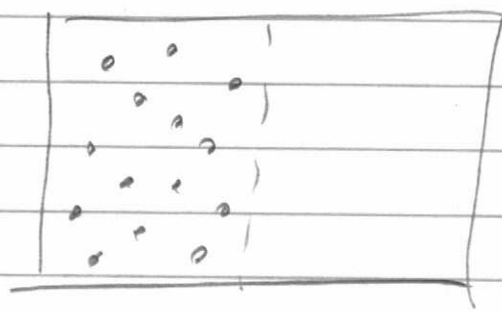
Mechanics - Lecture 10 - 4/5/2022

Addendum to the Lecture 9 notes: a thought-provoking example of Poincaré's recurrence thm involves a gas, modeled as many masses interacting by a pair potential (eg Lennard-Jones). Suppose a chamber has walls that reflect the particles, and they start initially in the left half of the chamber, at rest.



Soon, we expect that they'll be approx equally distributed in both halves! But recurrence thm says there's an initial state very similar to this to which system recurs in finite time! (Resolution of paradox offered by Arnold: recurrence time is far beyond any time we can observe.)

Let's turn now to studying why and how the Lagrangian + Hamiltonian viewpoints are related.

I mentioned already that in asserting such a connection, we need the structural hypothesis that

$$L(q,\dot q) \text{ is convex in } \dot q$$
$$(\text{with } q \text{ held fixed})$$

Let's spend a moment observing mystrice of convexity for $\min \int_{t_1}^{t_2} L(q,\dot q)\,dt$:

(a) when convexity fails, $\dot q$ can easily be discontinuous; the elementary example

$$\min \int_0^1 (u_x^2-1)^2\,dx$$

shows this (in fact any u st $u_x = \pm 1$ is a global min). Write this as

$$\int_0^1 W(u_x)\,dx \quad ;$$

then EL eqn is $\frac{\partial}{\partial x} W'(u_x) = 0$. If W is convex then W' is monotone, so u_x jumps $\Rightarrow$ $W'(u_x)$ jumps $\Rightarrow$

$\frac{d}{dx} W'(u_x)$ would have a pt mass at the location of the jump.

But if W' is non-monotone then a jump in u_x is possible, as the example shows.

In fact, if $L(q,\dot{q})$ is strictly convex with

$$\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \geq c\|\dot{q}\|^2$$

then solns of the EL eqn are as smooth as L is. The formal "proof" is easy:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i}$$

$$\Rightarrow \quad \sum_j \frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \ddot{q}_j = \frac{\partial L}{\partial q_i} - \sum_j \frac{\partial^2 L}{\partial \dot{q}_i \partial q_j} \dot{q}_j$$

Inverting the matrix on the left gives a "formula" for $\ddot{q}_j$. Differentiating further wrt t gives "formulas" for $\dddot{q}$ etc. [This calculation can be justified.]

(b) With convexity in $\dot{q}$ (and some growth conds), trajectories over short enough time intervals will

$$\text{minimize} \int_{t_1}^{t_2} L(q,\dot{q})\,dt$$

for given values of $q(t_1)$, $q(t_2)$.
[For example: when a particle is constrained to the sphere as in Lecture 8, its path is a constant-speed geodesic. Sufficiently short geodesics are minimizers of arc length, and minimizers of $\int L(q,\dot{q})$.]

Formal justification: changing vars to $s = t/\varepsilon$,

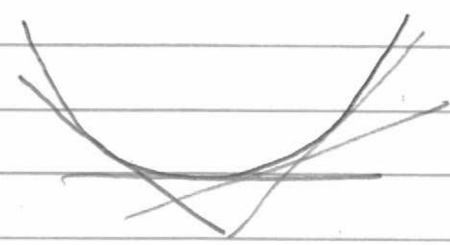
$$\int_0^{\varepsilon} L(q, \tfrac{dq}{dt})\,dt = \varepsilon \int_0^1 L(q, \tfrac{1}{\varepsilon}\tfrac{dq}{ds})\,ds.$$

As $\varepsilon \to 0$, we can expect the (strict) convexity in $\dot{q}$ to become dominant.

OK, so you're hopefully convinced that convexity might be useful. But how will it be used? Answer lies in the <u>Fenchel transform</u> (also known as Legendre transform).

