Introduction to the Calculus of Variations
Lecture Notes
Lecture 3

Swarnendu Sil

Spring 2021, IISc

Classical Methods

Hamiltonian formuation

We begin with a preliminary study of convex functions and in particular, the notion of convex duality known as the Legendre transform.

Legendre transform

Convex analysis

Definition 1. ***(i)*** *A set* $\Omega \subset \mathbb{R}^n$ *is said to be* convex *if for every* $x, y \in \Omega$ *and every* $\lambda \in [0,1]$, *then* $\lambda x + (1-\lambda) y \in \Omega$.

(ii) *Let* $\Omega \subset \mathbb{R}^n$ *be convex. A function* $f : \Omega \to \mathbb{R}$ *is said to be* convex *if for every* $x, y \in \Omega$ *and every* $\lambda \in [0,1]$, *the following inequality holds*

$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y).$$

(iii) *Let* $\Omega \subset \mathbb{R}^n$ *be convex. A function* $f : \Omega \to \mathbb{R}$ *is said to be* strictly convex *if for every* $x, y \in \Omega$, $x \neq y$, *and every* $\lambda \in (0,1)$, *the following strict inequality holds*

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y).$$

We now give some criteria equivalent to the convexity.

Theorem 2. *Let* $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(\mathbb{R}^n)$ *and denote the scalar product in* $\mathbb{R}^n$ *by* $\langle \cdot, \cdot \rangle$. *The following assertions are then equivalent.*

(i) f *is convex.*

(ii) *For every* $x, y \in \mathbb{R}^n$, *the following inequality holds*

$$f(x) \ge f(y) + \langle \nabla f(y), x - y \rangle.$$

(iii) *For every* $x, y \in \mathbb{R}^n$, *the following inequality is valid*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \ge 0.$$

If, moreover, $f \in C^2(\mathbb{R}^n)$, *then the above statements are equivalent to*

(iv) *for every* $x, v \in \mathbb{R}^n$, *the following inequality holds*

$$\langle \nabla^2 f(x) v, v \rangle \ge 0.$$

Example 3. *Let $n = 1$, $1 \leq p < \infty$, $f(x) = |x|^p$ and*

$$v_f^*(y) = \begin{cases} p\,|y|^{p-2}\,y & \text{if } 1 < p < \infty \\ +1 & \text{if } p = 1 \text{ and } y > 0 \\ 0 & \text{if } p = 1 \text{ and } y = 0 \\ -1 & \text{if } p = 1 \text{ and } y < 0. \end{cases}$$

It follows, trivially if $p = 1$ and from (ii) of the theorem otherwise, that, for every $x, y \in \mathbb{R}$,

$$|x|^p \geq |y|^p + v_f^*(y)(x - y).$$

Note that, when $p = 1$, we could have chosen $v_f^*(0)$ arbitrarily in $[-1, 1]$. Moreover, the quantity $v_f^*(y)$ is called, in convex analysis, the subgradient *of f at y. Draw a sketch of the graph of the function and draw the* set of subgradients. *See the exercises for more on subgradients of a convex function.*

The following is an important inequality for convex functions.

Theorem 4 (Jensen inequality)**.** *Let $\Omega \subset \mathbb{R}^n$ be open and bounded, $u = (u^1, \cdots, u^N) \in L^1(\Omega; \mathbb{R}^N)$ and $f : \mathbb{R}^N \to \mathbb{R}$ be convex, then*

$$f(u_\Omega) \leq \frac{1}{|\Omega|} \int_\Omega f(u(x))\ \mathrm{d}x$$

where

$$u_\Omega = (u_\Omega^1, \cdots, u_\Omega^N) \quad \text{with} \quad u_\Omega^i = \frac{1}{|\Omega|} \int_\Omega u^i(x)\ \mathrm{d}x.$$

Legendre transform

We now need to introduce the notion of duality, also known as Legendre transform, for convex functions. It is convenient to accept, in the definition, functions that are allowed to take the value $+\infty$ (a function that takes only finite values is called finite).

Definition 5 (Legendre transform)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ (or $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$).*

(i) *The* Legendre transform*, or* dual*, of f is the function $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ defined by*

$$f^*(x^*) = \sup_{x \in \mathbb{R}^n} \{\langle x; x^* \rangle - f(x)\}$$

where $\langle .; . \rangle$ denotes the scalar product in $\mathbb{R}^n$.

