Draft September 25, 2020

NEWTON, LAGRANGE, HAMILTON, NOETHER

A *VERY* BRIEF LOOK AT CLASSICAL MECHANICS

LARRY SUSANKA

Contents

1. **Variables**

There are three types of quantities we will use to define the state of a physical system:

- **time**, a real number denoted by t

Date: September 25, 2020.

- **geometrical variables** generally denoted by individual coordinates $q^1, \dots, q^N$, or simply the **configuration vector**[1] $q \in \mathbb{R}^N$.
- **dynamical variables**[2] denoted by $v^1, \dots, v^N$, or simply v, also in $\mathbb{R}^N$.

q may represent the configuration of a system of n particles in which case $N = 3n$ and the position of particle k with mass m_k is $\left(q^{3k-2},\, q^{3k-1},\, q^{3k}\right)$.

These variables are imagined to represent real things in the world, real locations or times: they are mappings between "reality" and coordinate space.

A prediction of the evolution of a physical system will correspond to a time-dependent path $q(t)$ through various "configuration vector" values.

When $\mathbb{R}^N$ is conceived of as the set of configuration vectors, it is called the **configuration-space**. $\mathbb{R} \times \mathbb{R}^N$ is called the **extended configuration-space** when, in addition, the first coordinate is interpreted as time.

So the time-evolution of a physical system is a path in configuration-space, and its graph is a subset of extended configuration-space.

Associated with this path will be tangent vectors $\frac{dq}{dt}(t)$ which will correspond to "dynamical vectors" $v(t) = \dot{q}(t)$.

We will also consider a function called here the **lagrangian**

$$\boldsymbol{L} \colon \mathbb{R} \times \mathbb{R}^{\boldsymbol{N}} \times \mathbb{R}^{\boldsymbol{N}} \to \mathbb{R}.$$

All functions in this note will be presumed to have whatever degree of continuity/differentiability required for a given calculation.

The second time through this material you are free to see how far this assumption can be relaxed in various places. Those primarily interested in physical applications have fewer worries on this score: there are no infinitely rigid walls or discontinuous forces, even if idealized models do posit such fictions.

In many examples L has units of energy, the joule, $[\mathrm{kg\,m^2\,s^{-2}}] = [\mathrm{J}]$.

2. **Action**

We will let $\mathcal{S}$ be a set of configuration-vector-valued functions of the form

$$q \colon [a, b] \to \mathbb{R}^N$$

where $[a, b]$ is a generic (i.e. not fixed for all members of $\mathcal{S}$) interval of times.

Members of $\mathcal{S}$ are potential configuration paths[3] and for the simplest examples the coordinates of q might have units of [m].

---

[1] In more general situations than those considered here q could correspond to a point in a coordinate patch of an N-dimensional manifold M, the configuration manifold. Solutions of the type we consider would be pieced together as the physical system moves out of one patch into another.

[2] Such a v could represent a vector in the tangent space at a point q of a configuration manifold M. In that context a pair (q, v) would correspond to a point in the tangent manifold TM.

[3] These configuration paths are not restricted to interpretation as rectangular coordinate paths. Kinetic energy, for instance, *cannot be presumed* to be given in terms of $\dot{q} \cdot \dot{q}$!

The derivatives of the functions in $\mathcal{S}$, of the form

$$\dot{q}\colon [a,b] \to \mathbb{R}^N,$$

would then have units $[\mathrm{m\,s^{-1}}]$.

$\mathcal{S}$ must be a pretty rich class of functions, containing many vector spaces of functions.[4]

We need a word about notation. $L = L(A,B,C)$ is just a function of $2N+1$ variables, and in our usage we normally put time in the "A-slot" and a configuration vector value q or path $q = q(t)$ in the "B-slot" and the derivative function $\dot{q} = \dot{q}(t)$ or more general velocity value v in the "C-slot."

When we indicate a partial derivative

$$\boldsymbol{\frac{\partial L}{\partial t}} \quad \textbf{or} \quad \boldsymbol{\frac{\partial L}{\partial q^i}} \quad \textbf{or} \quad \boldsymbol{\frac{\partial L}{\partial v^i} = \frac{\partial L}{\partial \dot{q}^i}}$$

there is no implication that there is actually a symbol t in place of A, or a symbol q^i in the ith place[5] in L, or a symbol $\dot{q}^i$ in the $N+i$ place in L, the i-th coordinate of C. Any symbol or function can go in those places! Using a more modern subscript notation for partial derivatives we have

$$\frac{\partial L}{\partial t}(A,B,C) = \partial_0 L(A,B,C) \quad \text{or} \quad \frac{\partial L}{\partial \dot{q}^i}(A,B,C) = \partial_{N+i} L(A,B,C)$$

and no t or $\dot{q}^i$ is in sight on the right-hand sides. The t or $\dot{q}^i$ indicated in a partial derivative just tells you to calculate the partial derivative with respect to the "slot" ***traditionally occupied by t or $\dot{q}^i$ in the type of calculation at hand.*** The "$\dot{q}^i$-slot" is the same slot that, in some different context, might be occupied by v^i. The presence of an actual path q is not implied, necessarily, by the notation $\frac{\partial L}{\partial \dot{q}^i}$, and the number or function in the $N+i$ slot of L ***must be taken from the surrounding context.***

After the partial derivative is found, a path $q(t)$ and its derivative $\dot{q}(t)$, ***ar any function with appropriate range***, may be substituted to create a function of t.

In contrast, the indication of a total derivative such as

$$\boldsymbol{\frac{dL}{dt} = \dot{L}}$$

will mean that some path, ***which might be indicated as $q(t)$ or some other symbol $f(t)$***, has been selected (even if it is not shown explicitly) and fed to L before differentiating the resulting composite function with respect to time:

$$\dot{L}(t) = \frac{d}{dt} L(t, f(t), \dot{f}(t)).$$

This total derivative would be calculated, then, by the chain rule:

$$\dot{L} = \frac{dL}{dt} = \frac{\partial L}{\partial t} + \frac{\partial L}{\partial q^i}\dot{f}^i + \frac{\partial L}{\partial \dot{q}^i}\ddot{f}^i$$

[4]For instance, the twice continuously differentiable functions on various domain time intervals.

[5]Here, (A,B,C) is identified with $(x_0, x_1, \ldots, x_{2N}) \in \mathbb{R}^{2N+1}$ where $A = x_0$ and $B_i = x_i$ and $C_i = x_{i+N}$ for $i = 1, \ldots, N$.

where we use the **Einstein summation convention** (as we will do throughout) to indicate summation over all values of an index variable in terms containing a product of exactly two factors with the same index.

We now define $\mathbf{\Gamma} \colon \mathcal{S} \to \mathbb{R}$ to be

$$\Gamma(q) = \int_a^b L(t, q(t), \dot{q}(t))\, dt.$$

The time dependence of q and $\dot{q}$, and even a reminder that there is an explicit path from $q(a)$ to $q(b)$, may be suppressed as in

$$\Gamma = \int_a^b L(t, q, \dot{q})\, dt = \int_a^b L\, dt.$$

Γ is called the **action functional** and $\Gamma(q)$ is sometimes called the **action** of q on the interval $[a, b]$.

In our simplest example, action[6] has units of joule-second, $[\mathrm{kg\, m^2\, s^{-1}}] = [\mathrm{J\, s}]$.

Hamilton's Principle states that in many physical situations the lagrangian[7] is to be $T - V$ where T is the kinetic energy and V is potential energy (a function whose gradient is the force) and the actual path taken during the evolution of a physical system described by this lagrangian is one that minimizes the action, at least locally, among "nearby" paths.

We do not defend this principle[8] here, but mention it to motivate the investigation that follows now.

3. The Euler-Lagrange System of Differential Equations

Suppose q is a path, a potential description of a physical evolution from time a to time b , and ζ is a path with $\zeta(a) = 0 = \zeta(b)$.

Define, for small real ϵ, the neighboring path Q_ϵ to be $q + \epsilon\,\zeta$.

We can define

$$\boldsymbol{\Gamma_\epsilon = \int_a^b L(t, Q_\epsilon, \dot{Q}_\epsilon)\, dt = \int_a^b L(t, q + \epsilon\,\zeta, \dot{q} + \epsilon\,\dot{\zeta})\, dt.}$$

So $q = Q_0$, and of course then $\Gamma = \Gamma_0$.

We say q is a **stationary path with respect to** $\boldsymbol{\zeta}$ when $\boldsymbol{\frac{d}{d\epsilon}\Big|_{\epsilon=0}\Gamma_\epsilon = 0}$, and

[6]The unit of action, the joule-second, is also the unit of angular momentum (moment of inertia times angular velocity) and the unit of Planck's constant (energy of a wave divided by frequency of that wave).

[7]In the manifold setting, the lagrangian is a (time-dependent) function on the tangent manifold TM of the configuration manifold M. As far as the ideas and conclusions are concerned, there is no real benefit to this generality in a first-pass through the material—even for people who know about manifolds, in my view.

[8]Sometimes also called the **Principle of Least Action**.

q is astationary path if it is stationary with respect to every ζ.

Assuming differentiability everywhere, we have by **Leibniz's Integral Rule**

$$\begin{aligned} \frac{d}{d\epsilon}\bigg|_{\epsilon=\varepsilon_0} \Gamma_\epsilon &= \frac{d}{d\epsilon}\bigg|_{\epsilon=\varepsilon_0} \int_a^b L(t, Q_\epsilon,\, \dot{Q}_\epsilon)\, dt \\ &= \int_a^b \frac{d}{d\epsilon}\bigg|_{\epsilon=\varepsilon_0} L(t, Q_\epsilon,\, \dot{Q}_\epsilon)\, dt \end{aligned}$$

and this equality will hold whether the derivative is 0 or not, and whether ε_0 is 0 or not, though the case $\varepsilon_0 = 0$ is actually the one that concerns us.

We need to calculate the interior derivative by the **chain rule** and, eventually, find out what it implies if that derivative is 0 for this (and every) ζ.

Let's change notation briefly. (It can't hurt to be double-clear about what this notation means, *at least once.*)

Define

$$\begin{aligned} x &= (x^0, x^1, \ldots, x^{2N}) = (t, Q_\epsilon,\, \dot{Q}_\epsilon) \\ &= (t, q^1 + \epsilon\zeta^1, \ldots, q^N + \epsilon\zeta^N,\ \dot{q}^1 + \epsilon\dot{\zeta}^1, \ldots, \dot{q}^N + \epsilon\dot{\zeta}^N). \end{aligned}$$

t (and so x^0) is independent of ϵ, but the other coordinates *do* depend on ϵ. Note that when $\epsilon = 0$ we have $x = (t, q, \dot{q})$.

Let's fix t and suspend the Einsten summation convention for the next few lines, and indicate partial differentiation with respect to the "ith slot" by $\boldsymbol{\partial_i}$.

$$\begin{aligned} \frac{dL \circ x}{d\epsilon}(\varepsilon) = [\partial_0 L]\,(x(\varepsilon))\frac{dx^0}{d\epsilon}(\varepsilon) \\ + \sum_{i=1}^{N} [\partial_i L](x(\varepsilon))\frac{dx^i}{d\epsilon}(\varepsilon) \\ + \sum_{i=1}^{N} [\partial_{N+i} L]\,(x(\varepsilon))\frac{dx^{N+i}}{d\epsilon}(\varepsilon). \end{aligned}$$

First of all, $\frac{dx^0}{d\epsilon}(\varepsilon) = 0$ for every ε. *So ignore the first term in the expansion.*

Also, for $i = 1, \ldots, N$

$$\frac{dx^i}{d\epsilon}(\varepsilon) = \frac{d}{d\epsilon}(q^i + \epsilon\zeta^i)(\varepsilon) = \zeta^i$$

and

$$\frac{dx^{N+i}}{d\epsilon}(\varepsilon) = \frac{d}{d\epsilon}(\dot{q}^i + \epsilon\dot{\zeta}^i)(\varepsilon) = \dot{\zeta}^i.$$

So these derivatives don't actually depend on ε.

Substituting those into the derivative expression gives ......

$$\frac{dL\circ x}{d\epsilon}(\varepsilon)=\sum_{i=1}^{N}\left[\partial_i L\right](x(\varepsilon))\zeta^i+\sum_{i=1}^{N}\left[\partial_{N+i}L\right](x(\varepsilon))\dot{\zeta}^i.$$

And then evaluating at $\varepsilon=0$ we have

$$\left.\frac{dL\circ x}{d\epsilon}\right|_{\varepsilon=0}=\sum_{i=1}^{N}\left[\partial_i L\right](t,q,\dot{q})\zeta^i+\sum_{i=1}^{N}\left[\partial_{N+i}L\right](t,q,\dot{q})\dot{\zeta}^i$$

and in the traditional shorthand notation we have, for each t,

$$\left.\frac{dL\circ x}{d\epsilon}\right|_{\varepsilon=0}=\sum_{i=1}^{N}\left(\frac{\partial L}{\partial q^i}\zeta^i+\frac{\partial L}{\partial\dot{q}^i}\dot{\zeta}^i\right).$$

This is the integrand of the derivative of the action on functions that vary from q by small multiples of ζ.

Reverting to the compact Einstein summation convention we have

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0}\Gamma_\epsilon=\int_a^b\left(\frac{\partial L}{\partial q^i}\zeta^i+\frac{\partial L}{\partial\dot{q}^i}\dot{\zeta}^i\right)dt.$$

Note: *The partial derivative expressions themselves do not depend on ζ!*

We want to rewrite some of the terms using **integration by parts.**

$$\int_a^b\frac{\partial L}{\partial\dot{q}^i}\dot{\zeta}^i\,dt=\left.\frac{\partial L}{\partial\dot{q}^i}\zeta^i\right|_a^b-\int_a^b\zeta^i\frac{d}{dt}\frac{\partial L}{\partial\dot{q}^i}\,dt.$$

Since $\zeta(a)=\zeta(b)=0$ the first term on the right is discarded and we have

$$\int_a^b\frac{\partial L}{\partial\dot{q}^i}\dot{\zeta}^i\,dt=-\int_a^b\zeta^i\frac{d}{dt}\frac{\partial L}{\partial\dot{q}^i}\,dt.$$

We then substitute this into the formula for $\left.\frac{d}{d\epsilon}\right|_{\epsilon=0}\Gamma_\epsilon$.

$$\begin{aligned}\left.\frac{d}{d\epsilon}\right|_{\epsilon=0}\Gamma_\epsilon&=\int_a^b\left(\frac{\partial L}{\partial q^i}\zeta^i+\frac{\partial L}{\partial\dot{q}^i}\dot{\zeta}^i\right)dt=\int_a^b\left(\frac{\partial L}{\partial q^i}\zeta^i-\zeta^i\frac{d}{dt}\frac{\partial L}{\partial\dot{q}^i}\right)dt\\&=\int_a^b\zeta^i\left(\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial\dot{q}^i}\right)dt.\end{aligned}$$

If q is a stationary path then this derivative is 0 for every ζ. Since each component of ζ is taken, independently of the other components, from a dense subset of the integrable functions, a standard result from integration theory[9] tells us that the factors

$$\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial\dot{q}^i}\qquad i=1,\dots,N$$

must be the zero function for each i.