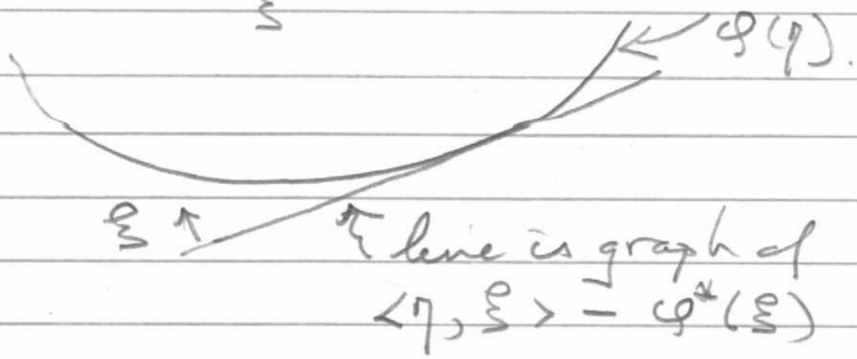
What is that?

- A convex fn is a sup of linear fns. Thus, for each slope we can adjust the intercept so the assoc plane just touches the graph. Envelope of these planes must be the graph of our convex fn.



- Eqn-based version of this: if φ is convex, then

$$\varphi(\eta) = \max_{\xi} \langle \eta, \xi \rangle - \varphi^*(\xi)$$



What is $\varphi^*(\xi)$? Well, by convexity

$$\varphi(\eta) \geq \langle \eta, \xi \rangle - \varphi^*(\xi)$$

with equality for each ξ at some η

so

$$\varphi^*(\xi) \geq \langle \eta, \xi \rangle - \varphi(\eta)$$

with equality for each ξ at some η

so

$$\varphi^*(\xi) = \sup_\eta \langle \eta, \xi \rangle - \varphi(\eta)$$

To be sure all slopes occur, one needs $\varphi(\eta)/|\eta| \to \infty$ as $|\eta| \to \infty$ ("φ has super linear growth").

Relevance of this to Hamiltonian vs Lagrangian mechanics: given $L(q, \dot{q})$, the assoc Hamiltonian is

$$H(q,p) = \max_{\dot{q}} \langle \dot{q}, p \rangle - L(q, \dot{q})$$

= Fenchel transform of L wrt $\dot{q}$ (holding q fixed)

We claim:

① $p_i = \frac{\partial L}{\partial \dot{q}_i}$ determines a well-defined change of coords in phase space: $(q, \dot{q}) \to (q, p)$

② we can recover the Lagrangian from the Hamiltonian as

$$L(q, \dot{q}) = \max_p \langle \dot{q}, p \rangle - H(q, p)$$

③ we can get $\dot{q}$ as fn of p and q by

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

④ a soln of

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i},$$

when viewed in the (q, p) coords, solves

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

If we accept ① - ③, ④ is just calculus: taking differentials,

(*) $$dH = \sum \frac{\partial H}{\partial p_i} dp_i + \frac{\partial H}{\partial q_i} dq_i$$

(where $\frac{\partial H}{\partial p_i}$ is calculated with q_i held fixed, etc.)

Now observe from ① together with

$$H(q,p) = \max_{\xi} \langle \xi, p \rangle - L(q, \xi)$$

with equality when $p_i = \frac{\partial L}{\partial \xi_i}$

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

$$H(q,p) = \langle p, \dot{q} \rangle - L(q, \dot{q})$$

Differentiating this:

(**) $$dH = \sum \dot{q}_i dp_i + \cancel{p_i d\dot{q}_i}$$
$$- \sum \frac{\partial L}{\partial q_i} dq_i + \cancel{\frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i}$$

(where $\frac{\partial L}{\partial q_i}$ is calculated with $\dot{q}_i$ fixed, etc.)

Comparing (*) with (**) we get

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

as functions on phase space.

Finally, let's consider how p and q change along a sol'n of $\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i}$: we get

$$\frac{d}{dt} p_i = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i} = -\frac{\partial H}{\partial q_i}$$

while

$$\frac{d}{dt} q_i = \dot{q}_i = \frac{\partial H}{\partial p_i}$$

as expected.

