Summary of SI2360 Analytical Mechanics and Classical Field Theory

Yashar Honarmandi
yasharh@kth.se

13 december 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

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

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 and the structure of the Lie algebra 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.

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

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

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

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

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):

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

Representations A representation of a group on a vector space is a homomorphism $D : G \to \mathrm{GL}(V)$, where $\mathrm{GL}(V)$ is the group of all invertible linear transformations on V. The dimension of the representation is defined as the dimension of V.

Representations on Tensor Product Spaces Suppose that there exists representations ρ_V and ρ_W of G on V and W respectively. The representation $\rho_{V\otimes W}$ of G on $V \otimes W$ is then defined by

$$\rho_{V\otimes W}(a)(\mathbf{v} \otimes \mathbf{w}) = (\rho_V(a)(\mathbf{v})) \otimes (\rho_W(a)(\mathbf{w})).$$

Representations on Direct Product Spaces Suppose that there exists representations ρ_V and ρ_W of G on V and W respectively. The representation $\rho_{V\oplus W}$ of G on $V \oplus W$ is then defined by

$$\rho_{V\oplus W}(a)(\mathbf{v}, \mathbf{w}) = (\rho_V(a)(\mathbf{v}), \rho_W(a)(\mathbf{w})).$$

Reducible and Irreducible Representations Two representations are equivalent if they satisfy $U\rho(a)U^{-1} = \rho'(a)$ for any group element a and some linear operator U. ρ is reducible if it is equivalent to a representation of the form

$$\begin{bmatrix} \rho_1 & \sigma \\ 0 & \rho_2 \end{bmatrix},$$

where ρ_1 and ρ_2 are representations of dimension n_1 and n_2 respectively and σ is an $n_1 \times n_2$ block satisfying $\sigma(ab) = \rho_1(a)\sigma(b) + \sigma(a)\rho_1(b)$. If this is not possible, the representation is irreducible.

Example: SO(3) As we know, SO(3) may be represented by rotations in three dimensions. In particular, we may choose a basis for $\mathbb{R}^3$ such that

$$\rho(R^3_{\mathbf{e}_z}) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The upper left block is a representation of SO(2), and the lower right block is the trivial representation $\rho(a) = 1$. Hence ρ is reducible.

Note that the upper right block is zero. For finite-dimensional groups, one can apparently always choose a basis such that this is possible. In this case we may write $\rho(R^3_{\mathbf{e}_z}) = \rho_2 \oplus 1$, where ρ_2 is the representation of SO(2).

Example: Tensor Product Representations Given some representation ρ on V, we consider the representation $\rho_{V\otimes V} = \rho \otimes \rho$. Next, we split a tensor T into its symmetric components $T_{\{ij\}}$ and antisymmetric components $T_{[ij]}$ such that $T_{ij} = T_{\{ij\}} + T_{[ij]}$. Next, we have for a symmetric tensor S

$$(\rho \otimes \rho S)_{ij} = \rho_{ik}\rho_{jl}S_{kl} = \rho_{jl}\rho_{ik}S_{lk} = (\rho \otimes \rho S)_{ji},$$

i.e. the representation preserves the symmetry of S. The same holds for antisymmetric tensors. Denoting the symmetric and antisymmetric subspaces as $(V \otimes V)_\pm$ we have $(V \otimes V)_+ \oplus (V \otimes V)_-$, we thus have

$$\rho \otimes \rho = \begin{bmatrix} \rho_{(V\otimes V)_+} & 0 \\ 0 & \rho_{(V\otimes V)_-} \end{bmatrix},$$

and $\rho \otimes \rho$ has been reduced to $\rho_{(V\otimes V)_+} \oplus \rho_{(V\otimes V)_-}$, which is the direct sum of its symmetric and antisymmetric representation.

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^b_a.

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.

More General Tensors To expand on the above, we first define tensors as we have discussed them of $(0, n)$ tensors. We now phrase the above as follows: Given a $(0, n+m)$ tensor T and a $(0, m)$ tensor $\mathbf{w}_1 \otimes \dots \otimes \mathbf{w}_m$ we may construct the tensor

$$(\tilde{T}(\mathbf{w}_1 \otimes \dots \otimes \mathbf{w}_m))(\mathbf{1}, \dots, \mathbf{n}) = T(\mathbf{w}_1, \dots, \mathbf{w}_m, \mathbf{1}, \dots, \mathbf{n}),$$

which is $(0, m)$. These must needs be linear.

Explicitly in terms of the contravariant tensor components we may construct such tensors as

$$T = T^{a_1\dots a_n}_{b_1\dots b_m} \bigotimes_{k=1}^{n} \mathbf{E}_{a_k} \bigotimes_{l=1}^{m} \mathbf{E}^{b_l}.$$

The components of such a tensor are

$$T(e^{a_1\dots a_n})(\mathbf{E}_{b_1}, \dots, \mathbf{E}_{b_m}) = T^{a_1\dots a_n}_{b_1\dots b_m} e^{a_1\dots a_n} \left(\bigotimes_{k=1}^{n} \mathbf{E}_{c_k}\right) \bigotimes_{l=1}^{m} \mathbf{E}_{b_l} \left(\mathbf{E}^{d_1}, \dots, \mathbf{E}^{d_m}\right) = T^{a_1\dots a_n}_{b_1\dots b_m}.$$

Recalling the transformation rules

$$(\mathbf{E}_b)' = L^a_b \mathbf{E}_a, \; L^a_b = \partial'_b \chi^a$$

and

$$(\mathbf{E}^b)' = K^b_a \mathbf{E}^a, \; K^b_a = \partial_a (\chi')^b$$

we may obtain general transformation rules for a (n, m) tensor. We first have

$$(e')^{a_1\dots a_n} = \bigotimes_{k=1}^{n} (\mathbf{E}^{a_k})' = \bigotimes_{k=1}^{n} \partial_{c_k} (\chi')^{a_k} \mathbf{E}^{c_k} = \left(\prod_{k=1}^{n} \partial_{c_k} (\chi')^{a_k}\right) \bigotimes_{k=1}^{n} \mathbf{E}^{c_k} = \left(\prod_{k=1}^{n} \partial_{c_k} (\chi')^{a_k}\right) e^{c_1\dots c_n}.$$

This yields

$$\begin{aligned}
(T')^{a_1\dots a_n}_{b_1\dots b_m} &= T((e')^{a_1\dots a_n}) \left((\mathbf{E}_{b_1})', \dots, (\mathbf{E}_{b_m})'\right) \\
&= T\left(\left(\prod_{k=1}^{n} \partial_{c_k} (\chi')^{a_k}\right) e^{c_1\dots c_n}\right) \left(\partial'_{b_1} \chi^{d_1} \mathbf{E}_{d_1}, \dots, \partial'_{b_m} \chi^{d_m} \mathbf{E}_{d_m}\right) \\
&= \left(\prod_{k=1}^{n} \partial_{c_k} (\chi')^{a_k}\right) \left(\prod_{l=1}^{m} \partial'_{b_l} \chi^{d_l}\right) T(e^{c_1\dots c_n})(\mathbf{E}_{d_1}, \dots, \mathbf{E}_{d_m}) \\
&= \left(\prod_{k=1}^{n} \partial_{c_k} (\chi')^{a_k}\right) \left(\prod_{l=1}^{m} \partial'_{b_l} \chi^{d_l}\right) T^{c_1\dots c_n}_{d_1\dots d_m}.
\end{aligned}$$

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

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

In component form:

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

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 $g_{ab} g^{bc} = \delta_a^c$.

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? Clearly they do not. One simple counterexample is when converting from Cartesian coordinates to any non-trivial coordinate system. In Cartesian coordinates all Christoffel symbols are zero, and no linear combination of these could possibly produce non-zero values. There is a transformation rule, however. To find it, we study

$$\begin{aligned}
(\Gamma')^a{}_{bc} &= (\mathbf{E}^a)' \cdot \partial'_b(\mathbf{E}_c)' \\
&= \partial_d(\chi')^a \mathbf{E}^d \cdot \partial'_b(\partial'_c \chi^f \mathbf{E}_f) \\
&= \partial_d(\chi')^a \mathbf{E}^d \cdot (\mathbf{E}_f \partial'_b \partial'_c \chi^f + \partial'_c \chi^f \partial'_b \mathbf{E}_f) \\
&= \partial_d(\chi')^a (\delta^d_f \partial'_b \partial'_c \chi^f + \partial'_c \chi^f \mathbf{E}^d \cdot \partial'_b \chi^g \partial_g \mathbf{E}_f) \\
&= \partial_d(\chi')^a (\partial'_b \partial'_c \chi^d + \partial'_c \chi^f \partial'_b \chi^g \Gamma^d_{gf}) \\
&= \partial_d(\chi')^a \partial'_c \chi^f \partial'_b \chi^g \Gamma^d_{gf} + \partial_d(\chi')^a \partial'_b \partial'_c \chi^d.
\end{aligned}$$

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 Next we study the derivatives of a tensor field

$$\tau = \tau^{a_1\ldots a_n}_{b_1\ldots b_m} e^{b_1\ldots b_m}_{a_1\ldots a_n}.$$

The tensor basis element differentiates according to the product rule, but with multiplication replaced by the tensor product. Hence

$$\begin{aligned}
\partial_a \tau &= e^{b_1\ldots b_m}_{a_1\ldots a_n} \partial_a \tau^{a_1\ldots a_n}_{b_1\ldots b_m} + \tau^{a_1\ldots a_n}_{b_1\ldots b_m} \partial_a e^{b_1\ldots b_m}_{a_1\ldots a_n} \\
&= e^{b_1\ldots b_m}_{a_1\ldots a_n} \partial_a \tau^{a_1\ldots a_n}_{b_1\ldots b_m} + \tau^{a_1\ldots a_n}_{b_1\ldots b_m} \left(\sum_{k=1}^{n} \Gamma^{c_k}_{aa_k} e^{b_1\ldots b_m}_{c_1\ldots c_n} - \sum_{l=1}^{m} \Gamma^{b_l}_{ad_l} e^{d_1\ldots d_m}_{a_1\ldots a_n} \right) \\
&= e^{b_1\ldots b_m}_{a_1\ldots a_n} \partial_a \tau^{a_1\ldots a_n}_{b_1\ldots b_m} + \sum_{k=1}^{n} \tau^{c_1\ldots c_n}_{b_1\ldots b_m} \Gamma^{a_k}_{ac_k} e^{b_1\ldots b_m}_{a_1\ldots a_n} - \sum_{l=1}^{m} \tau^{a_1\ldots a_n}_{d_1\ldots d_m} \Gamma^{d_l}_{ab_l} e^{b_1\ldots b_m}_{a_1\ldots a_n} \\
&= e^{b_1\ldots b_m}_{a_1\ldots a_n} \left(\partial_a \tau^{a_1\ldots a_n}_{b_1\ldots b_m} + \sum_{k=1}^{n} \tau^{c_1\ldots c_n}_{b_1\ldots b_m} \Gamma^{a_k}_{ac_k} - \sum_{l=1}^{m} \tau^{a_1\ldots a_n}_{d_1\ldots d_m} \Gamma^{d_l}_{ab_l} \right).
\end{aligned}$$

We thus define

$$\vec{\boldsymbol{\nabla}}_a \tau^{a_1\ldots a_n}_{b_1\ldots b_m} = \partial_a \tau^{a_1\ldots a_n}_{b_1\ldots b_m} + \sum_{k=1}^{n} \tau^{c_1\ldots c_n}_{b_1\ldots b_m} \Gamma^{a_k}_{ac_k} - \sum_{l=1}^{m} \tau^{a_1\ldots a_n}_{d_1\ldots d_m} \Gamma^{d_l}_{ab_l}.$$

The Gradient of a Tensor Field Based on the above, the directional derivative of a tensor field is

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

This is equal to the contraction of $\mathbf{n}$ with the object

$$\vec{\boldsymbol{\nabla}} \tau = \mathbf{E}^c \otimes \partial_c \tau,$$

which is defined as the gradient of τ. More explicitly, we have

$$\vec{\boldsymbol{\nabla}} \tau = e^{cb_1\ldots b_m}_{a_1\ldots a_n} \vec{\boldsymbol{\nabla}}_c \tau^{a_1\ldots a_n}_{b_1\ldots b_m}.$$

It has the interesting property of having a rank one higher than τ, which matches what we know - for instance, the gradient transforms a scalar into a vector.