[9]To see this directly, suppose that $\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial\dot{q}^i}\neq 0$ at some time t_0. $\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial\dot{q}^i}$ is continuous so there is an interval containing t_0 upon which the sign of $\frac{\partial L}{\partial q^i}-\frac{d}{dt}\frac{\partial L}{\partial\dot{q}^i}$ is constant. Let ζ^k be identically 0 when $k\neq i$. Create a non-negative smooth function ζ^i which is zero off this interval and for which $\zeta^i(t_0)>0$. For this ζ we then have $\left.\frac{d}{d\epsilon}\right|_{\epsilon=0}\Gamma_\epsilon\neq 0$.

With this stationarity condition we have the **Euler-Lagrange System of Differential Equations**, or "**ELS**" for short:

$$\frac{\partial L}{\partial q^i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} \qquad i = 1, \ldots, N.$$

Expanding $\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}$ we have the equivalent and explicit second order Euler-Lagrange System of N ordinary differential equations:

$$\frac{\partial L}{\partial q^i} = \frac{\partial^2 L}{\partial t \partial \dot{q}^i} + \frac{\partial^2 L}{\partial q^\mu \partial \dot{q}^i}\dot{q}^\mu + \frac{\partial^2 L}{\partial \dot{q}^\mu \partial \dot{q}^i}\ddot{q}^\mu \qquad i = 1, \ldots, N.$$

This system can, with appropriate boundary conditions, be solved for q numerically or, in a few very-special but often-presented cases, exactly.

All these steps are reversible, so if the ELS holds q will be stationary with respect to lagrangian L, establishing the equivalence of this particular system of second order *ordinary* differential equations with the stationarity condition on the integral of the lagrangian.

Hamilton's principle declares that this is a *necessary* condition for q to be the solution path of a physical system.[10]

So in one sense, the business of analytical mechanics is done. We can create stationary configuration paths to any desired accuracy, or exactly in some cases.

This is lagrangian mechanics.

As a final topic in this section we mention that there are many different lagrangians that produce the same stationary paths—and a physical scenario, for instance, cannot determine which of these you should use.

If the values of $\tilde{L}(t, q(t), \dot{q}(t))$ were to **differ** from $L(t, q(t), \dot{q}(t))$ by an **additive constant** F the action integral on every path used to create the ELS would change by $(b - a)F$, where a and b are the initial and final times under consideration, so this difference will not change which paths are stationary and which are not. Also if the new lagrangian is a **numerical multiple of the old one** the stationary paths do not change.

But even more differences yield identical solution paths.

Suppose $G = G(t, q)$ is differentiable and consider the two lagrangians

$$L(t, q, v) \qquad \text{and} \qquad L(t, q, v) + \frac{\partial G}{\partial t} + \frac{\partial G}{\partial q^i}v^i.$$

The additional term corresponds to a **directional derivative**: it is $\nabla G \cdot (1, v)$, the rate of change[11] of G at domain location (t, q) when you are moving in the domain of G at velocity $(1, v)$.

[10]A stationary solution must satisfy this DE system. The questions of sufficiency and uniqueness have not been addressed here. For instance if L is constant all paths are stationary. If the physicist has done his or her job while creating the lagrangian this is not a problem. However from a purely mathematical standpoint these are important issues, related to the specific form of the lagrangian. Legendre and Jacobi have created sufficient conditions for the existence of a unique minimizer of the action functional for paths connecting two domain points starting at a specific time.

[11]∇G is the **gradient** of function G.

If $q = q(t)$ is any path these two lagrangians differ, *along this path*, by a total time derivative $\dot{G} = \frac{d}{dt}G(t, q(t)) = \frac{\partial G}{\partial t} + \frac{\partial G}{\partial q^i}\dot{q}^i$.

So the actions of q calculated using these two lagrangians differ by the constant $G(b, q(b)) - G(a, q(a))$.

So if lagrangians $\tilde{L}$ and L differ by terms that can be recognized as

$$\boldsymbol{\nabla G \cdot (1, v) = \frac{\partial G}{\partial t} + \frac{\partial G}{\partial q^i} v^i}$$

where $\boldsymbol{\nabla G = \left(\frac{\partial G}{\partial t}, \frac{\partial G}{\partial q^1}, \ldots, \frac{\partial G}{\partial q^N}\right)}$ for some function G they produce the same stationary paths.

This method of creating a new lagrangian with the same solution paths as the old one is called a **gauge transformation.**

4. **The ELS and Newton's Laws**

The lagranian methods produce a system of differential equations and, often, so does an application of **Newton's Laws**. Hamilton's principle, that the lagrangian should be the difference between kinetic and potential energies and the solution path is the one that minimizes the action, connects the two methods.

This connection is deep and there are *numerous* special cases where equivalence can be made explicit, but we will only consider the situation of a several-particle system undergoing evolution subject to a (possibly time-dependent) potential as an example.

Suppose $x = (x^1, x^2, \ldots, x^{3n})$ represents the configuration of n particles in a **Cartesian coordinate system**[12] evolving under the influence of forces corresponding to potential $V = V(t, x)$.

These particles have mass vector $m = (m^1, m^2, \ldots, m^{3n})$ where, of course, the mass coordinates are identical for the "x, y and z" coordinates of each specific particle.

The force on a particle with coordinate x^i in the x^i-direction is (by definition of potential) equal to $-\frac{\partial V}{\partial x^i}$.

Thus, according to Newton's second law we have system of differential equations

$$m_i \ddot{x}^i = -\frac{\partial V}{\partial x^i} \qquad i = 1, \ldots, 3n. \qquad \text{(no summation here!)}$$

Hamilton's principle states we should use lagrangian

$$L = T - V = \frac{1}{2} m_i (\dot{x}^i)^2 - V(t, x).$$

The ELS states that $\frac{\partial L}{\partial x^i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}$ for $i = 1, \ldots, 3n$. Let's do this calculation.

$$\frac{\partial L}{\partial x^i} = -\frac{\partial V}{\partial x^i} \quad \text{and} \quad \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} = \frac{d}{dt} m_i \dot{x}^i = m_i \ddot{x}^i. \qquad \text{(no summation here!)}$$

[12]In these coordinate systems the square of distance in space is given by $(\Delta x^1)^2 + \cdots + (\Delta x^{3n})^2$.

So these two methods do produce the same system of differential equations and there really is no reason to prefer one approach over another in this case.

Lagrangian methods however are superior in many situations where Newton's laws are awkward to apply, such as problems with constraints, and also suggest approaches to problems where it is not obvious how Newton's laws can be made to apply at all.

Many physical systems feature **constraints**: the motion is confined to a subset of the configuration-space and is not all of $\mathbb{R}^N$. One way of thinking of this is to imagine we are confined to a configuration manifold, $M \subset \mathbb{R}^N$.

These constraints are often (usually) specified by one or more *equations* (not inequality) of the form $h(t, q) = 0$ for smooth h.[13]

These may be handled, without dealing with the configuration manifold directly, by the introduction of extra terms to produce a new lagrangian. We "add multiples of zero" to the old lagrangian:

$$\boldsymbol{L_c = L + \lambda_k h^k.}$$

These extra terms involve **Lagrange multipliers**, one for each constraint equation. These multipliers can be solved for to find additional equations linking the geometrical variables. Numerically, the Lagrange multiplier values are related to the forces required to enforce the constraint.

For instance consider a bead constrained to slide on a frictionless wire in the shape of a parabola, $y = x^2$.

The unconstrained lagrangian for a bead under the influence of gravity is

$$L\left(x, y, \dot{x}, \dot{y}\right) = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2\right) - mgy.$$

Adding the constraint $x^2 - y = 0$ produces constrained lagrangian

$$L_c\left(x, y, \dot{x}, \dot{y}, \lambda\right) = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2\right) - mgy + \lambda(x^2 - y).$$

$$\frac{\partial L_c}{\partial x} = \frac{d}{dt}\frac{\partial L_c}{\partial \dot{x}} \qquad \Longrightarrow \qquad 2\lambda x = m\ddot{x} \qquad \Longrightarrow \qquad \frac{\lambda}{m} = \frac{\ddot{x}}{2x}.$$

$$\frac{\partial L_c}{\partial y} = \frac{d}{dt}\frac{\partial L_c}{\partial \dot{y}} \qquad \Longrightarrow \qquad -\lambda - mg = m\ddot{y} \qquad \Longrightarrow \qquad \frac{\lambda}{m} = -g - \ddot{y}.$$

$$\ddot{x} = 2x\left(-g - \ddot{y}\right).$$

$$\frac{\partial L_c}{\partial \lambda} = \frac{d}{dt}\frac{\partial L_c}{\partial \dot{\lambda}} = 0 \qquad \Longrightarrow \qquad x^2 - y = 0 \qquad \Longrightarrow \qquad \ddot{y} = 2\dot{x}^2 + 2x\ddot{x}.$$

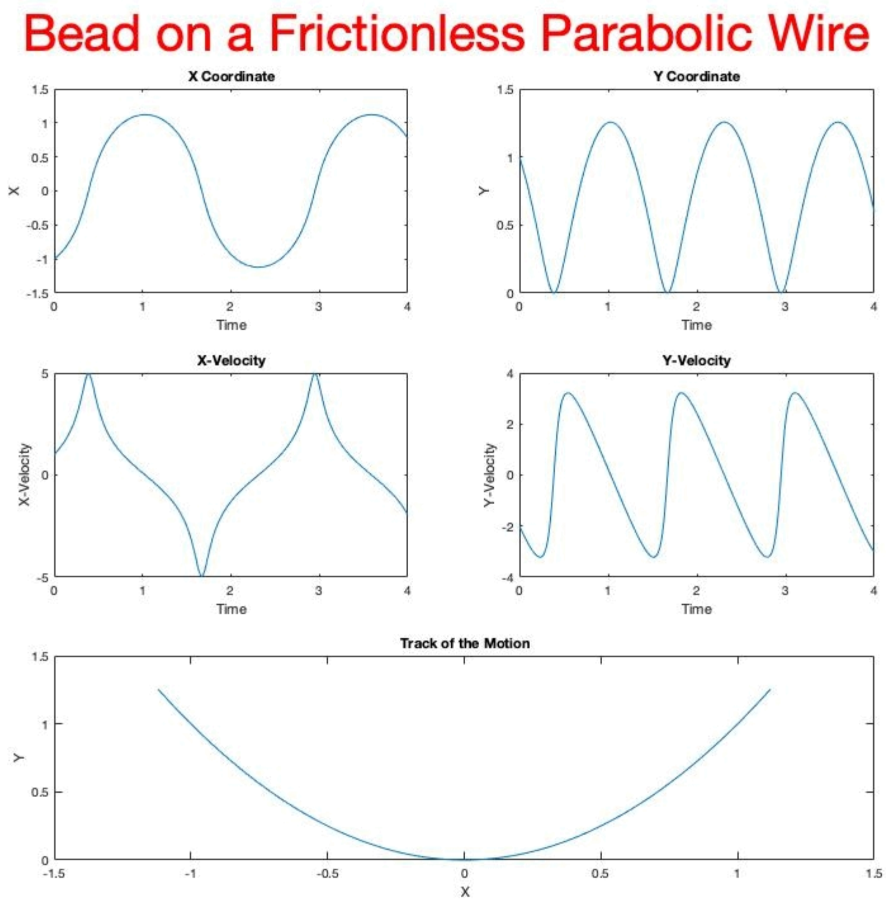
$$\ddot{x} = 2x\left(-g - \ddot{y}\right) \qquad \Longrightarrow \qquad \ddot{x} = \frac{-4x\dot{x}^2 - 2xg}{1 + 4x^2}.$$

[13]Constraints of this kind are called **holonomic**. Constraints *not* of this kind, perhaps one involving an *inequality*, are called **nonholonomic**. If time is *not* involved a constraint is called **scleronomous.** If time *is* involved the constraint is called **rheonomous.**

This produces the system

$$\dot{x} = w$$
$$\dot{w} = \frac{-4xw^2 - 2xg}{1+4x^2}$$

which can be solved, with initial x and $\dot{x}$ values, to whatever level of accuracy is required by numerical (Runge-Kutta, or even Euler) methods.



5. The ELS in "H-dot" and "p-dot" Form

The equations

$$\frac{\partial L}{\partial q^i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} \qquad i = 1, \dots, N$$

suggest that the rate of change of

$$\boldsymbol{p_i = \frac{\partial L}{\partial \dot{q}^i}}$$

is important and we call this quantity, which in our simplest example case has units $[\mathrm{kg\,m\,s^{-1}}]$, a **generalized momentum.** Here p_i is a function of time, associated with a generic path.

Its rate of change, $\frac{\partial L}{\partial q^i} = \dot{p}_i$, is called a **generalized force.**[14]

$\boldsymbol{p_i}$ is also said to be a **canonical momentum** and is referred to as **conjugate** to the **generalized coordinate** $\boldsymbol{q^i}$.

With this definition, the ELS in "$\boldsymbol{p}$ **dot**" form is

$$\boldsymbol{\frac{\partial L}{\partial q^i} = \dot{p}_i \qquad i = 1, \ldots, N.}$$

This equation holds *along a solution path.*

The total time derivative of the composition $L(t, q, \dot{q})$ on a solution path q is given by

$$\begin{aligned} \frac{dL}{dt} =& \frac{\partial L}{\partial t} + \frac{\partial L}{\partial q^i}\dot{q}^i + \frac{\partial L}{\partial \dot{q}^i}\ddot{q}^i \\ =& \frac{\partial L}{\partial t} + \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right)\dot{q}^i + p_i\ddot{q}^i \\ =& \frac{\partial L}{\partial t} + \dot{p}_i\dot{q}^i + p_i\ddot{q}^i \end{aligned}$$

Isolating $\frac{\partial L}{\partial t}$ we have

$$\frac{\partial L}{\partial t} = \frac{dL}{dt} - \dot{p}_i\dot{q}^i - p_i\ddot{q}^i = -\frac{d}{dt}\left(p_i\dot{q}^i - L\right).$$

So the time derivative of

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

seems to be important.

H, as a function of time, can be defined using coordinates of *any* path. But we can see that this number is preserved on a *solution* path precisely when L is independent of t on that path.

