare the two, you could say that this version only contains symmetries.

Consider the action of a symmetry on a given system. The requirement that the action be unchanged can be written as

$$\int_{\Omega'} \mathrm{d}(x')^\mu\, \mathcal{L}' - \int_{\Omega} \mathrm{d}x^\mu\, \mathcal{L} = 0.$$

The functional form of the Lagrangian density is unchanged, which also carries the consequence that the integration variables may be renamed. This yields

$$\int\limits_{\Omega'} \mathrm{d}x^\mu \, \mathcal{L}((\phi')^a, x^\mu) - \int\limits_{\Omega} \mathrm{d}x^\mu \, \mathcal{L}(\phi^a, x^\mu) = 0.$$

To illustrate what happens to this, consider the analogous scenario in one dimension, for which we have

$$\int\limits_{a+\delta a}^{b+\delta b} \mathrm{d}x \, f + \delta f - \int\limits_{a}^{b} \mathrm{d}x \, f = 0.$$

This can be rewritten as

$$\int\limits_{a+\delta a}^{b+\delta b} \mathrm{d}x \, \delta f + \int\limits_{b}^{b+\delta b} \mathrm{d}x \, f - \int\limits_{a}^{a+\delta a} \mathrm{d}x \, f = 0.$$

The terms containing an integral of a variation over an interval containing a variation will be of order greater than one, and may thus be ignored. In addition, the two latter integrals may be linearized to obtain

$$\int\limits_{a}^{b} \mathrm{d}x \, \delta f + f(b) \, \delta b - f(a) \, \delta a = \int\limits_{a}^{b} \mathrm{d}x \, \delta f + \frac{\mathrm{d}}{\mathrm{d}x}(f \, \delta x) = 0.$$

Returning to the case which we wanted to study, we obtain

$$\int\limits_{\Omega} \mathrm{d}x^\mu \, \mathcal{L}((\phi')^a, x^\mu) - \mathcal{L}(\phi^a, x^\mu) + \int\limits_{S} \mathrm{d}S_\mu \, \mathcal{L}(\phi^a, x^\mu) \, \delta x^\mu = \int\limits_{\Omega} \mathrm{d}x^\mu \, \mathcal{L}((\phi')^a, x^\mu) - \mathcal{L}(\phi^a, x^\mu) + \frac{\partial}{\partial x^\nu}(\mathcal{L}(\phi^a, x^\mu) \, \delta x^\nu) = 0,$$

where S is the boundary of Ω and we have made (hopefully proper) use of the n-dimensional divergence theorem. The difference in the first two terms can be expanded to first order as

$$\mathcal{L}((\phi')^a, x^\mu) - \mathcal{L}(\phi^a, x^\mu) = \frac{\partial \mathcal{L}}{\partial \phi^a} \bar{\delta} \phi^a + \frac{\partial \mathcal{L}}{\partial (\partial_\nu \phi^a)} \bar{\delta}(\partial_\nu \phi^a).$$

The use of the variation at a specific point is due to the fact that both Lagrangians are now evaluated at the same points. This is significant because while the total variation does not commute with the differentiation operators, this one does. Using the equations of motion, we additionally obtain

$$\frac{\partial \mathcal{L}}{\partial \phi^a} \bar{\delta} \phi^a + \frac{\partial \mathcal{L}}{\partial (\partial_\nu \phi^a)} \partial_\nu \bar{\delta} \phi^a = \bar{\delta} \phi^a \frac{\partial}{\partial x^\nu} \frac{\partial \mathcal{L}}{\partial (\partial_\nu \phi^a)} + \frac{\partial \mathcal{L}}{\partial (\partial_\nu \phi^a)} \partial_\nu \bar{\delta} \phi^a = \frac{\partial}{\partial x^\nu} \left(\bar{\delta} \phi^a \frac{\partial \mathcal{L}}{\partial (\partial_\nu \phi^a)} \right).$$

