Sammanfattning av SI2360 Analytisk mekanik och klassisk fältteori

Yashar Honarmandi
yasharh@kth.se

6 november 2019

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. But to introduce the involved techniques, we will first be studying a problem 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 that satisfy the boundary conditions. Given this, we parametrize this family of functions 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.

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

[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.

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 derivatives are computed the way they are.

What is a derivative with respect to a function? Mathematically, the functions are just numbers, and the derivatives show how much the integrand changes when we vary the given number. That is all. The derivatives are partial to signify that when we try varying one number, we keep all the others fixed.

And the solution... 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, for instance by setting them to zero (the need for this arises due to the problem in question being second-order, and thus requiring two conditions).

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

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

2 Group theory

Definition of a group A grou 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.
- infinite.
- discrete, i.e. all elements in the group can be labeled with some index.
- continuous.
- commutative, i.e. $ab = ba$ for all elements in the group.

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

Generators 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. In this context, we will use the generators in a wider context - for instance, the matrix J that is used to create rotation matrices is said to be a generator of the group.

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.

Homomorphisms and isomorphisms A homomorphisms is a map $f : G \to H$ such that $f(g_1)f(g_2) = f(g_1g_2)$. If the map is bijective, f is called an isomorphism.

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.

Combined with certain discrete translation, these are 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 Formally, a Lie group is a group containing a manifold and the group operation and inverse operation being smooth maps on the manifold. Its elements are $g(\theta)$, where $g(\mathbf{0}) = 1$. We can expand the map as $g(\theta) \approx 1 + A$, where

$$A = i\theta_a T_a,$$

where the T_a are the generators. That means that close to the identity, the non-commutativity of such maps is captured by the commutators, or Lie brackets:

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

The generators are sel-adjoint, so the constants $f_{a,b,c}$ are real.

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

$$D(g_1)D(g_2)v = D(g_1g_2)v, \ v \in V.$$

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

Small and large 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

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

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

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

Exponentiating yields

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

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

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

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

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

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

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

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

which implies

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

It can be shown that

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

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

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

Symmetries in classical mechanics

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

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

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

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

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

We see that these symmetries define a group.

Example: A particle in a moving potential

3 Differential Geometry

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

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

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

The first is the tangent basis of vectors

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

The second is the dual basis

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

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

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

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

The tangent basis vectors are thus

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

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

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

The solutions to this are

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

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

Orthogonality We can now compute the scalar product

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

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

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

Vector Components Any vector can now be written as

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

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

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

Coordinate Transformations Suppose that a vector can be expressed in two different tangent bases (corresponding to different choices of coordinates). How do we transform between these? A single 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, we try to transform between dual bases. We obtain

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

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 fomr the nomenvlature.

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

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

Normeringsvillkoret ger då direkt

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

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.

Volume Integrals We need to cheat with this, but we finally obtain

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

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

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 Suppose that we transform the basis according to

$$(\mathbf{E}')^{a_i} = L^{a_i}_{b_i} \mathbf{E}_{b_i}.$$

The components of the tensor then transform according to

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

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:

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

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

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

to be the tensor that satisfies

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

Then any tensor can be written as

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

where the $T^{a_1 \ldots a_n}$ 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 $\tilde{T}(\mathbf{w}_1 \otimes \ldots \otimes \mathbf{w}_m)$ such that

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

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 $T(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ such that

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

This is a linear rank n tensor.