H is called the **hamiltonian of the path** in the scenario determined by L.[15]

We have the following "$\boldsymbol{H}$ **dot**" implication of the ELS, applied to the hamiltonian and evaluated along a solution path:

$$\boldsymbol{\frac{\partial L}{\partial t} = -\dot{H}.}$$

[14] In the situation of holonomic constraints in the last section we have, evaluated along the constrained solution path, $\frac{d}{dt}\frac{\partial L_c}{\partial \dot{q}^i} = \frac{\partial L}{\partial q^i} + \lambda_k \frac{\partial h^k}{\partial q^i}$. The last terms can be construed as the generalized forces required to enforce the constraints, while $\frac{\partial L}{\partial q^i}$ represents the force acting from all other sources (gravity, springs etc.) as the configuration moves along the constrained solution.

[15] Here, the hamiltonian is defined *only* as a function of time along a path. Later we will expand on this definition to create a stand-alone hamiltonian.

The conditions under which H or one of the p^i are preserved during the evolution of a physical system constitute a special case of Noether's Theorem, which we will consider in more detail soon.

6. **Covariance: The Form of the ELS After Coordinate Changes**

The ELS and a resulting stationary solution (given initial conditions) is calculated using a specific—though generic—coordinate system, the lagrangian and a time-integral.

It is not inconceivable that some part of this calculation is dependent on the coordinate system beyond a simple translation of the solution to new coordinates.

This would be a problem for a primary user group: the evolution of a physical system cannot depend on the coordinates used to describe it. We need to verify that our predictions, made using these methods, are consistent.

Let's suppose geometrical variable q and Q describe the same physical system in different coordinates. Rectangular versus polar coordinates would be a typical example.

We suppose $q = W(t,Q)$ gives old geometrical vector q as a function of time t and new geometrical vector Q, and $Q = K(t,q)$ is the inverse relation.

$$(t,q) = (t, W(t,Q)) \qquad \longleftrightarrow \qquad (t,Q) = (t, K(t,q)).$$

Given any path $Q = Q(t)$ it follows that the corresponding path, simply translated to old coordinates, is $q(t) = W(t, Q(t))$. And any path $q = q(t)$ in old coordinates corresponds to new-coordinate path $Q(t) = K(t, q(t))$.

Let's consider a path $Q(t)$ in new coordinates, which may or may not satisfy the ELS. We can calculate $\dot{q}(t)$ using the chain rule and W, and that calculation gives

$$\dot{q}^j(t) = \left.\frac{\partial W^j}{\partial t}\right|_{(t,Q(t))} + \dot{Q}^k(t)\left.\frac{\partial W^j}{\partial Q^k}\right|_{(t,Q(t))} \qquad j = 1, \dots, N.$$

Thus, we express $\dot{q}$ at each time in terms of t, Q and $\dot{Q}$ and, of course, the coordinate translator function W which, itself, depends only on t and Q.

This is true for any path, and paths can go in any direction going through any point at any time. So we can create a vector-valued function $G = G(t, Q, \dot{Q})$ defined on all of $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$ (not just points along some specific path) whose coordinate functions are given by

$$\dot{q}^j = G^j(t, Q, \dot{Q}) = \left.\frac{\partial W^j}{\partial t}\right|_{(t,Q)} + \dot{Q}^k \left.\frac{\partial W^j}{\partial Q^k}\right|_{(t,Q)} \qquad j = 1, \dots, N$$

and which reflects the dependence of individual dynamical variable values $\dot{q}$ on generic $(t, Q, \dot{Q})$ under this coordinate transformation. The creation of this function G utilizes only W, the coordinate translator, independent of any path.

So we find that

$$\frac{\partial \dot{q}^j}{\partial \dot{Q}^k} = \frac{\partial W^j}{\partial Q^k}$$

and since $q = W(t, Q)$ these last expressions are traditionally written[16] as

$$(1) \qquad \dot{q}^j = \frac{\partial q^j}{\partial t} + \dot{Q}^k \frac{\partial q^j}{\partial Q^k} \text{ for } j = 1, \ldots, N \qquad \text{and} \qquad \frac{\partial \dot{q}^j}{\partial \dot{Q}^k} = \frac{\partial q^j}{\partial Q^k} \text{ for all } j, k.$$

We will continue to use G and W as a reminder of the functional relationships but express, in the end, all results in terms of q and Q in the traditional manner. Be aware that in all classical presentations G and W are suppressed entirely and other notational contrivances that suppress superscripts and subscripts are used. While these shorten the calculations substantially they do nothing to clarify them.

We use a symbol such as q^i or $\dot{q}^i$ in four distinct ways. First, it can be a generic component of a vector. Second, it can be a generic component functionally related to *a different* generic vector and time by $q^i = W(t, Q)$. Third, it can be a real function of time, $q^i = q^i(t)$. Fourth, it can be a composite function such as $\dot{q}^i = G^i(t, Q(t), \dot{Q}(t))$.

So you may see, either implied or expressed directly, (true) expressions of the form

$$q(t) = q(t, Q(t))$$

which make no sense on the face of it. A modern mathematician would interpose a function W and claim that $q(t) = W(t, Q(t))$. Suppressing W we have an assertion that q is a function of time and dependent on another function Q, so it makes sense to calculate $\frac{\partial q^i}{\partial Q^j}$.

There is a good reason why mathematicians avoid conflating a bunch of different—but closely related—things in this way. But theoretical mechanics has a history going back to the 18th century or before, and the user groups (i.e. physicists) are dedicated to this practice. We just have to be careful and deal with it.

Here is an example. We will need to recognize $\frac{d}{dt}\frac{\partial q^j}{\partial Q^i}$ as $\frac{\partial \dot{q}^j}{\partial \dot{Q}^i}$ but to do that we must sort out the multiple meanings of the parts of these expressions. It looks like we have an exchange of the order of differentiation, which given our liberal differentiability assumptions ought to be legitimate.

Since we see a total derivative indicated, we must have a specific path selected and are working with a path $q(t)$ in old coordinates which is the composite function $q(t) = W(t, Q(t))$ with path $Q(t)$ in new coordinates, and what we are trying to show is the equality of two values which can be calculated at each time and place along that path.

$$\begin{aligned} \frac{d}{dt}\frac{\partial q^j}{\partial Q^i} &= \frac{d}{dt}\frac{\partial W^j}{\partial Q^i}(t, Q(t)) = \frac{\partial}{\partial t}\frac{\partial W^j}{\partial Q^i} + \left(\frac{\partial}{\partial Q^k}\frac{\partial W^j}{\partial Q^i}\right)\frac{dQ^k}{dt} \\ &= \frac{\partial^2 W^j}{\partial t \partial Q^i} + \dot{Q}^k \frac{\partial^2 W^j}{\partial Q^k \partial Q^i}. \end{aligned}$$

[16]For each time t the matrix $\left(\frac{\partial q^j}{\partial Q^k}\right) = \frac{\partial q}{\partial Q}$ is called the **Jacobian matrix** of the coordinate transformation.

On the other hand we have $\dot{q}^j = \frac{\partial W^j}{\partial t} + \dot{Q}^k \frac{\partial W^j}{\partial Q^k}$ and each dynamical variable $\dot{Q}^k$ is independent of each geometrical variable Q^i so

$$\frac{\partial \dot{q}^j}{\partial Q^i} = \frac{\partial}{\partial Q^i}\frac{\partial W^j}{\partial t} + \dot{Q}^k \frac{\partial}{\partial Q^i}\frac{\partial W^j}{\partial Q^k} = \frac{\partial^2 W^j}{\partial Q^i \partial t} + \dot{Q}^k \frac{\partial^2 W^j}{\partial Q^i \partial Q^k}.$$

So by **Clairaut's theorem** on the equality of mixed partials we have numerical equality of these values for each t. That is,

$$\frac{d}{dt}\frac{\partial q^j}{\partial Q^i} = \frac{\partial \dot{q}^j}{\partial Q^i}. \qquad (2)$$

We will create a new lagrangian $\tilde{L}$ in these new coordinates which will be integrated along paths such as this to create the action in these new coordinates.

Each value $\tilde{L}(t, Q, \dot{Q})$ will be the same as $L(t, q, \dot{q})$ when $q = W(t, Q)$ and $\dot{q}$ is the dynamical vector calculated above, dependent on t, Q and $\dot{Q}$. Specifically, we have

$$\tilde{L}(t, Q, \dot{Q}) = L(t, q, \dot{q}) = L(t, W(t, Q), G(t, Q, \dot{Q})).$$

Consider the expression involved in the ELS in new coordinates:

$$\frac{\partial \tilde{L}}{\partial Q^i} - \frac{d}{dt}\frac{\partial \tilde{L}}{\partial \dot{Q}^i} \qquad i = 1, \ldots, N.$$

We will examine its parts.

$$\begin{aligned}\frac{\partial \tilde{L}}{\partial Q^i} &= \frac{\partial L}{\partial q^j}\frac{\partial W^j}{\partial Q^i} + \frac{\partial L}{\partial \dot{q}^j}\frac{\partial G^j}{\partial Q^i} \\ &= \frac{\partial L}{\partial q^j}\frac{\partial q^j}{\partial Q^i} + \frac{\partial L}{\partial \dot{q}^j}\frac{\partial \dot{q}^j}{\partial Q^i}.\end{aligned}$$

Calculating $\frac{\partial \tilde{L}}{\partial \dot{Q}^i}$ is somewhat easier since neither t nor W depend on $\dot{Q}^i$:

$$\frac{\partial \tilde{L}}{\partial \dot{Q}^i} = \frac{\partial L}{\partial \dot{q}^j}\frac{\partial G^j}{\partial \dot{Q}^i} = \frac{\partial L}{\partial \dot{q}^j}\frac{\partial \dot{q}^j}{\partial \dot{Q}^i} = \frac{\partial L}{\partial \dot{q}^j}\frac{\partial q^j}{\partial Q^i}. \qquad \text{(By Equation (1).)}$$

So $\frac{d}{dt}\frac{\partial \tilde{L}}{\partial \dot{Q}^i}$ is

$$\begin{aligned}\frac{d}{dt}\frac{\partial \tilde{L}}{\partial \dot{Q}^i} &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^j}\frac{\partial q^j}{\partial Q^i}\right) = \frac{\partial q^j}{\partial Q^i}\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^j} + \frac{\partial L}{\partial \dot{q}^j}\frac{d}{dt}\frac{\partial q^j}{\partial Q^i} \\ &= \frac{\partial q^j}{\partial Q^i}\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^j} + \frac{\partial L}{\partial \dot{q}^j}\frac{\partial \dot{q}^j}{\partial Q^i}. \qquad \text{(By Equation (2).)}\end{aligned}$$

Now we have

$$\begin{aligned}\frac{\partial \tilde{L}}{\partial Q^i} - \frac{d}{dt}\frac{\partial \tilde{L}}{\partial \dot{Q}^i} &= \frac{\partial L}{\partial q^j}\frac{\partial q^j}{\partial Q^i} + \frac{\partial L}{\partial \dot{q}^j}\frac{\partial \dot{q}^j}{\partial Q^i} - \left(\frac{\partial q^j}{\partial Q^i}\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^j} + \frac{\partial L}{\partial \dot{q}^j}\frac{\partial \dot{q}^j}{\partial Q^i}\right) \\ &= \frac{\partial q^j}{\partial Q^i}\left(\frac{\partial L}{\partial q^j} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^j}\right).\end{aligned}$$

It follows that if the Euler-Lagrange Equations hold for path $q(t)$ in the old coordinates then they hold for the related path in the new coordinates.

Switching the positions of "old" and "new" yields the conclusion that if the Euler-Lagrange Equations hold for path $Q(t)$ in the new coordinates then they hold for the related path in the old coordinates.

This behavior is sometimes referred to as **covariance under general geometrical coordinate transformations**[17] The evolution of a physical system should not depend on the coordinates used to describe it, and we have shown that paths calculated by the "least action" procedure enjoy this property.

7. Invariance of the Lagrangian With Respect to Parameterized Families of Transformations

First, we will consider **general invertible transformations**

$$\boldsymbol{T_\varepsilon :} \mathbb{R}\times\mathbb{R}^{\boldsymbol{N}} \to \mathbb{R}\times\mathbb{R}^{\boldsymbol{N}} \qquad \boldsymbol{(t,q) \longrightarrow (t_\varepsilon, q_\varepsilon)}$$

defined for real parameter ε, any one of all values in some (possibly tiny) interval around 0 and with the property that T_0 is the identity map on $\mathbb{R}\times\mathbb{R}^N$ and

$$\lim_{\varepsilon\to 0} T_\varepsilon(t,q) = (t,q) \quad \text{for all } (t,q)\in\mathbb{R}\times\mathbb{R}^N.$$

These are transformations on extended configuration-space.

As we did with L and the members of $\mathcal{S}$, we presume as much differentiability of T_ε as we require for a calculation, *including* differentiability with respect to ε.

Any path q from $\mathcal{S}$ has a graph which is a collection of points $(t,q(t))$ in $\mathbb{R}\times\mathbb{R}^N$ and T_ε shifts each such point to a new point

$$T_\varepsilon(t,q(t)) = (t_\varepsilon(t,q(t)),\, q_\varepsilon(t,q(t))) \in \mathbb{R}\times\mathbb{R}^N.$$

These points form the *graph* of a new path defined on the interval $[a',b'] = [t_\varepsilon(a,q(a)), t_\varepsilon(b,q(b))]$ and the new function that *has* this graph

$$Q_\varepsilon \colon [a',b'] \to \mathbb{R}^N$$

has values $Q_\varepsilon(w_\varepsilon) = q_\varepsilon(t,q(t))$ whenever $w_\varepsilon = t_\varepsilon(t,q(t))$.

In other words, $Q_\varepsilon \circ w_\varepsilon(t) = q_\varepsilon(t,q(t))$.

For each ε we define functional Φ_ε on our family of paths $\mathcal{S}$ by[18]

$$\boldsymbol{\Phi_\varepsilon(q) = \int_{a'}^{b'} L\left(w_\varepsilon, Q_\varepsilon(w_\varepsilon), \frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon)\right) dw_\varepsilon}$$

where $a' = t_\varepsilon(a,q(a))$ and $b' = t_\varepsilon(b,q(b))$.

For each ε, each specific path q creates a new path Q_ε with a different time parameter given by w_ε whose values vary from a' to b'.

The two time parameters

$$w_\varepsilon \longleftrightarrow t$$

---

[17]These are also called **point transformations**.

[18]In this section *none* of the paths are assumed to be stationary unless that is specified!

are related to each other via invertible and differentiable functions. When regarded as a function of t, we have

$$w_\varepsilon(t) = t_\varepsilon(t, q(t)).$$

$\Phi_0(q) = \Gamma(q)$, the action integral on path q.

$\Phi_\varepsilon(q)$ will be close to $\Gamma(q)$ for small ε, given our smoothness assumptions on everything in sight, but *how* close?

