Lecture 7

Hamiltonian mechanics on the cotangent bundle.

Shlomo Sternberg

July 21, 2010
1. Review - The cotangent bundle.
2. More on Hamiltonians on the cotangent bundle.
   - Conservation of energy.
   - Kinetic energy.
3. The Euler-Lagrange equations.
   - The Euler-Lagrange equations.
   - The principle of mechanical similarity.
4. The calculus of variations on the cotangent bundle.
5. Some Riemannian geometry.
   - The exponential map.
   - Gauss’s lemma.
   - Geodesics locally minimize arc-length.
   - Geodesics also locally minimize the energy.
7. Appendix: The Legendre transformation as a Lagrangian submanifold.
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 $T^*_xQ$ 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 $T^*_xQ$ to $x$ is smooth.
Review: The canonical one form on a cotangent bundle.

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 $p \in T^*_xQ$ to its base point $x$. If $v \in T_p(T^*Q)$ is a tangent vector to $T^*Q$ at $p$, then

$$d\pi_p v$$

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

$$\langle \alpha_p, v \rangle := \langle p, d\pi_p v \rangle \quad \text{if} \quad v \in T_p(T^*Q). \quad (1)$$
The canonical two form on the cotangent bundle.

This is defined as

$$\omega_Q = -d\alpha_Q.$$  \hfill (2)

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

$$\omega = dq \wedge dp = dq_1 \wedge dp_1 + \cdots + dq_n \wedge dp_n.$$  \hspace{1cm} (3)
Hamiltonian vector fields.

If $H$ is a $C^\infty$ function on $T^*Q$ then we get the corresponding vector field $X_H$ determined by

$$i(X_H)\omega = dH.$$
Hamilton’s equations on the cotangent bundle.

If $H$ is a $C^\infty$ function on $T^*Q$ and $X_H$ is the corresponding vector field, then in terms of the local expression (3) finding the trajectories of $X_H$ amounts to solving 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

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

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

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

The particle moves along a straight line with constant velocity (Galileo’s law) and $p = m\dot{q}$.
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 \( \{ H, H \} = 0. \)
Conservation of energy.

Noether’s theorem.

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

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

So

$$X_FH = 0 \Rightarrow X_HF = 0. \quad (4)$$

This argument applies, of course, to any symplectic manifold, not merely to the cotangent bundle.
Conservation of energy.

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 the pull-back of some smooth function on $Q$ via $\pi^*$. 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 $V$:

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

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

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

\[
K(\xi) = \frac{1}{2}(\xi, \xi).
\]
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 $T_x Q$ with $Q$ at each $x \in Q$ and so get coordinates $\dot{q}_1, \dot{q}_2, \dot{q}_3$ on each $T_x Q$. Suppose that we choose our metric to be $m \times$ (the Euclidean metric). Then, for each $x$

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

and hence

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

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 $\mathcal{L}$ is given by

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

where

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

\[ p_i = \sum_j g_{ij}(q_1, \ldots, q_n) \dot{q}_j. \quad (5) \]

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

\[ K(q_1, \ldots, q_n, p_1, \ldots, p_n) = \frac{1}{2} \sum_{ij} g^{ij}(q_1, \ldots, q_n)p_ip_j. \quad (6) \]
Kinetic energy.

\[ p_i = \sum_j g_{ij}(q_1, \ldots, q_n) \dot{q}_j. \quad (5) \]

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

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

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

\[ p_i = \frac{\partial L}{\partial \dot{q}_i}. \quad (7) \]
Kinetic energy.

The Legendre transformation.

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 transformation**.

In the Appendix to this lecture we show that the “right way” to think of a Legendre transformation is to think of it as a map from a vector space $V$ to its dual $V^*$ whose graph is a Lagrangian submanifold of $V \oplus V^*$. Here, at each point $q \in Q$ we take $V = T_qQ$.

For the time being, we compute in local coordinates.
Kinetic energy.