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.

7 Differential Geometry on Manifolds

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.

More formally, a manifold is described by a number of sets $U_i \subset \mathbb{R}^n$ called charts. To each chart belongs a set of coordinate functions χ_i which map from a subset $M_i \subset M$ to U_i such that χ_i is a smooth bijection. A set of charts such that every point $p \in M$ is found in at least one chart is called an atlas.

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 manifolds, 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

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

These live in the tangent space T_pM of the manifold M at the point p.

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. 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 $\mathbf{E}^a = d\chi^a$. These live in the dual space $T_p^\star M$ of the manifold M at the point p.

The dual basis satisfies satisfies

$$d\chi^a(\partial_b) = \partial_b \chi^a = \delta^a_b,$$

as expected. Using this, we obtain

$$df = (\partial_a f) d\chi^a,$$

which at least looks like the differential of a function.

The components transform according to

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

which is the transformation rule for covariant vector components.

Tensors Having identified a basis for the tangent and dual spaces, we may now construct tensors similarly to what we have previously done. Note now that as the tangent and dual vectors belong to different vector spaces, the notion of type (n, m) tensors is more clear. This also explains why we needed to be careful with indices being up or down when studying Euclidean space, as the difference is huge for manifolds.

Flow of Vector Fields The tangent bundle of a manifold is defined as $TM = \bigcup_p T_p M$. A vector field is a map $X : M \to TM$ such that $X(p) \in T_p M$. Given this, we may define the flow of a vector field as a collection of curves γ_X which given some starting point p satisfy

$$\left.\frac{\mathrm{d}\gamma_X}{\mathrm{d}\tau}\right|_{p,s} = \left. X \right|_{\gamma_X(p,s)}.$$

We may for a fixed s define the function $\gamma_{sX}(p) = \gamma_X(p, s)$, which maps M to itself.

Pushforwards and Pullbacks Consider some function f which maps a manifold M_1 to another manifold M_2, as well as a function $g : M_2 \to \mathbb{R}$. We then define the pullback of g to M_1 by f as $f^\star g = g \circ f$. We also define the pushforward of a vector $V \in T_p M$ as $(f_\star V)\phi = V(f^\star \phi)$.

Tangents and the Pushforward f maps the cooordinates χ^a of M_1 to the coordinates η^μ of M_2. By definition we have

$$\begin{aligned}(f_\star V)\phi &= V(f^\star \phi) \\ &= V^a \partial_a (\phi \circ f) \\ &= V^a \partial_\mu \phi \partial_a \eta^\mu,\end{aligned}$$

meaning $f_\star V = V^a \partial_a \eta^\mu \partial_\mu$.

How do we interpret this? Consider some curve γ in M_1 which is mapped to a curve α M_2 by f. If V is the tangent of γ at some particular point, we have

$$\dot{\alpha} = \dot{\eta^\mu} \partial_\mu = \partial_a \eta^\mu \dot{\chi^a} \partial_\mu = \partial_a \eta^\mu V^a \partial_\mu,$$

meaning that the pushforward of a tangent by f is the tangent of the curve produced by f.

The Pullback of Tensors We can now define the pullback of a $(0, m)$ tensor on M_2 according to

$$f^\star \omega(V_1, \ldots, V_m) = \omega(f_\star V_1, \ldots, f_\star V_m).$$

If f is a bijection we may also define the more general pullback of a (n, m) tensor on M_2 as

$$f^\star T(V_1, \ldots, V_m, \omega_1, \ldots, \omega_n) = T(f_\star V_1, \ldots, f_\star V_m, (f^{-1})^\star \omega_1, \ldots, (f^{-1})^\star \omega_n).$$

The Lie Derivative For a tensor field T we define the Lie derivative in the X-direction as

$$\mathcal{L}_X T = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\gamma_{\varepsilon X}^\star T - T \right).$$

An Expression for the Lie Derivative As the definition of the Lie derivative is similar to that of the usual derivative, it follows that it is linear in its argument and satisfies the product rule, where the product in question is now the tensor product. This means that we can find a general expression for the Lie derivative of a tensor field by considering its action on tangent and dual vectors.

First, for scalars we simply have $\mathcal{L}_X f = X^a \partial_a f$.

When studying the effect on tangent vectors we will have to be extra careful. Let $\gamma_{\varepsilon X}$ map p to q, and let $Y\big|_p(\omega) = Y^a\big|_p \, \partial_a \omega\big|_p$ to clarify at what point we are working. To first order in ε we then have

$$\begin{aligned}
\gamma_{\varepsilon X}^\star Y\Big|_q(\omega) &= Y\Big|_q(\gamma_{-\varepsilon X}^\star \omega) \\
&= Y^a\Big|_q \, \partial_a(\omega \circ \gamma_{-\varepsilon X})\Big|_q \\
&\approx \left(Y^a\Big|_p + \varepsilon \, X^b \partial_b Y^a\Big|_p \right) \partial_c \omega\Big|_{\gamma_{-\varepsilon X}(q)} \partial_a(\gamma_{-\varepsilon X})^c\Big|_q \\
&\approx \left(Y^a\Big|_p + \varepsilon \, X^b \partial_b Y^a\Big|_p \right) (\delta_a^c - \varepsilon \partial_a X^c) \, \partial_c \omega\Big|_p \\
&\approx \left(\delta_a^c \, Y^a\Big|_p + \varepsilon \delta_a^c \, X^b \partial_b Y^a\Big|_p - \varepsilon \partial_a X^c \, Y^a\Big|_p \right) \partial_c \omega\Big|_p \\
&= Y^a\Big|_p \, \partial_a \omega\Big|_p + \varepsilon \, X^b \partial_b Y^a\Big|_p \, \partial_a \omega\Big|_p - Y^a\Big|_p \, \varepsilon \partial_a X^c \, \partial_c \omega\Big|_p,
\end{aligned}$$

and thus $\mathcal{L}_X Y = [X, Y]$. In particular, we have $\mathcal{L}_X \partial_a = -(\partial_a X^b)\partial_b$.

To find the Lie derivative of dual vectors, we use the fact that the product rule (somehow) respects contraction, meaning

$$\begin{aligned}
\mathcal{L}_X \omega(Y) &= (\mathcal{L}_X \omega)(Y) + \omega(\mathcal{L}_X Y) \\
&= (\mathcal{L}_X \omega)_a Y^a + \omega_a (X^b \partial_b Y^a - Y^b \partial_b X^a).
\end{aligned}$$

As the left-hand side is simply equal to $X^a \partial_a(\omega_b Y^b)$, we obtain

$$\begin{aligned}
(\mathcal{L}_X \omega)_a Y^a &= X^a \partial_a(\omega_b Y^b) - \omega_b X^a \partial_a Y^b + \omega_a Y^b \partial_b X^a \\
&= Y^a X^b \partial_b \omega_a + \omega_b Y^a \partial_a X^b,
\end{aligned}$$

and thus

$$(\mathcal{L}_X \omega)_a = X^b \partial_b \omega_a + \omega_b \partial_a X^b.$$

In particular, we have $\mathcal{L}_X d\chi^a = \partial_c X^a d\chi^c$.

We are now ready to compute the Lie derivative of a tensor. Writing

$$T = T^{a_1 \ldots a_n}_{b_1 \ldots b_m} e^{b_1 \ldots b_m}_{a_1 \ldots a_n},$$

the product rule implies

$$\begin{aligned}
\mathcal{L}_X T &= (\mathcal{L}_X T^{a_1 \ldots a_n}_{b_1 \ldots b_m}) e^{b_1 \ldots b_m}_{a_1 \ldots a_n} + T^{a_1 \ldots a_n}_{b_1 \ldots b_m} \sum_{i=1}^n e^{b_1 \ldots b_m}_{a_1 \ldots a_{i-1}} \otimes \mathcal{L}_X e_{a_i} \otimes e_{a_{i+1} \ldots a_n} + T^{a_1 \ldots a_n}_{b_1 \ldots b_m} \sum_{j=1}^m e^{b_1 \ldots b_{j-1}} \otimes \mathcal{L}_X e^{b_j} \otimes e^{b_{j+1} \ldots b_m}_{a_1 \ldots a_n} \\
&= X^a \partial_a T^{a_1 \ldots a_n}_{b_1 \ldots b_m} e^{b_1 \ldots b_m}_{a_1 \ldots a_n} - T^{a_1 \ldots a_n}_{b_1 \ldots b_m} \sum_{i=1}^n e^{b_1 \ldots b_m}_{a_1 \ldots a_{i-1}} \otimes \partial_{a_i} X^a e_a \otimes e_{a_{i+1} \ldots a_n} + T^{a_1 \ldots a_n}_{b_1 \ldots b_m} \sum_{j=1}^m e^{b_1 \ldots b_{j-1}} \otimes \partial_a X^{b_j} e^a \otimes e^{b_{j+1} \ldots b_m}_{a_1 \ldots a_n} \\
&= X^a \partial_a T^{a_1 \ldots a_n}_{b_1 \ldots b_m} e^{b_1 \ldots b_m}_{a_1 \ldots a_n} - \sum_{i=1}^n T^{a_1 \ldots a_n}_{b_1 \ldots b_m} \partial_{a_i} X^a e^{b_1 \ldots b_m}_{a_1 \ldots a_{i-1}} \otimes e_a \otimes e_{a_{i+1} \ldots a_n} + \sum_{j=1}^m T^{a_1 \ldots a_n}_{b_1 \ldots b_m} \partial_a X^{b_j} e^{b_1 \ldots b_{j-1}} \otimes e^a \otimes e^{b_{j+1} \ldots b_m}_{a_1 \ldots a_n}.
\end{aligned}$$

