HAMILTON JACOBI THEORY IN CALCULUS OF VARIATION

LONG CHEN

ABSTRACT.

CONTENTS

1. CONJUGATE OF CONVEX FUNCTIONS

In this section, we provide an 'outside' characterization of convex sets and convex functions. It is a coincise version of the section in Convex Analysis.

1.1. **Convex Sets.** A subset K of a vector space V is convex if

$$x, y \in K, \alpha \in [0,1] \Longrightarrow (1-\alpha)x + \alpha y \in K.$$

For a linear subspace u, it requires $\alpha x + \beta y \in S$ for any $\alpha, \beta \in \mathbb{R}$. Convexity, on the other hand, restricts the coefficients such that $\alpha + \beta = 1$ and $\alpha, \beta \geq 0$. Geometrically, convexity implies that the line segment connecting x and y is contained within the set, while a linear subspace requires that the plane spanned by vectors x and y is contained in the subspace.

Date: June 6, 2023.

In particular, 0 must be in a subspace, while a convex set may not necessarily contain the origin.

Convexity is translation invariant, meaning that if K is convex, then $K + z := \{x + z : x \in K\}$ is also convex for any $z \in V$, and $K + z$ is isomorphic to K. The intrinsic properties of a convex set are thus translation invariant. Depending on the context, the assumption that $0 \in K$ or $0 \notin K$ can be easily satisfied through translation.

1.2. **Dual Representation.** Consider an example where we want to cut a polygon out of a sheet of paper. One way is to use scissors to carefully trim along the boundary. However, what if we don't have scissors and only have a knife at our disposal? In this case, we can use the knife to make a series of straight cuts, dividing the paper into two sections with each cut, and keeping the portion containing the convex polygon. This method demonstrates an 'outside' approach to handling convex shapes by focusing on their properties related to straight lines and planes.

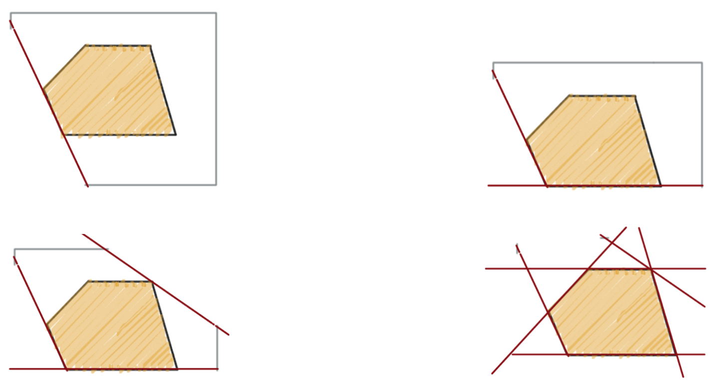

FIGURE 1. Dual representation of a convex set

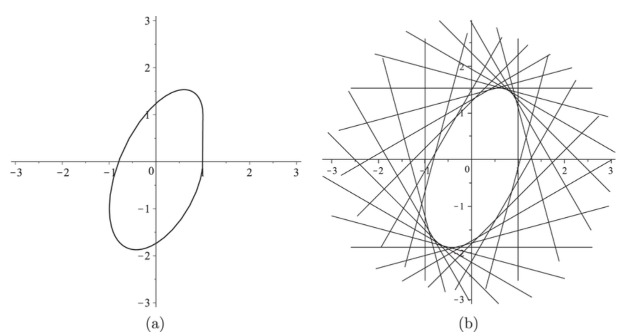

FIGURE 2. Supporting hyperplane of a convex set.

Now we present a mathematical description. Let V be a Hilbert space endowed with an inner product $(\cdot, \cdot)$. Given a vector $n \in V$ and a number $b \in \mathbb{R}$, we can define the closed half-planes

$$H^+_{n,b} := \{u \in V \mid (n, u) \geq b\}, \quad H^-_{n,b} := \{u \in V \mid (n, u) \leq b\},$$

and the hyperplane

$$H_{n,b} := H^+_{n,b} \cap H^-_{n,b} = \{u \in V \mid (n, u) = b\}.$$

The inner product (n, u) represents the signed distance of u along the normal direction n. This formalization allows us to describe geometric objects, such as half-planes and hyperplanes, and their relationships with convex sets in a Hilbert space.

