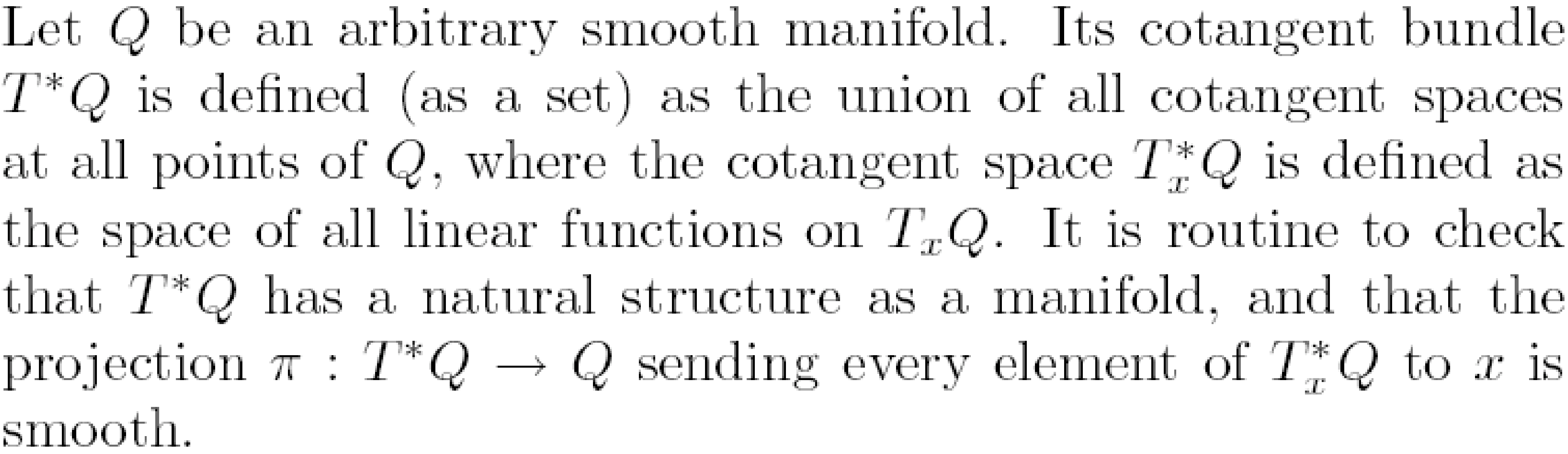
Symplectic Geometry
Lecture 7

Hamiltonian mechanics on the cotangent bundle.

Review: The cotangent bundle.



Let Q be an arbitrary smooth manifold. Its cotangent bundle T^*Q is defined (as a set) as the union of all cotangent spaces at all points of Q, where the cotangent space T_x^*Q is defined as the space of all linear functions on T_xQ. It is routine to check that T^*Q has a natural structure as a manifold, and that the projection $\pi : T^*Q \to Q$ sending every element of T_x^*Q to x is smooth.

Review: the canonical one form.

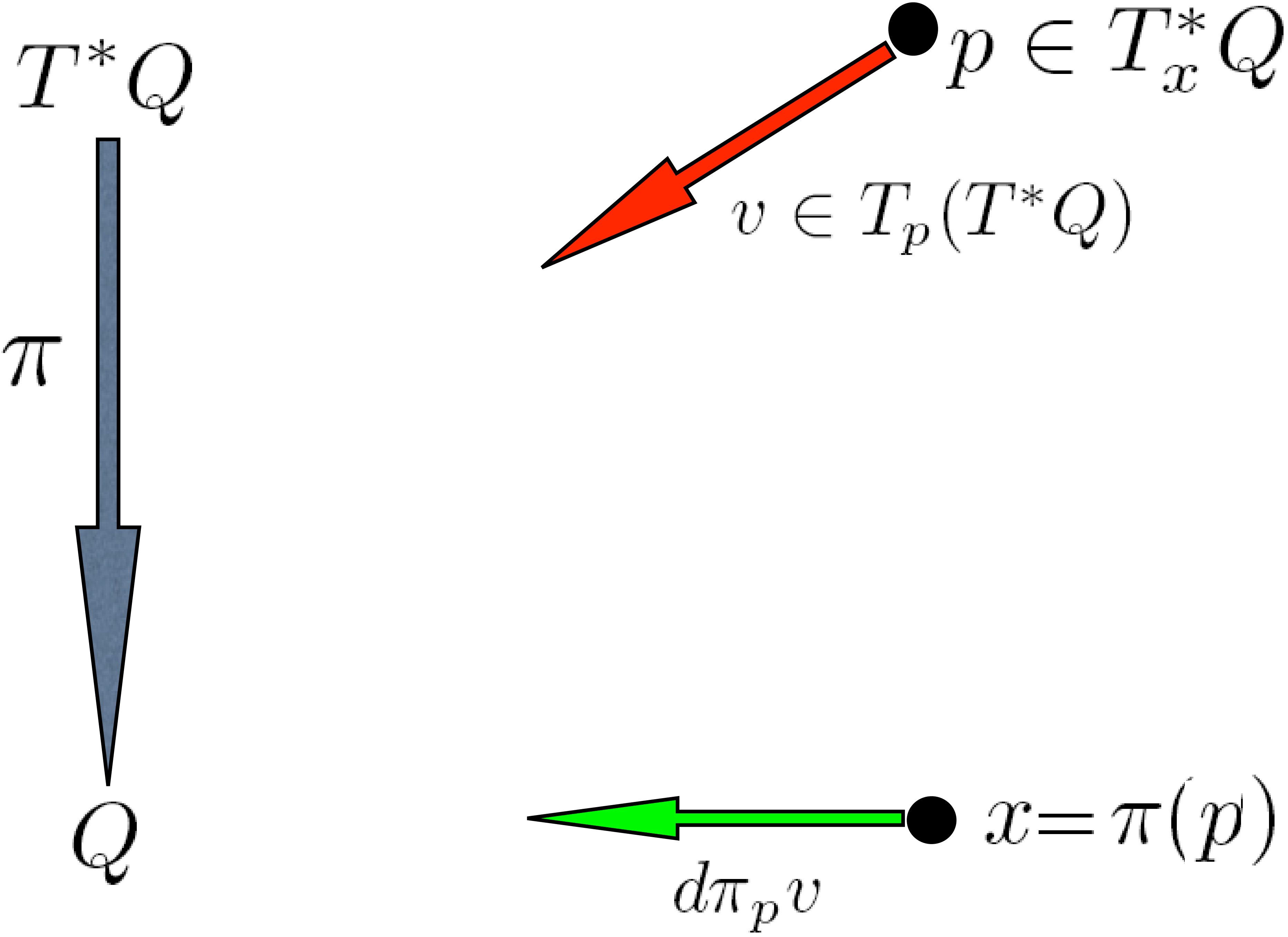
If Q is a differentiable manifold, then its cotangent bundle T^*Q carries a **canonical one form** $\alpha = \alpha_Q$ defined as follows: Let

$$\pi : T^*Q \to Q$$

be the projection sending any covector $p \in T_x^*Q$ to its base point x. If $v \in T_p(T^*Q)$ is a tangent vector to T^*Q at p, then