Tensor contraction Given a complete set of vectors $\mathbf{v}_i$ and their dual $\mathbf{v}^i$ such that $\mathbf{v}_i \cdot \mathbf{v}^i = \delta_i^j$, 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:

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

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

The metric tensor The metric tensor g is a rank 2 tensor defined by $g(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$. It is a symmetric tensor. 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

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

We note that

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

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

6 Geometry

Covariant derivatives and Christoffel symbols When computing a derivative, one must account both for the change in the quantity itself and the change of basis. We have

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

where the $\Gamma^c_{\ ba}$ are called Christoffel symbols. These satisfy

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

Note that

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

which implies

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

Using this, we can compute the partial derivate of $\mathbf{v} = v^a \mathbf{E}_a$ with respect to χ^a as

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

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. We now define the covariant derivative of the contravariant components of $\mathbf{v}$ as

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

We would also like to define the covariant derivative of the covariant components of a vector field. 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

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

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

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

A minor simplification yields

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

Expanding the time derivative yields

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

This may be written as

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

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 $\mathcal{L}^2$. Expanding the derivative yields

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

Multiplying this by $-g^{cd}\mathcal{L}$ yields

$$g_{ac}g^{cd}\ddot{\chi}^a + \frac{1}{2}\dot{\chi}^a\dot{\chi}^b g^{cd}(2\partial_b g_{ac} - \partial_c g_{ab}) = g_{ac}g^{cd}\ddot{\chi}^a + \frac{1}{2}\dot{\chi}^a\dot{\chi}^b g^{cd}(\partial_b g_{ac} + \partial_b g_{ac} - \partial_c g_{ab}) = 0.$$

The a and b indices are summed over, and may thus be swapped. Combined with the symmetry of the metric tensor, this yields

$$g_{ac}g^{cd}\ddot{\chi}^a + \frac{1}{2}\dot{\chi}^a\dot{\chi}^b g^{cd}(\partial_b g_{ac} + \partial_a g_{cb} - \partial_c g_{ab}) = 0.$$

Summation of the first term over a gives $g_{ac}g^{cd}\ddot{\chi}^a = g^{cd}\ddot{\chi}_c$, and summation over c gives $g^{cd}\ddot{\chi}_c = \ddot{\chi}^d$. This thus yields

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

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

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

Comparison to the chain rule yields

$$\frac{\mathrm{d}\dot{\gamma}}{\mathrm{d}t} = (\ddot{\chi}^a + \dot{\chi}^a\dot{\chi}^c\Gamma^d_{\ ac})\mathbf{E}_d.$$

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 have no idea how to do that.

The geometry of curved space In curved space, we face the restriction that there is no position vector. All vectors in curved space are instead restricted to the tangent space. It turns out that tangent vectors at a point have coordinates $\dot{\chi}^a$ and that the tangent vectors consist of the tangent vectors to the coordinate lines, i.e. partial derivatives.

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

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

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

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

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

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

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

7 Classical mechanics

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

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

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

where $\mathcal{L}$ is the Lagrangian (discussed below). Hamilton's principle states that for the motion of the system in configuration space, $\delta S = 0$. This can be expressed as

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

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

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

How to form 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.

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

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

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

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

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

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

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

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

At this point I feel it only reasonable to discuss how I (or rather, Goldstein) arrived at this particular Lagrangian. To the trained eye it is very clear that the given equation of motion would result from that Lagrangian, but surely the core idea cannot be to just guess Lagrangians based on what we already know? Well, yes it can. The goal of physics is to describe reality, so if someone told you that their measurements

implied the given equations of motion, is it not your job as a theorist to construct a theory which describes that experiment? And how do you that, if not by constructing an appropriate Lagrangian? It might seem somewhat ad hoc to construct theories based on desired results in this fashion, but the most important check for a theory to satisfy is its compliance with the reality around us. If that is ad hoc to you, then perhaps mathematics will feel more satisfactory to you than physics.

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

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

Kinetic energy Kinetic energy is defined by a rank 2 tensor as

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

where the dot now really represents the time derivative.

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 define T_{ab} as the kinetic metric, with corresponding Christoffel symbols.

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

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

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

which is a generalization of Newton's second law.

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

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

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

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

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

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 the 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

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

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, t). 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.