Theorem 1.1 (Dual representation of a convex set)**.** *A non-empty and closed convex set u is the intersection of all closed half space H^+ containing it, i.e.,*

$$S = \cap_{H^+, S \subseteq H^+} H^+.$$

If $S \subseteq H^+_{n,b_2}$, then $S \subseteq H^+_{n,b_1}$ provided $b_1 < b_2$. So we can define

$$b^* = b^*(n) = \sup_{S \subseteq H^+_{n,b}} b.$$

Given the normal vector n, consider the signed distance from u to the normal direction, which is $\inf_{u \in S}(n, u)$. Then we claim

$$\inf_{u \in S}(n, u) = \sup_{S \subseteq H^+_{n,b}} b, \tag{1}$$

which can be proved easily by definition. Indeed, as u is convex and closed, we have $u^* \in S$ to achieve the infimum, and $b^* = (n, u^*)$ as illustrated by the following figure. This example illustrates the duality of optimization problems, which involves switching the roles of variables u and b, as well as the optimization problems $\inf$ and $\sup$.

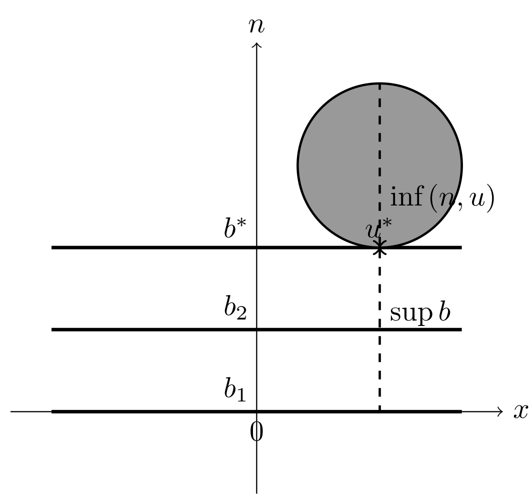

FIGURE 3. Duality.

The hyperplane H_{n,b^*} is known as the supporting plane of the set u at point u^*. When the boundary of u is smooth, the supporting plane coincides with the tangent plane of u at u^*. We can refine the characterization of a convex set from an external perspective as follows:

$$S = \bigcap_{n \in V} H^+_{n,b^*}. \tag{2}$$

The function $b^*(n)$, as a function of n, determines the convex set. A helpful analogy to consider is that n represents a test, while $b^*(n)$ corresponds to the test score. Evaluating a convex set can be thought of as assessing its properties through a comprehensive, 360-degree examination.

1.3. **Epigraph.** We associate a convex set with a convex function and use the geometric characterization of convex sets to study convex functions.

Given a function $f : V \to \mathbb{R} \cup \{+\infty\}$, its *effective domain* is

$$\mathrm{dom}(f) = \{x \in V : f(x) < +\infty\}.$$

The *graph* of f is a surface

$$G(f) = \{(x, f(x)), x \in \mathrm{dom}(f)\} \subset V \times \mathbb{R}.$$

The *epigraph* (or supergraph) of f is the set

$$\mathrm{epi}(f) = \{(x, t) \in V \times \mathbb{R}, x \in \mathrm{dom}(f), f(x) \le t\}. \tag{3}$$

Given a point $(x, t) \in \mathrm{epi}(f) \subset V \times \mathbb{R}$, there is a vertical line passing through that point, and $f(x)$ will be the lowest value for the part in $\mathrm{epi}(f)$. That is,

$$f(x) = \inf_{(x,t)\in \mathrm{epi}(f)} t. \tag{4}$$

A continuous function $f : X \to \mathbb{R} \cup \{+\infty\}$ is *convex* if

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y) \quad \forall\, x, y \in X, \forall\, \alpha \in [0, 1]. \tag{5}$$

Theorem 1.2. *f and $\mathrm{dom}(f)$ are convex if and only if $\mathrm{epi}(f)$ is convex.*

1.4. **Conjugate Function.** For a convex function f, we will use the dual representation of the convex set $\mathrm{epi}(f)$ to characterize f and introduce the conjugate function of f.

Consider an affine function $L(x; a, b) := \langle a, x\rangle - b$. Its graph defines a hyperplane $H_{a,b} = \{(x, t), b = \langle a, x\rangle - t\}$ in $V \times \mathbb{R}$. Here, the notation is slightly abused as $n = (-a, 1)$ in $V \times \mathbb{R}$. Then $\mathrm{epi}(f) \subset H^+_{a,b}$ is equivalent to

$$t \ge \langle a, x\rangle - b, \quad \forall (x, t) \in \mathrm{epi}(f). \tag{6}$$

This is the basic inequality connecting the four variables $\{x, t, a, b\}$.