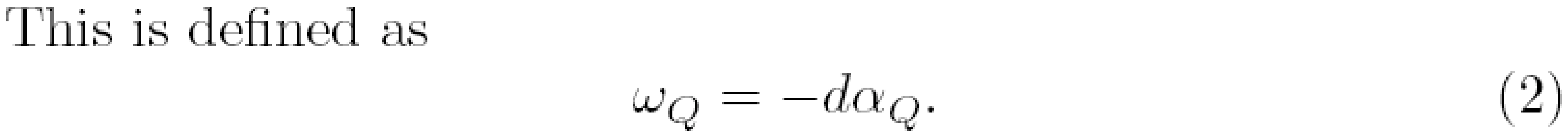
is a tangent vector to Q at x. In other words, $d\pi_p v \in T_xQ$. But $p \in T_x^*Q$ is a linear function on T_xQ, and so we can evaluate p on $d\pi_p v$. The canonical linear differential form α is defined by

$$\langle \alpha_p, v \rangle := \langle p, d\pi_p v \rangle \quad \text{if} \quad v \in T_p(T^*Q). \tag{1}$$



$$\langle \alpha_p, v \rangle := \langle p, d\pi_p v \rangle$$

Review: the canonical two form.


This is defined as

$$\omega_Q = -d\alpha_Q. \tag{2}$$

Review: Expressions in local coordinates.

Let $q^1, \ldots, q^n$ be local coordinates on Q. Then $dq^1, \ldots, dq^n$ are differential forms which give a basis of T_x^*Q at each x in the coordinate neighborhood U. In other words, the most general element of T_x^*Q can be written as $p_1(dq^1)_x + \cdots + p_n(dq^n)_x$. Thus $q^1, \ldots, q^n, p_1, \ldots, p_n$ are local coordinates on

$$\pi^{-1}U \subset T^*Q.$$

In terms of these coordinates the canonical one-form is given by

$$\alpha = p \cdot dq = p_1 dq^1 + \cdots + p_n dq^n$$

Hence the canonical two-form has the standard local expression

$$\omega = dq \wedge \cdot dp = dq^1 \wedge dp_1 + \cdots + dq^n \wedge dp_n. \qquad (3)$$

$$\omega = dq \wedge \cdot dp = dq^1 \wedge dp_1 + \cdots + dq^n \wedge dp_n. \qquad (3)$$

If H is a C^∞ function on T^*Q then we get the corresponding vector field X_H which in terms of the local expression (3) takes on the form of Hamilton's differential equations:

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}.$$

Suppose that $Q = \mathbb{R}^n$ so that these coordinates are in fact global, corresponding to (the standard, say) linear coordinates on Q. If

$$H(q,p) = \frac{1}{2m}\sum_i p_i^2$$

where $m > 0$ is the "mass", these equations become

$$\dot{q}_i = \frac{1}{m}p_i, \quad \dot{p}_i = 0.$$

Rectilinear motion.

$$\dot{q}_i = \frac{1}{m} p_i, \quad \dot{p}_i = 0.$$

So if we write $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ so that $\dot{\mathbf{q}}$ is the velocity then the solution curves of this system are

$$\mathbf{q}(t) = \mathbf{q}(0) + t\dot{\mathbf{q}}(0), \quad \mathbf{p}(t) \equiv m\dot{\mathbf{q}}(0).$$

The particle moves along a straight line with constant velocity (Galileo's law) and $\mathbf{p} = m\mathbf{q}$ is the momentum.

Conservation of energy.

On general principles we have

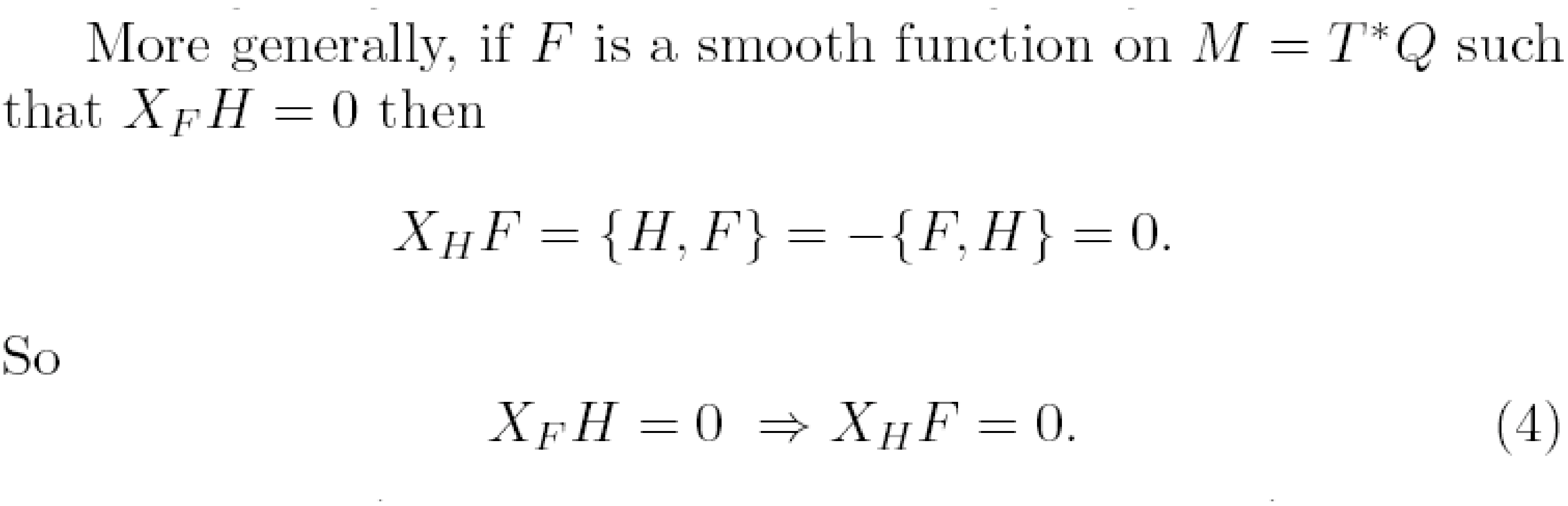
$$X_H H = 0$$

since

$$X_H H = D_{X_H} H = i(X_H) dH = i(X_H) i(X_H) \omega = 0.$$

The function H is sometimes called the energy, so the law $X_H H = 0$ is known as the conservation of energy. We can also derive this fact from the anti-symmetry of the Poisson brackets since $X_H H = \{H, H\}$ and by anti-symmetry $\{H, H\} = 0$.

Noether’s theorem.



More generally, if F is a smooth function on $M = T^*Q$ such that $X_F H = 0$ then

$$X_H F = \{H, F\} = -\{F, H\} = 0.$$

So

$$X_F H = 0 \ \Rightarrow X_H F = 0. \tag{4}$$

Kinetic and potential energy.