So 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).
Kinetic energy.

The Legendre transformation in local coordinates.

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

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

Suppose that the map \( \mathcal{L} \) is a diffeomorphism. Then the inverse map

\[
\mathcal{L}^{-1} : T^*Q \to TQ
\]

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

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

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

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

(8)
Then

\[
\frac{\partial H}{\partial p_i} = \dot{q}_i \circ \mathcal{L}^{-1} + \sum_j p_j \frac{\partial \dot{q}_j \circ \mathcal{L}^{-1}}{\partial p_i} - \sum_j \frac{\partial L}{\partial \dot{q}_j} \circ \mathcal{L}^{-1} \times \frac{\partial \dot{q}_j \circ \mathcal{L}^{-1}}{\partial p_i}.
\]

Since

\[p_i = \frac{\partial L}{\partial \dot{q}_i}\]

the last two terms cancel and we get

\[
\frac{\partial H}{\partial p_i} = \dot{q}_i.
\]
Kinetic energy.

For example, suppose that $Q$ is a Riemannian 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 $\nu \mapsto \frac{1}{2} (\nu, \nu)$ by $K_T$ (the kinetic energy expressed in terms of the tangent bundle). The local expression for $K_T$ is, as before,

$$K_T(q_1, \ldots, q_n, \dot{q}_1 \ldots \dot{q}_n) = \frac{1}{2} \sum_{ij} g_{ij}(q_1, \ldots, q_n) \dot{q}_i \dot{q}_j.$$
Finally, we let \( K \) be the function on \( T^*Q \) given by

\[
K(\xi) = \frac{1}{2}(\xi, \xi) = \frac{1}{2}\langle \xi, \xi^{-1}\xi \rangle.
\]

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

\[
K(q_1, \ldots, q_n, p_1, \ldots, p_n) = \frac{1}{2} \sum_{ij} g^{ij}(q_1, \ldots, q_n)p_ip_j.
\]

Now let \( L \) be the function on \( TQ \) given by

\[
L = K_T - U.
\] \hspace{1cm} (10)

Then

\[ H(\xi) = \langle \xi, \xi^{-1}(\xi) \rangle - L \circ \xi^{-1}(\xi) \]

or

\[ H = K + U. \]
The first half of Hamilton’s equations.

We return to general considerations. We stay with the assumption that $\mathcal{L}$ gives an isomorphism from $TQ$ to $T^*Q$.

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 $\mathcal{L}^{-1}$ then become

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

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 \( C_T \).
The second half of Hamilton’s equations.

We have

\[ \frac{\partial H}{\partial q_i} = \frac{\partial}{\partial q_i} \left[ \sum_j p_j \left( \dot{q}_j \circ \mathcal{L}^{-1} \right) - L \circ \mathcal{L}^{-1} \right] \]

\[ = \sum_j p_j \frac{\partial \dot{q}_j \circ \mathcal{L}^{-1}}{\partial q_i} - \frac{\partial L \circ \mathcal{L}^{-1}}{\partial q_i} - \sum_j \frac{\partial L}{\partial q_j} \frac{\partial \dot{q}_j \circ \mathcal{L}^{-1}}{\partial q_i}. \]

The first and last terms cancel since

\[ p_j = \frac{\partial L}{\partial \dot{q}_j} \]

as a function on \( TQ \). So

\[ \frac{\partial H}{\partial q_i} = - \frac{\partial L \circ \mathcal{L}^{-1}}{\partial q_i}. \] (12)
So the second half of Hamilton’s equations read

\[ \frac{dp_i}{dt} = \frac{\partial L \circ \mathcal{L}^{-1}}{\partial q_i}. \]

If we substitute \( p_i = \frac{\partial L}{\partial q_i} \), we obtain

\[ \frac{d(\partial L/\partial \dot{q}_i)}{dt} - \frac{\partial L}{\partial q_i} = 0. \] (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

\[ \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = 0 \]  

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 \( \alpha \), and suppose that the \( g_{ij} \) are constant under the diffeomorphism given by 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 $\beta$, replacing $t$ by $s = \beta t$. Then

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

so

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

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

So

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

and hence

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

So replacing $q_i$ by $\alpha q_i$ and $t$ by $\beta t$ carries solutions of the Euler-Lagrange equations into solutions.
Kepler’s third law.