Now, we will eliminate two of these four variables by optimizing two of them. First, fix x; among many t satisfying (6), choosing the smallest one

$$\inf_{t\in \mathrm{epi}(f)} t, \quad t \ge \langle a, x\rangle - b,$$

which is $f(x) = \inf_{(x,t)\in \mathrm{epi}(f)} t$. Rewrite the inequality as $b \ge \langle a, x\rangle - f(x)$. Next, for a fixed a, we choose the closest b by selecting

$$\inf_{b\in\mathbb{R}} b, \quad b \ge \langle a, x\rangle - f(x).$$

By the duality (1), this motivates the definition of the conjugate of f, also called the dual of f or Legendre transformation, $f^* : V^* \to \mathbb{R} \cup \{+\infty\}$:

$$f^*(a) := \sup_{x\in \mathrm{dom}(f)} \left[\langle a, x\rangle - f(x)\right]. \tag{7}$$

By definition, we have the Fenchel inequality.

Lemma 1.3 (Fenchel inequality)**.** *For a convex function, define the conjugate $f^*(a)$ by (7). Then for any $x \in \mathrm{dom}(f)$ and any $a \in V^*$*

$$f(x) + f^*(a) \ge \langle a, x\rangle. \tag{8}$$

1.5. **Smooth and Strongly Convex Functions.** Note that $f^*(a) = \infty$ is possible for convex but not strongly convex functions. For example, $f(x) = e^{-x}$, then, for any positive p, $f^*(p) = +\infty$. If restricted to smooth convex function class $\mathcal{S}^{1,1}_{\mu,L}$, where is the class of μ-strongly convex function and ∇f is Lipschitz continuous with constant L, there exists a unique maximizer for the $\sup$ problem in the definition (7). More specifically for a given $p \in V^*$, $f^*(p) < \infty$ and there exists a unique $x(p)$ s.t.

$$\nabla f(x(p)) = p, \text{ and} \tag{9}$$

$$f^*(p) = \langle p, x(p)\rangle - f(x(p)). \tag{10}$$

In view of (9), the function $x(p)$ from $V^* \to V$ can be thought of as the inverse of $\nabla f : V \to V^*$. It is indeed the gradient of the dual of f, i.e., $x(p) = \nabla f^*(p)$.

In view of (10), equality in (8) is obtained when $x = x(p)$.

Lemma 1.4. *For $f \in \mathcal{S}^{1,1}_{\mu,L}$, let $x(p) := \arg\sup\left[\langle a, x\rangle - f(x)\right]$, we have*

(1) $f^* \in \mathcal{C}^1$ and $\nabla f^*(p) = x(p), \nabla f(x(p)) = p,$.

(2) f^* is $1/L$-strongly convex, i.e., $f^* \in \mathcal{S}^1_{1/L}$.

(3) ∇f^* is $1/\mu$-continuous. Consequently $f^* \in \mathcal{S}^{1,1}_{1/L,1/\mu}$.

As an illustration example, consider $f(x) = \frac{1}{2}(Ax, x)$ where A is an SPD operator with $\mu = \lambda_{\min}(A)$ and $L = \lambda_{\max}(A)$. Then $f^*(p) = \frac{1}{2}(A^{-1}x, x)$.

The above result shows that the dual of a strongly convex and Lipschitz smooth function f remains in this class with switched and reversed parameters. Therefore the dual operator can be applied to f^* and the dual operator is reflective in the sense that.

$$f(x) = f^{**}(x) := \sup_{a \in V^*}\left[\langle a, x\rangle - f^*(a)\right]. \tag{11}$$

2. Hamiltonian Systems

In this section we shall apply the Legendre transformation to the Lagrangian to define Hamiltonian. The dual formulation of the Euler-Lagrange equation is the Hamitonian system.

2.1. **Hamiltonian system.** Recall that $I(x) = \int_0^1 L(t, x(t), \dot{x}(t))\,\mathrm{d}t$, which can be considered as the restriction of the functional

$$I(x, v) = \int_0^1 L(t, x(t), v(t))\,\mathrm{d}t \tag{12}$$

to the curve $(x(t), \dot{x}(t))$ in the space (x, v). Assume $L(t, x, \cdot)$ is convex in v. We apply Legendre transformation to L and define Hamiltonian

$$H(t, q, p) := L^*(t, q, \cdot) = \sup_v \left\{(p, v) - L(t, q, v)\right\}. \tag{13}$$

The $\sup$ is achieved at $v(p, q, t)$ s.t.