There are no free indices, so we may switch the places of the numbered indices and a to obtain

$$\begin{aligned}\mathcal{L}_X T =& X^a \partial_a T^{a_1\ldots a_n}_{b_1\ldots b_m} e^{b_1\ldots b_m}_{a_1\ldots a_n} - \sum_{i=1}^{n} T^{a_1\ldots a_{i-1} a a_{i+1}\ldots a_n}_{b_1\ldots b_m} \partial_a X^{a_i} e^{b_1\ldots b_m}_{a_1\ldots a_{i-1}} \otimes e_{a_i} \otimes e_{a_{i+1}\ldots a_n} \\ &+ \sum_{j=1}^{m} T^{a_1\ldots a_n}_{b_1\ldots b_{i-1} a b_{i+1}\ldots b_m} \partial_{b_j} X^a e^{b_1\ldots b_{j-1}} \otimes e^{b_j} \otimes e^{b_{j+1}\ldots b_m}_{a_1\ldots a_n} \\ =& \left(X^a \partial_a T^{a_1\ldots a_n}_{b_1\ldots b_m} - \sum_{i=1}^{n} T^{a_1\ldots a_{i-1} a a_{i+1}\ldots a_n}_{b_1\ldots b_m} \partial_a X^{a_i} + \sum_{j=1}^{m} T^{a_1\ldots a_n}_{b_1\ldots b_{i-1} a b_{i+1}\ldots b_m} \partial_{b_j} X^a \right) e^{b_1\ldots b_m}_{a_1\ldots a_n},\end{aligned}$$

allowing us to identify

$$(\mathcal{L}_X T)^{a_1\ldots a_n}_{b_1\ldots b_m} = X^a \partial_a T^{a_1\ldots a_n}_{b_1\ldots b_m} - \sum_{i=1}^{n} T^{a_1\ldots a_{i-1} a a_{i+1}\ldots a_n}_{b_1\ldots b_m} \partial_a X^{a_i} + \sum_{j=1}^{m} T^{a_1\ldots a_n}_{b_1\ldots b_{i-1} a b_{i+1}\ldots b_m} \partial_{b_j} X^a.$$

Connections A connection is an operator on a tensor space that satisfies the following:

- $\vec{\boldsymbol{\nabla}}_X f = Xf = X^a \partial_a f$ for a scalar field f.
- $\vec{\boldsymbol{\nabla}}_{X+Y} T = \vec{\boldsymbol{\nabla}}_X T + \vec{\boldsymbol{\nabla}}_Y T$.
- $\vec{\boldsymbol{\nabla}}_{fX} T = f \vec{\boldsymbol{\nabla}}_X T$.
- $\vec{\boldsymbol{\nabla}}_X (TS) = (\vec{\boldsymbol{\nabla}}_X T) S + T \vec{\boldsymbol{\nabla}}_X S$.

Connections on Manifolds On a manifold, a connection is specified by choosing n independent vectors X_i and defining

$$\vec{\boldsymbol{\nabla}}_{X_i} X_j = \Gamma^k_{ij} X_k,$$

where the expansion coefficients are called connection coefficients or Christoffel symbols. There is no unique way to do this, as the connection then depends on the choice of vectors.

The Connection of a Tensor Field Specify the connection according to $\vec{\boldsymbol{\nabla}}_{\partial_a} \partial_b = \Gamma^c_{ab} \partial_c$. For a tensor

$$T = T^{a_1\ldots a_n}_{b_1\ldots b_m} e^{b_1\ldots b_m}_{a_1\ldots a_n}$$

we find using the product rule

$$\vec{\boldsymbol{\nabla}}_X T = (\vec{\boldsymbol{\nabla}}_X T^{a_1\ldots a_n}_{b_1\ldots b_m}) e^{b_1\ldots b_m}_{a_1\ldots a_n} + T^{a_1\ldots a_n}_{b_1\ldots b_m} \sum_{i=1}^{n} e^{b_1\ldots b_m}_{a_1\ldots a_{i-1}} \otimes \vec{\boldsymbol{\nabla}}_X e_{a_i} \otimes e_{a_{i+1}\ldots a_n} + T^{a_1\ldots a_n}_{b_1\ldots b_m} \sum_{j=1}^{m} e^{b_1\ldots b_{j-1}} \otimes \vec{\boldsymbol{\nabla}}_X e^{b_j} \otimes e^{b_{j+1}\ldots b_m}_{a_1\ldots a_n}.$$

To proceed we will need to require that the connection respect contraction, yielding

$$\begin{aligned}\vec{\boldsymbol{\nabla}}_X \omega(Y) &= \vec{\boldsymbol{\nabla}}_X \omega(Y) + \omega(\vec{\boldsymbol{\nabla}}_X Y) \\ &= (\vec{\boldsymbol{\nabla}}_X \omega)_a Y^a + \omega_a (X^b \vec{\boldsymbol{\nabla}}_b (Y^c \partial_c))^a \\ &= (X^b \vec{\boldsymbol{\nabla}}_b \omega)_a Y^a + \omega_a (X^b (\partial_b Y^c \partial_c + Y^c \Gamma^d_{bc} \partial_d))^a \\ &= X^b (\vec{\boldsymbol{\nabla}}_b \omega)_a Y^a + \omega_a X^b \partial_b Y^a + \omega_a X^b Y^c \Gamma^a_{bc}.\end{aligned}$$

On the other side we have

$$\vec{\boldsymbol{\nabla}}_X \omega(Y) = X^a (Y^b \partial_a \omega_b + \omega_b \partial_a Y^b),$$

implying

$$\begin{aligned}Y^a X^b (\vec{\boldsymbol{\nabla}}_b \omega)_a &= X^a (Y^b \partial_a \omega_b + \omega_b \partial_a Y^b) - \omega_a X^b \partial_b Y^a - \omega_a X^b Y^c \Gamma^a_{bc} \\ &= Y^a X^b \partial_b \omega_a - Y^a X^b \omega_c \Gamma^c_{ba},\end{aligned}$$

and as this must be true for all X and Y we have

$$\vec{\boldsymbol{\nabla}}_b\omega = (\partial_b\omega_a - \omega_c\Gamma^c_{ba})d\chi^a.$$

We now proceed according to

$$\begin{aligned}\vec{\boldsymbol{\nabla}}_X T =& X^a\partial_a T^{a_1\ldots a_n}_{b_1\ldots b_m} e^{b_1\ldots b_m}_{a_1\ldots a_n} + T^{a_1\ldots a_n}_{b_1\ldots b_m}\sum_{i=1}^{n} X^a\Gamma^b_{aa_i} e^{b_1\ldots b_m}_{a_1\ldots a_{i-1}}\otimes e_b\otimes e_{a_{i+1}\ldots a_n} - T^{a_1\ldots a_n}_{b_1\ldots b_m}\sum_{j=1}^{m} X^a\Gamma^{b_j}_{ab} e^{b_1\ldots b_{j-1}}\otimes e^b\otimes e^{b_{j+1}\ldots b_m}_{a_1\ldots a_n}\\ =& X^a\left(\partial_a T^{a_1\ldots a_n}_{b_1\ldots b_m} + \sum_{i=1}^{n}\Gamma^{a_i}_{ab}T^{a_1\ldots a_{i-1}ba_{i+1}\ldots a_n}_{b_1\ldots b_m} - \sum_{j=1}^{m}\Gamma^b_{ab_j}T^{a_1\ldots a_n}_{b_1\ldots b_{j-1}bb_{j+1}\ldots b_m}\right)e^{b_1\ldots b_m}_{a_1\ldots a_n},\end{aligned}$$

and we recognize

$$(\vec{\boldsymbol{\nabla}}_X T)^{a_1\ldots a_n}_{b_1\ldots b_m} = X^a\left(\partial_a T^{a_1\ldots a_n}_{b_1\ldots b_m} + \sum_{i=1}^{n}\Gamma^{a_i}_{ab}T^{a_1\ldots a_{i-1}ba_{i+1}\ldots a_n}_{b_1\ldots b_m} - \sum_{j=1}^{m}\Gamma^b_{ab_j}T^{a_1\ldots a_n}_{b_1\ldots b_{j-1}bb_{j+1}\ldots b_m}\right).$$

Parallel Transport A vector X is termed parallel if $\vec{\boldsymbol{\nabla}}_a X = 0$. This system is overdetermined, and generally has no solution on a manifold. We may, however, define X to be parallel along a curve γ if

$$\vec{\boldsymbol{\nabla}}_{\dot\gamma} X = 0.$$

This allows us to define the parallel transport as the vector field that solves the above equation with the vector X as its initial condition. This system of equations is solvable.

In particular, using the properties of the connection we find

$$\begin{aligned}\vec{\boldsymbol{\nabla}}_{\dot\gamma}X &= \vec{\boldsymbol{\nabla}}_{\dot\chi^a\partial_a}X^c\partial_c\\ &= \dot\chi^a\vec{\boldsymbol{\nabla}}_a X^c\partial_c\\ &= \dot\chi^a((\vec{\boldsymbol{\nabla}}_a X^c)\partial_c + X^c\vec{\boldsymbol{\nabla}}_a\partial_c)\\ &= \dot\chi^a(\partial_a X^b + X^c\Gamma^b_{ac})\partial_b.\end{aligned}$$

Geodesics and the Geodesic Equation A geodesic is defined as a curve with a tangent vector that is parallel along itself.

By definition a geodesic satisfies

$$\vec{\boldsymbol{\nabla}}_{\dot\gamma}\dot\gamma = \dot\chi^a(\partial_a\dot\chi^b + \dot\chi^c\Gamma^b_{ac})\partial_b = (\ddot\chi^b + \dot\chi^a\dot\chi^c\Gamma^b_{ac})\partial_b = 0.$$

This is the geodesic equation. Given a starting point and a tangent vector, it is solvable.

Torsion The torsion tensor is a $(1,2)$ tensor defined as

$$T(X,Y) = \vec{\boldsymbol{\nabla}}_X Y - \vec{\boldsymbol{\nabla}}_Y X - [X,Y].$$

To find its components, we note that

$$\begin{aligned}T(\partial_a,\partial_b) &= \vec{\boldsymbol{\nabla}}_a\partial_b - \vec{\boldsymbol{\nabla}}_b\partial_a - [\partial_X a, \partial_Y b]\\ &= (\Gamma^c_{ab} - \Gamma^c_{ba})\partial_c.\end{aligned}$$

Curvature Consider some vector Z parallel transported along a small closed loop. The parallel transport is linear, so the result of this process must be connected to some $(1,1)$ tensor. Supposing that the loop is spanned by X and Y, we have

$$Z' - Z = R(X,Y)Z = \vec{\boldsymbol{\nabla}}_X\vec{\boldsymbol{\nabla}}_Y Z - \vec{\boldsymbol{\nabla}}_Y\vec{\boldsymbol{\nabla}}_X Z - \vec{\boldsymbol{\nabla}}_{[X,Y]}Z.$$

We define $R(X,Y)Z$ as the Riemann curvature tensor. It is a $(1,3)$ tensor. Its components are defined by

$$\begin{aligned} R(\partial_a,\partial_b)\partial_c &= R^d{}_{cab}\partial_d \\ &= \vec{\boldsymbol{\nabla}}_a\vec{\boldsymbol{\nabla}}_b\partial_c - \vec{\boldsymbol{\nabla}}_b\vec{\boldsymbol{\nabla}}_a\partial_c \\ &= \vec{\boldsymbol{\nabla}}_a\Gamma^f_{bc}\partial_f - \vec{\boldsymbol{\nabla}}_b\Gamma^f_{ac}\partial_f \\ &= (\vec{\boldsymbol{\nabla}}_a\Gamma^f_{bc})\partial_f + \Gamma^f_{bc}\vec{\boldsymbol{\nabla}}_a\partial_f - (\vec{\boldsymbol{\nabla}}_b\Gamma^f_{ac})\partial_f - \Gamma^f_{ac}\vec{\boldsymbol{\nabla}}_b\partial_f \\ &= (\partial_a\Gamma^d_{bc} + \Gamma^f_{bc}\Gamma^d_{af} - \partial_b\Gamma^d_{ac} - \Gamma^f_{ac}\Gamma^d_{bf})\partial_d, \end{aligned}$$

and thus

$$R^d{}_{cab} = \partial_a\Gamma^d_{bc} - \partial_b\Gamma^d_{ac} + \Gamma^f_{bc}\Gamma^d_{af} - \Gamma^f_{ac}\Gamma^d_{bf}.$$

Note the placements of the indices.