In many cases (such as the one we studied last time) the energy is the sum of two terms,

$$H = K + U$$

where K is called the **kinetic energy** and U is called the **potential energy**.

The potential energy is as before, it is the pull-back of some smooth function on Q via π^*. We will now give a more general version of the kinetic energy.

Kinetic energy.

Suppose that Q is a Riemannian manifold. This means that each tangent space T_xQ is endowed with a (positive definite) scalar product $(\cdot,\cdot)_x$ and that the $(\cdot,\cdot)_x$ vary smoothly with x.

If V is a vector space with a scalar product $(\cdot,\cdot)_V$ then $(\cdot,\cdot)_V$ induces a linear isomorphism $V \to V^*$ where $v \in V$ goes into the linear function on V given by scalar product with V:

$$\mathfrak{L} : V \to V^*, \qquad v \mapsto (v,\cdot).$$

This in turn induces a scalar product $(\cdot,\cdot)_{V^*}$ given by

$$(\ell, m)_{V^*} = \langle \ell, \mathfrak{L}^{-1} m\rangle = (\mathfrak{L}^{-1}\ell, \mathfrak{L}^{-1}m)_V .$$

So if Q is a Riemannian manifold then we get a scalar product $(\cdot,\cdot)_{T_xQ^*}$ on each cotangent space.

This in turn induces a scalar product $(\cdot,\cdot)_{V^*}$ given by

$$(\ell, m)_{V^*} = \langle \ell, \mathfrak{L}^{-1} m \rangle = (\mathfrak{L}^{-1}\ell, \mathfrak{L}^{-1} m)_V .$$

So if Q is a Riemannian manifold then we get a scalar product $(\cdot,\cdot)_{T_xQ^*}$ on each cotangent space. Sacrificing precision for notational simplicity, I will drop the subscript T_xQ^* unless it is absolutely necessary. The kinetic energy is then defined as the function K which is quadratic in the cotangent variables given by

$$K(\xi) = \frac{1}{2}(\xi, \xi).$$

$$K(\xi) = \frac{1}{2}(\xi, \xi).$$

For example, suppose that $Q = \mathbb{R}^3$ and that q_1, q_2, q_3 are the coordinates relative to an orthonormal coordinate system on Q. We identify T_xQ with Q at each $x \in Q$ and so get coordinates $\dot{q}_1, \dot{q}_2, \dot{q}_3$ on each T_xQ. Suppose that we choose our metric to be $m\times$ the Euclidean metric. Then, for each x

$$\mathfrak{L}(\dot{q}_1, \dot{q}_2, \dot{q}_3) = (m\dot{q}_1, m\dot{q}_2, m\dot{q}_3)$$

and hence

$$K(q_1, q_2, q_3, p_1, p_2, p_3) = \frac{1}{2m}\left(p_1^2 + p_2^2 + p_3^2\right).$$

This is our old expression for the energy of a free classical three dimensional particle.

More generally, if Q is a Riemannian manifold and $q_1, \ldots, q_n$ is a system of local coordinates, we can write the quadratic form associated with the metric as

$$\sum g_{ij}(q_1, \ldots, q_n)\dot{q}_i\dot{q}_j$$

where the $q_1, \ldots, q_n, \dot{q}_1 \ \ldots \dot{q}_n$ are the corresponding local coordinates on TQ. Then $\mathfrak{L}$ is given by

$$\mathfrak{L}(q_1, \ldots, q_n, \dot{q}_1 \ \ldots \dot{q}_n) = (q_1, \ldots, q_n, p_1 \ \ldots p_n)$$

where

$$p_i = \sum_j g_{ij}(q_1, \ldots, q_n)\dot{q}_j. \tag{5}$$

$$p_i = \sum_j g_{ij}(q_1, \ldots, q_n)\dot{q}_j. \tag{5}$$

We can solve these equations for the $\dot{q}$ since the matrix $(g_{ij}(q_1, \ldots, q_n$
is invertible for all $(q_1, \ldots, q_n)$. Let

$$(g^{ij}) = (g^{ij}((q_1, \ldots, q_n))$$

denote the inverse matrix. Then the kinetic energy is given by

$$K(q_1, \ldots, q_n, p_1, \ldots, p_n) = \frac{1}{2}\sum_{ij} g^{ij}(q_1, \ldots, q_n)p_i p_j \tag{6}$$

We can write the equations (5) in a more instructive form which has an important generalization. Let L be the function on TQ which assigns to each $v \in T_xQ$ the value

$$L(v) = \frac{1}{2}(v, v)_x.$$

$$p_i = \sum_j g_{ij}(q_1, \ldots, q_n)\dot{q}_j. \tag{5}$$

Let

$$L(v) = \frac{1}{2}(v, v)_x.$$

In terms of the local coordinates $q_1, \ldots, q_n, \dot{q}_1 \ldots \dot{q}_n$ we have

$$L(q_1, \ldots, q_n, \dot{q}_1 \ldots \dot{q}_n) = \frac{1}{2}\sum_{ij} g_{ij}(q_1, \ldots, q_n)\dot{q}_i\dot{q}_j.$$

Then we can write (5) as

$$p_i = \frac{\partial L}{\partial \dot{q}_i}. \tag{7}$$

This map from TQ to T^*Q makes sense for any smooth function L defined on TQ as we shall soon explain, not necessarily for the specific function L given above and is known as the Legendre transform.

This map from TQ to T^*Q makes sense for any smooth function L defined on TQ as we shall soon explain, not necessarily for the specific function L given above and is known as the Legendre transform. Before giving the symplectic geometry (and invariant) description of what is going on, let us compute in local coordinates where $q_1, \ldots, q_n, \dot{q}_1 \ldots \dot{q}_n$ are coordinates on (a neighborhood) in TQ and $q_1, \ldots, q_n, p_1, \ldots, p_n$ are coordinates on T^*Q both corresponding to a choice of coordinates $q_1, \ldots, q_n$ on (a neighborhood) in Q. For ease of notation we will assume that these coordinates are valid on all of TQ, T^*Q and Q (by restriction if necessary).

The Legendre transform in local coordinates.