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”.
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 (or extremizes) $I[C]$. 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

$$\bar{C}(t) := \mathcal{L}(C_T(t))$$

is a solution of Hamilton's equations (at least if $\mathcal{L} : TQ \rightarrow T^*Q$ 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

$$I[C] = \int_C \alpha - \int_{t_1}^{t_2} H(C(t)) dt$$  \hspace{1cm} (14)$$

where $\alpha = \alpha_Q$ is the fundamental one form on $T^* Q$. 
To prove: \( I[C] = \int_C \alpha - \int_{t_1}^{t_2} H(C(t))dt \). (14)

**Proof.**

In local coordinates

\[
\int_C \alpha = \sum_i \int_C p_i dq_i = \sum_i \int_{C_T} \frac{\partial L}{\partial \dot{q}_i} dq_i = \sum_i \int_{t_1}^{t_2} \frac{\partial L}{\partial q_i} \dot{q}_i(t) dt
\]

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

Now

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

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

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

will be defined, and at any $t$ we have

$$(\phi_s \circ C)'(t) = d(\phi_s)_C(t)(C'(t)).$$

Therefore, by (14),

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

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

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

Shlomo Sternberg

Lecture 7
\[ I[\phi_s \circ C] = \int_{\mathcal{L} \circ (d\phi_s) \circ C_T} \alpha - \int_{t_1}^{t_2} H((\mathcal{L} \circ (d\phi_s)(C_T(t))) \, dt. \quad (15) \]

We want 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 \( \mathcal{L} \). That is, we can consider the flow \( \mathcal{L} \circ (d\phi_s) \circ \mathcal{L}^{-1} \) on \( T^*Q \). Clearly

\[ \pi \circ (\mathcal{L} \circ (d\phi_s) \circ \mathcal{L}^{-1}) = \phi_s. \]

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

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

\[ \int_C DZ_* \alpha - \int_{t_1}^{t_2} (DZ_* H)(\overline{C}(t)) \, dt. \]
Let us apply Weil’s formula and the definition of the fundamental form $\alpha$: We have $\text{D}_{Z^*} \alpha = d\langle \alpha, Z^* \rangle + i(Z^*)d\alpha$ and at any $\xi \in T^* Q$ we have $\langle \alpha, Z^* \rangle = \langle \xi, Z(\pi(\xi)) \rangle$. So

$$\int_{C} \text{D}_{Z^*} \alpha = \int_{C} i(Z^*)d\alpha + \int_{C} d\langle \overline{C}(t), Z_C(t) \rangle$$

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

Finally, we obtain the formula

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

$$-\int_{C} i(Z^*)\omega - \int_{t_1}^{t_2} \langle \text{D}_{Z^*} H)(\overline{C}(t))dt + \langle \overline{C}(t_2), Z(C(t_2)) \rangle - \langle \overline{C}(t_1), Z(C(t_1)) \rangle$$

(16)
\[ \frac{d}{ds} I[\phi_s \circ C]_{s=0} = \]
\[ = - \int_C i(Z_*) \omega - \int_{t_1}^{t_2} (DZ_* H)(\overline{C}(t)) dt + \langle \overline{C}(t_2), Z(C(t_2)) \rangle - \langle \overline{C}(t_1), Z(C(t_1)) \rangle. \]

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

\[ \int_C i(Z_*) \omega + \int_{t_1}^{t_2} (DZ_* H)(\overline{C}(t)) dt = 0. \] (17)
Let us choose the flow $\phi_s$ to be the identity outside a coordinate neighborhood on $Q$ and on the coordinate neighborhood look like