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

What is a quasi-symmetry, then? 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 $\tau(t, 0) = t, \;\; Q^a(q, 0) = q^a$ and for small $s = \varepsilon$ that $t \to t + \varepsilon \delta t, \;\; q^a \to q^a + \varepsilon \delta q^a$. This is assumed to be normalized such that δt is either 0 or 1. How? Don't ask. A quasi-symmetry of a system with Lagrangian $\mathcal{L}$ is a transformation such that

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

for some F. The variation of the Lagrangian can be written as

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

The total time derivative of the Lagrangian is given by

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

which yields

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

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

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

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

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

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

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

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

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

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

Its general momentum is

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

The Hamiltonian is

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

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

A first attempt is keeping $\delta\mathbf{x}$ constant and not varying thiime(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.

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 $\mathrm{d}V = \mathrm{d}q\,\mathrm{d}p$. 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 (q, p) and after $\mathrm{d}t$ as (q', p'). To first order in time we have

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

The volume element is given by

$$\begin{aligned}\mathrm{d}V' &= \left(\mathrm{d}q + \left(\partial_q\partial_p\mathcal{H}\,\mathrm{d}q + \partial_p^2\mathcal{H}\,\mathrm{d}p\right)\mathrm{d}t\right)\left(\mathrm{d}p - \left(\partial_q^2\mathcal{H}\,\mathrm{d}q + \partial_p\partial_q\mathcal{H}\,\mathrm{d}p\right)\mathrm{d}t\right)\\ &= \mathrm{d}q\,\mathrm{d}p + \left(-\,\mathrm{d}q\left(\partial_q^2\mathcal{H}\,\mathrm{d}q + \partial_p\partial_q\mathcal{H}\,\mathrm{d}p\right) + \mathrm{d}p\left(\partial_q\partial_p\mathcal{H}\,\mathrm{d}q + \partial_p^2\mathcal{H}\,\mathrm{d}p\right)\right)\mathrm{d}t\\ &= \mathrm{d}q\,\mathrm{d}p + \left(-\partial_q^2\mathcal{H}(\mathrm{d}q)^2 + \left(\partial_q\partial_p\mathcal{H} - \partial_p\partial_q\mathcal{H}\right)\mathrm{d}q\,\mathrm{d}p + \partial_p^2\mathcal{H}(\mathrm{d}p)^2\right)\mathrm{d}t\,.\end{aligned}$$

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 implies

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

Poisson brackets Consider a function $f(q, p, t)$. 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.

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

We have

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

implying

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

This is an equivalent statement of Liouville's theorem.

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

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

Canonical transformations and Poisson brackets It turns out that a transformation in phase space is canonical if and only if they preserve the following equations:

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

To show this, we apply the symplectic approach. Consider some point $\mathbf{x}$ in phase space. Under the canonical transformation, it transforms to y. Suppose now that the relation

$$\dot{x}_i = J_{ij}\partial_{x_j}\mathcal{H}$$

applies. According to the equations of motion, this would imply

$$J_{ij} = \begin{cases} 1, & j = i + n, i = 1, 2, \ldots, n, \\ -1, & j = i - n, i = n + 1, \ldots, 2n. \end{cases}$$

If the transformation is canonical, then the same should be true after the transformation. On the other hand, the chain rule yields

$$\dot{y}_i = \partial_{x_j} y_i J_{ik} \partial_{x_k} y_m \partial_{y_m} H.$$

Comparing this with the Jacobian $\mathcal{J}$ yields

$$J = \mathcal{J} J \mathcal{J}^T.$$

Generators of canonical transformations Consider a 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

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

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. We dub this a generating function, and say that G generates the transformation.

We now reinsert this into the coordinate transformations. By considering the transformation as a map onto the same phase space, we obtain

$$\partial_\alpha q_i = \partial_{p_i} G,\ \partial_\alpha p_i = -\partial_{q_i} G.$$

We notice the strong analogy with Hamilton's equations. This kind of transformations can be thought of as flows in phase space.