The Metric Tensor The metric tensor will be taken as the $(0,2)$ tensor that defines inner products on manifolds. The inner product, and therefore also the metric tensor, is a map from $T_pM \times T_pM$ that is symmetric and positive definite. Using this we may extend more of the previously performed work, for instance on curve length.

Metric Compatibility A connection is metric compatible if $\vec{\boldsymbol{\nabla}}_X g = 0$ for all vectors X.

The Geodesic Equation Revisited As the metric defines length, we define the curve length as

$$l_\gamma = \int_\gamma \mathrm{d}s = \int_0^1 \mathrm{d}t\,\sqrt{g_{ab}\dot{\chi}^a\dot{\chi}^b}.$$

Defining $\sqrt{\mathcal{L}} = g_{ab}\dot{\chi}^a\dot{\chi}^b$, the curve that minimizes the distance between the start and end points satisfies

$$\partial_a\sqrt{\mathcal{L}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\sqrt{\mathcal{L}}}{\partial\dot{\chi}^a} = \frac{1}{2\sqrt{\mathcal{L}}}\left(\partial_a\mathcal{L} - \sqrt{\mathcal{L}}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{\sqrt{\mathcal{L}}}\frac{\partial\mathcal{L}}{\partial\dot{\chi}^a}\right)\right) = 0.$$

One can always choose a parametrization such that $\sqrt{\mathcal{L}} = 0$ (the arc length parametrization is one example), yielding

$$\frac{1}{2\sqrt{\mathcal{L}}}\left(\partial_a\mathcal{L} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{\chi}^a}\right) = 0,$$

which is equivalent to extremizing the integral of $\mathcal{L}$. In terms of the coordinate functions we thus have

$$\partial_a g_{ab}\dot{\chi}^a\dot{\chi}^b - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{\chi}^a}$$

The Levi-Civita Connection For any manifold with some metric there exists a unique connection that is both metric compatible and torsion free. This connection is termed the Levi-Civita connection.

The Levi-Civita Connection and Geodesics The connection coefficients (or Christoffel symbols) defined by the Levi-Civita connection are symmetric due to the torsion being zero. This implies

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

The Induced Metric Given some immersion f of M_1 into M_2 and supposing that the metric g exists on M_2, this induces a metric $\tilde{g} = f^\star g$ on M_1.

Curvature and the Metric Using a Levi-Civita connection, we introduce the covariant curvature with components

$$R_{abcd} = \frac{1}{2}\left(\partial_a\partial_d g_{bc} + \partial_b\partial_c g_{ad} - \partial_a\partial_c g_{bd} - \partial_b\partial_d g_{ac}\right) + g_{fh}(\Gamma^f_{bc}\Gamma^h_{ad} - \Gamma^f_{bd}\Gamma^h_{ac}).$$

We find $R_{abcd} = R_{cdab} = -R_{abdc} = -R_{bacd}$. In addition, the curvature satisfies the Bianchi identity $R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$, implying $R_{abcd} + R_{acdb} + R_{adbc} = 0$, meaning that the total number of independent components is $\frac{1}{12}n^2(n^2-1)$.

The Ricci Tensor The Ricci tensor is defined as $R_{ab} = R^c{}_{acb}$.

The Ricci Scalar The Ricci scalar is defined as the trace of the Ricci scalar: $\mathcal{R} = g^{ab}R_{ab} = g^{ab}R^c{}_{acb}$.

The Einstein Tensor The Einstein tensor is defined as $G_{ab} = R_{ab} - \frac{1}{2}g_{ab}\mathcal{R}$.

Killing Fields K is a Killing field (be careful looking this up on the internet) if $\mathcal{L}_K g = 0$.

The Lie Derivative with Killing Fields Let K be a Killing field. We then obtain

$$\begin{aligned}
\mathcal{L}_K g_{ab} &= K^a\partial_c g_{ab} + g_{ac}\partial_b K^c + g_{cb}\partial_a K^c \\
&= K^d(\Gamma^c_{da}g_{cb} + \Gamma^c_{db}g_{ac}) + g_{ac}\partial_b K^c + g_{cb}\partial_a K^c \\
&= g_{ac}(\partial_b K^c + \Gamma^c_{db}K^d) + g_{cb}(\partial_a K^c + K^d\Gamma^c_{da}) \\
&= g_{ac}\vec{\boldsymbol{\nabla}}_b K^c + g_{cb}\vec{\boldsymbol{\nabla}}_a K^c \\
&= \vec{\boldsymbol{\nabla}}_b K_a + \vec{\boldsymbol{\nabla}}_a K_b = 0,
\end{aligned}$$

where we are now using the proper covariant derivative.

Differential Forms The set of p-forms, or differential forms, is the set of $(0,p)$ tensors that are completely antisymmetric. They are constructed using the wedge product, defined as

$$\bigwedge_{k=1}^{p} d\chi^{a_k} = \sum_{\sigma\in S_p}\mathrm{sgn}(\sigma)\bigotimes_{k=1}^{p} d\chi^{a_{\sigma(k)}}.$$

Here S_p is the set of permutations of p elements. There exists

$$n_p^N = \binom{N}{k}$$

basis elements. We note that the wedge product is antisymmetric under the exchange of two basis elements. Hence, once an ordering of indices has been chosen, any permutation will simply create a linearly dependent map.

Consider now some antisymmetric tensor ω. Introducing the antisymmetrizer

$$\bigotimes_{k=1}^{p} d\chi^{[a_{\sigma(k)}]} = \frac{1}{p!}\sum_{\sigma\in S_p}\mathrm{sgn}(\sigma)\bigotimes_{k=1}^{p} d\chi^{a_{\sigma(k)}},$$

the symmetry yields

$$\omega = \omega_{a_1\ldots a_p}\bigotimes_{k=1}^{p} d\chi^{a_{\sigma(k)}} = \omega_{a_1\ldots a_p}\bigotimes_{k=1}^{p} d\chi^{[a_{\sigma(k)}]} = \frac{1}{p!}\omega_{a_1\ldots a_p}\bigwedge_{k=1}^{p} d\chi^{a_k}.$$

The Exterior Derivative We define the exterior derivative of a differential form according to

$$d\omega = \frac{1}{p!}\partial_{a_1}\omega_{a_2\ldots a_{p+1}}\bigwedge_{k=1}^{p+1} d\chi^{a_k},$$

which is a $p+1$-form. This notation makes sense, as at least in the case of a 0-form, we obtain

$$d\omega = \partial_a\omega d\chi^a,$$

which is exactly the form of a dual vector. Somehow this transforms as a tensor.

Integration of Differential Forms Consider a set of p tangent vectors X_i. The corresponding coordinate displacements are $\mathrm{d}\chi_i^a = X_i^a\,\mathrm{d}t_i$, with no sum over i. We would now like to compute the p-dimensional volume defined by the X_i and $\mathrm{d}t_i$. We expect that if any of the X_i are linearly dependent the volume should be zero. We also expect that the volume be linear in the X_i. This implies

$$\mathrm{d}V_p = \omega(X_1, \ldots, X_p)\,\mathrm{d}t_1 \ldots \mathrm{d}t_p$$

for some differential form ω. We now define the integral over the p-volume S over the p-form ω as

$$\int\limits_S \omega = \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_p\, \omega(\dot\gamma_1, \ldots, \dot\gamma_p).$$

Here the γ_i are the set of curves that span S, the dot symbolizes the derivative with respect to the individual curve parameters and the right-hand integration is performed over the appropriate set of parameter values.

Stokes' Theorem Stokes' theorem relates the integral of a differential form $d\omega$ over some subset V of a manifold to an integral over ∂V of another differential form.

To derive it, consider a $p+1$-volume parametrized such that all t_i range from 0 to 1 and such that for any fixed t_{p+1}, the remaining t_i parametrize a p-dimensional surface V_p with a boundary independent of t_{p+1}. This construction is somewhat restrictive, but only necessary in the derivation.

For some p-form ω and $p+1$-volume V we have

$$\begin{aligned}
\int\limits_V d\omega &= \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1}\, d\omega(\dot\gamma_1, \ldots, \dot\gamma_{p+1}) \\
&= \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1} \left(\partial_{a_1} \omega_{a_2 \ldots a_{p+1}} \sum_{\sigma \in S_{p+1}} \mathrm{sgn}(\sigma) \bigotimes_{k=1}^{p+1} d\chi^{a_{\sigma(k)}} \right) (\dot\gamma_1, \ldots, \dot\gamma_{p+1}) \\
&= \sum_{\sigma \in S_{p+1}} \frac{\mathrm{sgn}(\sigma)}{p!} \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1} \left(\partial_{a_1} \omega_{a_2 \ldots a_{p+1}} \bigotimes_{k=1}^{p+1} d\chi^{a_{\sigma(k)}} \right) (\dot\gamma_1, \ldots, \dot\gamma_{p+1}) \\
&= \sum_{\sigma \in S_{p+1}} \frac{\mathrm{sgn}(\sigma)}{p!} \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1} \left(\partial_{a_{p+1}} \omega_{a_1 \ldots a_p} \bigotimes_{k=1}^{p+1} d\chi^{a_{\sigma(k)}} \right) (\dot\gamma_1, \ldots, \dot\gamma_{p+1}),
\end{aligned}$$

where the latter follows from the cyclicity imposed by the summation. We have

$$d\chi^{a_{\sigma(k)}} \dot\gamma_k = \frac{\mathrm{d}\chi^a}{\mathrm{d}t_k} \partial_a \chi^{a_{\sigma(k)}} = \frac{\mathrm{d}\chi^{a_{\sigma(k)}}}{\mathrm{d}t_k},$$

and thus

$$\begin{aligned}
\int\limits_V d\omega &= \sum_{\sigma \in S_{p+1}} \frac{\mathrm{sgn}(\sigma)}{p!} \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1}\, \partial_{a_{p+1}} \omega_{a_1 \ldots a_p} \prod_{k=1}^{p+1} \partial_{t_k} \chi^{a_{\sigma(k)}} \\
&= \sum_{\sigma \in S_{p+1}} \frac{\mathrm{sgn}(\sigma)}{p!} \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1}\, \partial_{a_{p+1}} \omega_{a_1 \ldots a_p} \prod_{k=1}^{p+1} \partial_{t_{\sigma(k)}} \chi^{a_k},
\end{aligned}$$

where we have once again utilized the cyclicity. Denote the integral inside the sum as $I(\sigma, \omega)$. We have

$$\begin{aligned}
I(\sigma, \omega) &= \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1}\, \partial_{a_{p+1}} \omega_{a_1 \ldots a_p} \prod_{k=1}^{p+1} \partial_{t_{\sigma(k)}} \chi^{a_k} \\
&= \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1}\, \partial_{t_{\sigma(p+1)}} \omega_{a_1 \ldots a_p} \prod_{k=1}^{p} \partial_{t_{\sigma(k)}} \chi^{a_k}.
\end{aligned}$$

To proceed, we integrate by parts and obtain

$$I(\sigma, \omega) = \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1}\, \omega_{a_1 \ldots a_p} \prod_{k=1}^{p} \partial_{t_{\sigma(k)}} \chi^{a_k} \Bigg|_{t_{\sigma(p+1)} = t^-_{\sigma(p+1)}}^{t_{\sigma(p+1)} = t^+_{\sigma(p+1)}} - \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_{p+1}\, \omega_{a_1 \ldots a_p} \partial_{t_{\sigma(p+1)}} \prod_{k=1}^{p} \partial_{t_{\sigma(k)}} \chi^{a_k},$$

where the former terms contains no integration over $t_{\sigma(p+1)}$, and is instead evaluated at the maximal and minimal values of $t_{\sigma(p+1)}$ given the values of the other parameters.