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

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

- $q_j \mapsto q_j, \ j \neq i,$
- $q_i \mapsto q_i + s\psi$
- $\dot{q}_j \mapsto \dot{q}_j, \ j \neq i,$
- $\dot{q}_i \mapsto \dot{q}_i + s \sum_j \frac{\partial \psi}{\partial q_j} \dot{q}_j.$
So the vector field $Z_T$ generating this flow on $TQ$ is

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

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

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

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

\[ D_{Z^*} H = \psi \frac{\partial H}{\partial q_i} + \sum_j B_j \frac{\partial H}{\partial p_j}. \]

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

Now we know that the curve \( \bar{C} \) 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 \( \psi \) 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

**Proposition**

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

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. 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 $N_2$. 
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 (inverse of the) Legendre transform, a vector field, call it $Y$ on $TQ$. Let $x \in Q$. For each $v \in T_xQ$ there is a unique (local) trajectory $C_v$ of $Y$ such that

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

This is guaranteed by the fundamental existence and uniqueness theorem for ordinary differential 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

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

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

and

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

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

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 \( x \) but initial vector \( sv \).
The uniqueness theorem of ordinary differential equations tells us that

\[ C_v(st) = C_{sv}(t). \] (18)
The exponential map.

The map \( \exp : T_x Q \to Q \) is defined by

\[ v \mapsto C_v(1). \]

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

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

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 \( C_u(t) \) is defined for all \( |t| < \epsilon \) and then \( \exp \) is defined for all \( \|v\| < \epsilon \) (with \( \exp(0) = x \)).
The map \( \exp \) is a differentiable map from some neighborhood of the origin in the vector space \( T_xQ \) into \( Q \). I want to compute its derivative at the origin. In fact, I claim that if we identify the tangent space \( T_0(T_xQ) \) with \( T_xQ \) (which we can do since \( T_xQ \) is a vector space), then the derivative of \( \exp \) at 0 is the identity. To see this, let \( v \in T_xQ \) 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

\[
d(\exp_0) = \text{id}.
\]
The exponential map.

\[ d(\exp_0) = \text{id}. \]

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 \textbf{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

$$I[C_v(\cdot)] = \frac{1}{2} \|v\|^2.$$
Let $\beta_s$ be a one parameter group of rotations about the origin in $T_x Q$. 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 $T_p Q$. 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 $\phi_s$. As $\beta$ varies over all rotations in $T_p Q$, 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**

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

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

Let $y \in O$, $y \neq x$ and $w \in T_y Q$. I claim that

$$\|w\| \geq |\langle dr(y), w \rangle|$$

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

(19)
Geodesics locally minimize arc-length.

Proof.

Write $y = \exp v$, $v \in T_p Q$. 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 $w_2 = 0$. $\square$
Geodesics locally minimize arc-length.

Now let $D$ be any curve joining $x$ to $y = \exp v \in O$. 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

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

Furthermore, equality holds only if $D'(t)$ is a non-negative multiple of a tangent vector to a fixed geodesic through $x$. We have proved:
Geodesics locally minimize arc-length.

**Theorem**

Let $\epsilon > 0$ be so small that the exponential map is a diffeomorphism of the set $\|v\| < \epsilon$ in $T_x Q$ onto an open set $O$ about $x$ in $Q$. Let $y = \exp v \in O$. Then the geodesic $C_v$ joining $x$ to $y$ has length $\|v\|$ and any other curve $D$ joining $x$ to $y$ is strictly longer unless $D$ differs 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

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

with equality holding only if $\| D'(t) \|$ is a constant. If $D$ joins $x$ to $y = \exp v \in \mathcal{O}$ 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

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 $x$ to $y$. 