Hence we have

$$\int\limits_{\Omega} \mathrm{d}x^\mu \, \frac{\partial}{\partial x^\nu} \left(\bar{\delta} \phi^a \frac{\partial \mathcal{L}}{\partial (\partial_\nu \phi^a)} + \mathcal{L} \, \delta x^\nu \right) = 0,$$

which is already in the form of a conservation law for the quantities

$$\bar{\delta} \phi^a \frac{\partial \mathcal{L}}{\partial (\partial_0 \phi^a)} + \mathcal{L} \, \delta x^0$$

and the corresponding currents

$$\bar{\delta} \phi^a \frac{\partial \mathcal{L}}{\partial (\partial_i \phi^a)} + \mathcal{L} \, \delta x^i \, .$$

To obtain a neater form, we specify the form of the transformation as

$$\delta x^\mu = \alpha^r X^\mu_r, \;\; \delta \phi^a = \alpha^r \Phi^a_r$$

where the α^r are a set of transformation parameters. To first order, we have

$$\delta\phi^a = \bar{\delta}\phi^a + \partial_\mu\phi^a\,\delta x^\mu\,,$$
$$\bar{\delta}\phi^a = \alpha^r(\Phi^a_r - \partial_\mu\phi^a X^\mu_r).$$

For this kind of transformation, one thus obtains

$$\int\limits_\Omega \mathrm{d}x^\mu\,\frac{\partial}{\partial x^\nu}\left(\alpha^r(\Phi^a_r - \partial_\sigma\phi^a X^\sigma_r)\frac{\partial\mathcal{L}}{\partial(\partial_\nu\phi^a)} + \alpha^r X^\nu_r\mathcal{L}\right) = \int\limits_\Omega \mathrm{d}x^\mu\,\alpha^r\frac{\partial}{\partial x^\nu}\left(\left(\mathcal{L}\delta^\nu_\sigma - \partial_\sigma\phi^a\frac{\partial\mathcal{L}}{\partial(\partial_\nu\phi^a)}\right)X^\sigma_r + \Phi^a_r\frac{\partial\mathcal{L}}{\partial(\partial_\nu\phi^a)}\right).$$

Setting this equal to zero and changing the sign of the integrand yields

$$\int\limits_\Omega \mathrm{d}x^\mu\,\alpha^r\frac{\partial}{\partial x^\nu}\left(\left(\partial_\sigma\phi^a\frac{\partial\mathcal{L}}{\partial(\partial_\nu\phi^a)} - \mathcal{L}\delta^\nu_\sigma\right)X^\sigma_r - \Phi^a_r\frac{\partial\mathcal{L}}{\partial(\partial_\nu\phi^a)}\right) = 0,$$

which is the final form of Nöether's theorem. It states that

$$\frac{\partial}{\partial x^\nu}\left(\left(\partial_\sigma\phi^a\frac{\partial\mathcal{L}}{\partial(\partial_\nu\phi^a)} - \mathcal{L}\delta^\nu_\sigma\right)X^\sigma_r - \Phi^a_r\frac{\partial\mathcal{L}}{\partial(\partial_\nu\phi^a)}\right) = 0,$$

which is a conservation law for the quantities corresponding to $\nu = 0$.

A more general proof can be performed for a transformation such that

$$\delta\mathcal{L} = \frac{\mathrm{d}}{\mathrm{d}\alpha}\frac{\mathrm{d}V^\mu}{\mathrm{d}x^\mu},$$

i.e. a divergence, but I will leave this case for the future me.

Example: Charge Conservation

Poisson Brackets Poisson brackets of two functionals on phase space are defined as

$$\{F, G\} = \int \mathrm{d}^D\mathbf{r}\,\partial_\phi F\partial_\pi G - \partial_\pi F\partial_\phi G$$

We can somehow show that

$$\{\phi(x), \phi(y)\} = \{\pi(x), \pi(y)\} = 0,\ \{\phi(x), \pi(y)\} = \delta(x - y).$$