$$p = L_v(t, q, v). \tag{14}$$

That is we can solve (14) to get v as a function of (p, q, t) as L is convex in v. Then

$$H(t, q, p) = (p, v) - L(t, q, v), \text{ when } p = L_v(t, q, v). \tag{15}$$

By the reflexive of the conjugate operator, we can recover Lagrangian from Hamiltonian by

$$L(t, x, v) = H^*(t, q, \cdot) = \sup_p \left\{(p, v) - H(t, q, p)\right\}. \tag{16}$$

We can write

$$L(t,x,v) = (p,v) - H(t,q,p), \text{ when } v = H_p(t,q,p). \tag{17}$$

When both (p,q) are functions of t, i.e., $(p(t),q(t))$, we define

$$F(p,q) = \int_a^b (p(t),\dot{q}(t)) - H(t,q(t),p(t))\,\mathrm{d}t. \tag{18}$$

Then by (16), we obtain the equivalence of the problem

$$\min_x I(x,\dot{x}) = \inf_q \sup_p F(p,q).$$

By computing the first variation of the saddle point system $\inf_q \sup_p F(p,q)$, we obtain the Hamiltonian system

Hamilton system

$$\begin{cases} \dot{p} = -H_q(t,q,p), \\ \dot{q} = H_p(t,q,p). \end{cases} \tag{19}$$

The Hamilton system can be written as an anti-symmetric gradient flow

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -J\nabla H(y). \tag{20}$$

where $y = (p,q)^{\intercal}$ and $J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ is an skew-symmetric matrix.

The matrix J is skew-symmetric and unitary satisfying

$$J^{\intercal} = -J = J^{-1}, \quad -J^{\intercal}J = J^2 = -I.$$

Geometrically J is a rotation of 90° clockwise.

For Euler-Lagrange equation, two boundary conditions are provided while for Hamiltonian system (19), two initial condition is given.

As comparison, another major flow considered in the physics is the so-called gradient flow:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\nabla E(y), \tag{21}$$

where E is called energy. The two flow (20) and (21) will define different dynamic system.

2.2. Equivalence to Euler-Lagrange Equation.

We write

$$H(t,p,q) = (p,v) - L(t,q,v)$$

and apply total differentiation to both sides. LHS is

$$\mathrm{d}H = H_p\,\mathrm{d}p + H_q\,\mathrm{d}q + H_t\,\mathrm{d}t$$

and RHS is

$$p\,\mathrm{d}v + v\,\mathrm{d}p - L_t\,\mathrm{d}t - L_q\,\mathrm{d}q - L_v\,\mathrm{d}v = v\,\mathrm{d}p - L_q\,\mathrm{d}q - L_t\,\mathrm{d}t + (p - L_v)\,\mathrm{d}v.$$

Compare the coefficients, we obtain the relation:

$$v = H_p \tag{22}$$

$$H_q = -L_q \tag{23}$$

$$H_t = -L_t \tag{24}$$

$$p = L_v. \tag{25}$$

We take the total differentiation by treating all variables (p, q, v, t) are independent. The relation (25) indeed implies these variable should be related.

Now we further impose conditions on these four variables (p, q, v, t). Assume $q(t)$ solves the Euler-Lagrange equation $\dot{L}_v(t, q, \dot{q}) = L_q(t, q, \dot{q})$. Then (22)-(23) becomes the Hamiltonian system (19) by substituting $v = \dot{q}$ and $L_q = \dot{L}_v = \dot{p}$. That is we verified the solution of the Euler-Lagrange equation will give the solution to the Hamiltonian system by the change of variable $q = x$ and $p = L_v$. On the other hand, suppose $(p(t), q(t))$ solves the Hamiltonian system (19). Then $\dot{L}_v(t, q, \dot{q}) = \dot{p} = -H_q = L_q$, i.e., q solves the Euler-Lagrange equation.

On the solution curve $(p(t), q(t))$ of the Hamiltonian system or $x(t)$ of the E-L equation, as the relation $p = L_v(t, q, \dot{q})$ or $\dot{q} = H_p(t, q, p)$ holds, we can safely switch between Hamiltonian and Lagrangian

$$H(t, q(t), p(t)) = (p, \dot{q}) - L(t, q, \dot{q}) \tag{26}$$

$$L(t, q(t), \dot{q}(t)) = (p, \dot{q}) - H(t, q(t), p(t)). \tag{27}$$

2.3. **Conservation of Hamiltonian.** When Hamiltonian $H(p, q)$ is independent of t, for solutions to the Hamiltonian system, we obtain the conservation of Hamiltonian, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t} H(p(t), q(t)) = H_p \dot{p} + H_q \dot{q} = 0.$$

Using the matrix form (20), the calculation is

$$\frac{\mathrm{d}}{\mathrm{d}t} H(y(t)) = (\nabla H(y), \dot{y}) = (\nabla H(y), -J \nabla H(y)) = 0.$$

An important example is $L(x, v) = T(v) - U(x)$. For $T(v) = \frac{1}{2} m|v|^2$, the variable $p = mv$ is the momentum. Then

$$H = p\dot{q}(p) - L = pv - T(v) - U(q) = \frac{1}{2} m|v|^2 + U(q) = T + U$$

is the total energy.