Suppose that L is a smooth function on TQ and define the map

$$\mathfrak{L} : TQ \to T^*Q, \quad (q_1, \ldots, q_n, \dot{q}_1 \ \ldots \dot{q}_n) \mapsto \left(q_1, \ldots, q_n, \frac{\partial L}{\partial \dot{q}_1}, \ldots \frac{\partial L}{\partial \dot{q}_n} \right)$$

Suppose that the map $\mathfrak{L}$ is a diffeomorphism. Then the inverse map

$$\mathfrak{L}^{-1} : T^*Q \to TQ$$

is of the same form. Indeed, define the function H on T^*Q by

$$H(q,p) := \sum p_i \dot{q}_i - L(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$$

$$H(q,p) := \sum p_i \dot{q}_i - L(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$$

where, in this expression, the $\dot{q}_i$ are regarded as functions on T^*Q via $\mathfrak{L}^{-1}$, that is, in the more precise (but uglier) expression

$$H(q,p) = \sum_i p_i \dot{q}_i \circ \mathfrak{L}^{-1} - L \circ \mathfrak{L}^{-1}. \tag{8}$$

Then

$$\frac{\partial H}{\partial p_i} = \dot{q}_i + \sum_j p_j \frac{\partial \dot{q}_j \circ \mathfrak{L}^{-1}}{\partial p_i} - \sum_j \frac{\partial L}{\partial \dot{q}_j} \circ \mathfrak{L}^{-1} \times \frac{\partial \dot{q}_j \circ \mathfrak{L}^{-1}}{\partial p_i},$$

Since

$$p_i = \frac{\partial L}{\partial \dot{q}_i}$$

the last two terms cancel and we get

$$\frac{\partial H}{\partial p_i} = \dot{q}_i. \tag{9}$$

For example, suppose that Q is a Riemmanian manifold, and suppose that we are given a function U on Q. By abuse of notation, we will use the same letter U to denote the pull-back of U to TQ under the projection of $TQ \to Q$ and the pull-back of U to T^*Q under the projection $\pi : T^*Q \to Q$. In all three cases we have the local expression

$$U = U(q_1, \ldots, q_n).$$

Let us now denote the function $v \mapsto \frac{1}{2}(v, v)$ by K_T (the kinetic energy expressed in terms of the tangent bundle). The local expression for K_T is, as before,

$$K_T(q_1, \ldots, q_n, \dot{q}_1 \ldots \dot{q}_n) = \frac{1}{2} \sum_{ij} g_{ij}(q_1, \ldots, q_n) \dot{q}_i \dot{q}_j.$$

$$U = U(q_1, \ldots, q_n)$$

Finally, we let K be the function on T^*Q given by

$$K(\xi) = \frac{1}{2}(\xi, \xi) = \frac{1}{2}\langle \xi, \mathfrak{L}^{-1}\xi \rangle.$$

The local expression for K is given by (6):

$$K(q_1, \ldots, q_n, p_1, \ldots, p_n) = \frac{1}{2}\sum_{ij} g^{ij}(q_1, \ldots, q_n)p_i p_j.$$

Now let L be the function on TQ given by

$$L = K_T - U. \tag{10}$$

Then

$$H = \langle \xi, \mathfrak{L}^{-1} \rangle - L \circ \mathfrak{L}^{-1}$$

or

$$H = K + U.$$

The Euler-Lagrange equations.

The first half of Hamilton's equations, the equations $\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}$ when translated back to TQ via $\mathfrak{L}^{-1}$ then become

$$\frac{dq_i(t)}{dt} = \dot{q}_i(t). \tag{11}$$

To understand the meaning of these equations consider the following: Let $t \mapsto C(t)$ be a differentiable curve on Q. Then at each t there is a well defined tangent vector $C'(t) \in T_{C(t)}Q$. Then we get a curve, let us call it $t \mapsto C_T(t) = (C(t), C'(t))$ on TQ. Of course not all curves on TQ are of this form. Equation (11) says that if $t \mapsto (q(t), p(t))$ satisfies the first half of Hamilton's equations then the corresponding curve on TQ is of the form C_T.

We turn to the second half of Hamilton's equations: We have

$$\frac{\partial H}{\partial q_i} = \frac{\partial}{\partial q_i}\left[\sum_j p_i \dot{q}_j \circ \mathfrak{L}^{-1} - L \circ \mathfrak{L}^{-1}\right]$$

$$= \sum_j p_j \frac{\partial \dot{q}^j \circ \mathfrak{L}^{-1}}{\partial q_i} - \frac{\partial L \circ \mathfrak{L}^{-1}}{\partial q_i} - \sum_j \frac{\partial L}{\partial \dot{q}_j} \frac{\partial \dot{q}_j \circ \mathfrak{L}^{-1}}{\partial q_i}.$$

The first and last terms cancel since

$$p_j = \frac{\partial L}{\partial \dot{q}_j}$$

as a function on TQ. So

$$\frac{\partial H}{\partial q_i} = -\frac{\partial L \circ \mathfrak{L}^{-1}}{\partial q_i}. \tag{12}$$

$$\frac{\partial H}{\partial q_i} = -\frac{\partial L \circ \mathfrak{L}^{-1}}{\partial q_i}. \tag{12}$$

So the second half of Hamilton's equations read

$$\frac{dp_i}{dt} = \frac{L \circ \mathfrak{L}^{-1}}{\partial q_i}.$$

If we substitute $p_i = \frac{\partial L}{\partial \dot{q}_i}$, we obtain

$$\frac{d(\partial L/\partial \dot{q}_i)}{dt} - \frac{\partial L}{\partial q_i} = 0. \tag{13}$$

These are the famous Euler-Lagrange equations for variational problems. I will discuss the calculus of variations from the point of view of the cotangent bundle below.

The principle of mechanical similarity.

The equations (13) are unchanged if we replace L by cL where c is any non-zero constant. Suppose that Q is an open subset of a vector space which is stable under all multiplications by positive numbers α, and suppose that the g_{ij} are constant under this multiplication. For example, Q might be $\mathbb{R}^3 - \{0\}$ and the metric used is the Euclidean metric. Suppose that U is homogeneous of degree p, i.e. that

$$U(\alpha q_1, \ldots, \alpha q_n) = \alpha^p U(q_1, \ldots, q_n).$$

Let us change our time scale by a factor of β, replacing t by $s = \beta t$. Then

$$\frac{d\alpha q_i}{ds} = \frac{\alpha}{\beta}\frac{dq_i}{dt}$$

so

$$\frac{1}{2}\sum_{ij} g_{ij}\frac{d\alpha q_i}{ds}\frac{d\alpha q_j}{ds} = \frac{\alpha^2}{\beta^2}\sum_{ij}\frac{dq_i}{dt}\frac{dq_j}{dt}.$$

$$\frac{1}{2}\sum_{ij} g_{ij}\frac{d\alpha q_i}{ds}\frac{d\alpha q_j}{ds} = \frac{\alpha^2}{\beta^2}\sum_{ij}\frac{dq_i}{dt}\frac{dq_j}{dt}.$$

Let us choose β so that

$$\frac{\alpha^2}{\beta^2} = \alpha^p, \quad \text{i.e.} \quad \beta = \alpha^{1-\frac{1}{2}p}.$$

So

$$\frac{1}{2}\sum_{ij} g_{ij}\frac{d\alpha q_i}{ds}\frac{d\alpha q_j}{ds} = \alpha^p\sum_{ij}\frac{dq_i}{dt}\frac{dq_j}{dt}$$

and hence

$$L\left(\alpha q_1,\ldots,\alpha q_n,\frac{d\alpha q_1}{ds},\ldots\frac{d\alpha q_1}{ds}\right) = \alpha^p L\left(q_1,\ldots,q_n\frac{dq_1}{dt},\ldots,\frac{dq_n}{dt}\right).$$

Kepler’s third law.

$$L\left(\alpha q_1, \ldots, \alpha q_n, \frac{d\alpha q_1}{ds}, \ldots \frac{d\alpha q_1}{ds}\right) = \alpha^p L\left(q_1, \ldots, q_n \frac{dq_1}{dt}, \ldots, \frac{dq_n}{dt}\right).$$

So replacing q_i by αq_i and t by βt carries solutions of the Euler-Lagrange equations into solutions. For example, if U is homogeneous of degree -1 as in the inverse square law, we must take

$$\beta = \alpha^{\frac{3}{2}}.$$

In particular, the period of an periodic orbit is proportional to the $\frac{3}{2}$-power of its linear dimension - Kepler’s “third law”.