It looks like

$$\Phi_\varepsilon(q) = \int_{a'}^{b'} L\left(w_\varepsilon, Q_\varepsilon(w_\varepsilon), \frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon)\right) dw_\varepsilon \quad \text{and} \quad \Phi_0(q) = \int_a^b L\left(t, q(t), \dot{q}(t)\right) dt$$

might be simply related by some kind of a "change-of-variable" formula, but *this is not the case.*

Consider a simple example where $w_\varepsilon(t) = t$ and $Q_\varepsilon(t) = q(t) + \varepsilon v$ for a fixed vector v.

$$\Phi_\varepsilon(q) = \int_a^b L\left(t, q(t) + \varepsilon v, \dot{q}(t)\right)\right) dt.$$

The path $q(t) + \varepsilon v$ is shifted away from $q(t)$ and there is no reason, with our conditions on L, that the two integrands, though close numerically for tiny ε, should be otherwise related.

The difference $\Phi_\varepsilon(q) - \Phi_0(q)$ can be calculated as

$$\begin{aligned}
&\Phi_\varepsilon(q) - \Phi_0(q) \\
&\quad = \int_{a'}^{b'} L\left(w_\varepsilon, Q_\varepsilon(w_\varepsilon), \frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon)\right) dw_\varepsilon - \int_a^b L\left(t, q(t), \dot{q}(t)\right) dt \\
&\quad = \int_a^b \left[L\left(w_\varepsilon(t), Q_\varepsilon(w_\varepsilon(t)), \frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon(t))\right) \frac{dw_\varepsilon}{dt} - L\left(t, q(t), \dot{q}(t)\right)\right] dt \\
&\quad = \int_a^b \left(L_\varepsilon \frac{dw_\varepsilon}{dt} - L\right) dt
\end{aligned}$$

where we define L_ε as the specific indicated function of time—that is, L composed with the shifted path as indicated.

Later we will need to look more carefully at L_ε, so let's begin that process now.

$L_\varepsilon = L_\varepsilon(t)$ was created using, and depends specifically on, ε and q.

Noting $Q_\varepsilon(w_\varepsilon) = q_\varepsilon(t, q(t))$ when $w_\varepsilon = t_\varepsilon(t, q(t))$ we have

$$\begin{aligned}
L_\varepsilon(t) &= L\left(w_\varepsilon(t), Q_\varepsilon(w_\varepsilon(t)), \frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon(t))\right) \\
&= L\left(t_\varepsilon(t, q(t)), q_\varepsilon(t, q(t)), \frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon(t))\right).
\end{aligned}$$

Calculating $\frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon(t))$ will be tricky, and we will need to do that. When we get more structure in our transformations we will get at it by noting that

$$\frac{dQ_\varepsilon}{dw_\varepsilon} = \frac{d}{dw_\varepsilon} Q_\varepsilon \circ w_\varepsilon \circ t = \frac{dQ_\varepsilon \circ w_\varepsilon}{dt} \frac{dt}{dw_\varepsilon} = \frac{dQ_\varepsilon \circ w_\varepsilon}{dt} \Big/ \frac{dw_\varepsilon}{dt}. \tag{3}$$

Back to our main track, we can calculate the difference $\Phi_\varepsilon(q) - \Phi_0(q)$ as

$$\Phi_\varepsilon(q) - \Phi_0(q) = \int_a^b \left(L_\varepsilon \frac{dw_\varepsilon}{dt} - L \right) dt.$$

We say the action integral $\Gamma = \Phi_0$ is **invariant under this family of transformations** provided

$$\lim_{\varepsilon \to 0} \frac{\Phi_\varepsilon(q) - \Phi_0(q)}{\varepsilon} = 0.$$

or, equivalently, if for *every* path q

$$\left. \frac{d\Phi_\varepsilon(q)}{d\varepsilon} \right|_{\varepsilon=0} = 0.$$

Because q—and hence the interval $[a, b]$—is arbitrary[19], the condition on

$$\int_a^b \left(L_\varepsilon \frac{dw_\varepsilon}{dt} - L \right) dt$$

is equivalent to the same restriction applied pointwise to the integrand (that is, for each t) and this condition is easier to apply so we adopt it.

Γ is therefore said to be invariant under this family of transformations provided the **invariance condition** is satisfied:

$$\lim_{\varepsilon \to 0} \frac{L_\varepsilon \frac{dw_\varepsilon}{dt} - L}{\varepsilon} = 0$$

for each q at each time in the domain of q.

We now get more specific about our family of transformations by positing that t_ε and q_ε take a certain form.

We assume

$$\boldsymbol{t_\varepsilon = t + \varepsilon\tau + \varepsilon^2 f} \quad \text{and} \quad \boldsymbol{q_\varepsilon = q + \varepsilon\zeta + \varepsilon^2 g}$$

where $\tau = \tau(t, q)$ and $f = f(t, q)$ are real functions and $\zeta = \zeta(t, q)$ and $g = g(t, q)$ are functions into $\mathbb{R}^N$, and all four are fixed functions—that is, independent of ε—and therefore characterize the whole family of transformations.

We will focus on τ and ζ, and say they **generate this family of transformations.**[20]

f and g will be irrelevant in our subsequent calculations.

Suppose given a family of transformations with generators τ and ζ.

We will find an explicit condition on the generators and q that is equivalent to the invariance condition

$$\lim_{\varepsilon \to 0} \frac{L_\varepsilon \frac{dw_\varepsilon}{dt} - L}{\varepsilon} = 0$$

for the family of transformations applied to this lagrangian and each time.

[19]The unlimited differentiability condition on T_ε implies that $L_\varepsilon \frac{dw_\varepsilon}{dt}$ has continuous derivative with respect to ε and also allows us to apply Leibniz's integral rule and differentiate past the integral sign.

[20]In fact, the main result to follow (Noether's First Theorem) only requires consideration of coordinate transformations that move paths by "infinitesimal" amounts, corresponding to a "tiny" interval of ε around 0, and that only in some small neighborhood of a path.

Note that for each **fixed time** this is the derivative with respect to ε of the function $L_\varepsilon \frac{dw_\varepsilon}{dt} - L$ evaluated at $\varepsilon = 0$.

The fact we aim to prove next is the equivalence of the invariance condition to an equation called **The Invariance Identity**.

$$\lim_{\varepsilon \to 0} \frac{L_\varepsilon \frac{dw_\varepsilon}{dt} - L}{\varepsilon} = 0 \qquad \text{(The Invariance Condition)}$$

if and only if

$$\frac{\partial L}{\partial q^i}\zeta^i + p_i \dot{\zeta}^i + \frac{\partial L}{\partial t}\tau - H\dot{\tau} = 0 \qquad \text{(The Invariance Identity)}$$

for *every* path q, where $\dot{\tau} = \frac{d\tau}{dt}$ and $\dot{\zeta}^i = \frac{d\zeta^i}{dt}$ and $p_i = \frac{\partial L}{\partial \dot{q}^i}$ and $H = p_i \dot{q}^i - L$.

Proof:

Recall that for each **fixed time** the invariance condition is that the derivative with respect to ε of the function $L_\varepsilon \frac{dw_\varepsilon}{dt} - L$ evaluated at $\varepsilon = 0$ is 0.

L itself is independent of ε and so has derivative 0.

Differentiating the first term by the product rule gives

$$\frac{d}{d\varepsilon}\left(L_\varepsilon \frac{dw_\varepsilon}{dt} - L\right) = \frac{dw_\varepsilon}{dt}\frac{d}{d\varepsilon}L_\varepsilon + L_\varepsilon \frac{d}{d\varepsilon}\frac{dw_\varepsilon}{dt}. \tag{4}$$

Let's begin identifying these four factors evaluated at $\varepsilon = 0$.

$$\frac{dw_\varepsilon}{dt}\frac{d}{d\varepsilon}L_\varepsilon + L_\varepsilon \frac{d}{d\varepsilon}\frac{dw_\varepsilon}{dt} \quad \text{is to be evaluated at } \varepsilon = 0.$$

First, when L_ε is evaluated at $\varepsilon = 0$ we have

$$L_0 = L. \tag{5}$$

Next, as a function of time w_ε is given as $w_\varepsilon = t + \varepsilon\tau + \varepsilon^2 g$ so

$$\frac{dw_\varepsilon}{dt} = 1 + \varepsilon\dot{\tau} + \varepsilon^2 \dot{g} \tag{6}$$

which is 1 when evaluated at $\varepsilon = 0$.

Also

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \frac{dw_\varepsilon}{dt} = \dot{\tau}. \tag{7}$$

This gives us 3 out of the 4 factors when evaluated at $\varepsilon = 0$.

All that remains is to calculate the fourth factor $\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} L_\varepsilon$, and we will do that by applying the chain rule.

For each fixed t (remember, ε will be the variable here) we have

$$\begin{aligned} L_\varepsilon(t) &= L\left(w_\varepsilon(t), Q_\varepsilon(w_\varepsilon(t)), \frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon(t))\right) \\ &= L(A, B, C). \end{aligned}$$

$$A = t + \varepsilon\tau(t, q(t)) + \varepsilon^2 f(t, q(t)) \Rightarrow \left.\frac{dA}{d\varepsilon}\right|_{\varepsilon=0} = \tau. \tag{8}$$

$$B = q(t) + \varepsilon\zeta(t, q(t)) + \varepsilon^2 g(t, q(t)) \Rightarrow \left.\frac{dB}{d\varepsilon}\right|_{\varepsilon=0} = \zeta. \tag{9}$$

Going back to Equation (4) we have

$$C = \frac{dQ_\varepsilon}{dw_\varepsilon}(w_\varepsilon(t)) = \frac{\frac{dQ_\varepsilon \circ w_\varepsilon}{dt}}{\frac{dw_\varepsilon}{dt}} = \frac{\frac{d}{dt} q_\varepsilon(t, q(t))}{\frac{dw_\varepsilon}{dt}} = \frac{\dot{q} + \varepsilon\dot{\zeta} + \varepsilon^2 \dot{g}}{1 + \varepsilon\dot{\tau} + \varepsilon^2 \dot{f}}. \tag{10}$$

We still need to calculate $\left.\frac{dC}{d\varepsilon}\right|_{\varepsilon=0}$ which we do by the quotient rule applied to

$$C = \frac{\dot{q} + \varepsilon\dot{\zeta} + \varepsilon^2 \dot{g}}{1 + \varepsilon\dot{\tau} + \varepsilon^2 \dot{f}}.$$

Any term not involving ε will differentiate to 0 as that rule is applied. And any term possessing an ε factor *after* the differentiation procedure will evaluate to 0 when ε is finally set to 0.

So we have

$$\left.\frac{dC}{d\varepsilon}\right|_{\varepsilon=0} = \frac{\dot{\zeta} \cdot 1 - \dot{q}\dot{\tau}}{1^2} = \dot{\zeta} - \dot{q}\dot{\tau}. \tag{11}$$

Using Equations (8), (9) and (11) we can calculate the fourth term as

$$\begin{aligned}
\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} L_\varepsilon &= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} L(A, B, C) \\
&= \frac{\partial L}{\partial t}\frac{dA}{d\varepsilon} + \frac{\partial L}{\partial q^i}\frac{dB^i}{d\varepsilon} + \frac{\partial L}{\partial \dot{q}^i}\frac{dC^i}{d\varepsilon} \\
&= \frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\zeta^i + \frac{\partial L}{\partial \dot{q}^i}\left(\dot{\zeta}^i - \dot{q}^i\dot{\tau}\right) \\
&= \frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\zeta^i + p_i\left(\dot{\zeta}^i - \dot{q}^i\dot{\tau}\right).
\end{aligned} \tag{12}$$

Resurrecting the four factors, evaluated at $\varepsilon = 0$, using Equations (5), (6), (7) and (12) we have

$$\begin{aligned}
&\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0}\left(L_\varepsilon \frac{dw_\varepsilon}{dt} - L\right) \\
&\quad = \frac{dw_\varepsilon}{dt}\frac{d}{d\varepsilon}L_\varepsilon + L_\varepsilon \frac{d}{d\varepsilon}\frac{dw_\varepsilon}{dt} \quad \text{(evaluated at } \varepsilon = 0) \\
&\quad = 1 \cdot \left(\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} L_\varepsilon\right) + L_0\dot{\tau} \\
&\quad = \frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\zeta^i + p_i(\dot{\zeta}^i - \dot{q}^i\dot{\tau}) + L\dot{\tau} \\
&\quad = \frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\zeta^i + p_i\dot{\zeta}^i - \left(p_i\dot{q}^i - L\right)\dot{\tau} \\
&\quad = \frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\zeta^i + p_i\dot{\zeta}^i - H\dot{\tau}.
\end{aligned}$$

When

$$0 = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0}\left(L_\varepsilon \frac{dw_\varepsilon}{dt} - L\right) = \frac{dw_\varepsilon}{dt}\frac{d}{d\varepsilon}L_\varepsilon + L_\varepsilon \frac{d}{d\varepsilon}\frac{dw_\varepsilon}{dt} \quad (\text{at } \varepsilon = 0)$$

we have

$$0 = \frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\zeta^i + p_i\dot{\zeta}^i - H\dot{\tau}.$$

This is the **invariance identity.**

These terms can be untangled to produce the 4 terms of the product rule applied to $L_\varepsilon \frac{dw_\varepsilon}{dt} - L$, so if the invariance identity holds so too does the invariance condition and the theorem is proved. □

In many applications the invariance condition is not exactly satisfied but a related condition *is* satisfied that is sufficient to create a conservation law, which is our ultimate goal.

Instead of

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0}\left(L_\varepsilon \frac{dt_\varepsilon}{dt} - L\right) = 0 \quad \text{for every } q \text{ and every time}$$

suppose given a function F defined on $\mathbb{R} \times \mathbb{R}^N$ and consider the terms

$$\nabla F = \left(\frac{\partial F}{\partial t}, \frac{\partial F}{\partial q^1}, \ldots, \frac{\partial F}{\partial q^N}\right).$$

∇F can be used to produce a function

$$(t, q, v) \to \nabla F \cdot (1, v) = \frac{\partial F}{\partial t} + \frac{\partial F}{\partial q^i}v^i$$

defined on all of $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$, and if $q = q(t)$ is an arbitrary path we have

$$\dot{F} = \frac{d}{dt}F(t, q(t)) = \frac{\partial F}{\partial t} + \frac{\partial F}{\partial q^i}\dot{q}^i = \nabla F \cdot (1, \dot{q}).$$

So suppose we have a function F of this kind producing, when composed with any path in this way, a total time derivative. Suppose also that

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0}\left(L_\varepsilon \frac{dt_\varepsilon}{dt} - L\right) = \dot{F} \quad \text{for every path } q \text{ and every time.}$$

We say L is **divergence-invariant** in this case, or **invariant up to the divergence term F**.