Let's return now to justify ①–③: they follow from the assertion that if φ is C^1 + convex and

$$\frac{\varphi(\eta)}{|\eta|} \longrightarrow \infty \quad \text{as } |\eta| \longrightarrow \infty$$

[which clearly implies finiteness of $\varphi^*(\xi) = \max_\eta \langle \xi, \eta \rangle - \varphi(\eta)$] Then

(A) φ^* is convex, and $\frac{\varphi^*(\xi)}{|\xi|} \longrightarrow \infty$ as $|\xi| \longrightarrow \infty$

(B) $\varphi^{**} = \varphi$

Proof of (A): φ^* = max of linear fns, so it is certainly convex. Taking $\eta = \lambda\,\xi/|\xi|$ as a test choice gives

$$\varphi^*(\xi) \geq \lambda|\xi| - \varphi(\lambda\,\xi/|\xi|)$$

$$\Rightarrow \frac{\varphi^*(\xi)}{|\xi|} \geq \lambda - \underbrace{\frac{\max \text{ of } \varphi \text{ on } B_\lambda}{|\xi|}}_{\to 0 \text{ as } |\xi| \to \infty}$$

so

$$\liminf_{|\xi|\to\infty} \frac{\varphi^*(\xi)}{|\xi|} \geq \lambda \quad \text{for any } \lambda \in \mathbb{R}.$$

Proof of (B): Clearly $\varphi^*(\xi) + \varphi(\eta) \geq \langle \xi, \eta\rangle$ for all ξ, η. So

$$\varphi(\eta) \geq \langle \xi, \eta \rangle - \varphi^*(\xi).$$

whence

$$\varphi \geq \varphi^{**}$$

For the reverse, observe that

$$\varphi^*(\xi) = \langle \xi, \eta\rangle - \varphi(\eta) \quad \text{when } \nabla_\eta \varphi = \xi.$$

So

$$\varphi(\eta) = \langle \xi, \eta \rangle - \varphi^*(\xi) \quad \text{when } \nabla_\eta \varphi = \xi.$$

Thus

$$\varphi^{**}(\eta) = \max_{\xi} \langle \xi, \eta \rangle - \varphi^{*}(\xi)$$

$$\geq \underbrace{\langle \nabla_\eta \varphi, \eta \rangle - \varphi^{*}(\nabla_\eta \varphi)}_{\varphi(\eta)}$$

so $\varphi^{**} = \varphi$, as claimed.

<u>Rmk:</u> note that if a_{ij} is symmetric + pos def then

$$\varphi(\xi) = \sum \tfrac{1}{2} a_{ij} \xi_i \xi_j = \tfrac{1}{2}\langle A\xi, \xi \rangle$$

$$\Rightarrow \varphi^{*}(\eta) = \tfrac{1}{2}\langle A^{-1}\eta, \eta \rangle$$

The discussion thus far is correct, but I don't like it very much; it doesn't really explain (to my view) <u>why</u> there should be such a connection.

Here is a different viewpoint that's more conceptual, and therefore more attractive. Returning to the Lagrangian var'l pbm, consider its min value as a fn of the <u>final</u> <u>time</u> <u>and</u> <u>position</u> (this is known as "action minimization")

$$u(t_2, x_2) = \min_{q(t_2) = x_2} \int_{t_1}^{t_2} L(q, \dot{q})\,dt$$

[Note that $q(t_1) = x_1$ is arbitrary here.]
The optimizer will of course be a Lagrangian trajectory (i.e. it solves the EL eqn).

By "principle of dynamic programming",

$$u(t, x) \simeq u(t - \Delta t, x - \alpha \Delta x) + L(x, \alpha)\Delta t$$

by taking paths whose last little bit has $\dot{q} = \alpha$. Proceeding formally:

$$u(t, x) \simeq \min_\alpha u(t, x) - \Delta t \cdot \left\{ \begin{matrix} u_t + \alpha \cdot \nabla u \\ - L(x, \alpha) \end{matrix} \right\}$$

Cancel u + divide by Δt to get

$$\begin{aligned} u_t &= \min_\alpha \{ L(x, \alpha) - \alpha \cdot \nabla u \} \\ &= - \max_\alpha \{ \alpha \cdot \nabla u - L(x, \alpha) \}. \\ &= - H(x, \nabla u) \end{aligned}$$

Thus: $u(t,x)$ solves $u_t + H(x, \nabla u) = 0$ for $t > t_1$, with $u = 0$ at $t = t_1$. (Note: t_1 was fixed throughout the preceding discussion.)