The calculus of variations.

Let L be a function on TQ. For any curve interval $[t_1, t_2] \subset \mathbb{R}$ and any smooth curve $C : [t_1, t_2] \to Q$ define

$$I[C] := \int_{t_1}^{t_2} L(C_T(t))dt = \int_{t_1}^{t_2} L(C(t), C'(t))dt.$$

Let p and q be points of Q. The problem posed in the calculus of variations is: Among all curves C with $C(t_1) = p$ and $C(t_2) = q$ find that curve which minimizes $I[C]$. The standard answer is that a necessary condition is that C must solve the Euler-Lagrange equations, which are second order ordinary differential equations. From what we know, this is the same as saying that the curve

$$\overline{C}(t) := \mathfrak{L}(C_T(t))$$

is a solution of Hamilton's equations (at least if $\mathfrak{L} : TQ \to T^*Q$ is a diffeomorphism). I would like to establish this result directly on the cotangent bundle.

For this I need a purely cotangent bundle description of $I[C]$. I claim that

$$I[C] = \int_{\overline{C}} \alpha - \int_{t_1}^{t_2} H(\overline{C}(t))dt \tag{14}$$

where $\alpha = \alpha_Q$ is the fundamental one form on T^*Q. Indeed, in local coordinates

$$\int_{\overline{C}} \alpha = \sum_i \int_{\overline{C}} p_i dq_i = \sum_i \int_{C_T} \frac{\partial L}{\partial \dot{q}_i} dq_i = \sum_i \int_{t_1}^{t_2} \frac{\partial L}{\partial \dot{q}_i} \dot{q}_i(t) dt$$

since on the curve C_T we have $\qquad \dot{q}_i(t) = \dfrac{dq_i}{dt}.$

Now

$$\sum_i \int_{t_1}^{t_2} \frac{\partial L}{\partial \dot{q}_i} \dot{q}_i(t) dt = \sum_i \int_{t_1}^{t_2} (p_i \circ \mathfrak{L}(t)) \dot{q}_i(t) dt$$

by the definition of the Legendre transform. Also $-H = -\sum_i p_i \dot{q}_i + L$ proving (14).

Let Z be a vector field on Q which generates a flow $s \mapsto \phi_s$. We then get a flow

$$s \mapsto d(\phi_s)$$

on TQ. For all sufficiently small s the curve

$$\phi_s \circ C(\cdot) : t \mapsto (\phi_s(C(t))$$

Therefore, by (14),

$$I[\phi_s \circ C] = \int_{t_1}^{t_2} L(d(\phi_s)(C_T(t))dt$$

$$= \int_{\mathfrak{L}\circ(d(\phi_s))\circ C_T} \alpha - \int_{t_1}^{t_2} H\left((\mathfrak{L} \circ (d\phi_s)(C_T(t))\right) dt. \qquad (15)$$

$$\int_{\mathfrak{L}\circ(d(\phi_s))\circ C_T}\alpha-\int_{t_1}^{t_2}H\,((\mathfrak{L}\circ(d\phi_s)(C_T(t))\,dt. \qquad (15)$$

I would like to compute the derivative of this expression with respect to s.

We can transfer the flow $d\phi_s$ on TQ to T^*Q using $\mathfrak{L}$. That is, we can consider the flow $\mathfrak{L}\circ(d\phi_s)\circ\mathfrak{L}^{-1}$ on T^*Q. Clearly

$$\pi\circ\left(\mathfrak{L}\circ(d\phi_s)\circ\mathfrak{L}^{-1}\right)=\phi_s.$$

Let Z_* denote the vector field on T^*Q which generates the flow $\mathfrak{L}\circ(d\phi_s)\circ\mathfrak{L}^{-1}$. Then we have

$$d\pi(Z_*)=Z.$$

If we differentiate (15) with respect to s and set $s=0$ we obtain

$$\int_{\overline{C}}D_{Z_*}\alpha-\int_{t_1}^{t_2}(D_{Z_*}H)(\overline{C}(t))dt.$$

$$\int_{\overline{C}} D_{Z_*}\alpha - \int_{t_1}^{t_2} (D_{Z_*}H)(\overline{C}(t))dt.$$

Let us apply Weil's formula and the definition of the fundamental form α: We have

$$D_{Z_*}\alpha = d\langle \alpha, Z_* \rangle + i(Z_*)d\alpha$$

and at any $\xi \in T^*Q$ we have $\langle \alpha, Z_* \rangle = \langle \xi, Z(\pi(\xi)) \rangle$. So

$$\int_{\overline{C}} D_{Z_*}\alpha = \int_{\overline{C}} i(Z_*)d\alpha + \int_{\overline{C}} d\langle \overline{C}(t), Z_{C(t)} \rangle$$

$$= \int_{\overline{C}} i(Z_*)d\alpha + \langle \overline{C}(t_2), Z(C(t_2)) \rangle - \langle \overline{C}(t_1), Z(C(t_1)) \rangle.$$

Finally, we obtain the formula

$$\frac{d}{ds} I[\phi_s \circ C]_{|s=0} =$$

$$-\int_{\overline{C}} i(Z_*)\omega - \int_{t_1}^{t_2} (D_{Z_*}H)(\overline{C}(t))dt + \langle \overline{C}(t_2), Z(C(t_2))\rangle - \langle \overline{C}(t_1), Z(C(t_1))\rangle. \tag{16}$$

In particular, if the vector field Z vanishes at p and q the last two terms vanish, and the condition

$$\frac{d}{ds} I[\phi_s \circ C]_{|s=0} = 0$$

becomes

$$\int_{\overline{C}} i(Z_*)\omega + \int_{t_1}^{t_2} (D_{Z_*}H)(\overline{C}(t))dt = 0. \tag{17}$$

$$\int_{\overline{C}} i(Z_*)\omega + \int_{t_1}^{t_2} (D_{Z_*} H)(\overline{C}(t))dt = 0.$$