Let us proceed to simplify this. For starters, if $\sigma(p+1) \neq p+1$, the remaining integration in the first term is done over the boundary of V_p. Furthermore, there exists a k such that $\sigma(k) = p+1$, and as we are at the boundary of V_p we must have

$$\partial_{t_{\sigma(k)}} \chi^{a_k} = 0.$$

Otherwise, the remaining integration domain is V_p. Next, the latter integral contains a sequence of terms proportional to

$$\partial_{t_{\sigma(p+1)}} \partial_{t_{\sigma(k)}} \chi^{a_k},$$

which is symmetric with respect to exchanging $p+1$ and k. These terms therefore cancel in the sum, and we are left with

$$\begin{aligned}\int_V d\omega &= \sum_{\sigma \in S_p} \frac{\mathrm{sgn}(\sigma)}{p!} \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_p \, \omega_{a_1 \ldots a_p} \prod_{k=1}^{p} \partial_{t_{\sigma(k)}} \chi^{a_k} \Bigg|_{t_{\sigma(p+1)}=0}^{t_{\sigma(p+1)}=1} \\ &= \int \mathrm{d}t_1 \ldots \int \mathrm{d}t_p \, \omega \Bigg|_{t_{\sigma(p+1)}=0}^{t_{\sigma(p+1)}=1}\end{aligned}$$

where the condition that $\sigma(p+1) = p+1$ restricts the summation to S_p.

The remaining integration domain is, as stated before, a parametrization of V_p, which at the extremal values of t_{p+1} must be at the boundary of V. The minus sign from the integral evaluation tells us that the integrals are taken with opposite orientation, meaning that together they indeed form an integration over the (closed) boundary of V. We thus arrive at Stokes' theorem,

$$\int_V d\omega = \oint_{\partial V} \omega.$$

Example: Reobtaining Familiar Theorems Many familiar integration theorems are in fact consequences of Stokes' theorem. Let us rederive them.

We start with a 1-form $d\omega$, which will be integrated over a 1-dimensional volume, i.e. a curve. We have

$$\int_\gamma d\omega = \omega(p) - \omega(q),$$

where q and p are the start and end points of γ. This is the analogue of integrating a vector field along a curve.

Next, we consider a 2-form written as the exterior derivative of a 1-form. Writing $\omega = \omega_i d\chi^i$ we have

$$d\omega = \partial_j \omega_i d\chi^j \wedge d\chi^i.$$

Using Cartesian coordinates and restricting ourselves to two dimensions we have

$$\begin{aligned}d\chi^j \wedge d\chi^i(\dot\gamma_s, \dot\gamma_t)\,\mathrm{d}s\,\mathrm{d}t &= (d\chi^j \otimes d\chi^i - d\chi^i \otimes d\chi^j)(\dot\gamma_s, \dot\gamma_t)\,\mathrm{d}s\,\mathrm{d}t \\ &= (\partial_s \chi^j \partial_t \chi^i - \partial_s \chi^i \partial_t \chi^j)\,\mathrm{d}s\,\mathrm{d}t \\ &= (\delta_{jk}\delta_{im} - \delta_{ik}\delta_{jm})\partial_s \chi^k \partial_t \chi^m \,\mathrm{d}s\,\mathrm{d}t \\ &= \varepsilon_{jin}\varepsilon_{nkm}\partial_s \chi^k \partial_t \chi^m \,\mathrm{d}s\,\mathrm{d}t \\ &= \varepsilon_{jik}\,\mathrm{d}S_k\,.\end{aligned}$$

Thus we have

$$\int_S d\omega = \int_S \mathrm{d}S_k\, \varepsilon_{ijk} \partial_i \omega_j = \int_S \mathrm{d}\mathbf{S} \cdot \vec{\boldsymbol{\nabla}} \times \boldsymbol{\omega}.$$

At the same time we have

$$\begin{aligned}\int_{\partial S} \omega &= \int_{\partial S} \mathrm{d}t\, \omega_a d\chi^a(\dot{\gamma}_t) \\ &= \int_{\partial S} \mathrm{d}t\, \omega_a \frac{\mathrm{d}\chi^a}{\mathrm{d}t} \\ &= \int_{\partial S} \mathrm{d}\mathbf{r} \cdot \boldsymbol{\omega},\end{aligned}$$

hence

$$\int_S \mathrm{d}\mathbf{S} \cdot \vec{\boldsymbol{\nabla}} \times \boldsymbol{\omega} = \int_{\partial S} \mathrm{d}\mathbf{r} \cdot \boldsymbol{\omega},$$

which is the more boring version of Stokes' theorem.

Next we consider a 2-form and its exterior derivative. We have

$$\int_V d\omega = \frac{1}{2} \int_V \partial_a \omega_{bc} d\chi^a \wedge d\chi^b \wedge d\chi^c.$$

As the components of ω may be chosen such that $\omega_{ab} = -\omega_{ba}$ we may write $\omega_{ab} = \varepsilon_{abc}\omega_c$ for some suitable (and arbitrary) choice of ω_c. We thus have

$$\begin{aligned}\int_V d\omega &= \frac{1}{2} \int_V \varepsilon_{bcd} \partial_a \omega_d d\chi^a \wedge d\chi^b \wedge d\chi^c \\ &= \frac{1}{2} \int \mathrm{d}x^1 \int \mathrm{d}x^2 \int \mathrm{d}x^3\, \varepsilon_{bcd}\varepsilon_{abc} \partial_a \omega_d \\ &= \frac{1}{2} \int \mathrm{d}x^1 \int \mathrm{d}x^2 \int \mathrm{d}x^3\, (\delta_{ad}\delta_{bb} - \delta_{ab}\delta_{bd}) \partial_a \omega_d \\ &= \int \mathrm{d}x^1 \int \mathrm{d}x^2 \int \mathrm{d}x^3\, \partial_a \omega_a \\ &= \int \mathrm{d}x^1 \int \mathrm{d}x^2 \int \mathrm{d}x^3\, \vec{\boldsymbol{\nabla}} \cdot \boldsymbol{\omega}.\end{aligned}$$

At the same time we have

$$\begin{aligned}\int_{\partial V} \omega &= \frac{1}{2} \int_{\partial V} \varepsilon_{ijk} \omega_i d\chi^j \wedge d\chi^k \\ &= \frac{1}{2} \int_{\partial V} \mathrm{d}S_m\, \varepsilon_{ijk}\varepsilon_{jkm}\omega_i \\ &= \frac{1}{2} \int_{\partial V} \mathrm{d}S_m\, (\delta_{im}\delta_{jj} - \delta_{ij}\delta_{jm})\omega_i \\ &= \int_{\partial V} \mathrm{d}S_i\, \omega_i \\ &= \int_{\partial V} \mathrm{d}\mathbf{S} \cdot \boldsymbol{\omega}.\end{aligned}$$

Hence we have

$$\int \mathrm{d}x^1 \int \mathrm{d}x^2 \int \mathrm{d}x^3\, \vec{\boldsymbol{\nabla}} \cdot \boldsymbol{\omega} = \int_{\partial V} \mathrm{d}\mathbf{S} \cdot \boldsymbol{\omega},$$

which is the divergence theorem.

Example: The n-Dimensional Divergence Theorem Let us generalize the latter to an n-dimensional case. Consider a $n-1$-form and its exterior derivative. We have

$$\int_V d\omega = \frac{1}{(n-1)!} \int_V \partial_{a_1} \omega_{a_2 \dots a_n} \bigwedge_{k=1}^{n} d\chi^{a_k}.$$

We may once again choose coordinates such that the components of ω are completely antisymmetric, yielding

$$\begin{aligned}
\int_V d\omega &= \frac{1}{(n-1)!} \int_V \varepsilon_{a_2 \dots a_n a_{n+1}} \partial_{a_1} \omega_{a_{n+1}} \bigwedge_{k=1}^{n} d\chi^{a_k} \\
&= \frac{1}{(n-1)!} \int \mathrm{d}^n x \, \varepsilon_{a_2 \dots a_n a_{n+1}} \varepsilon_{a_1 \dots a_n} \partial_{a_1} \omega_{a_{n+1}} \\
&= \frac{1}{(n-1)!} \int \mathrm{d}^n x \, \varepsilon_{a_2 \dots a_{n+1}} \varepsilon_{a_2 \dots a_n a_1} \partial_{a_1} \omega_{a_{n+1}}.
\end{aligned}$$

Summing over all but the first and last index will give a factor $(n-1)!$, as this is the number of permutations of $n-1$ unique indices. The remaining product is non-zero only when the last indices are equal, and is in this case equal to 1, hence

$$\begin{aligned}
\int_V d\omega &= \int \mathrm{d}^n x \, \delta_{a_{n+1} a_1} \partial_{a_1} \omega_{a_{n+1}} \\
&= \int \mathrm{d}^n x \, \partial_a \omega_a.
\end{aligned}$$

At the same time we have

$$\begin{aligned}
\int_{\partial V} \omega &= \frac{1}{(n-1)!} \int_{\partial V} \varepsilon_{a_1 \dots a_n} \omega_{a_n} \bigwedge_{k=1}^{n-1} d\chi^{a_k} \\
&= \frac{1}{(n-1)!} \int_{\partial V} \mathrm{d}S_m \, \varepsilon_{a_1 \dots a_n} \varepsilon_{a_1 \dots a_{n-1} m} \omega_{a_n} \\
&= \int_{\partial V} \mathrm{d}S_a \, \omega_a.
\end{aligned}$$

Hence we have

$$\int \mathrm{d}^n x \, \partial_a \omega_a = \int_{\partial V} \mathrm{d}S_a \, \omega_a,$$

which is a generalization of the divergence theorem. Note that this only applies if the coordinates of the manifold are akin to Cartesian coordinates.