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

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}\alpha} = \{\mathcal{H}, G\}.$$

We have seen that a symmetry of the Hamiltonian is a transformation such that $\delta\mathcal{H} = 0$. Supposing this to be true, we have $\dot{G} = \{\mathcal{H}, G\} = 0$, and G is conserved. In other words, if G generates a symmetry, then it is conserved.

Generators from Hamilton's principle We return to the principle of least action. The principle of least action can be stated as a variational problem on phase space instead, where we seek the extrema of

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

Recalling that the equations of motion must be preserved under canonical transformations, the action under a canonical transformation can be written as

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

and this is still extremal under the transformation. One way for this to be true is if

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

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 $F = F_1(q, Q, t)$.

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 $F_2 = F_2(q, P, t)$.

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 $F_3 = F_3(p, Q, t)$.

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

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

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 is not always possible. For instance, it might not be possible to find a generator of a certain kind, in which case performing a transformation to or from that kind is meaningless.

Example: A failed generator transform 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.

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 system is trivial. The equation

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

thus defines a differential equation for the generator of the transformation we seek. Calling, for no apparent reason, the generator F_2 the action S yields the Hamilton-Jacobi equation

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

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

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

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

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

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

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

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

Finally, using the Hamilton-Jacobi equation, we obtain

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

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

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

We start with the familiar Hamiltonian

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

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

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

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

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

Inserting this into the Hamilton-Jacobi equation yields

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

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

Solving the time part first, we have

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

Returning to the space part, we are left with

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

Solving this for the derivative yields

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

X is now given by integrating. The action is thus

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

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

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

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

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

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

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

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

This yields

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

where we have defined

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

We then obtain

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

which is the familiar solution of the harmonic oscillator.

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

Example: Hamilton-Jacobi theory in spherical coordinates

The Schrödinger equation from Hamilton-Jacobi theory

Quantum mechanics and the action

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

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

Structures of theory A theory in physics contain

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

Example: Hamiltonian mechanics

8 Relativity

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Inserted into the defining equation, we obtain

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

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

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

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

The defining equation now yields

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

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

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

Expanding the bracket yields

$$(1-\phi K^T)(\sigma_z - \phi\sigma_z K) = \sigma_z,$$
$$\sigma_z - \phi\sigma_z K - \phi K^T\sigma_z + \phi^2 K^T\sigma_z K = \sigma_z.$$

Ignoring higher-order terms yields

$$\sigma_z K + K^T\sigma_z = 0,$$
$$(\sigma_z K)^T = -\sigma_z K.$$

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

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

The transformation can now be written as

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

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

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

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

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

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

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

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

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

We can now define the four velocity

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

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

9 Classical field theory

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

Lagrangian dynamics 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

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

Alternatively, by defining $x^0 = t$ and extending the summation over spatial coordinates to include $i = 0$, we can write

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

Solving models To solve models, we usually allow for periodic boundary conditions. The field is then expanded as a Fourier series, or a Fourier transform in the limit of a large domain or small lattice constant. We will in any case find that the system is compact in Fourier space, i.e. there are only non-zero contributions within some compact region.

Nöether's theorem In this context Nöether's theorem states that symmetries of a system are associated with conservative current. In field theory, a symmetry is a transformation $\phi \to \phi_a$, where a is some continuous transformation parameter, such that for the quantity

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}a} = \frac{\mathrm{d}V^\mu}{\mathrm{d}x^\mu}$$

there are quantities j^μ such that

$$\frac{\mathrm{d}j^\mu}{\mathrm{d}x^\mu} = 0.$$

Hamiltonian dynamics In Hamiltonian dynamics, 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

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

where the functional derivative is given by

$$\frac{\delta H}{\delta\pi} = \partial_\pi\mathcal{H} - \partial_x\partial_{\pi_x}\mathcal{H} + \ldots$$

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.

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

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