Let us choose the flow ϕ_s to be the identity outside a coordinate neighborhood on Q and on the coordinate neighborhood look like

$$\phi_s(q_1, \dots, q_n) = (q_1, \dots, q_i + s\psi, \dots, q_n).$$

In other words we are only varying the i-th coordinate by sending $q_i \mapsto q_i + s\psi$. Here ψ is a smooth function (of all the variables). The flow induced on TQ is then

$$\begin{aligned} q_j &\mapsto q_j, \quad j \neq i, \\ q_i &\mapsto q_i + s\psi \\ \dot{q}_j &\mapsto \dot{q}_j, \quad j \neq i, \\ \dot{q}_i &\mapsto \dot{q}_i + s\sum_j \frac{\partial \psi}{\partial q_j}\dot{q}_j. \end{aligned}$$

So the vector field Z_T generating this flow on TQ is

$$Z_T = \psi \frac{\partial}{\partial q_i} + \sum_j \frac{\partial \psi}{\partial q_j} \frac{\partial}{\partial \dot{q}_j}.$$

Finally, the vector field Z_* on T^*Q will have the form

$$Z_* = d\mathfrak{L}(Z_T) = \psi \frac{\partial}{\partial q_i} + \sum_j B_j \frac{\partial}{\partial p_j}$$

where the B_j are some functions on T^*Q which depend linearly on Z. (This is, of course, a consequence of the fact that $d\pi(Z_*) = Z$.) So

$$i(Z)\omega = \psi dp_i - \sum_j B_j dq_j$$

and

$$D_{Z_*} H = \psi \frac{\partial H}{\partial q} + \sum_j B_j \frac{\partial H}{\partial p_j}.$$

$$i(Z)\omega = \psi dp_i - \sum_j B_j dq_j \qquad D_{Z_*}H = \psi \frac{\partial H}{\partial q} + \sum_j B_j \frac{\partial H}{\partial p_j}.$$

$$\int_{\overline{C}} i(Z_*)\omega + \int_{t_1}^{t_2} (D_{Z_*}H)(\overline{C}(t))dt = 0. \tag{17}$$

So if the curve $\overline{C}$ is given by

$$t \mapsto (q_1(t), \ldots, q_n(t), p_1(t), \ldots, p_n(t))$$

the condition (17) becomes

$$\int_{t_1}^{t_2} \left[\sum_j B_j \left(\frac{dq}{dt} - \frac{\partial H}{\partial p_j} \right) - \psi \left(\frac{dp_i}{dt} + \frac{\partial H}{\partial q_i} \right) \right] dt = 0.$$

$$\int_{t_1}^{t_2} \left[\sum_j B_j \left(\frac{dq}{dt} - \frac{\partial H}{\partial p_j} \right) - \psi \left(\frac{dp_i}{dt} + \frac{\partial H}{\partial q_i} \right) \right] dt = 0.$$

Now we know that the curve $\overline{C}$ satisfies the first half of Hamilton's equations, so the sum in the integrand vanishes, and condition (17) reduces to the condition

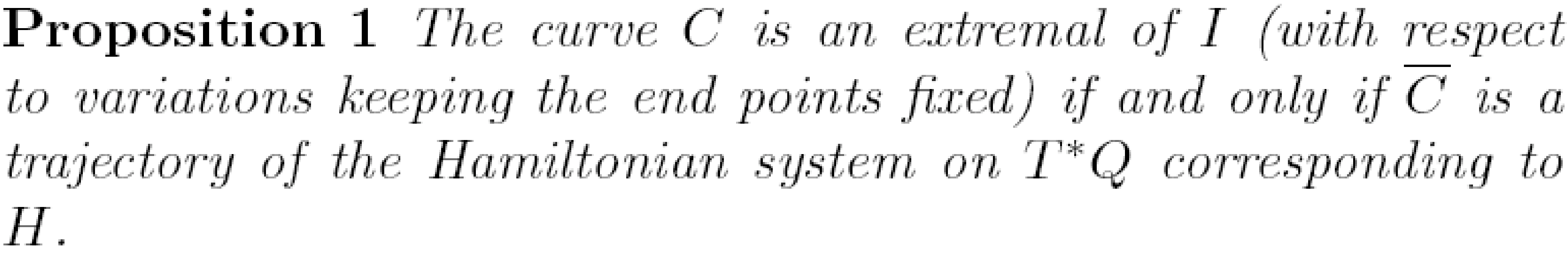
$$\int_{t_1}^{t_2} \psi \left(\frac{dp_i}{dt} + \frac{\partial H}{\partial q_i} \right) dt = 0.$$

This must hold for all ψ whose support lies in a coordinate neighborhood and which vanishes at p and q. This can only happen if

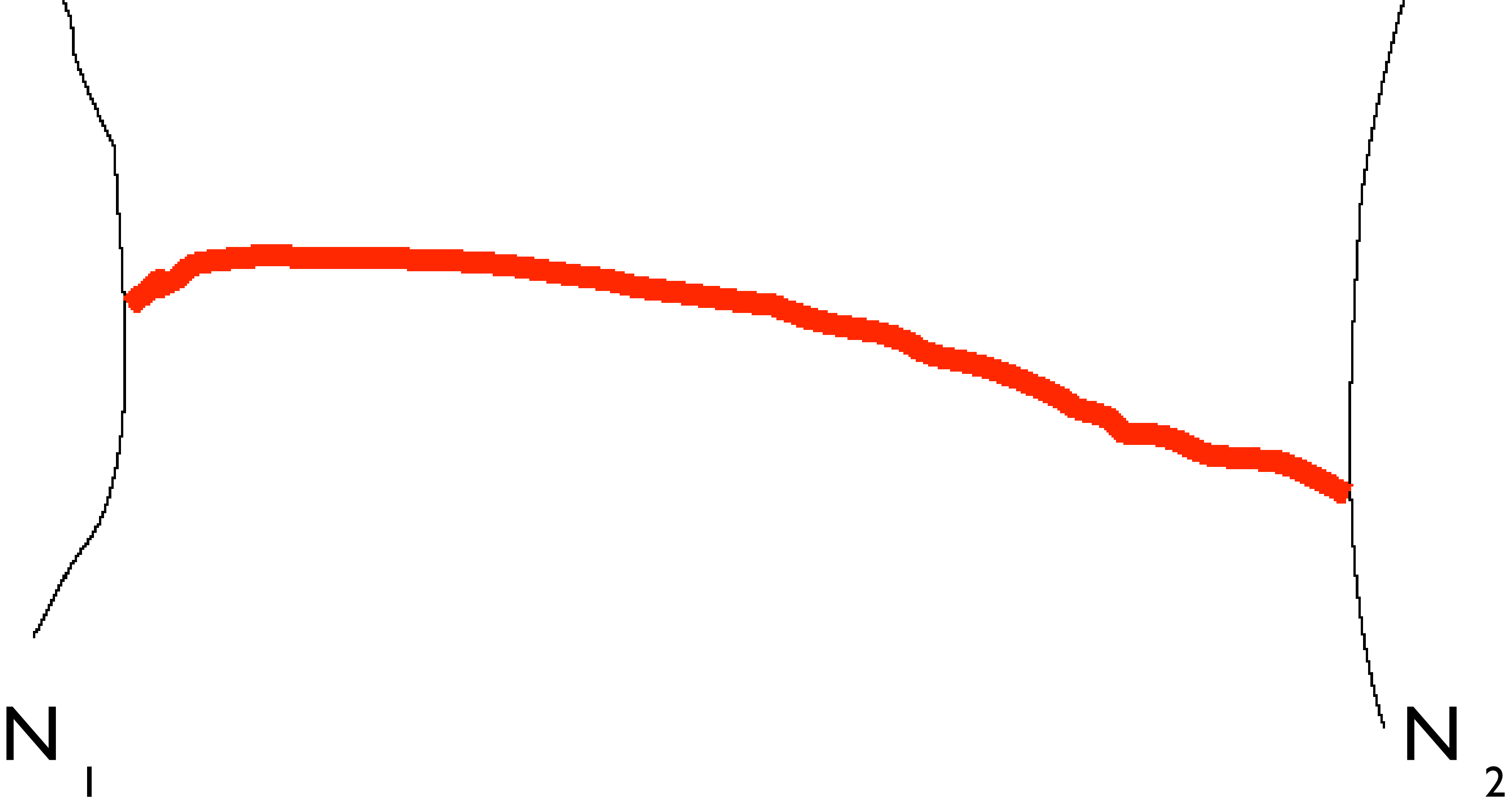
$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}.$$