8 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 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, $\delta S = 0$.

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

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

Invariances of Lagrangian Dynamics Lagrangian dynamics are invariant if the Lagrangian is changed by a total time derivative. This is because such a term adds only two boundary terms to the action, which should not be affected by variations of the coordinates.

Lagrangian dynamics are also invariant under coordinate changes. To realize this, suppose that the Lagrangian is expressed in coordinates q^a such that the equations of motion are satisfied. Now write these as $q^a = q^a(Q^a, t)$. This implies

$$\begin{aligned}
\dot{q}^a &= \frac{\partial q^a}{\partial Q^b}\dot{Q}^b + \partial_t q^a, \\
\frac{\partial\mathcal{L}}{\partial q^a} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{q}^a} &= 0.
\end{aligned}$$

We would like to show that

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

Using the chain rule, we obtain

$$\begin{aligned}
&\frac{\partial\mathcal{L}}{\partial q^b}\frac{\partial q^b}{\partial Q^a} + \frac{\partial\mathcal{L}}{\partial\dot{q}^b}\frac{\partial\dot{q}^b}{\partial Q^a} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}^b}\frac{\partial\dot{q}^b}{\partial\dot{Q}^a}\right) = 0, \\
&\frac{\partial\mathcal{L}}{\partial q^b}\frac{\partial q^b}{\partial Q^a} + \frac{\partial\mathcal{L}}{\partial\dot{q}^b}\frac{\partial}{\partial Q^a}\left(\frac{\partial q^b}{\partial Q^c}\dot{Q}^c + \partial_t q^b\right) - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}^b}\frac{\partial q^b}{\partial Q^a}\right) = 0, \\
&\frac{\partial\mathcal{L}}{\partial q^b}\frac{\partial q^b}{\partial Q^a} + \frac{\partial\mathcal{L}}{\partial\dot{q}^b}\left(\frac{\partial^2 q^b}{\partial Q^a\partial Q^c}\dot{Q}^c + \frac{\partial\partial_t q^b}{\partial Q^a}\right) - \frac{\partial\mathcal{L}}{\partial\dot{q}^b}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial q^b}{\partial Q^a} - \frac{\partial q^b}{\partial Q^a}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{q}^b} = 0, \\
&\frac{\partial\mathcal{L}}{\partial q^b}\frac{\partial q^b}{\partial Q^a} + \frac{\partial\mathcal{L}}{\partial\dot{q}^b}\left(\frac{\partial^2 q^b}{\partial Q^a\partial Q^c}\dot{Q}^c + \frac{\partial\partial_t q^b}{\partial Q^a}\right) - \frac{\partial q^b}{\partial Q^a}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{q}^b} - \frac{\partial\mathcal{L}}{\partial\dot{q}^b}\left(\frac{\partial^2 q^b}{\partial Q^a\partial Q^c}\dot{Q}^c + \partial_t\frac{\partial q^b}{\partial Q^{ax}}\right) = 0, \\
&\frac{\partial q^b}{\partial Q^a}\left(\frac{\partial\mathcal{L}}{\partial q^b} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{q}^b}\right) = 0.
\end{aligned}$$

As the coordinate change is invertible, the equations of motion are still satisfied.

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 has velocity dependence, 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 the other way round. 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

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

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

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

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

Great - we did it!

As a side note, how does this Lagrangian change under a gauge transform? Such a transform adds a partial time derivative of a function $-f$ to the electric potential and the gradient f to the magnetic potential. Under such a transform we have

$$\begin{aligned} \mathcal{L}' &= \frac{1}{2}mv^2 - q(\phi - \partial_t f - \mathbf{v}\cdot(\mathbf{A} + \vec{\boldsymbol{\nabla}}f))\\ &= \mathcal{L} + q(\partial_t f + \mathbf{v}\cdot\vec{\boldsymbol{\nabla}}f)\\ &= \mathcal{L} + \frac{\mathrm{d}f}{\mathrm{d}t}. \end{aligned}$$

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

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

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

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

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

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

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 the Lagrangian does not change its functional form and the action is changed by some constant depending on the boundary conditions. Symmetries are the transformations such that this constant is 0.

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

To prove it, we must first convert the definitions we have to local form. The new coordinates are given by $Q_i(\tau) = q_i(t) + \delta q_i(\tau)$, and thus mix up the time, which we have to fix. By the definition of a quasi-symmetry, we have

$$\int_{t_1+\delta t_1}^{t_2+\delta t_2} \mathrm{d}\tau\, \mathcal{L}(Q_i,\tau) - \int_{t_1}^{t_2} \mathrm{d}t\, \mathcal{L}(q_i,t) = \int_{t_1}^{t_2} \mathrm{d}t\, \frac{\mathrm{d}F}{\mathrm{d}t},$$

for some function F. This is of the form

$$\int_{a+\delta a}^{b+\delta b} \mathrm{d}x'\, f + \delta f - \int_a^b \mathrm{d}x\, f = c.$$

This can be rewritten as

$$\int_{a+\delta a}^{b+\delta b} \mathrm{d}x\, \delta f + \int_b^{b+\delta b} \mathrm{d}x\, f - \int_a^{a+\delta a} \mathrm{d}x\, f = c.$$

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

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

Hence, we have

$$\int_{t_1}^{t_2} \mathrm{d}\tau\, \mathcal{L}(Q_i,t) - \mathcal{L}(q_i,t) + \frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{L}(q_i,t)\,\delta t) - F = 0.$$

This completes the conversion of time.

Next, the two first terms can be expanded to first order as

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

where we have introduced $\bar{\delta}q^a = \delta q^a(t)$. Using the equations of motion, we further have

$$\mathcal{L}(Q_i,t) - \mathcal{L}(q_i,t) = \bar{\delta}q^a \frac{\mathrm{d}}{\mathrm{d}t}\partial_{\dot{q}^a}\mathcal{L} + \partial_{\dot{q}^a}\mathcal{L}\frac{\mathrm{d}}{\mathrm{d}t}\bar{\delta}q^a = \frac{\mathrm{d}}{\mathrm{d}t}\left(\bar{\delta}q^a \partial_{\dot{q}^a}\mathcal{L}\right).$$

Hence we have

$$\int_{t_1}^{t_2} \mathrm{d}\tau \, \frac{\mathrm{d}}{\mathrm{d}t} \left(\mathcal{L}(q_i, t) \, \delta t + \bar{\delta} q^a \partial_{\dot{q}^a} \mathcal{L} \right) - F = 0.$$

Finally, we have to first order

$$\delta q^a = \bar{\delta} q^a + \dot{q^a} \, \delta t \, ,$$

yielding

$$\int_{t_1}^{t_2} \mathrm{d}\tau \, \frac{\mathrm{d}}{\mathrm{d}t} \left(\mathcal{L} \, \delta t + (\delta q^a - \dot{q^a} \, \delta t) \partial_{\dot{q}^a} \mathcal{L} - F \right) = 0.$$

The time interval in consideration may now be taken to be arbitrarily small, hence we must have that

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

is a conserved quantity for this system. We can introduce the general momenta and the Hamiltonian to rewrite this as

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

This is the statement of Nöether's theorem. 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.

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 Consider a particle in a potential $V = V(\mathbf{x} - \mathbf{v}t)$, where $\mathbf{v}$ is constant. The symmetry $\delta\mathbf{x} = \mathbf{v}\,\delta t$ with δt as a constant seems like it should be a symmetry of such a system. The transformed action is

$$\begin{aligned}
S' &= \int_{t_0+\delta t}^{t_1+\delta t} \mathrm{d}t'\, \frac{1}{2} m(\dot{\mathbf{x}}')^2 - V(\mathbf{x}' - \mathbf{v}t') \\
&= \int_{t_0+\delta t}^{t_1+\delta t} \mathrm{d}t'\, \frac{1}{2} m\dot{\mathbf{x}}^2 - V(\mathbf{x} + \mathbf{v}\,\delta t - \mathbf{v}(t+\delta t)) \\
&= \int_{t_0}^{t_1} \mathrm{d}t\, \frac{1}{2} m\dot{\mathbf{x}}^2 - V(\mathbf{x} - \mathbf{v}t) \\
&= S,
\end{aligned}$$

hence the transformation is indeed a symmetry of the system. The corresponding conserved quantity is

$$\begin{aligned}
J &= \mathcal{H}\,\delta t - p_i\,\delta x_i \\
&= \left(\frac{1}{2} m\dot{\mathbf{x}}^2 + V(\mathbf{x} - \mathbf{v}t)\right)\delta t - m\dot{\mathbf{x}}\cdot\mathbf{v}\,\delta t \\
&= \left(\frac{1}{2} m(\dot{\mathbf{x}} - \mathbf{v})^2 + V(\mathbf{x} - \mathbf{v}t) - \frac{1}{2} m\mathbf{v}^2\right)\delta t\,,
\end{aligned}$$

which is equal to the total energy in the rest frame of the potential plus an extra term due to the Galilei boost relative to the lab frame.

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

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

Hence there is an equivalence between G being a quasi-symmetry of the Hamiltonian and being conserved.

Example: The Central Force Problem Consider a particle in two dimensions moving under the influence of a central force. The Lagrangian is

$$\mathcal{L} = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\phi}^2\right) - V(r).$$

The transformation $\delta\phi = 1$ is a symmetry of the Lagrangian, with corresponding conserved quantity

$$J = -p_\phi.$$

The symmetry generated by p_ϕ is indeed $\delta\phi = 1$, demonstrating the equivalence.

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

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

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

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

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

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

Solving this for the derivative yields

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

X is now given by integrating. The action is thus

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

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

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

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

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

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

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

This set of coordinates and momenta is called action-angle variables.

Action-Angle Variables The action-angle variables of a system is the set of generalized coordinates θ^i and generalized momenta I_i such that the Hamiltonian only depends on the generalized momenta. The claim is that the generalized momenta are

$$I_i = \frac{1}{2\pi} \int_{\gamma_i} \mathrm{d}q^j \, p_j,$$

where the integration is performed twice over the set of available coordinates in phase space, i.e. over the period of the motion.

Example: The Harmonic Oscillator Consider a harmonic oscillator with energy E. In this case we have the one action-angle momentum

$$I = \frac{1}{2\pi}\int \mathrm{d}q\, p.$$

The momentum can be re-expressed in terms of the coordinate to obtain

$$I = \frac{1}{2\pi}\int \mathrm{d}q \sqrt{2m\left(E - \frac{1}{2}m\omega^2 q^2\right)}.$$

The extremal values of the coordinate are $\pm\sqrt{\frac{2E}{m\omega^2}}$, yielding

$$\begin{aligned}
I &= \frac{1}{\pi}\int_{-\sqrt{\frac{2E}{m\omega^2}}}^{\sqrt{\frac{2E}{m\omega^2}}} \mathrm{d}q \sqrt{2m\left(E - \frac{1}{2}m\omega^2 q^2\right)} \\
&= \frac{\sqrt{2mE}}{\pi}\int_{-\sqrt{\frac{2E}{m\omega^2}}}^{\sqrt{\frac{2E}{m\omega^2}}} \mathrm{d}q \sqrt{1 - \frac{m\omega^2}{2E}q^2} \\
&= \frac{\sqrt{2mE}}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}x \sqrt{\frac{2E}{m\omega^2}}\cos x\sqrt{1-\sin^2(x)} \\
&= \frac{2E}{\pi\omega}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}x \cos^2(x) \\
&= \frac{E}{\pi\omega}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}x\, 1 + \cos 2x \\
&= \frac{E}{\omega}.
\end{aligned}$$

In other words, the Hamiltonian in the new variables is $H = \omega I$, and the frequency of the corresponding coordinate is simply ω.