(ii) *The* bidual *of f is the function $f^{**} : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ defined by*

$$f^{**}(x) = \sup_{x^* \in \mathbb{R}^n} \{\langle x; x^* \rangle - f^*(x^*)\}.$$

Remark 6. ***(i)*** *In general, f^* takes the value $+\infty$, even if f takes only finite values. See the problem sheet for an example.*

(ii) *If $f \not\equiv +\infty$, then $f^* > -\infty$. (Prove this!)*

Example 7. *Let us try to calculate the Legendre transform of*

$$f(x) = \frac{1}{p}|x|^p \qquad \text{for all } x \in \mathbb{R}^n \qquad \text{and } 1 < p < \infty.$$

By definition, we have

$$f^*(x^*) = \sup_{x \in \mathbb{R}^n} \{\langle x^*, x\rangle - f(x)\}. \tag{1}$$

If the supremum is achieved at a point $y \in \mathbb{R}^n$ (i.e. a maxima) for the following function

$$g(x) := \langle x^*, x\rangle - f(x),$$

then we must have

$$0 = \nabla g(y) = x^* - \nabla f(y) = x^* - |y|^{p-2} y.$$

But

$$x^* = |y|^{p-2} y \Rightarrow |x^*|^{\frac{1}{p-1}} = |y| \text{ and } y = \frac{x^*}{|y|^{p-2}}.$$

Combining the last two identities, we deduce

$$y = |x^*|^{\frac{2-p}{p-1}} x^* = |x^*|^{\frac{p-2(p-1)}{p-1}} x^* = |x^*|^{\left(\frac{p}{p-1}-2\right)} x^* = |x^*|^{p'-2} x^*,$$

where $p' = \frac{p}{p-1}$ is the Hölder conjugate of p, i.e

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

So plugging it in (1), we deduce

$$f^*(x^*) = \langle x^*, y\rangle - f(y) = \langle x^*, y\rangle - \frac{1}{p}|y|^p = \frac{1}{p'}|x^*|^{p'}.$$

We now gather some properties of the Legendre transform (for a proof see the Problem sheet).

Theorem 8. *Let $f : \mathbb{R}^n \to \mathbb{R}$ (or $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$).*

(i) *The function f^* is convex (even if f is not).*

(ii) *The function f^{**} is convex and $f^{**} \leq f$. If, furthermore, f is convex, bounded below and finite then $f^{**} = f$. More generally, if f is bounded below and finite but not necessarily convex, then f^{**} is its* convex envelope *(which means that it is the largest convex function that is smaller than f).*

(iii) *The following identity always holds: $f^{***} = f^*$.*

(iv) *If $f \in C^1(\mathbb{R}^n)$, convex and finite, then*

$$f(x) + f^*(\nabla f(x)) = \langle \nabla f(x); x\rangle, \quad \forall x \in \mathbb{R}^n.$$

(v) *If $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex and if*

$$\lim_{|x|\to\infty} \frac{f(x)}{|x|} = +\infty$$

then $f^* \in C^1(\mathbb{R}^n)$. Moreover, if $f \in C^1(\mathbb{R}^n)$ and

$$f(x) + f^*(x^*) = \langle x^*; x\rangle$$

then

$$x^* = \nabla f(x) \quad \text{and} \quad x = \nabla f^*(x^*).$$

Hamiltonian formulation

So far we have investigated the stationary points of the following functional

$$u \mapsto I(u) := \int_a^b f(t, u(t), \dot{u}(t))\ \mathrm{d}t, \qquad \text{for } u \in X.$$

We showed that C^2 stationary points satisfy the Euler-Lagrange equations. Now we are going to show that in some cases, these C^2 stationary points are also the stationary points of another functional whose EL equations are going to be systems of $2N$ first order ODEs instead of the system of N second order ODEs we obtained.

The functional is, for $u, v \in C^2\left([a,b]; \mathbb{R}^N\right)$,

$$(u,v) \mapsto J(u,v) := \int_a^b \left[\langle \dot{u}(t), v(t)\rangle - H(t, u(t), v(t))\right]\ \mathrm{d}t.$$

where the function $H : [a,b] \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is called the **Hamiltonian** and it is the **Legendre transform** (strictly speaking, a partial Legendre transform) of the **Lagrangian** f, i.e.

$$H(x,u,v) = \sup_{\xi\in\mathbb{R}^N} \left\{\langle v, \xi\rangle - f(x,u,\xi)\right\}.$$

Now since the Hamiltonian is the Legendre transform of the Lagrangian, we should be able to infer some regularity of the Hamiltonian if the Lagrangian is nice enough. This is the spirit of the following lemma.