This must hold for all i. So we have proved

Main result.



Proposition 1 *The curve C is an extremal of I (with respect to variations keeping the end points fixed) if and only if $\overline{C}$ is a trajectory of the Hamiltonian system on T^*Q corresponding to H.*



Let us go back to equation (16) and consider the following problem: Minimize I along all curves joining submanifolds N_1 and N_2, not merely along curves joining two points.

$$\frac{d}{ds}I[\phi_s \circ C]_{|s=0} =$$

$$-\int_{\overline{C}} i(Z_*)\omega - \int_{t_1}^{t_2} (D_{Z_*}H)(\overline{C}(t))dt + \langle \overline{C}(t_2), Z(C(t_2))\rangle - \langle \overline{C}(t_1), Z(C(t_1))\rangle. \tag{16}$$

$$\frac{d}{ds} I[\phi_s \circ C]_{|s=0} =$$

$$- \int_{\overline{C}} i(Z_*)\omega - \int_{t_1}^{t_2} (D_{Z_*} H)(\overline{C}(t))dt + \langle \overline{C}(t_2), Z(C(t_2)) \rangle - \langle \overline{C}(t_1), Z(C(t_1)) \rangle. \tag{16}$$

Certainly such a curve will have to solve the easier minimization problem, and so be a trajectory of the mechanical system. So in (16) the integrals will vanish, and we get the additional conditions:

$$\langle \overline{C}(t_1), v \rangle = 0 \quad \forall v \in T_p(N_1)$$

and

$$\langle \overline{C}(t_2), v \rangle = 0 \quad \forall v \in T_p(N_2).$$

In the case where $H = K$ is the kinetic energy of a Riemann metric, this says that the curve must be orthogonal to N_1 and N_2.

Some Riemannian geometry.

I now want to derive some results in Riemannian geometry which we will use as tools in our study of symplectic geometry.

Let Q be a manifold with a Riemann metric. This determines a Hamiltonian vector field on T^*Q and, via the Legendre transform, a vector field, call it Y on TQ. Let $p \in Q$. For each $v \in T_pQ$ there is a unique (local) trajectory C_v, such that

$$C_v(0) = p, \quad C_v'(0) = v.$$

This is guaranteed by the fundamental existence and uniqueness theorem for ordinary differential equations as applied to the Euler-Lagrange equations. We can regard $C_v(t)$ as a function of v and t. This is true for a general Lagrangian. But for the case of pure kinetic energy we can say more:

Let s be a real number. Consider the curves (for various v)

$$t \mapsto C_v(st).$$

In local coordinates we have $\dfrac{dq_i(st)}{dt} = s\dot{q}_i(st),$

$$\frac{\partial L}{\partial q_i}(q_1(st), \ldots, q_n(st), s\dot{q}_1(st), \ldots, s\dot{q}_n(st)) = \frac{1}{2}s^2 \sum_{k\ell} \frac{\partial g_{k\ell}}{\partial q_i}\dot{q}_k(st)\dot{q}_\ell(st),$$

and

$$\frac{d}{dt}\left[\frac{\partial L}{\partial \dot{q}_i}(q_1(st), \ldots, q_n(s), s\dot{q}_1(st), \ldots, \dot{q}_n(st))\right]$$
$$= \frac{d}{dt}s\sum_k g_{ik}(q_1(st), \ldots, q_n(st))\dot{q}_k(st)$$

Doing the differentiation with respect to t pulls out another factor of s.

$$\frac{dq_i(st)}{dt} = s\dot{q}_i(st),$$

$$\frac{\partial L}{\partial q_i}\left(q_1(st), \ldots, q_n(st), s\dot{q}_1(st), \ldots, s\dot{q}_n(st)\right) = \frac{1}{2}s^2 \sum_{k\ell} \frac{\partial g_{k\ell}}{\partial q_i}\dot{q}_k(st)\dot{q}_\ell(st),$$

and

$$\frac{d}{dt}\left[\frac{\partial L}{\partial \dot{q}_i}(q_1(st), \ldots, q_n(s), s\dot{q}_1(st), \ldots, \dot{q}_n(st))\right]$$

$$= \frac{d}{dt} s \sum_k g_{ik}(q_1(st), \ldots, q_n(st))\dot{q}_k(st)$$

Doing the differentiation with respect to t pulls out another factor of s. So we see that the curve $t \mapsto C_v(st)$ is a again a solution of the Euler-Lagrange equation, this time with initial point p but initial vector sv. The uniqueness theorem of ordinary differential equations tells us that

$$C_v(st) = C_{sv}(t). \tag{18}$$

The exponential map.

$$C_v(st) = C_{st}(v). \tag{18}$$

The map exp by

$$v \mapsto C_v(1).$$

This is defined in some neighborhood of the origin. Indeed, by (18)

$$C_v(1) = C_u(\|v\|) \quad \text{where } u = \frac{1}{\|v\|}v$$

is the unit vector in the v direction (if $v \neq 0$). Since the unit sphere is compact, there is some $\epsilon > 0$ such that $C_t(u)$ is defined for all $|t| < \epsilon$ and then exp is defined for all $\|v\| < \epsilon$ (with $\exp(0) = p$).

The map exp is a differentiable map from some neighborhood of the origin in the vector space T_pQ into Q. I want to compute its derivative at the origin. In fact, I claim that if we identify the tangent space $T_0(T_pQ))$ with T_pQ (which we can do since T_PQ is a vector space, then the derivative of exp at 0 is the identity. To see this, let $v \in T_pQ$ and consider the line

$$t \mapsto \ell_v(t) := tv.$$

The tangent vector to this line at 0 (under our identification) is just v. But

$$\exp(\ell_v(t)) = \exp(tv) = C_{tv}(1) = C_v(t)$$

also has v as its tangent vector at the origin. So we have proved that

$$d(\exp_0) = \mathrm{id}\,.$$

$$d(\exp_0) = \mathrm{id}\,.$$

The implicit function theorem then tells us that exp is a diffeomorphism in some neighborhood of the origin, and that exp carries straight lines through the origin into trajectories. These trajectories are called **geodesics** for reasons which will soon become apparent.