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

9 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 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

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

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.

10 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

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

We will also later be using the metric g, which is diagonal in spacetime and has a value of 1 for time and -1 for space.

Example: A String A string has the Lagrangian density

$$\mathcal{L} = \frac{1}{2}\left(\rho(\partial_t\phi)^2 - \frac{1}{2}T(\vec{\boldsymbol{\nabla}}\phi)^2\right).$$

The equation of motion is

$$0 - \rho\partial_t^2\phi + T\partial_i^2\phi = 0,$$

which is the wave equation.

Example: Klein-Gordon Theory The Klein-Gordon Lagrangian density is

$$\mathcal{L} = \frac{1}{2}\left(\frac{1}{c^2}\rho(\partial_t\phi)^2 - (\vec{\boldsymbol{\nabla}}\phi)^2 - m^2\phi^2\right) = .$$

The equation of motion is

$$-m^2\phi - \frac{1}{c^2}\partial_t^2\phi + \partial_i^2\phi = 0,$$

or alternatively

$$\partial_\mu\partial^\mu\phi + m^2\phi = 0.$$

Example: Galilei Non-Invariance of the Wave Equation While the wave equation appears in many contexts in classical physics, it turns out that it is generally not invariant under Galilei boosts. Such a boost is of the form $\delta\phi = 0,\ \delta x^\mu = \frac{v}{c}x^0\delta^\mu_1$. The new derivatives are

$$\partial_0 = (\partial')_0 - \frac{v}{c}(\partial')_1,\ \partial_i = (\partial')_i,$$

with squares

$$\partial_0\partial^0 = (\partial')_0(\partial')^0 + \left(\frac{v}{c}\right)^2(\partial')_1(\partial')^1 - 2\frac{v}{c}(\partial')_0\partial^1,\ \partial_i\partial^i = (\partial')_i(\partial')^i.$$

The wave equation now reads

$$\begin{aligned}\partial_\mu\partial^\mu\phi &= \left((\partial')_0(\partial')^0 + \left(\frac{v}{c}\right)^2(\partial')_1\partial^1 - 2\frac{v}{c}(\partial')_0(\partial')^1 + (\partial')_i(\partial')^i\right)\phi \\ &= (\partial')_\mu(\partial')^\mu\phi + \frac{v}{c}\left(\frac{v}{c}(\partial')_1\partial^1 - 2(\partial')_0(\partial')^1\right)\phi,\end{aligned}$$

which is generally different. The wave equation itself thus implies the existence of a special frame of reference in which it is satisfied, namely the rest frame of the medium.

Example: Complex Klein-Gordon Theory Klein-Gordon theory can be extended to complex fields by considering the original field ϕ and its complex conjugate as two separate fields. The Lagrangian density is

$$\mathcal{L} = \partial_\mu\phi\partial^\mu\phi^* - V(\phi^*\phi),\ V(\alpha) = m^2\alpha.$$

Example: The Schrödinger Equation Wave mechanics can be formulated as a field theory with the Lagrangian density

$$\mathcal{L} = \frac{\hbar^2}{2m}\partial_i\psi^*\partial^i\psi + V\psi^*\psi + \frac{\hbar c}{2i}(\psi^*\partial_0\psi - \psi\partial_0\psi^*).$$

Example: The Electromagnetic Field We will try to formulate a Lagrangian density which reproduces Maxwell's equations as its equations of motion. To do this, we introduce the fields $A^0 = \frac{1}{c}\phi, A^i = A^i$, the metric $g_{\mu\nu}$, which is 1 if $\mu = \nu = 0$, -1 if μ and ν refer to spatial coordinates and 0 otherwise, the object j^μ, which is $c\rho$ for $\mu = 0$ and j^i otherwise, and the tensor

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

The Lagrangian density we propose as the correct one is

$$\mathcal{L} = -\frac{1}{4\mu_0}F_{\mu\nu}F^{\mu\nu} - A_\mu j^\mu.$$

To show that this is the correct one, we first use the fact that $g_{\mu\nu} = g^{\mu\nu}$ to obtain

$$F^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}.$$

Next we write the Lagrangian density as

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4\mu_0}g^{\mu\alpha}g^{\nu\beta}F_{\alpha\beta}F_{\mu\nu} - A_\mu j^\mu \\ &= -\frac{1}{4\mu_0}g^{\mu\alpha}g^{\nu\beta}(\partial_\alpha A_\beta - \partial_\beta A_\alpha)(\partial_\mu A_\nu - \partial_\nu A_\mu) - A_\mu j^\mu.\end{aligned}$$

To clear things up further, we should remember that in the formal sense, the derivatives of the fields are all independent variables, yielding

$$\frac{\partial(\partial_\alpha A_\beta)}{\partial(\partial_\nu A_\mu)} = \delta_{\alpha\nu}\delta_{\beta\mu}.$$

The equations of motion are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial A_\sigma} - \partial_\kappa \frac{\partial \mathcal{L}}{\partial(\partial_\kappa A_\sigma)} &= 0, \\
-j^\sigma + \partial_\kappa \frac{1}{4\mu_0} g^{\mu\alpha} g^{\nu\beta} \left((\delta_{\kappa\alpha}\delta_{\sigma\beta} - \delta_{\kappa\beta}\delta_{\sigma\alpha})(\partial_\mu A_\nu - \partial_\nu A_\mu) + (\delta_{\kappa\mu}\delta_{\sigma\nu} - \delta_{\kappa\nu}\delta_{\sigma\mu})(\partial_\alpha A_\beta - \partial_\beta A_\alpha) \right) &= 0, \\
-j^\sigma + \partial_\kappa \frac{1}{4\mu_0} \left((g^{\mu\kappa} g^{\nu\sigma} - g^{\mu\sigma} g^{\nu\kappa}) F_{\mu\nu} + (g^{\kappa\alpha} g^{\sigma\beta} - g^{\sigma\alpha} g^{\kappa\beta}) F_{\alpha\beta} \right) &= 0, \\
-j^\sigma + \partial_\kappa \frac{1}{4\mu_0} \left((F^{\kappa\sigma} - F^{\sigma\kappa}) + (g^{\kappa\alpha} g^{\sigma\beta} F_{\kappa\sigma} - g^{\sigma\alpha} g^{\kappa\beta} F_{\alpha\beta}) \right) &= 0, \\
\partial_\kappa F^{\kappa\sigma} &= \mu_0 j^\sigma.
\end{aligned}$$

Let's examine the equations of motion more carefully. The first is

$$\partial_\kappa F^{\kappa 0} = \mu_0 j^0 = \frac{1}{\varepsilon_0 c}\rho,$$

which is a Maxwell equation if $F^{i0} = \frac{1}{c}E^i$. Let's try to show that. We have

$$\begin{aligned}
F^{i0} &= g^{i\alpha} g^{0\beta} (\partial_\alpha A_\beta - \partial_\beta A_\alpha) \\
&= g^{i\alpha} (\frac{1}{c}\partial_\alpha \phi - \partial_0 A_\alpha) \\
&= \frac{1}{c}\partial^i \phi - \partial_0 A^i \\
&= \frac{1}{c}\left(\partial^i \phi - \partial_t A^i \right) \\
&= \frac{1}{c} E^i.
\end{aligned}$$

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

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

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

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.

Example: A String

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 transformation changes the action by a constant value

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

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

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

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

Note that the second requirement in the definition implies that

$$\delta\mathcal{L} = \partial_\mu V^\mu.$$

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 on the action can be written as

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

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

$$\int_{\Omega'} \mathrm{d}x^\mu\,\mathcal{L}((\phi')^a, x^\mu) - \int_\Omega \mathrm{d}x^\mu\,\mathcal{L}(\phi^a, x^\mu) = \int_\Omega \mathrm{d}x^\mu\,\partial_\nu V^\nu.$$

We reuse the argument from the previous proof, generalizing it (somehow) to multiple dimensions to obtain

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

and thus

$$\int_\Omega \mathrm{d}x^\mu\,\mathcal{L}((\phi')^a, x^\mu) - \mathcal{L}(\phi^a, x^\mu) + \frac{\partial}{\partial x^\nu}(\mathcal{L}(\phi^a, x^\mu)\,\delta x^\nu) = \int_\Omega \mathrm{d}x^\mu\,\partial_\nu V^\nu,$$

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

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

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

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

and the corresponding currents

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

Next, we use the expansion

$$\delta\phi^a = \bar{\delta}\phi^a + \partial_\mu \phi^a \, \delta x^\mu$$

to obtain

$$\int_\Omega \mathrm{d}x^\mu \, \frac{\partial}{\partial x^\nu} \left((\delta\phi^a - \partial_\mu \phi^a \, \delta x^\mu) \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} + \mathcal{L}\, \delta x^\nu - V^\nu \right) = 0.$$

The argument extends arbitrarily to the interior of the integration domain, hence we must have

$$\partial_\nu J^\nu = 0,$$

where

$$J^\nu = V^\nu - \delta\phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} + \left(\partial_\mu \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_\mu \mathcal{L} \right) \delta x^\mu \,.$$

This is the final form of Nöether's theorem.

Example: Energy Conservation Consider a transformation where only time is varied (normalized to 1) acting on a Lagrangian density with no explicit time dependence. The variation of the Lagragian is zero under this transformation as it has no time dependence. Nöether's theorem gives the currents as

$$J^\nu = \left(\partial_\mu \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_\mu \mathcal{L} \right) \delta^\mu_0 = \partial_0 \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_0 \mathcal{L}.$$

Its time component is

$$J^0 = \partial_i \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_0 \phi^a)} - \mathcal{L} = \mathcal{H},$$

hence this symmetry corresponds to the conservation of total energy.

Example: Momentum Conservation Consider a transformation where only a single coordinate is varied (normalized to 1) acting on a Lagrangian density with no explicit dependence on that coordinate. The variation of the Lagragian is zero under this transformation. The corresponding Nöether current is

$$J^\nu = \left(\partial_\mu \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_\mu \mathcal{L}\right) \delta^\mu_i = \partial_i \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_i \mathcal{L}.$$

Its time component is

$$J^0 = c\partial_i \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_t \phi^a)} = c\pi_a \partial_i \phi^a,$$

which can be shown to be a momentum density in the i-direction.

Example: The Energy-Momentum Tensor Consider a transformation akin to the ones above, where a single coordinate is varied and the Lagrangian is independent of that coordinate. Denoting its index as κ, the Nöether current is

$$J^\nu = \left(\partial_\mu \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_\mu \mathcal{L}\right) \delta^\mu_\kappa = \partial_\kappa \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_\kappa \mathcal{L} = \Theta^\nu_\kappa.$$

These are the components of the energy-momentum tensor. If the theory is invariant under space-time translations corresponding to index ν, it satisfies

$$\partial_\mu \Theta^\mu_\nu = 0$$

according to Nöether's theorem.