Lemma 9 (Regularity of the Hamiltonian)**.** *Let $f \in C^2\left([a,b] \times \mathbb{R}^N \times \mathbb{R}^N\right)$, $f = f(t,u,\xi)$ be such that*

(convexity) $f_{\xi\xi}(t,u,\xi)$ ***positive definite*** *, for every $(t,u,\xi) \in [a,b]\times\mathbb{R}^N\times\mathbb{R}^N$,*

(coercivity) $f(t,u,\xi) \geq \omega(|\xi|) + g(t,u)$, *for every $(t,u,\xi) \in [a,b]\times\mathbb{R}^N\times\mathbb{R}^N$*

where ω is nonnegative, continuous and increasing with $\lim_{t\to\infty} \omega(t)/t = \infty$ and $g : [a,b] \times \mathbb{R}^N \to \mathbb{R}$ is continuous.
Then the Hamiltonian $H \in C^2\left([a,b] \times \mathbb{R}^N \times \mathbb{R}^N\right)$ and we have

$$H_t(t,u,v) = -f_t(t,u,H_v(t,u,v))$$
$$H_u(t,u,v) = -f_u(t,u,H_v(t,u,v)),$$

$$H(t,u,v) = \langle v, H_v(t,u,v)\rangle - f(t,u,H_v(t,u,v))$$
$$\text{and} \qquad v = f_\xi(t,u,\xi) \quad \text{if and only if} \quad \xi = H_v(t,u,v).$$

Proof. Note that the coercivity assumptions imply

$$\lim_{|\xi|\to\infty} \frac{f(t,u,\xi)}{|\xi|} = +\infty \qquad \text{for every } (t,u) \in [a,b] \times \mathbb{R}^N.$$

Thus given any $t \in [a,b]$ and $u, v \in \mathbb{R}^N$, the the supremum in the definition of H, i.e.

$$\sup_{\xi\in\mathbb{R}^N} \{\langle v,\xi\rangle - f(t,u,\xi)\}$$

is achieved at some $\xi = \xi(t,u,v) \in \mathbb{R}^N$. Thus the function $g : \mathbb{R}^N \to \mathbb{R}$ defined by

$$g(y) := \langle v, y\rangle - f(t,u,y)$$

achieves a maxima at $y = \xi$. Hence we must have $\nabla g(y)|_{y=\xi} = 0$. i.e.

$$v = f_\xi(t,u,\xi).$$

So far we have established that $H(t,u,v)$ is finite everywhere and

$$v = f_\xi(t,u,\xi).$$

One can actually establish the continuity of H as well already. Let ξ be the maximizer

$$H(t,u,v) = \langle v,\xi\rangle - f(t,u,\xi)$$

By definition of H, for some other point $(\bar{t},\bar{u},\bar{v})$, we have

$$H(\bar{t},\bar{u},\bar{v}) \geq \langle \bar{v},\xi\rangle - f(\bar{t},\bar{u},\xi).$$

Thus, we obtain

$$H(t,u,v) - H(\bar{t},\bar{u},\bar{v}) \leq \langle v - \bar{v},\xi\rangle + [f(\bar{t},\bar{u},\xi) - f(t,u,\xi)].$$

Now continuity of H follows from the continuity of $(t,u) \mapsto f(t,u,\xi)$ for every ξ. Now we want to invert the equation

$$v = f_\xi(t,u,\xi)$$

and express ξ as a function of t, u, v. But since $f \in C^2$ and $f_{\xi\xi}$ is positive definite and hence invertible, inverse function theorem implies that $\xi = \xi(t, u, v)$ is C^1. Now the equation

$$H(t, u, v) = \langle v, \xi(t, u, v)\rangle - f(t, u, \xi(t, u, v))$$

immediately implies H is C^1. Furthermore, we deduce

$$\begin{aligned} H_t &= \langle v - f_\xi, \xi_t\rangle - f_t = -f_t, \\ H_u &= \langle v - f_\xi, \xi_u\rangle - f_u = -f_u, \\ H_v &= \xi + \langle v - f_\xi, \xi_v\rangle = \xi. \end{aligned}$$