Divergence-invariance produces a modified equivalence which we state now.

The Divergence-Invariance Identity:

$$\lim_{\varepsilon \to 0} \frac{L_\varepsilon \frac{dw_\varepsilon}{dt} - L}{\varepsilon} = \dot{F}$$

if and only if

$$\frac{\partial L}{\partial q^i}\zeta^i + p_i\dot{\zeta}^i + \frac{\partial L}{\partial t}\tau - H\dot{\tau} = \dot{F}$$

for *every* path q, where $\dot{\tau} = \frac{d\tau}{dt}$ and $\dot{\zeta}^i = \frac{d\zeta^i}{dt}$ and $p_i = \frac{\partial L}{\partial \dot{q}^i}$ and $H = p_i\dot{q}^i - L$.

The original invariance identity is a special case where $F = 0$.

As a side note there is an interesting relationship between Φ_ε and F. You may recall that before we began considering

$$\frac{d}{d\varepsilon}\left(L_\varepsilon \frac{dw_\varepsilon}{dt} - L\right)$$

at individual times along a path q we were concerned with the global condition

$$\Phi_\varepsilon(q) - \Phi_0(q) = \int_a^b \left(L_\varepsilon \frac{dw_\varepsilon}{dt} - L\right) dt$$

on the whole path q.

If L is invariant up to divergence term F then

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \Phi_\varepsilon(q) = \int_a^b \frac{d}{dt} F(t, q(t))\, dt = F(b, q(b)) - F(a, q(a)).$$

So the derivative of $\Phi_\varepsilon(q)$ at $\varepsilon = 0$, though defined as an integral using the entire curve q, depends only on the endpoints of this path.

8. Conservation Laws: Noether's First Theorem for One Independent Variable

We are now ready to prove Noether's Theorem for one independent variable. This is merely a matter of rearranging the divergence-invariance identity and making an observation.

We assume

$$\frac{\partial L}{\partial q^i}\zeta^i + p_i \dot{\zeta}^i + \frac{\partial L}{\partial t}\tau - H\dot{\tau} = \dot{F}$$

and that q is a stationary path so the Euler-Lagrange identities hold:

$$\frac{\partial L}{\partial t} = -\dot{H} \quad \text{and} \quad \frac{\partial L}{\partial q^i} = \dot{p}_i.$$

This turns the divergence-invariance identity into:

$$\dot{p}_i \zeta^i + p_i \dot{\zeta}^i - \dot{H}\tau - H\dot{\tau} = \dot{F}.$$

Noether's First Theorem For One Independent Variable:

If L is invariant up to the divergence term F for the family of transformations with generators τ and ζ, and if q is a stationary path then

$$H\tau + F - p_i\zeta^i \quad \text{is conserved (i.e. is constant) on path } q.$$

Proof:

Given:

$$\dot{p}_i \zeta^i + p_i \dot{\zeta}^i - \dot{H}\tau - H\dot{\tau} = \dot{F}.$$

we have:

$$0 = -\dot{p}_i \zeta^i - p_i \dot{\zeta}^i + \dot{H}\tau + H\dot{\tau} + \dot{F} \Longrightarrow 0 = \frac{d}{dt}\left(H\tau + F - p_i\zeta^i\right).$$

The theorem follows. □

9. The Killing Equations

Known constants-of-an-interaction put conditions on the type of lagrangian that governs the physics, and so are useful if you are creating a new theory: that is, creating a lagrangian from scratch.

On the other hand, if you are given a known lagrangian it is not always clear how to find generators of symmetries for it, or even if there *are any* symmetries.

We will consider this aspect now, introducing the **Killing Equations.**[21].

The underlying idea is fairly simple, and will produce *necessary* conditions on generators that quite often are sufficient for invariance of the lagrangian with respect to transformations you actually produce, and to rule out the possibility of invariant families of certain types.

The invariance identity

$$\frac{\partial L}{\partial q^i}\zeta^i + p_i\dot{\zeta}^i + \frac{\partial L}{\partial t}\tau - H\dot{\tau} = \dot{F}$$

must hold for *any* path q including, for instance, those paths for which $\dot{q}^i = 1$ for a particular i and $q^k = 0$ for $k \neq i$.

Replacing H by $p_i\dot{q}^i - L$ and p_i by $\frac{\partial L}{\partial \dot{q}^i}$ we have

$$\frac{\partial L}{\partial q^i}\zeta^i + \frac{\partial L}{\partial \dot{q}^i}\dot{\zeta}^i + \frac{\partial L}{\partial t}\tau - \left(\frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L\right)\dot{\tau} = \dot{F}.$$

Recall $\tau = \tau(t,q)$ and $\zeta = \zeta(t,q)$ so

$$\dot{\tau} = \frac{\partial \tau}{\partial t} + \frac{\partial \tau}{\partial q^k}\dot{q}^k \quad \text{and} \quad \dot{\zeta}^i = \frac{\partial \zeta^i}{\partial t} + \frac{\partial \zeta^i}{\partial q^k}\dot{q}^k.$$

Substituting these into the invariance equation produces

$$\begin{aligned}\dot{F} = \frac{\partial L}{\partial q^i}\zeta^i &+ \frac{\partial L}{\partial \dot{q}^i}\left(\frac{\partial \zeta^i}{\partial t} + \frac{\partial \zeta^i}{\partial q^k}\dot{q}^k\right)\\ &+ \ \frac{\partial L}{\partial t}\tau - \left(\frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L\right)\left(\frac{\partial \tau}{\partial t} + \frac{\partial \tau}{\partial q^k}\dot{q}^k\right).\end{aligned}$$

Expanding and re-ordering the eight terms gives

$$\begin{aligned}\dot{F} = L\frac{\partial \tau}{\partial t} + L\frac{\partial \tau}{\partial q^k}\dot{q}^k + \frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\zeta^i + \frac{\partial L}{\partial \dot{q}^i}\frac{\partial \zeta^i}{\partial t}\\ + \frac{\partial L}{\partial \dot{q}^i}\frac{\partial \zeta^i}{\partial q^k}\dot{q}^k - \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i\frac{\partial \tau}{\partial t} - \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i\frac{\partial \tau}{\partial q^k}\dot{q}^k.\end{aligned}$$

This provides numerous necessary relationships that must be satisfied by τ and the ζ^i, revealed by particular choices of q in conjunction with guesses about generators τ and ζ, such as $\tau = 0$ or $\tau = 1$.

These necessary relationships can suggest a path to generators of families of transformations for which L is invariant.

[21]Named after Wilhem Killing (1847-1923).

Let's examine an example from classical physics, a Newtonian particle subject to lagrangian with time-dependent potential

$$L(t,x,\dot{x}) = \frac{1}{2}m\dot{x}^2 - U(t,x).$$

So $\frac{\partial L}{\partial x} = -\frac{\partial U}{\partial x}$ and $\frac{\partial L}{\partial \dot{x}} = m\dot{x}$ and $\frac{\partial L}{\partial t} = -\frac{\partial U}{\partial t}$.

Feeding this into the invariance identity with $\dot{F} = 0$ yields

$$\begin{aligned}0 &= L\frac{\partial \tau}{\partial t} + L\frac{\partial \tau}{\partial q^k}\dot{q}^k + \frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\zeta^i + \frac{\partial L}{\partial \dot{q}^i}\frac{\partial \zeta^i}{\partial t} + \frac{\partial L}{\partial \dot{q}^i}\frac{\partial \zeta^i}{\partial q^k}\dot{q}^k - \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i\frac{\partial \tau}{\partial t} - \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i\frac{\partial \tau}{\partial q^k}\dot{q}^k \\ &= \left(\frac{1}{2}m\dot{x}^2 - U\right)\frac{\partial \tau}{\partial t} + \left(\frac{1}{2}m\dot{x}^2 - U\right)\frac{\partial \tau}{\partial x}\dot{x} - \frac{\partial U}{\partial t}\tau - \frac{\partial U}{\partial x}\zeta + m\dot{x}\frac{\partial \zeta}{\partial t} + m\dot{x}\frac{\partial \zeta}{\partial x}\dot{x} - m\dot{x}\dot{x}\frac{\partial \tau}{\partial t} - m\dot{x}\dot{x}\frac{\partial \tau}{\partial x}\dot{x} \\ &= \left(-U\frac{\partial \tau}{\partial t} - \frac{\partial U}{\partial t}\tau - \frac{\partial U}{\partial x}\zeta\right) + \left(-U\frac{\partial \tau}{\partial x} + m\frac{\partial \zeta}{\partial t}\right)\dot{x} + \left(m\frac{\partial \zeta}{\partial x} - \frac{1}{2}m\frac{\partial \tau}{\partial t}\right)\dot{x}^2 + \left(-\frac{1}{2}m\frac{\partial \tau}{\partial x}\right)\dot{x}^3.\end{aligned}$$

Since path x is arbitrary, each coefficient of this polynomial in $\dot{x}$ must be the zero function. Setting these four coefficients to 0, individually, yields the four **Killing equations** for this lagrangian.

$$0 = \left(-U\frac{\partial \tau}{\partial t} - \frac{\partial U}{\partial t}\tau - \frac{\partial U}{\partial x}\zeta\right) + \left(-U\frac{\partial \tau}{\partial x} + m\frac{\partial \zeta}{\partial t}\right)\dot{x} + \left(m\frac{\partial \zeta}{\partial x} - \frac{1}{2}m\frac{\partial \tau}{\partial t}\right)\dot{x}^2 + \left(-\frac{1}{2}m\frac{\partial \tau}{\partial x}\right)\dot{x}^3.$$

The third degree term tells us τ must be a function of t alone. The first degree term requires, then, that ζ depends on x alone.

The second degree term tells us, then, that both $\frac{d\zeta}{dx}$ and $\frac{d\tau}{dt}$ must be constant with $\frac{d\tau}{dt}$ twice as big as $\frac{d\zeta}{dx}$.

Let's suppose $\frac{d\zeta}{dx} = C$. Then $\zeta = Cx + B$ and $\frac{d\tau}{dt} = 2C$ and $\tau = 2Ct + A$, for constants A and B.

The first Killing equation

$$\begin{aligned}-U\frac{\partial \tau}{\partial t} - \frac{\partial U}{\partial t}\tau - \frac{\partial U}{\partial x}\zeta &= 0 \\ -U2C - \frac{\partial U}{\partial t}(2Ct + A) - \frac{\partial U}{\partial x}(Cx + B) &= 0\end{aligned}$$

puts severe restriction on the form of U—remember, we *have* U in this scenario. If $C = 0$ we must have

$$\frac{\partial U}{\partial t}A = -\frac{\partial U}{\partial x}B$$

for constants A and B. And if $C \neq 0$ we must have

$$U = -\frac{\partial U}{\partial t}\left(\frac{2Ct + A}{2C}\right) - \frac{\partial U}{\partial x}\left(\frac{Cx + B}{2C}\right)$$

for constants A and B.

These conditions can either be satisfied for certain constant A, B and C or not. If not there are no invariant families for this lagrangian. If yes these constants will be easy to determine and we have generators for an invariant family.

The modification to allow for divergence-invariance requires modifying the first Killing equation to

$$-U\frac{\partial \tau}{\partial t} - \frac{\partial U}{\partial t}\tau - \frac{\partial U}{\partial x}\zeta = \dot{F}$$

$$-U2C - \frac{\partial U}{\partial t}(2Ct + A) - \frac{\partial U}{\partial x}(Cx + B) = \dot{F}$$

for some divergence term F.

We may be able to identify the left-hand-side as a total derivative on arbitrary path x of some function $F = F(t, x(t))$ for certain constant A, B and C.

If we can do that, we then proceed to find conserved quantities via Noether's Theorem in the divergence-invariance setting too.

10. **The Hamiltonian**

In preparation for later efforts we are going to convert, here and there, to a somewhat more explicit and/or modern notation.

First of all, we may use $\boldsymbol{\partial_0}$ in place of $\frac{\partial}{\partial t}$ and $\boldsymbol{\partial_i}$ in place of $\frac{\partial}{\partial q^i}$ for $i = 1, \ldots, N$.

With a function like the lagrangian which traditionally operates on a dynamical variable symbol we will feel free to use $\boldsymbol{\partial_{N+i}}$ in place of $\frac{\partial}{\partial \dot{q}^i}$ for $i = 1, \ldots, N$.

Soon we will define functions which uses a conjugate momentum symbol in their last arguments and we may use $\boldsymbol{\partial_{N+i}}$ to replace $\frac{\partial}{\partial p_i}$ for $i = 1, \ldots, N$ as well when calculating partial derivatives at these last "slots."

Thus the subscript or superscript on ∂ refers to the location of the variable in the list of arguments, not the symbol for that variable, which can vary from place to place.

The conjugate momenta and hamiltonian

$$p_i = \partial_{N+i} L \quad \text{and} \quad H = p_i \dot{q}^i - L$$

were defined as functions of time along a path q, and the symbol $\dot{q}^i$ in the formula for H referred to the time derivative of path q. When the path is a *solution* path physicists say the calculation is "**on-shell**," and by that they mean they are free to use the Euler-Lagrange equations

$$\partial_i L = \frac{d}{dt}\left(\partial_{N+i} L\right) \qquad i = 1, \ldots, N$$

which, in fact, are used to *calculate* stationary solutions. On-shell we have the ELS "H dot" and "p dot" equations, for instance:

$$\partial_0 L = -\dot{H} \qquad \text{and} \qquad \partial_i L = \dot{p}_i \quad i = 1, \ldots, N$$

which tell us when the hamiltonian or a conjugate momentum will be conserved along a solution path.

However the *formulas* for p_i and H can be evaluated as functions of (t, q, v) on all of $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$ without regard to any path, a situation referred to as "**off-shell,**" where v is a generic dynamical vector and unrelated to q.

The momentum covector

$$p = (p_1, \ldots, p_N) = \left(\left[\partial_{N+1} L\right](t, q, v), \ldots, \ \left[\partial_{2N} L\right](t, q, v)\right)$$

is given through L in terms of (t, q, v).

Dependence of this off-shell interpretation of H on v can often, at least *in principle*, be eliminated in favor of p and we discuss that now.

For each fixed t and q we have a "momentum function" K defined for generic dynamical vector v which produces a vector in $\mathbb{R}^N$:

$$K(t,q,\ \cdot\ ) \colon \mathbb{R}^N \to \mathbb{R}^N \qquad \text{(fixed } t \text{ and } q\text{)}$$

which is given by

$$K(t,q,v) = \left([\partial_{N+1} L]\,(t,q,v), \ldots,\ [\partial_{2N} L]\,(t,q,v) \right). \qquad \text{(fixed } t \text{ and } q\text{)}$$

So we have $p_i = K_i(t,q,v) = [\partial_{N+i} L]\,(t,q,v)$ for $i = 1, \ldots, N$.