Geodesics also locally minimize the energy.
We have seen that if $L$ is a function on $TQ$ and if

$$I[C] = \int_C L(C(t), C'(t))dt$$

then

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

Here $\alpha$ is the canonical one form on $T^*Q$, $\overline{C}$ is the image in $T^*Q$ of the curve $C_T = (C, C')$ in $TQ$ via the Legendre transformation

$$\mathcal{L} : TQ \rightarrow T^*Q$$

determined by $L$, and $H$ is the Hamiltonian function on $T^*Q$ associated to $L$. 
From this we derived the fact that a curve $\gamma$ in $T^*Q$ is a trajectory of the vector field $X_H$ if and only if two conditions are satisfied:

- $\gamma = \mathcal{L}(C_T) = \mathcal{L}(C, C')$ and
- $C$ is an extremal of $I$ (relative to variations with fixed end points).

In the proof of this assertion we used the fact that a curve $\gamma$ in $T^*Q$ was of the form $\gamma = \mathcal{L}(C_T)$ if and only if the first half of Hamilton’s equations were satisfied, and then verified the second assertion above by varying the curve $C$ on $Q$. 
But we can consider a different variational problem - where we vary among \textit{all} possible smooth curves on $T^*Q$, not necessarily curves on $T^*Q$ arising from curves on $Q$.

We will formulate this in the slightly more general setting of exact symplectic manifolds. We recall that a symplectic manifold $(M, \omega)$ is called \textbf{exact} if there is a one form $\alpha$ on $M$ such that $d\alpha = -\omega$, and that we have chosen such a one form.
Hamilton’s principle.

Let $H$ be a smooth function on an exact symplectic manifold $M = (M, \alpha)$. For any smooth curve $\gamma : [a, b] \to M$ define

$$A_H(\gamma) := \int_\gamma \alpha - \int_a^b H \circ \gamma(t) dt.$$

Let $\Lambda_1$ and $\Lambda_2$ be submanifolds of $M$ such that $\alpha = 0$ when pulled back to $\Lambda_1$ or $\Lambda_2$. 
Theorem

For a curve $\gamma : [a, b] \rightarrow M$ with $\gamma(a) \in \Lambda_1$ and $\gamma(b) \in \Lambda_2$ the following are equivalent:

- $\gamma$ is an extremal of $A_H$ relative to variations with $\gamma(a) \in \Lambda_1$ and $\gamma(b) \in \Lambda_2$.
- $\gamma$ is an extremal of $A_H$ relative to variations with fixed end points.
- $\gamma$ is a trajectory of $X_H$. 
Proof, part 1.

Let $Y$ be a vector field defined in a neighborhood of $\gamma$ with $Y(\gamma(a)) \in T_{\gamma(a)}\Lambda_1$ and $Y(\gamma(b)) \in T_{\gamma(b)}\Lambda_2$. Let $\phi_s$ be the flow generated by $Y$. Then

$$\frac{d}{ds}A_H(\phi_s \circ \gamma)|_{s=0} = \int_\gamma D_Y \alpha - \int_a^b [(D_Y H) \circ \gamma](t)dt.$$

By Weil's formula the first integral is

$$\int_\gamma i(Y)d\alpha + \int_\gamma d\alpha(Y) = \int_\gamma i(Y)d\alpha + \alpha(Y(\gamma(b))) - \alpha(Y(\gamma(a))).$$

Under either of the first two assumptions,

$$\alpha(Y(\gamma(b))) = \alpha(Y(\gamma(a))) = 0.$$
Proof, part 2.

So the vanishing of

\[
\frac{d}{ds} A_H(\phi_s \circ \gamma)|_{s=0} = \int_{\gamma} D_Y \alpha - \int_{a}^{b} [(D_Y H) \circ \gamma](t) dt
\]

for all \( Y \) implies and is equivalent to

\[
i(Y)d\alpha(\dot{\gamma}(t)) - (i(Y)dH)(\gamma(t)) = 0
\]

for all \( t \) and all \( Y \) which is the same as saying that

\[
i(\dot{\gamma}(t))d\alpha(Y) = -i(Y)dH(\gamma(t))
\]

and since \( \omega = -d\alpha \) this is the same as saying that \( \gamma \) is a trajectory of \( X_H \). \( \square \)
The Legendre transformation as a Lagrangian submanifold.