The last equation also proves $\xi = H_v$ if and only if $v = f_\xi$. But since f is C^2 and ξ is C^1, we deduce that the maps

$$(t, u, v) \mapsto f_t(t, u, \xi(t, u, v)) \text{ and } (t, u, v) \mapsto f_u(t, u, \xi(t, u, v))$$

are both C^1 as well. Thus the equations (which we deduced on last slide)

$$\nabla H = \begin{pmatrix} H_t \\ H_u \\ H_v \end{pmatrix} = \begin{pmatrix} -f_t \\ -f_u \\ \xi \end{pmatrix}$$

implies that H is C^2. □

Hamilton's equations

The Euler-Lagrange equations for the functional

$$J(u, v) := \int_a^b [\langle \dot{u}(t), v(t)\rangle - H(t, u(t), v(t))] \, dt, \qquad \text{for } u, v \in C^2([a, b]; \mathbb{R}^N)$$

are called the **Hamilton's equations** and sometimes also called the **canonical form** of the Euler-Lagrange equation of the Lagrangian formulation. The equations are the following $2N$ first order ODEs.

$$\begin{cases} \dot{u}(t) = H_v(t, u(t), v(t)), \\ \dot{v}(t) = -H_u(t, u(t), v(t)). \end{cases} \tag{$\mathbf{H}$}$$

Indeed, letting $w := (w_1, w_2) = (u, v) \in \mathbb{R}^N \times \mathbb{R}^N = \mathbb{R}^{2N}$ and letting $\zeta := (\zeta_1, \zeta_2) = (\dot{u}, \dot{v})$ stand for the derivatives variable, we can write the Lagrangian density as a function $\mathcal{H} : [a, b] \times \mathbb{R}^{2N} \times \mathbb{R}^{2N}$ given by

$$\mathcal{H}(t, w, \zeta) := \langle \zeta_1, w_2\rangle - H(t, w_1, w_2).$$

Thus we obtain

$$\mathcal{H}_\zeta = \begin{pmatrix} \mathcal{H}_{\zeta_1} \\ \mathcal{H}_{\zeta_2} \end{pmatrix} = \begin{pmatrix} w_2 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \mathcal{H}_w = \begin{pmatrix} \mathcal{H}_{w_1} \\ \mathcal{H}_{w_2} \end{pmatrix} = \begin{pmatrix} -H_{w_1} \\ \zeta_1 - H_{w_2} \end{pmatrix}$$

Recall that the Euler-Lagrange equation is

$$\frac{d}{dt}\mathcal{H}_\zeta = \mathcal{H}_w.$$

So we deduce

$$\frac{d}{dt}w_2 = -H_{w_1} \qquad \text{and} \qquad 0 = \zeta_1 - H_{w_2}.$$

This yields

$$\dot{v} = -H_u \qquad \text{and} \qquad 0 = \dot{u} - H_v,$$

respectively, which is nothing but (**H**).

Equivalence of Lagrangian and Hamiltonian formulation

Now we are going to show that under suitable assumptions, the Hamilton's equations are equivalent to the Euler-Lagrange equations for the Lagrangian formulations for C^2 critical points.

Theorem 10 (Hamiltonian and Lagrangian formulation)**.** *Let f satisfy the hypotheses of lemma 9 and let H be its Hamiltonian. Let $u, v \in C^2\left([a,b];\mathbb{R}^N\right)$ satisfy,*

$$(\boldsymbol{H}) \qquad \begin{cases} \dot{u}(t) = H_v(t, u(t), v(t)), \\ \dot{v}(t) = -H_u(t, u(t), v(t)), \end{cases} \qquad \textit{for every } t \in [a,b].$$

Then u verifies

$$(\boldsymbol{EL}) \qquad \frac{d}{dt}\left[f_\xi(t, u(t), \dot{u}(t))\right] = f_u(t, u(t), \dot{u}(t)), \qquad \textit{for every } t \in (a,b).$$

*Conversely, if $u \in C^2\left([a,b];\mathbb{R}^N\right)$ satisfies (**EL**) then (u,v) are C^2 solutions of (**H**) where*

$$v(t) = f_\xi(t, u(t), \dot{u}(t)), \qquad \textit{for every } t \in [a,b].$$

Proof. Now that we have done all the hard work in proving lemma 9, the proof is easy. By lemma 9,

$$\dot{u} = H_v \quad \text{implies} \quad v(t) = f_\xi(t, u(t), \dot{u}(t)).$$

But then,

$$\frac{d}{dt}f_\xi = \dot{v} = -H_u = -\langle v - f_\xi, \xi_u\rangle + f_u = f_u,$$

which is the (**EL**) equations. Conversely, if $u \in C^2\left([a,b];\mathbb{R}^N\right)$ satisfies (**EL**) then

$$v(t) = f_\xi(t, u(t), \dot{u}(t)) \quad \text{implies} \quad \dot{u} = H_v.$$

Also,

$$\dot{v} = \frac{d}{dt}f_\xi = f_u = \langle v - f_\xi, \xi_u\rangle - H_u = -H_u,$$

verifying (**H**). □

First Integrals

We begin with a few definitions.