$K(t,q,v)$ is the momentum covector which would be produced if, at time t, you were on a path passing through point q with velocity v.

When the derivative of this map has nonzero determinant[22] it is invertible, and we will presume this to be true for each t and q.

The inverse will be a function

$$J(t,q,\ \cdot\ ) \colon \mathbb{R}^N \to \mathbb{R}^N. \qquad \text{(fixed } t \text{ and } q\text{)}$$

defined by

$$J(t,q,p) = v \text{ exactly when } p = \left(\partial_{N+1} L(t,q,v), \ldots,\ \partial_{2N} L(t,q,v) \right).$$

In other words, **$p_i = \partial_{N+i} L(t,q,J(t,q,p))$ for each i by *definition of* J.**

So here $v^i = J^i(t,q,p)$ for $i = 1, \ldots, N$.

$J(t,q,p)$ is the unique velocity that would be required to produce momentum p if you were passing through point q at time t.

In some special cases it may actually be possible to perform this inversion to produce a formula for J algebraically, but in any case *there exists* an inverse function and accurate numerical approximations to it can be obtained.

The collection of all coordinate pairs $(q,p) \in \mathbb{R}^N \times \mathbb{R}^N$ where p is to be interpreted as a generalized momentum conjugate to generalized coordinate q via a lagrangian L is called the **phase-space**[23] and with this interpretation on the last two factors $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$ is called the **extended phase-space**.

We can create a *new* function $\mathcal{L} = \mathcal{L}(t,q,p)$ defined on extended phase-space given by

$$\boldsymbol{\mathcal{L}(t,q,p) = L(t,q,J(t,q,p))}$$

closely related to the lagrangian called the **hamiltonian-lagrangian**.

With paths q and derivative $\dot{q}$ and *associated* conjugate momentum p as above we have

$$\boldsymbol{\mathcal{L}(t,q(t),p(t)) = L(t,q(t),\dot{q}(t))}$$

but in fact $\mathcal{L}$ is defined for any triple (t,q,p) whether there is a path involved or not, just as L is defined for any triple (t,q,v), where v may be a generic dynamical vector unrelated to q.

---

[22]This is a necessary condition on the lagrangian. To proceed further the determinant must be nonzero in a neighborhood of a solution path, at least. We assume here that there is a global inverse function.

[23]In more advanced treatments the conjugate momenta (the various p covectors) will correspond to members of the cotangent space at each point q of the configuration manifold.

We then can create a new function defined on the extended phase-space, called the **hamiltonian**,

$$\mathcal{H}\colon \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$$

which is frequently defined using the shorthand formula

$$\mathcal{H} = p_i v^i - \mathcal{L}$$

but is more properly given explicitly as

$$\mathcal{H}(t,q,p) = p_i J^i(t,q,p) - L(t,q,J(t,q,p)).$$

This is *not* the function we were calling H before, which was a function of time defined only along paths in geometrical space.

This new hamiltonian is a composit function, the formula for the old H composed with the new function J, which depends on t, q and p. It is defined both on-shell, where the Euler-Lagrange and the "dot" equations hold, but also off-shell too, for any (t,q,p) in extended phase-space $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$.

This new $\mathcal{H}$, as related off-shell to the old L by the above formula, is called a **Legendre transformation of L.**[24]

(13)

$$\begin{aligned}
\partial_{N+i}\mathcal{H}(t,q,p) &= \partial_{N+i}\left[p_j J^j(t,q,p) - L(t,q,J(t,q,p))\right] \\
&= J^i(t,q,p) + p_j\left[\partial_{N+i}J^j\right](t,q,p) - \partial_{N+i}L(t,q,J(t,q,p)) \\
&= J^i(t,q,p) + p_j\left[\partial_{N+i}J^j\right](t,q,p) - \left[\partial_{N+j}L\right](t,q,J(t,q,p))\left[\partial_{N+i}J^j\right](t,q,p) \\
&= J^i(t,q,p) + p_j\left[\partial_{N+i}J^j\right](t,q,p) - p_j\left[\partial_{N+i}J^j\right](t,q,p) \\
&= J^i(t,q,p) = v^i
\end{aligned}$$

where v is the velocity required to produce momentum p at time t and place q.

It is entirely possible—*if you can keep the dependencies straight*—to do the same calculation without mention of J, and most sources follow that route. You may prefer this too, for obvious reasons. We have:

$$\begin{aligned}
\frac{\partial \mathcal{H}}{\partial p_i} &= \frac{\partial}{\partial p_i}\left(p_j v^j - \mathcal{L}\right) = v^i + p_j\frac{\partial v^j}{\partial p_i} - \frac{\partial}{\partial p_i}L(t,q,v) \\
&= v^i + p_j\frac{\partial v^j}{\partial p_i} - \frac{\partial L}{\partial v^j}\frac{\partial v^j}{\partial p_i} = v^i + p_j\frac{\partial v^j}{\partial p_i} - p_j\frac{\partial v^j}{\partial p_i} = v^i.
\end{aligned}$$

We do not assume that v is derivative $\dot{q}$ of a path q. Though that situation will occur, we want to be careful not to make the presumption.

Examining either calculation we see that the Legendre transformation of $\mathcal{H}$ applied to the p variables will reproduce L.

$$\text{Lagrangian} = L \quad \longleftrightarrow \quad \text{Hamiltonian} = \mathcal{H}$$

is just one example of a pair related this way in applications.

[24] There is a different Legendre transformation for each choice of domain variables. The Legendre transform we work with here corresponds to choosing $\dot{q}$.

For instance, in thermodynamics the Helmholtz free energy and Gibbs energy are obtained by performing Legendre transforms on the internal energy and enthalpy, respectively.

Another calculation we can carry out off-shell from the raw definition of $\mathcal{H}$ is

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial q_i} &= \frac{\partial}{\partial q_i}\left(p_j v^j - \mathcal{L}\right) = p_j \frac{\partial v^j}{\partial q_i} - \frac{\partial}{\partial q_i} L(t,q,v) \\ &= p_j \frac{\partial v^j}{\partial q_i} - \frac{\partial L}{\partial q^i} - \frac{\partial L}{\partial v^j}\frac{\partial v^j}{\partial q^i} = -\frac{\partial L}{\partial q^i}. \end{aligned} \tag{14}$$

We learned earlier that on-shell (that is, along a solution path) the "p-dot" form of the ELS is $\frac{\partial L}{\partial q^i} = \dot{p}_i$. So we have a system of $2N$ *first order* ordinary differential equations to be solved for paths q and p:

$$\dot{q}^i = \frac{\partial \mathcal{H}}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} \qquad \text{for } i = 1, \ldots, N.$$

These are called **Hamilton's equations.** Solving these equations provides an alternative to the system of N *second order* differential equations given as the ELS.

Expanding on the first group of equations we have, by Equation (13)

$$\frac{dq^i}{dt}(t) = \partial_{N+i}\mathcal{H}(t,q,p) = J^i(t,q,p)$$

and $J(t,q,p)$ is the one and only velocity vector which will produce momentum vector p at time t and location $q(t)$. So the first group of equations tells us that path q is "compatible" in the sense that this path together with its derivative $\dot{q}$ will produce p at every time.

With that assurance in hand the second group of equations provides the ELS.

The major practical advantage in the hamiltonian formulation is that the first derivatives are isolated in Hamilton's equations, while in the original ELS the derivatives are second order and none are isolated. The downside is the necessity of writing the generalized velocities in terms of the conjugate momenta in order to set this all up. If the inversion is easy this method may be preferable. The real benefit in the hamiltonian approach lies in other areas.

Hamilton's equations have lots of consequences. For instance on-shell we have

$$\frac{d\mathcal{H}}{dt} = \frac{\partial \mathcal{H}}{\partial t} + \frac{\partial \mathcal{H}}{\partial q^i}\dot{q}^i + \frac{\partial \mathcal{H}}{\partial p_i}\dot{p}_i = \frac{\partial \mathcal{H}}{\partial t} - \dot{p}_i\dot{q}^i + \dot{q}^i\dot{p}_i = \frac{\partial \mathcal{H}}{\partial t}.$$

This is not surprising: we had a similar result obtained through the lagrangian from before. We saw there that the hamiltonian, defined on a solution path, satisfies $\dot{H} = -\frac{\partial L}{\partial t}$. Explicit time dependence of $\mathcal{H}$ happens through L so this is $\frac{\partial \mathcal{H}}{\partial t}$.

Analysis of a physical system using this hamiltonian and solutions provided by Hamilton's equations is called **hamiltonian mechanics**.

11. **The Poisson Bracket**

Given any two real-valued functions $F = F(t, q, p)$ and $G = G(t, q, p)$ defined on $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$, their **Poisson bracket** $\boldsymbol{\{F, G\}}$ is another function defined on extended phase space given as

$$\boldsymbol{\{F, G\} = \frac{\partial F}{\partial q^i}\frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial G}{\partial q^i}.}$$

Often we will be interested in the Poisson bracket calculated along a phase-space path $(q, p) = (q(t), p(t))$ and in that event the Poisson bracket $\{F, G\}$ composed with this path is a function of time.

Since this bracket does depend on coordinates, it may sometimes be convenient to honor this using the notation $\boldsymbol{\{F, G\} = \{F, G\}_{q,p}}$.

The Poisson bracket satisfies a number of identities. It is **anticommutative** and **bilinear**: that is, for constants a, b and functions E, F and G we have

$$\boldsymbol{\{F, G\} = -\{G, F\}} \quad \text{and} \quad \boldsymbol{\{E, aG + bF\} = a\{E, G\} + b\{E, F\}}.$$

It is easy to check that it also satisfies **Leibniz's rule** and the **Jacobi identity**:

$$\boldsymbol{\{EF, G\} = E\{F, G\} + F\{E, G\}}$$

$$\text{and} \qquad \boldsymbol{\{E, \{F, G\}\} + \{F, \{G, E\}\} + \{G, \{E, F\}\} = 0.}$$

If F is a vector-valued function we apply the bracket to each coordinate function separately, to produce a vector-valued Poisson bracket satisfying these same rules.

Components of the vectors p and q satisfy

$$\{q^m, p_k\} = \frac{\partial q^m}{\partial q^i}\frac{\partial p_k}{\partial p_i} - \frac{\partial q^m}{\partial p_i}\frac{\partial p_k}{\partial q^i} = \delta_{m,k}$$

where $\boldsymbol{\delta_{m,k}} = 1$ if $m = k$ and 0 otherwise. (This is the **Kronecker delta.**) Also

$$\{p_m, p_k\} = \{q^m, q^k\} = 0 \qquad \text{for all } k, m.$$

Time-dependent functions that synopsize information about a trajectory $q = q(t)$ (such as temperature) are important (of course) and the Poisson bracket against the hamiltonian provides a formula for the time-evolution of any function of this type.

We have, on-shell,

$$\{F, \mathcal{H}\} = \frac{\partial F}{\partial q^i}\frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial \mathcal{H}}{\partial q^i} = \frac{\partial F}{\partial q^i}\dot{q}^i + \frac{\partial F}{\partial p_i}\dot{p}_i = \dot{F} - \frac{\partial F}{\partial t}.$$

So if F has no explicit time-dependence then along a solution path we have

$$\dot{F} = \{F, \mathcal{H}\} \qquad \text{(on-shell)}.$$

The intriguing thing about this last relation is that it applies to *any* phase-space function F, and provides a total time derivative at points along a solution path without calculating a time derivative, only partial derivatives with respect to position and momentum. It is called **Hamilton's equation of motion.**

If $(q,p) = (q(t), p(t))$ is any path in phase-space satisfying Hamilton's equations we have

$$\{q^j, \mathcal{H}\} = \frac{\partial q^j}{\partial q^i}\frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial q^j}{\partial p_i}\frac{\partial \mathcal{H}}{\partial q^i} = \frac{\partial \mathcal{H}}{\partial p_j} = \dot{q}^j.$$

Also,

$$\{p_j, \mathcal{H}\} = \frac{\partial p_j}{\partial q^i}\frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial p_j}{\partial p_i}\frac{\partial \mathcal{H}}{\partial q^i} = -\frac{\partial \mathcal{H}}{\partial q_j} = \dot{p}_j.$$

So we have an alternative, and equivalent, form of **Hamilton's equations** expressed in terms of the Poisson bracket.

$$\dot{q} = \{q, \mathcal{H}\} \qquad \textbf{and} \qquad \dot{p} = \{p, \mathcal{H}\}.$$

Finally, suppose functions $f = f(q,p)$ and $g = g(q,p)$ are phase space functions which are constant on a solution path. Then on that path, according to the Jacobi identity,

$$\{f, \{g, \mathcal{H}\}\} + \{g, \{\mathcal{H}, f\}\} + \{\mathcal{H}, \{f, g\}\} = 0 = 0 + 0 + \{\mathcal{H}, \{f, g\}\}$$

so $\{f, g\}$ is *also* a constant of the motion. This procedure can sometimes produce new and *independent* (i.e. not a simple combination of f and g) constants of the motion.

12. **The Least Action Principle in Phase-space**

There is a least action principle in Hamiltonian mechanics, but unlike the lagrangian formulation, in the hamiltonian context it involves the entire phase-space.

We suppose given a path $(q,p) = (q(t), p(t))$ in phase-space and $(\zeta, \chi) = (\zeta(t), \chi(t))$ is a selected phase-space variation, a path with $(\zeta(a), \chi(a)) = (\zeta(b), \chi(b)) = (0, 0)$.

We define path (Q_ϵ, P_ϵ) by

$$(Q_\epsilon, P_\epsilon) = (q, p) + \epsilon(\zeta, \chi).$$

The action on path (q,p) is defined as an integral using the **phase-space lagrangian**

$$\Gamma = \int_a^b \mathcal{L}_{ph}\, dt = \int_a^b p_i \dot{q}^i - \mathcal{H}(t, q, p)\, dt.$$

$\mathcal{L}_{ph}$ is not the same as the hamiltonian-lagrangian $\mathcal{L}$. It is defined as indicated above as a function $\mathcal{L}_{ph} = \mathcal{L}_{ph}(t, q, p, v)$ with domain $\mathbb{R}^{3N+1}$

$$\boldsymbol{\mathcal{L}_{ph}(t, q, p, v, w) = p_i v^i - \mathcal{H}(t, q, p)}$$

and in the integral v is replaced by the time derivative $\dot{q}$ of geometrical path $q = q(t)$.

We define

$$\begin{aligned}\Gamma_\epsilon &= \int_a^b P_{\epsilon i} \dot{Q}_\epsilon^i - \mathcal{H}(t, Q_\epsilon, P_\epsilon)\, dt \\ &= \int_a^b (p_i + \epsilon \chi_i)\left(\dot{q}^i + \epsilon \dot{\zeta}^i\right) - \mathcal{H}(t, q + \epsilon\zeta, p + \epsilon\chi)\, dt\end{aligned}$$

so that as before $\Gamma = \Gamma_0$.