<u>More</u>: along the optimal paths we have

$$u_t = L(x,\alpha) - \alpha \cdot \nabla u \quad \text{with } \alpha = \dot{x}(t),$$

so we expect a connection to the <u>method of characteristics</u> (which is, by definition, a way of solving 1st order pde's by reducing them to ODE's). In fact, Hamilton's eqns are the <u>characteristic eqns</u> for $u_t + H(x, \nabla u) = 0$; more specifically, if

$$\frac{dx}{dt} = \nabla_p H \quad \text{and} \quad \frac{dp}{dt} = -\nabla_x H$$

then along the resulting curve

$$(\ast\ast\ast) \qquad \frac{d}{dt} u(x(t), t) = \langle p, \dot{x} \rangle - H(p, x(t))$$

(so that finding u along well-chosen curves is reduced to solving ODE's).

Explaining $(\ast\ast\ast)$: if $u_t + H(x, \nabla u) = 0$ then differentiation gives

$$\frac{\partial^2 u}{\partial x_i \partial t} + \frac{\partial H}{\partial x_i} + \sum_j \frac{\partial H}{\partial p_j}\frac{\partial^2 u}{\partial x_i \partial x_j}) = 0$$

Now,

$$\frac{d}{dt} \nabla_i u(x(t),t) = \frac{\partial^2 u}{\partial x_i \partial t} + \sum \frac{\partial^2 u}{\partial x_i \partial x_j}\frac{dx_j}{dt}$$

along any curve. If we choose $\frac{dx_j}{dt} = \frac{\partial H}{\partial p_j}$ then we get

$$\frac{d}{dt} \nabla_i u(x(t),t) = -\frac{\partial H}{\partial x_i}$$

(so we have $\nabla_i u(x(t),t) = p_i$) and

$$\frac{d}{dt} u(x(t),t) = \langle \nabla u, \dot{x} \rangle + u_t$$

$$= \langle p, \dot{x} \rangle - H$$

as asserted.

In Hamiltonian mechanics, the action

$$\int_{t_1}^{t_2} L(q,\dot{q})\,dt$$

is in general not minimized by trajectories when $t_2 - t_1$ is large (think of geodesics on a sphere, for example). If we insist on minimization (over any time interval) then we're not doing Hamiltonian mechanics but rather optimal control theory. This is a rich and useful subject (a good introduction is provided by L Craig Evans' notes "An introduction to mathematical optimal control theory", available at his website)

Let me try to capture the flavor of this topic through a classic example. Rather than keeping t_1 fixed, I'd rather keep $t_2 = T$ (the final time) fixed, and consider

$$u(x,t) = \min_{y(t)=x} \int_t^T \tfrac{1}{2}|\dot{y}|^2\,dt + g(y(T))$$

cost of travel from initial location x to final location $y(T)$

penalty at final time depends on location

Optimal paths have constant velocity (by Jensen) so in fact

$\boxed{\text{formula}}$. $$u(x,t) = \min_{z \in \mathbb{R}^N} \left\{ \frac{1}{2} \frac{|z-x|^2}{(T-t)} + g(z) \right\}.$$

But arguing as we did earlier gives a Hamilton-Jacobi eqn for u:

$$u(x,t) \simeq \min_{\alpha} u(t+\Delta t, x+\alpha\Delta t) + \tfrac{1}{2}|\alpha|^2 \Delta t$$
$$\simeq u(x,t) + \Delta t \left\{ u_t + \alpha \cdot \nabla u + \tfrac{1}{2}|\alpha|^2 \right\}$$

so (formally)

$$u_t + \min_{\alpha} \left\{ \langle \alpha, \nabla u \rangle + \tfrac{1}{2}|\alpha|^2 \right\} = 0.$$

The best α is $-\nabla u$, giving the "final-value problem"

$$u_t - \tfrac{1}{2}|\nabla u|^2 = 0 \quad \text{for } t < T$$
$$u = g \quad \text{at } t = T$$

But: when the optimal z in our "formula" for u is nonunique (which can easily happen!) $u(x,t)$ is not smooth (nor even differentiable). Therefore it is not so simple to justify our formal calculation.

The theory of <u>viscosity solutions</u> (of 1st order pde) provides a rigorous approach that works despite the nonsmoothness of u. Evans' pde book (chapters 3+10) is a good place to read about that.