Consider the curve $C_v(\cdot) = \exp(\cdot v)$ which is defined for $0 \leq t \leq 1$ so long as $\|v\| < \epsilon$. We have

$$I[C_v(\cdot)] = \int_0^1 L(C_v'(t))dt = \frac{1}{2}\int_0^1 \|C_v'(t)\|^2 dt.$$

By the conservation of (kinetic) energy, $\frac{1}{2}\|C_v'(t)\|^2$ is constant. Since $C_v'(0) = v$, we see that

$$I[C_v(\cdot)] = \frac{1}{2}\|v\|^2.$$

Gauss's lemma.

Let β_s be a one parameter group of rotations about the origin in T_pQ. Then

$$\phi_s := \exp \circ \beta_s \circ \exp^{-1}$$

defines a one parameter group on the open set of Q which is the image under exp of the set $\|v\| < \epsilon$ in T_pQ. Now

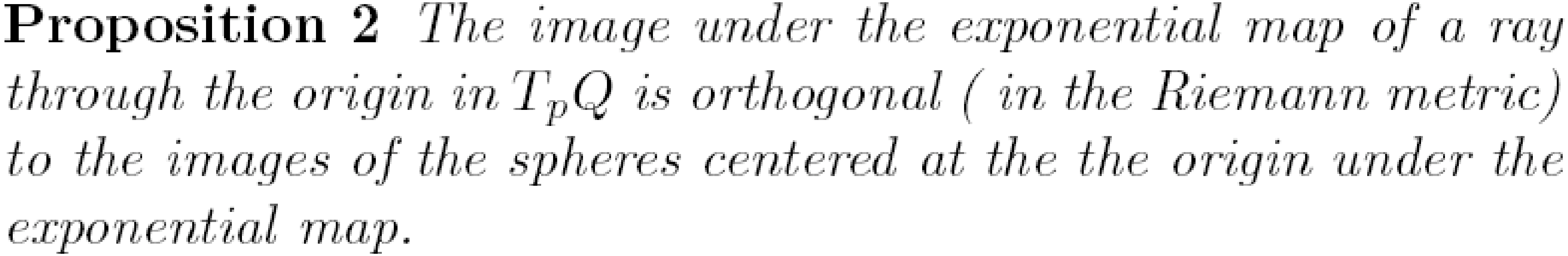
$$I[\phi_s \circ C_v(\cdot)] = \frac{1}{2}\|\beta_s v\|^2 = \frac{1}{2}\|v\|^2.$$

Hence from (16)

$$\langle C_v'(1), Z_{C(1)}\rangle = \frac{d}{ds} I[\phi_s \circ C_v(\cdot)]_{|s=0} = 0,$$

where Z is the vector field generating ϕ_s. As β varies over all rotations in T_pQ, the tangent vectors $Z_{C(1)}$ vary over all tangent vectors to the image of the sphere $S_{\|v\|}$ under exp. This is Gauss's lemma:

Gauss's lemma.



Proposition 2 *The image under the exponential map of a ray through the origin in* T_pQ *is orthogonal (in the Riemann metric) to the images of the spheres centered at the the origin under the exponential map.*

Geodesics locally minimize arc-length.

Let O be the image of $\|v\| < \epsilon$ under the exponential map, and let r be the function defined on O by

$$r(x) := \| \exp^{-1}(x)\|.$$

Let $x \in O, \quad x \neq p$ and $w \in T_xQ$. I claim that

$$\|w\| \geq |\langle dr(x), w\rangle| \tag{19}$$

with equality holding only if w is tangent to a geodesic through p.

$$\|w\| \geq |\langle dr(x), w\rangle| \tag{19}$$

with equality holding only if w is tangent to a geodesic through p.

Proof. Write $x = \exp v, \quad v \in T_pQ$. Decompose

$$w = w_1 + w_2$$

where w_1 is some multiple of $C'_v(1)$ and w_2 is tangent to the image of the sphere through v under exp. We know that this decomposition is orthogonal, and so

$$\|w\|^2 = \|w_1\|^2 + \|w_2\|^2.$$

The value of $dr(x)$ on $C'_v(1)$ is $\|v\|$, proving (19) with equality holding only if $w_2 = 0$.

Now let D be any curve joining p to $x = \exp v \in O$. The length of D is

$$\int_0^1 \|D'(t)\| dt$$

by definition. Let t_1 be the first time that $D(t) \in \exp S_{\|v\|}$ where $S_{\|v\|}$ is the sphere of radius $\|v\|$ about the origin in T_pQ. Then

$$\int_0^1 \|D'(t)\| dt \geq \int_0^{t_1} \|D'(t)\| dt.$$

By (19) this last integral is $\geq \int_0^{t_1} |\langle dr, D'(t)\rangle| dt$. But

$$\int_0^{t_1} |\langle dr, D'(t)\rangle| dt \geq \int_0^{t_1} \langle dr, D'(t)\rangle dt = r(D'(t)) = \|v\|.$$

Furthermore equality holds only if $D'(t)$ is a non-negative multiple of a tangent vector to a fixed geodesic through p. We have proved:

Theorem 1 *Let $\epsilon > 0$ be so small that the exponential map is a diffeomorphism of the set $\|v\| < \epsilon$ in T_pQ onto an open set O about p in Q. Let $x = \exp v \in O$. then the geodesic C_v joining p to x has length $\|v\|$ and any other curve D joining p to x is strictly longer unless D differes from C_v only in a monotone change of parameter.*

Geodesics also locally minimize the energy.

Let D be any smooth curve from $[0,1]$ to Q. The Cauchy-Schwarz inequality tells us that

$$\left(\int_0^1 \|D'(t)\| dt\right)^2 \leq \int_0^2 \|D'(t)\|^2 dt \int_0^1 1 dt = 2I[D],$$

with equality holding only if $\|D'(t)\|$ is a constant. If D joins p to $x = \exp v \in O$ we get

$$I[D] \geq \frac{1}{2}\left(\int_0^1 \|D'(t)\| dt\right)^2 \geq \frac{1}{2}\left(\int_0^1 \|C_v'(t)\| dt\right)^2 = \frac{1}{2}\|v\|^2.$$

Equality holds in the second inequality only if $D'(t)$ is proportional to $C_v'(t)$ for all t while equality holds in the first inequality only if $\|D'(t\|$ is a constant. We conclude that $D'(t) = C_v'(t)$. We have proved

Theorem 2 *Under the hypotheses of Theorem 1 the curve* C_v *is a strict absolute minimum for* $I[C]$ *among all curves* $C : [0,1] \to Q$ *joining* p *to* x.