We say (q,p) is stationary with respect to (ζ, χ) when $\frac{d}{d\epsilon}\big|_{\epsilon=0}\Gamma_\epsilon = 0$, and

(q, p) is a stationary path if it is stationary with respect to every (ζ, χ).

As before, we need to calculate this derivative.

$$\begin{aligned}\left.\frac{d}{d\epsilon}\right|_{\epsilon=\varepsilon_0}\Gamma_\epsilon =& \left.\frac{d}{d\epsilon}\right|_{\epsilon=\varepsilon_0}\int_a^b (p_i+\epsilon\chi_i)\left(\dot{q}^i+\epsilon\dot{\zeta}^i\right)-\mathcal{H}(t,q+\epsilon\zeta,p+\epsilon\chi)\,dt\\ &=\int_a^b \left.\frac{d}{d\epsilon}\right|_{\epsilon=\varepsilon_0}(p_i+\epsilon\chi_i)\left(\dot{q}^i+\epsilon\dot{\zeta}^i\right)-\mathcal{H}(t,q+\epsilon\zeta,p+\epsilon\chi)\,dt\\ &=\int_a^b \left.\left(\chi_i\left(\dot{q}^i+\epsilon\dot{\zeta}^i\right)+(p_i+\epsilon\chi_i)\,\dot{\zeta}^i-\left[\frac{\partial\mathcal{H}}{\partial q^i}\zeta^i+\frac{\partial\mathcal{H}}{\partial p_i}\chi_i\right]\right)\right|_{\epsilon=\varepsilon_0}dt\end{aligned}$$

which gives, when the evaluation at $\epsilon = 0$ is taken,

$$\int_a^b \chi_i\dot{q}^i+p_i\dot{\zeta}^i-\left[\frac{\partial\mathcal{H}}{\partial q^i}\zeta^i+\frac{\partial\mathcal{H}}{\partial p_i}\chi_i\right]dt$$

where the hamiltonian partial derivatives are evaluated at $(t, q(t), p(t))$ for each time inside the integral.

Note that, since $\zeta(a) = \zeta(b) = 0$ we have $\int_a^b p_i\dot{\zeta}^i\,dt = p_i\zeta^i|_a^b - \int_a^b \dot{p}_i\zeta^i\,dt$ so with this substitution the derivative of the action becomes

$$\int_a^b \chi_i\dot{q}^i-\dot{p}_i\zeta^i-\left[\frac{\partial\mathcal{H}}{\partial q^i}\zeta^i+\frac{\partial\mathcal{H}}{\partial p_i}\chi_i\right]dt=\int_a^b \chi_i\left(\dot{q}^i-\frac{\partial\mathcal{H}}{\partial p_i}\right)-\zeta^i\left(\dot{p}_i+\frac{\partial\mathcal{H}}{\partial q^i}\right)dt.$$

Since ζ and χ are arbitrary functions, subject only to the endpoint restrictions, if this is to be 0 for *any* ζ and χ we must have the parenthesized terms identically zero for each i, independent of time.

That is, Hamilton's equations are valid for any path (q, p) in phase-space which is stationary *as a phase-space path*. The steps are reversible, so if Hamilton's equations hold the path (q, p) is stationary.

We know that any path q that is stationary with respect to the original pre-Legendre-transform Lagrangian set-up produces a momentum function p for which Hamilton's equations hold for the phase-space path (q, p).

Suppose now that Hamilton's equations hold for some phase-space path (q, p). Could it be that q *is not stationary* in the old sense?

$v(t) = J(t, q(t), p(t))$ is the one and only velocity vector that will reproduce momentum $p(t)$ at time t and place $q(t)$: that is, at each time t

$$\left[\partial_{N+j}L\right](t,q(t),v(t))=p_j(t)\qquad j=1,\dots,N.$$

The first set of Hamilton's equations

$$\dot{q}^i=\frac{\partial\mathcal{H}}{\partial p_i}$$

assure us (via Equation (13)) that the time derivative $\dot{q}$ of solution path q is exactly that velocity vector at each time.

With that in hand the second set of Hamilton's equations and Equation (14) assert that at each time t

$$\boldsymbol{\dot{p}_i =} -\partial_i\mathcal{H}(t,q,p)=-\left[-\partial_iL\right](t,q,J(t,q,p))=\boldsymbol{[\partial_iL]\,(t,q,\dot{q}).}$$

This is the ELS, so q is stationary in the older sense when (q, p) is stationary in phase-space.

13. **The Routhian**

14. **Canonical Transformations: Covariance for Hamiltonian Solutions**

A generic point $(q, v) \in \mathbb{R}^N \times \mathbb{R}^N$ in the lagrangian description of a physical system is not associated with a specific path. Rather, we have associated it with *any*[25] path $q(t)$ for which there is a time t_0 with $q = q(t_0)$ and $\frac{dq}{dt}(t_0) = v$.

So there is a "rigidity" in the way we "coordinatize" velocities: a path traveling at unit speed and constant direction in geometrical space is associated with a parallel unit vector in the coordinates of dynamical space. This choice is implicit: it is reflected in our calculation in the lagrangian setup that

$$\text{Point transformation } Q \leftrightarrow q\text{:} \qquad \frac{\partial q^i}{\partial Q^k} = \frac{\partial \dot{q}^i}{\partial \dot{Q}^k}$$

for all j and k where Q represents new coordinates for the geometrical variable. This was used in the proof of covariance (i.e. consistency) of stationary paths for the lagrangian action.

Consider a change in geometrical coordinates (i.e. a point transformation) in the lagrangian universe. For the set $\mathbb{R}^N \times \mathbb{R}^N$ of configuration and velocity, the geometrical coordinate change is in the first N coordinates and determines the coordinate change in the last N coordinates.

A solution path $(q(t), p(t))$ in phase-space that has been "translated" from a solution path $(q(t), \dot{q}(t))$ in the lagrangian universe carries the same physical meaning: really, it is only $q(t)$ that matters.

However we will see that with the hamiltonian setup we have ***more freedom in our choice of coordinate changes*** for phase-space $\mathbb{R}^N \times \mathbb{R}^N$ while retaining the desirable covariance. We may ***use a hamiltonian that has either been translated directly from old coordinates or modified in certain ways.*** These modifications of coordinates and/or hamiltonian can potentially have big advantages in finding the solution, or thinking about the solution process at least.

In these new coordinates the old coordinates q and p, obtained through translation from the lagrangian universe, can be mixed in ways that are impossible to obtain via a point transformation of the geometrical variables.

The allowed invertible transformations of the hamiltonian extended phase-space $\mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N$ are called **canonical** or **contact transformations** and will be derived from precisely those coordinate changes which are "mutually covariant" in the sense

---

[25] This point of view makes it easy to translate to a manifold setting, where tangent vectors at point q can be construed as the family of paths through q with velocity v.

that a stationary path for one will transform to a stationary path calculated using the transformed, but possibly otherwise modified, phase-space lagrangian.

We will start by using a ***directly-transformed phase-space lagrangian*** with ***coordinate changes that are time-independent***, and enhance this to include more general situations later.

Our new coordinates are obtained by a coordinate change in extended phase-space

$$(x_1, \dots, x_{2N}) = (q, p) \longleftrightarrow (y_1, \dots, y_{2N}) = (Q, P)$$

for which Q and P are functions of q and p. So, of course, q and p are functions of Q and P.

We will, for now, use typical convention from multivariable calculus for which derivatives are matrices, columns represent vectors and rows represent covectors such as differentials. Columns are used to represent derivatives of phase-space paths.

The coordinate change suggested above corresponds to invertible maps relating extended phase-space members indicated[26] variously, as convenient, by

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_{2N} \end{pmatrix} = \begin{pmatrix} Q^t \\ P^t \end{pmatrix} = G(x)$$

with associated extended phase-space points

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_{2N} \end{pmatrix} = \begin{pmatrix} q^t \\ p^t \end{pmatrix} = F(y).$$

Let G' and F' denote the position-dependent derivative functions

$$G' = \left(\frac{\partial G^i}{\partial x_j}\right) = \begin{pmatrix} \frac{\partial G^1}{\partial x_1} & \cdots & \frac{\partial G^1}{\partial x_N} & \frac{\partial G^1}{\partial x_{N+1}} & \cdots & \frac{\partial G^1}{\partial x_{2N}} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial G^N}{\partial x_1} & \cdots & \frac{\partial G^N}{\partial x_N} & \frac{\partial G^N}{\partial x_{N+1}} & \cdots & \frac{\partial G^N}{\partial x_{2N}} \\ \frac{\partial G^{N+1}}{\partial x_1} & \cdots & \frac{\partial G^{N+1}}{\partial x_N} & \frac{\partial G^{N+1}}{\partial x_{N+1}} & \cdots & \frac{\partial G^{N+1}}{\partial x_{2N}} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial G^{2N}}{\partial x_1} & \cdots & \frac{\partial G^{2N}}{\partial x_N} & \frac{\partial G^{2N}}{\partial x_{N+1}} & \cdots & \frac{\partial G^{2N}}{\partial x_{2N}} \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}} & \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} \\ \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}} & \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} \end{pmatrix}$$

and

$$F' = \left(\frac{\partial F^i}{\partial y_j}\right) = \begin{pmatrix} \frac{\partial F^1}{\partial y_1} & \cdots & \frac{\partial F^1}{\partial y_N} & \frac{\partial F^1}{\partial y_{N+1}} & \cdots & \frac{\partial F^1}{\partial y_{2N}} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial F^N}{\partial y_1} & \cdots & \frac{\partial F^N}{\partial y_N} & \frac{\partial F^N}{\partial y_{N+1}} & \cdots & \frac{\partial F^N}{\partial y_{2N}} \\ \frac{\partial F^{N+1}}{\partial y_1} & \cdots & \frac{\partial F^{N+1}}{\partial y_N} & \frac{\partial F^{N+1}}{\partial y_{N+1}} & \cdots & \frac{\partial F^{N+1}}{\partial y_{2N}} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial F^{2N}}{\partial y_1} & \cdots & \frac{\partial F^{2N}}{\partial y_N} & \frac{\partial F^{2N}}{\partial y_{N+1}} & \cdots & \frac{\partial F^{2N}}{\partial y_{2N}} \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{Q}} & \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{P}} \\ \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{Q}} & \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{P}} \end{pmatrix}.$$

[26]The notation P^t denotes the transpose of P.

The Jacobians G' and F' are inverse matrices at corresponding points in phase-space.

$$F'G' = \begin{pmatrix} \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{Q}} & \frac{\partial \boldsymbol{q}}{\partial \boldsymbol{P}} \\ \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{Q}} & \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{P}} \end{pmatrix} \begin{pmatrix} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}} & \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} \\ \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}} & \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{Q}} + \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}}\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{Q}} & \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{P}} + \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}}\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{P}} \\ \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{Q}} + \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}}\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{Q}} & \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{P}} + \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}}\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{P}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

where $\mathbf{I}$ represents and $\mathbf{0}$ represents the identity and zero matrices of the appropriate size, in this case $N \times N$.

Suppose phase-space path $(Q, P) = (Q(t), P(t))$ is stationary, producing stationary value

$$\int_a^b P(t)\dot{Q}(t) - \mathcal{Z}(t, Q(t), P(t))\, dt$$

where $\mathcal{Z}$ is the transformed hamiltonian

$$\mathcal{Z}(t, Q, P) = \mathcal{H}(t, q(Q, P), p(Q, P))$$

created from hamiltonian $\mathcal{H} = \mathcal{H}(t, q, p)$.

The path is stationary exactly when, for each t, Hamilton's equations are valid for this phase-space path:

$$\dot{Q}^i = \frac{\partial \mathcal{Z}}{\partial P_i} \qquad \text{and} \qquad \dot{P}_i = -\frac{\partial \mathcal{Z}}{\partial Q^i} \qquad \text{for } i = 1, \ldots, N.$$

We want conditions on the transformation which will imply

$$\dot{q}^i = \frac{\partial \mathcal{H}}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q^i} \qquad \text{for } i = 1, \ldots, N$$

everywhere along the transformed phase-space path $(q, p) = (q(t), p(t))$.

Focus on a specific time t_0, and for this time define on phase-space

$$\widetilde{\mathcal{Z}}(y) = \mathcal{Z}(t_0, y) = \mathcal{H}(t_0, F(y)) \quad \text{and} \quad \widetilde{\mathcal{H}}(x) = \mathcal{H}(t_0, x).$$

$$\begin{aligned} d\widetilde{\mathcal{Z}} &= \left(\frac{\partial \mathcal{Z}}{\partial y_1}, \ldots, \frac{\partial \mathcal{Z}}{\partial y_{2N}}\right) = \left(\frac{\partial \mathcal{Z}}{\partial Q_1}, \ldots, \frac{\partial \mathcal{Z}}{\partial Q_N}, \frac{\partial \mathcal{Z}}{\partial P_1}, \ldots, \frac{\partial \mathcal{Z}}{\partial P_N}\right) \\ &= d\widetilde{\mathcal{H}}\, F' = \left(\frac{\partial \mathcal{H}}{\partial x_1}, \ldots, \frac{\partial \mathcal{H}}{\partial x_{2N}}\right) F' = \left(\frac{\partial \mathcal{H}}{\partial q_1}, \ldots, \frac{\partial \mathcal{H}}{\partial q_N}, \frac{\partial \mathcal{H}}{\partial p_1}, \ldots, \frac{\partial \mathcal{H}}{\partial p_N}\right) F'. \end{aligned}$$

$$\text{Let} \quad \boldsymbol{\psi} = \begin{pmatrix} 0 & \ldots & 0 & -1 & \ldots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & -1 \\ 1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \ldots & 1 & 0 & \ldots & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{-I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix}.$$

So $\boldsymbol{\psi}^{-1} = -\boldsymbol{\psi} = \boldsymbol{\psi}^t$, the transpose of $\boldsymbol{\psi}$.