Let \((M, \omega)\) be a symplectic manifold with

\[ \omega = -d\beta \quad \text{and} \quad \omega = d\gamma \]

for one forms \(\beta\) and \(\gamma\). So if \(M\) is simply connected, then

\[ \beta + \gamma = dh \]

for some function \(h\) on \(M\).
Let $\iota : \Lambda \to M$ be a Lagrangian submanifold, so both $\iota^* \beta$ and $\iota^* \gamma$ are closed. So if $\Lambda$ is simply connected, then there are functions $f$ and $g$ on $\Lambda$ such that

$$\iota^* \beta = df \quad \text{and} \quad \iota^* \gamma = dg.$$ 

Since $\iota^*(\beta + \gamma) = \iota^* dh$ we have

$$df + dg = d\iota^* h.$$ 

So starting with $f$ we can choose the local constants on $\Lambda$ so that

$$g = \iota^* h - f.$$
Suppose that $M = V \oplus V^*$ where $V$ is a vector space with linear coordinates $x = (x_1, \ldots, x_d)$ on $V$ and dual coordinates $y = (y_1, \ldots, y_d)$ on $V^*$. We take

$$\omega = dx \wedge dy = dx_1 \wedge dy_1 + \cdots + dy_d \wedge dx_d.$$ 

So $\omega$ is the canonical two form on $M$ if we think of $M$ as $T^*V$, and $\omega = -d\beta$ where

$$\beta = y \cdot dx$$

is the canonical one form. Similarly, if we think of $M$ as the cotangent bundle of $V^*$, we take

$$\gamma = x \cdot dy$$

so

$$\beta + \gamma = dh, \quad h = x \cdot y.$$
Now suppose that the projections $\text{pr}_1 : \Lambda \to V$ and $\text{pr}_2 : \Lambda \to V^*$ are diffeomorphisms. Let $L = f \circ \text{pr}_1^{-1}$. The equation $i^* \beta = df$ becomes

$$y \cdot dx = dL$$
on $V$, i.e.

$$y_i = \frac{\partial L}{\partial x_i}$$

which is just the Legendre transformation corresponding to $L$. In other words

$$\mathcal{L} = \text{pr}_2 \circ \text{pr}_1^{-1},$$

i.e. $\Lambda$ is the graph of $\mathcal{L}$. Interchanging the role of $V_1$ and $V_2$ and setting $H = g \circ \text{pr}_2^{-1}$ gives the inverse Legendre transformation where

$$H(y) = x \cdot y - L(x)$$

where $x$ is thought of as a function of $y$ via $\mathcal{L}^{-1}$. 
In short, we should think of the Legendre transformation as being the special case of a Lagrangian subspace $\Lambda \subset V \oplus V^*$ where the projections $\text{pr}_1 : \Lambda \to V$ and $\text{pr}_2 : \Lambda \to V^*$ are diffeomorphisms in which case

$$\mathcal{L} = \text{pr}_2 \circ \text{pr}^{-1}_1, \quad \text{and} \quad \mathcal{L}^{-1} = \text{pr}_1 \circ \text{pr}^{-1}_2.$$
In our application of the Legendre transformation to the passage from Lagrangian to Hamiltonian mechanics, we took $V = T_q Q$ and $V^* = T^*_q Q$, and the Legendre transformation depended on $q \in Q$ as a parameter. So in terms of the discussion of the past few slides we should make the substitutions

$$x_i \leftrightarrow \dot{q}_i \quad \text{and} \quad y_i \leftrightarrow p_i.$$