Example: Rotations Consider a theory subject to rotation about the z-axis by a small angle α. This rotation is described by $\delta x^i = \alpha \varepsilon^{3\ i}_{j}\, x^{\,j},\ \delta x^0 = 0$. If this is a symmetry of the Lagrangian, the corresponding Nöether current is

$$J^\nu = \left(\partial_\mu \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_\mu \mathcal{L}\right) \alpha \varepsilon^{3\ i}_{j}\, x^{\,j}.$$

$$J^\nu = \varepsilon^{3\ i}_{j}\, x^{\,j} \left(\partial_i \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_i \mathcal{L}\right) = \varepsilon^{3\ i}_{j}\, x^{\,j} \Theta^\nu_i.$$

Raising and unraising an index and defining $\delta^\nu_\mu = g_{\mu\alpha}\eta^{\alpha\nu}$, we write the Nöether current as

$$M^{\nu;12} = \varepsilon^{3\ i}_{j}\, x^{\,j} \Theta^{\nu i}.$$

One can convince oneself that the general case for a rotation about any axis is obtained by simply replacing the 1 and 2.

Example: Lorentz Boosts Lorentz boosts can be treated similarly by introducing $\gamma = \cosh(\lambda)$, which implies $\beta\gamma = \sinh(\lambda)$. Boosting along the x-direction, the transformation is given by

$$(x^0)' = x^0 \cosh(\lambda) - x^1 \sinh(\lambda),\ (x^1)' = x^1 \cosh(\lambda) - x^0 \sinh(\lambda).$$

For a small boost such that it is a symmetry of the Lagrangian, we have $\delta x^0 = -x^1,\ \delta x^1 = -x^0$ and the Nöether current

$$\begin{aligned}
J^\nu &= -\left(\partial_0 \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_0 \mathcal{L}\right) x^1 - \left(\partial_1 \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \delta^\nu_1 \mathcal{L}\right) x^0 \\
&= -\left(\partial^0 \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} - \eta^{\nu 0} \mathcal{L}\right) x^1 - \left(-\partial^1 \phi^a \frac{\partial \mathcal{L}}{\partial(\partial_\nu \phi^a)} + \eta^{\nu 1} \mathcal{L}\right) x^0 \\
&= \Theta^{\nu 1} x^0 - \Theta^{\nu 0} x^1 \\
&= M^{\nu;01},
\end{aligned}$$

where we have generalized our previous definition to

$$M^{\mu;\alpha\beta} = \Theta^{\nu\beta} x^\alpha - \Theta^{\nu\alpha} x^\beta.$$

Gauge Principles A gauge principle is the construction of a theory, starting with one which is invariant under some unitary group, and then constructing a new theory by adding a gauge field. This is better shown by example, because I really have no idea what it means.

Example: Maxwell Theory From Pure Arguments Suppose we did not now the Lagrangian density of Maxwell theory, but wanted a theory of the fields A^μ such that

1. The equations of motion are second-order.
2. The theory is relativistically invariant.
3. The theory is invariant under gauge transformations $A^\mu \to A^\mu + \partial_\mu \alpha$ for any function α.

Can this be done?

The first criterion implies that at most the first-order derivative of the fields may enter. The third implies that the derivative appears as

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

The second criterion implies that F appears a scalar, that is as $F_{\mu\nu}F^{\mu\nu}$. Hence the Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$

satisfies all the criterial. Higher-order terms may exist, but come with other issues. Other second-order terms may also be constructed, but can be shown to be total derivatives.

As a side note, what happens if the fields also enter into the Lagrangian? This would need to happen as

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + j_\mu A^\mu.$$

In this case, the gauge transformation yields

$$\mathcal{L} \to \mathcal{L} + j_\mu \partial^\mu \alpha = \mathcal{L} + \partial^\mu \alpha j_\mu - \alpha \partial^\mu j_\mu.$$

The middle term does not affect the equations of motion and thus can be ignored. For the gauge transformation to be a symmetry of the Lagrangian, we must therefore have $\partial^\mu j_\mu$, which was necessary for the Maxwell equations to be consistent to begin with.

Example: Gauge Transformations of Complex Klein-Gordon Theory and Connections to Maxwell Theory Complex Klein-Gordon theory is already invariant under multiplication of the field by a constant $e^{i\alpha}$. This cannot immediately be extended to α being a function, as the derivatives (when treated strictly as operators on the field) transform with it according to

$$\partial_\mu \to \partial_\mu + i \partial_\mu \alpha,$$

hence this cannot be a symmetry of the Lagrangian. What can be done, however, is to construct a new operator

$$D_\mu = \partial_\mu - i A_\mu$$

with functions A_μ which are transformed with the field according to $A_\mu \to A_\mu + \partial_\mu \alpha$. The Lagrangian

$$\mathcal{L} = D_\mu \phi D^\mu \phi^* - V(\phi^* \phi)$$

thus has such gauge transforms as a symmetry.

This Lagrangian can also be coupled to the functions A_μ by adding the first term from the Maxwell theory Lagrangian with a coupling parameter λ, netting a new Lagrangian

$$\mathcal{L} = D_\mu \phi D^\mu \phi^* - V(\phi^* \phi) - \frac{1}{4\lambda^2} F_{\mu\nu} F^{\mu\nu}.$$

This Lagrangian describes an electromagnetic field which is coupled to another field described by the Klein-Gordon equation.

Example: Coupling of Schrödinger Theory to the Electromagnetic Field The previously demonstrated method is typical for theories whose symmetries contain $U(1)$, and can be generalized to the field theory of wave mechanics. By repeating the same procedure as previously with a transformation

$$D_\mu = \partial_\mu + iqa_\mu,$$

we obtain the Lagrangian

$$\mathcal{L} = \frac{\hbar^2}{2m}(\partial_i\psi^* - iqa_i\psi^*)(\partial^i\psi + iqa^i\psi) + V\psi^*\psi + \frac{\hbar c}{2i}(\psi^*(\partial_0\psi + iqa_0\psi) - \psi(\partial_0\psi^* - iqa_0\psi^*)),$$

where we have changed the sign of the gauge field when the complex conjugate is involved. The equations of motion for ψ^* are

$$\begin{aligned}
0 &= -\frac{\hbar^2}{2m}iqa_i(\partial^i\psi - iqa^i\psi) + V\psi + \frac{\hbar c}{2i}(\partial_0\psi + iqa_0\psi + iqa_0\psi) - \partial_0\frac{\hbar c}{2i}(-\psi) - \partial_i\frac{\hbar^2}{2m}(\partial^i\psi + iqa^i\psi), \\
0 &= -\frac{\hbar^2}{2m}(\partial_i\partial^i\psi - \partial_i iqa^i\psi - iqa_i\partial^i\psi - q^2a_ia^i\psi) + V\psi + \frac{\hbar c}{i}\partial_0\psi + \hbar cqa_0\psi, \\
i\hbar\partial_t\psi &= -\frac{\hbar^2}{2m}(\partial_i\partial^i\psi + \partial_i iqa^i\psi + iqa_i\partial^i\psi + q^2a_ia^i\psi) + (V + \hbar cqa_0)\psi.
\end{aligned}$$

Introducing $p_i = \frac{\hbar}{i}\partial_i$ and identifying $A_i = \hbar a_i$, we have

$$\begin{aligned}
i\hbar\partial_t\psi &= \frac{1}{2m}(p_ip^i\psi - \hbar qp_ia^i\psi - \hbar qa_ip^i\psi + \hbar^2q^2a_ia^i\psi) + (V + \hbar cqa_0)\psi \\
&= \frac{1}{2m}(p_i - qA_i)(p^i - qA^i)\psi + (V + q\phi)\psi,
\end{aligned}$$

which, when compared to the Hamiltonian of a charged particle in a magnetic field, is exactly the correct equation of motion.

Non-Abelian Gauge Theories Similar principles can also be applied to field theories which depend on multiple fields, where the transformation of the fields consists of constructing linear combinations of the old fields by the action of some non-Abelian group.

To be somewhat more specific, consider a Lagrangian such as

$$\mathcal{L} = \partial_\mu(\phi^a)^*\partial^\mu\phi^a - V((\phi^a)^*\phi^a),$$

which is invariant under a transformation $\phi^a \to \psi^a = U^a_b\phi^b$. To obtain a criterion for the transformations such that this is invariant, note that

$$\begin{aligned}
\mathcal{L}' &= \partial_\mu(U^a_b\phi^b)^*\partial^\mu U^a_c\phi^c - V((U^a_b\phi^b)^*U^a_c\phi^c) \\
&= (U^a_b)^*U^a_c\partial_\mu(\phi^b)^*\partial^\mu\phi^c - V((U^a_b)^*U^a_c(\phi^b)^*\phi^c).
\end{aligned}$$

Introducing the Hermitian conjugate of the transformation matrix, we have

$$\mathcal{L}' = (U^\dagger)^b_aU^a_c\partial_\mu(\phi^b)^*\partial^\mu\phi^c - V((U^\dagger)^b_aU^a_c(\phi^b)^*\phi^c).$$

One way (and, indeed, the only way as far as I can see) for this to be equal to the old Lagrangian is if $(U^\dagger)^b_aU^a_c = \delta^b_c$, which is the definition of U being unitary. The group of unitary matrices is non-Abelian, somewhat complicating the proceeding matters compared to what we have previously seen.

As a side note, I include a vector notation for this Lagrangian, that being

$$\mathcal{L} = \partial_\mu(\boldsymbol{\phi})^\dagger\partial^\mu\boldsymbol{\phi} - V((\boldsymbol{\phi})^\dagger\boldsymbol{\phi}),$$

with which the symmetry under the action of U is more clear.

We would like to extend the set of symmetries to position-dependent unitary matrices. This will affect the derivative term according to

$$\partial^\mu U\boldsymbol{\phi} = U\partial^\mu\boldsymbol{\phi} + (\partial^\mu U)\boldsymbol{\phi} = U(\partial^\mu + U^\dagger(\partial^\mu U))\boldsymbol{\phi}.$$

To remove this new term, we add a gauge field A_μ, where each component is a matrix, and transform the derivative according to

$$D_\mu = \partial_\mu - iA_\mu.$$

$$D_\mu U\boldsymbol{\phi} = U(\partial_\mu + U^\dagger(\partial_\mu U))\boldsymbol{\phi} - iA_\mu U\boldsymbol{\phi} = U(\partial_\mu + U^\dagger(\partial_\mu U) - iU^\dagger A_\mu U)\boldsymbol{\phi},$$

and the transformation is thus a symmetry if the gauge field transforms according to

$$\begin{aligned} -iA'_\mu &= U^\dagger(\partial_\mu U) - iU^\dagger A_\mu U, \\ A'_\mu &= iU^\dagger(\partial_\mu U) + U^\dagger A_\mu U \end{aligned}$$

The interpretation of the gauge fields depend on the context, but this is a powerful tool for handling field theories.