We assume Hamilton's equations for (P,Q) and $\mathcal{Z}$ so translating that into matrix form we have, evaluated at each particular time,

$$\dot{y} = \left(d\widetilde{\mathcal{Z}}\,\boldsymbol{\psi}\right)^t = \boldsymbol{\psi}^t\, d\widetilde{\mathcal{Z}}^t$$

and we want conditions under which this implies

$$\dot{x} = \left(d\widetilde{\mathcal{H}}\,\boldsymbol{\psi}\right)^t = \boldsymbol{\psi}^t\, d\widetilde{\mathcal{H}}^t.$$

Using the chain rule and Hamilton's equations on $\dot{y}$ produces

$$G'\dot{x} = \dot{y} = \boldsymbol{\psi}^t d\widetilde{\mathcal{Z}}^t = \boldsymbol{\psi}^t \left(d\widetilde{\mathcal{H}}\,F'\right)^t = \boldsymbol{\psi}^t\, F'^t\, d\widetilde{\mathcal{H}}^t \quad \Longleftrightarrow \quad \dot{x} = G'^{-1}\boldsymbol{\psi}^t\, F'^t\, d\widetilde{\mathcal{H}}^t.$$

$G'^{-1} = F'$ so $\dot{x}$ satisfies Hamilton's equations for hamiltonian $\mathcal{H}$ exactly when

$$\boldsymbol{\psi}^t = G'^{-1}\boldsymbol{\psi}^t\, F'^t \quad \Longleftrightarrow \quad \boldsymbol{\psi} = G'\boldsymbol{\psi}\, G'^t \quad \Longleftrightarrow \quad F' = \boldsymbol{\psi}\, G'^t\boldsymbol{\psi}^t.$$

Products of matrices which satisfy this condition also do, and these matrices are invertible so they form a group.

Matrices that satisfy this relation are called **symplectic** and a **transformation, such as G, is called symplectic** if its jacobian is symplectic on the whole phase-space.

So the symplectic matrices form a group with matrix multiplication, and symplectic transformations form a group with composition.

Time-independent and symplectic transformations are canonical.

The argument outlined above is reversible, so ***transformed paths are stationary with respect to the transformed hamiltonian exactly when the (time-independent) transformation jacobian G' satisfies $\psi = G'\psi\, G'^t$ on all of its domain phase-space.***

Expanding $F' = \boldsymbol{\psi}\, G'^t\boldsymbol{\psi}^t$ yields the interesting $n\times n$ block relation between any symplectic matrix and its inverse, reminiscent of the relation between an orthogonal matrix and its inverse:

$$F' = \begin{pmatrix} \frac{\partial q}{\partial Q} & \frac{\partial q}{\partial P} \\ \frac{\partial p}{\partial Q} & \frac{\partial p}{\partial P} \end{pmatrix} = \boldsymbol{\psi}\, G'^t\boldsymbol{\psi}^t = \begin{pmatrix} \left(\frac{\partial P}{\partial p}\right)^t & -\left(\frac{\partial P}{\partial q}\right)^t \\ -\left(\frac{\partial Q}{\partial p}\right)^t & \left(\frac{\partial Q}{\partial q}\right)^t \end{pmatrix}.$$

Expanding the symplectic condition $\boldsymbol{\psi} = G'\boldsymbol{\psi}\, G'^t$ we have

$$\begin{pmatrix} \mathbf{0} & \textbf{-I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}} & \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} \\ \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}} & \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \textbf{-I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}} & \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} \\ \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}} & \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} \end{pmatrix}^{\boldsymbol{t}}$$

$$= \begin{pmatrix} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} & -\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}} \\ \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} & -\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}} \end{pmatrix} \begin{pmatrix} \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}} & \left(\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}} \\ \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}}\right)^{\boldsymbol{t}} & \left(\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}}\right)^{\boldsymbol{t}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}} - \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}} \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}}\right)^{\boldsymbol{t}} & \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} \left(\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}} - \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}} \left(\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}}\right)^{\boldsymbol{t}} \\ \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}} - \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}} \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}}\right)^{\boldsymbol{t}} & \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} \left(\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}} - \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}} \left(\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}}\right)^{\boldsymbol{t}} \end{pmatrix}.$$

This equation asserts that the block matrices along the diagonal are, each, the zero matrix so the condition is exactly that both $\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}} \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}}$ and $\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} \left(\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}}$ are to be symmetric.

The two equations involving the cross-diagonal terms are equivalent, and assert that $\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{p}} \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{q}}\right)^{\boldsymbol{t}} - \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{q}} \left(\frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{p}}\right)^{\boldsymbol{t}} = \mathbf{I}$.

In coordinate form we have

$$0 = \frac{\partial Q^m}{\partial q^i}\frac{\partial Q^k}{\partial p_i} - \frac{\partial Q^m}{\partial p_i}\frac{\partial Q^k}{\partial q^i} \quad \text{and} \quad 0 = \frac{\partial P_m}{\partial q^i}\frac{\partial P_k}{\partial p_i} - \frac{\partial P_m}{\partial p_i}\frac{\partial P_k}{\partial q^i}$$

$$\text{and} \quad \delta_{m,k} = \frac{\partial Q^m}{\partial q^i}\frac{\partial P_k}{\partial p_i} - \frac{\partial Q^i}{\partial p_i}\frac{\partial P_k}{\partial q^i}.$$

Translating that into Poisson bracket form we have

$$\boldsymbol{\{Q^m, Q^k\}_{q,p} = 0} \quad \text{and} \quad \boldsymbol{\{P_m, P_k\}_{q,p} = 0} \quad \text{and} \quad \boldsymbol{\{Q^m, P_k\}_{q,p} = \delta_{m,k}}.$$

We have, then, that time-independent phase-space transformation $\boldsymbol{G = (Q, P)}$ ***is canonical using the transformed hamiltonian if and only if these three Poisson bracket conditions are satisfied.***

15. **Symplectic Matrices**

Since $\det(\psi) = 1$ the relation $\boldsymbol{\psi} = G'\boldsymbol{\psi}\, G'^t$ implies that $\mathbf{det}\boldsymbol{(G') = \pm 1}$.

In fact, **symplectic matrices also preserve orientation**: i.e. $\mathbf{det}\boldsymbol{(G') = 1}$ so **this type of canonical transformation preserves oriented phase-space volume.** A complete proof of this requires some spadework, interesting in its own right, and we proceed with that now.

To clean up notation a tiny bit let M, below, be a generic $2N \times 2N$ symplectic matrix. Then $\psi = M\psi M^t$ which, since $\psi^{-1} = \psi^t = -\psi$, implies $\psi M^{-1} = M^t\psi$.

Also, note that $v^t\psi v = 0$ and $v^t\psi w = -w^t\psi v$ for any pair of vectors v and w.

Suppose α is a nonzero member of our field[27] and v a nonzero vector in $\mathbb{R}^{2N}$.

Define the **symplectic transvection $\boldsymbol{E_{\alpha,v}}$** to be the matrix

$$E_{\alpha,v} = I + \alpha v v^t \psi$$

where I is the identity matrix of the appropriate size.

$$\begin{aligned} E_{\alpha,v}\psi E_{\alpha,v}^t &= (I + \alpha v v^t \psi)\psi(I + \alpha v v^t \psi)^t = (I + \alpha v v^t \psi)\psi(I + \alpha \psi^t v v^t) \\ &= (I + \alpha v v^t \psi)(\psi + \alpha \psi \psi^t v v^t) = (I + \alpha v v^t \psi)(\psi + \alpha v v^t) \\ &= \psi + \alpha v v^t + \alpha v v^t \psi \psi + \alpha^2 v v^t \psi v v^t = \psi + \alpha v v^t - \alpha v v^t + 0 = \psi. \end{aligned}$$

So a symplectic transvection actually is symplectic.

$$E_{\alpha,v} E_{-\alpha,v} = (I + \alpha v v^t \psi)(I - \alpha v v^t \psi) = I$$

so the inverse of a symplectic transvection is also a symplectic transvection.

$(I - E_{\alpha,v})^2 = 0$ so the only eigenvalue of $E_{\alpha,v}$ is 1 and therefore (examine the Jordan canonical form) $\det(E_{\alpha,v}) = 1$.

If $M = I$ then $M = E_{\alpha,v}E_{-\alpha,v}$ is the product of two symplectic transvections for arbitrary nonzero α and v.

We will prove that any symplectic M, not just the identity, is the product of symplectic transvections and since any such product has determinant 1 we conclude that any symplectic matrix has determinant 1.

The idea of the proof is that if $M \neq I$ we can replace M by a product of the form $E_{\alpha^{-1},v}M$ or $E_{\alpha^{-1},q}E_{1,v}M$, and this new symplectic matrix acts as the identity on a subspace containing but strictly larger than the subspace upon which M acts as the identity.

We can replace M by this new symplectic matrix and iterate until arriving, after a finite number of steps, at a product $E_{\alpha_1,v_1} \dots E_{\alpha_k,v_k} M = I$. Then $M = E_{-\alpha_k,v_k} \dots E_{-\alpha_1,v_1}$ and the result is proved.

If $M \neq I$ there is a nonzero vector u for which $0 \neq (M - I)u$ and we assume in the discussion below that this situation pertains, and terminate the algorithm if/when we arrive at the identity matrix.

Let $v = (M - I)u$ so that $v^t = u^t M^t - u^t$.

Consider $w \in \ker(M - I) = \ker(M^{-1} - I)$. We have for any nonzero α

$$\begin{aligned} (E_{\alpha,v}M - I)w &= ((I + \alpha v v^t \psi)M - I)w = \alpha v v^t \psi w = \alpha v(u^t M^t - u^t)\psi w \\ &= \alpha v u^t M^t \psi w - \alpha v u^t \psi w = \alpha v u^t \psi M^{-1} w - \alpha v u^t \psi w = 0. \end{aligned}$$

Therefore $\boldsymbol{\ker(M - I) \subset \ker(E_{\alpha,v}M - I)}$. The only conditions required for this conclusion are that M is symplectic, $\alpha \neq 0$ and $v = Mu - u \neq 0$.

There are now two possible cases.

[27] The field is assumed here to be the real numbers, though the argument given goes through without change for complex vector spaces.

First, there might be a vector u for which $\alpha = u^t\psi Mu \neq 0$, in which case we are guaranteed that $u \neq Mu$ so $v = (M - I)u \neq 0$.

$$\begin{aligned} E_{\alpha^{-1},v}Mu &= (I + \alpha^{-1}vv^t\psi)Mu = Mu + \alpha^{-1}vv^t\psi Mu = Mu + \alpha^{-1}v((Mu)^t - u^t)\psi Mu \\ &= Mu + \alpha^{-1}v(Mu)^t\psi Mu - \alpha^{-1}vu^t\psi Mu = Mu - v = u. \end{aligned}$$

Letting $M' = E_{\alpha^{-1},v}M$ we have $M'u = u$ but $Mu \neq u$.

So we have $\ker(M - I) \oplus \mathbb{R}u \subset \ker(M' - I)$.

***M′* is symplectic and acts as the identity on a subspace *strictly* larger than the corresponding subspace for *M*.**

The second possibility to consider is that $p^t\psi Mp = 0$ for every p so our first case cannot proceed.

That implies the matrix ψM is skew-symmetric, since for any p and q

$$\begin{aligned} 0 &= (p + q)^t\psi M(p + q) = p^t\psi Mp + p^t\psi Mq + q^t\psi Mp + q^t\psi Mq \\ &= p^t\psi Mq + p^t\psi Mq. \end{aligned}$$

Then we have

$$-\psi M = (\psi M)^t = M^t\psi^t = -M^t\psi = -\psi M^{-1}$$

and therefore $M^2 = I$: i.e. M is an involution.

Remember, we are assuming here that M is not the identity, so we can select u for which $v = (M - I)u \neq 0$ and now define $M' = E_{1,v}M = (I + vv^t\psi)M$.

As before, the kernel of $M - I$ is contained in the kernel of $M' - I$: in other words, M' acts as the identity on a subspace containing, and possibly larger than, the subspace upon which M acts as the identity.

If M' is the identity we terminate this algorithm.

If M' is not the identity we proceed as follows.

Setting $w = M^{-1}\psi^{-1}v$ we have $\beta = v^t\psi Mw \neq 0$.

By the assumption of skew-symmetry we have $w^t\psi Mw = 0$.

Also $M^t\psi = \psi M^{-1} = \psi M$ and $0 = M^2 - I = (M - I)(M + I)$.

So the number α defined below as $w^t\psi M'w$ is seen, by the following calculation, to be nonzero.

We use in this calculation, as before, that $v = (M - I)u$ so $v^t = u^t(M^t - I)$. In addition, because M is symplectic and an involution, we have $(M^t - I)\psi = \psi(M^{-1} - I) = \psi(M - I)$.

$$\begin{aligned} \alpha &= w^t\psi M'w = w^t\psi(I + vv^t\psi)Mw = w^t\psi Mw + w^t\psi v(v^t\psi Mw) \\ &= \beta w^t\psi v = -\beta v^t\psi w = -\beta(v^t\psi(M + I)w - v^t\psi Mw) \\ &= -\beta v^t\psi(M + I)w + \beta^2 = -\beta u^t(M^t - I)\psi(M + I)w + \beta^2 \\ &= -\beta u^t\psi(M^{-1} - I)(M + I)w + \beta^2 = -\beta u^t\psi(M - I)(M + I)w + \beta^2 = \beta^2 > 0. \end{aligned}$$

In particular, $M'w \neq w$.

Applying the first case to M' with $q = (M' - I)w$ and this α we find that $M'' = E_{\alpha^{-1},q}M' = E_{\alpha^{-1},q}E_{1,v}M$ acts as the identity on a subspace strictly larger than does M.

If $M'' = I$ we terminate the procedure. If not we iterate.

In at most $2N$ steps the algorithm will terminate with the identity written as the product of M with a list of symplectic transvections and the result is proved.

16. **More General Canonical Transformations**

We have seen that stationary solutions for the lagrangian on configuration-space are covariant under *any* coordinate change, even time-dependent coordinate changes.

We have created a phase

References

[1] Emil Artin, *Geometric Algebra.* 1964: Interscience Publishers.

[2] Melvin Calkin, *Lagrangian and Hamiltonian Mechanics.* 1996: World Scientific.

[3] Patrick Hamill, *A Student's Guide to Lagrangians and Hamiltonians.* 2014: Cambridge University Press.

[4] Herbert Goldstein and Charles Poole and John Safko, *Classical Mechanics, 3rd Ed.* 2011: Pearson.

[5] Yvette Kosmann-Schwarzbach, *The Noether Theorems.* 2011: Springer.

[6] L. D. Landau and E. M. Lifshitz , *Mechanics 3rd Edition.* 1976: Pergamon Press.

[7] Mark Levi, *Classical Mechanics with Calculus of Variations and Optimal Control.* 2014: AMS.

[8] John David Logan, *Invariant Variational Principles.* 1977: Academic Press.

[9] Dwight Neuenschwander, *Emmy Noether's Wonderful Theorem 2nd Edition.* 2017: Johns Hopkins University Press.

[10] Gennadi Sardanashvily, *Noether's Theorems.* 2015: Atlantis Press.

[11] Jacob Schwichtenberg, *Physics From Symmetry 2nd Edition.* 2018: Springer.

[12] Jacob Schwichtenberg, *No-Nonesense Classical Mechanics.* 2019: No-Nonesense Books.

[13] Michael Spivak, *Physics For Mathematicians Mechanics I.* 2010: Publish or Perish Inc.

Index