Definition 11 (Integral Curves)**.** *An* integral curve *of the vector field is a curve which is tangent to the vector field at each point. Mathematically, given a vector field X on $\mathbb{R}^N$, the map $\phi : [a,b] \to \mathbb{R}^N$ is an* integral curve *of the vector field X if it satisfies*

$$\dot{\phi}(t) = X(\phi(t)) \qquad \text{for each } t \in [a,b]. \tag{2}$$

Clearly, (2) *is a system of ODEs. ϕ is also called an* integral curve *for the system of ODEs in this case as well.*

Definition 12 (First Integral)**.** *A* ***first integral*** *of a system of differential equations is a function which has a constant value along each integral curve of the system.*

Remark 13. *Note that it is perfectly allowed for a first integral to have two different constant value along two different integral curves and thus in particular, can be far from a constant function. Of course, all constant functions would automatically be a first integral of* any *system of differential equations. But these are of little use. So in spite of the definition, by first integral we almost always mean nontrivial (i.e. not globally constant) first integrals.*

Remark 14. *First integrals are usually called* ***conserved quantities*** *in physics literature, for obvious reasons. Finding first integrals often make solving the Hamilton's equations (or, at least formally equivalently, the Euler-Lagrange equations) simpler. So this is an extremely useful technique. See the assignments for seeing this technique being used to solve the Brachistochrone and the minimal surface of revolution problem.*

Now our task is to find a simple, algebraic, necessary and sufficient criterion for deciding when a given function is a first integral. The criterion is best expressed in the Hamiltonian formulation. In fact, one of the principal theoretical advantage of the Hamiltonian formulation is this criterion, which are not so crisp in terms of the Lagrangian formulation.

Theorem 15 (First Integral)**.** *A function $\Phi \in C^2\left(\mathbb{R}^N \times \mathbb{R}^N\right)$, $\Phi = \Phi(u,v)$, is a first integral of the Hamilton's equations with Hamiltonian $H = H(t,u,v) \in C^2\left([a,b] \times \mathbb{R}^N \times \mathbb{R}^N\right)$ if and only if the* ***Poisson Bracket***

$$\{\Phi, H\} := \langle \Phi_u, H_v \rangle - \langle \Phi_v, H_u \rangle := \sum_{i=1}^{N} \frac{\partial \Phi}{\partial u^i}\frac{\partial H}{\partial v^i} - \frac{\partial \Phi}{\partial v^i}\frac{\partial H}{\partial u^i}$$

vanishes identically.

Proof. Along each integral curve $(u(t), v(t))$ of the Hamilton's equations, we have

$$\dot{\Phi}(t) = \frac{d}{dt}\Phi = \langle \Phi_u, \dot{u} \rangle + \langle \Phi_v, \dot{v} \rangle = \{\Phi, H\}.$$

$\square$

Remark 16. *The Poisson bracket is an example of a* ***commutator bracket****. So it satisfies all the usual properties of commutator brackets like bilinearity, Leibnitz rule*[1] *and the Jacobi's identity*[2]*. Clearly, this is anticommutative. It is intimately related to another extremely useful bracket operation in mathematics, the* ***Lie Bracket*** *via the construction of* Hamiltonian vector fields *in symplectic geometry. In fact, the precise relationship is*

$$X_{\{f,g\}} = -[X_f, X_g],$$

where the notation $X_\cdot$ denotes the Hamiltonian vector field associated to a function and $[\cdot,\cdot]$ denotes the Lie Bracket of vector fields.

Note that the last theorem was about finding first integrals which does not depend explicitly on t. See Assignments for a general result which involves first integrals $\Phi = \Phi(t, u, v)$ with explicit t-dependence.

If the Hamiltonian H does not depend explicitly on t, the previous theorem tells us that H itself is a first integral, since

$$\{H, H\} \equiv 0.$$

In physics, this is usually stated as the fact that the Hamiltonian (i.e. the total energy) of a mechanical system is a conserved quantity.

This however, is not a coincidence! This is just a special instance of a profound general fact known as Noether's theorem, to which we shall trun our attention to in the next section.

[1]The Leibnitz rule means products distribute as derivatives do. More precisely,

$$\{fg, h\} = \{f, h\}\, g + f\, \{g, h\}.$$

[2]The Jacobi's identity is about the associative law and says that the sum of all cyclic permutation of three arguments vanishes. More precisely,

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0.$$