As a comparison, for the gradient flow (21), we have the energy dispersion

$$\frac{\mathrm{d}}{\mathrm{d}t} E(y(t)) = (\nabla E(y), \dot{y}) = -\|\nabla E(y)\|^2 \leq 0.$$

That is the energy $E(y(t))$ is non-increasing along the trajectory of the gradient flow.

2.4. **Liuville's theorem.** This is part of content in [1, Chapter 3, Section 16]. Define the phase flow $g^t : (p(0), q(0)) \to (p(t), q(t))$ where (p, q) is the solution of Hamiltonian system with initial condition $(p(0), q(0))$. It forms an one-parameter group of transformations.

Theorem 2.1 (Liuville's Theorem). *The phase flow preserves volume: for any region D we have* $\operatorname{vol}(g^t(D)) = \operatorname{vol}(D)$.

A general proposition: $g^t(x) = x + f(x)t + O\left(t^2\right)$ is the group of transformation corresponding to ODE

$$\dot{x} = f(x).$$

Let $D(0)$ be a region and $D(t) = g^t D(0)$. Define the function $V(t) = \mathrm{vol}(D(t))$.

Lemma 2.2. $\det(I + At) = 1 + t\,\mathrm{tr}(A) + O\left(t^2\right)$ *as* $t \to 0$

Proof. Use the characteristic polynomial of A: $\det(\lambda I - A)$ and let $t = -1/\lambda$. □

Lemma 2.3.

$$\left.\frac{\mathrm{d}V}{\mathrm{d}t}\right|_{t=0} = \int_{D(0)} \mathrm{div}\, f \,\mathrm{d}x.$$

Proof. We write

$$V(t) = \int_{D(t)} 1 \,\mathrm{d}y = \int_{D(0)} \det\left(\frac{\partial x(t)}{\partial x(0)}\right) \mathrm{d}x,$$

where we have applied change of variable $x(0) \to x(t)$. We expand $x(t)$ in terms of t as

$$x(t) = x(0) + \dot{x}(0)t + O(t^2) = x(0) + f(x(0))t + O(t^2).$$

Then the Jacobian matrix satisfies

$$\left(\frac{\partial x(t)}{\partial x(0)}\right) = I + \frac{\partial f}{\partial x} + O(t).$$

By Lemma 2.2, we have

$$\left.\frac{\mathrm{d}V}{\mathrm{d}t}\right|_{t=0} = \int_{D(0)} \det\left(I + \frac{\partial f}{\partial x}\right) \mathrm{d}x = \int_{D(0)} \mathrm{div}\, f \,\mathrm{d}x.$$

□

When applied to the Hamiltonian flow, the setting is $x = (p, q)^\intercal$ and $f(x) = (-H_q, H_p)$ so that

$$\mathrm{div}\, f = \partial_p(-H_q) + \partial_q H_p = 0.$$

Therefore the phase flow defined by the Hamiltonian system preserves volume. A famous illustration is Arnold's cat in [1].

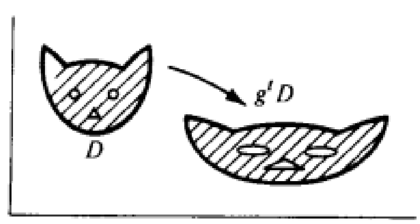

Figure 48 Conservation of volume

FIGURE 4. Conservation of volume of the Hamiltonian flow.

An application of Liouville's theorem is the following Poincaré's recurrence theorem.

Theorem 2.4 (Poincaré's recurrence theorem)**.** *Let g be a volume-preserving continuous one-to-one mapping which maps a bounded region D of euclidean space onto itself: $gD = D$. Then in any neighborhood U of any point of D there is a point $x \in U$ which returns to U, i.e., $g^n x \in U$ for some $n > 0$.*

Proof. We consider the images of the neighborhood U:

$$U, gU, g^2U, \dots, g^nU, \dots$$

All of these have the same volume. If they never intersected, D would have infinite volume. Therefore, for some $k \geq 0$ and $l \geq 0$, with $k > l$,

$$g^kU \cap g^lU \neq \varnothing$$

Therefore, $g^{k-l}U \cap U \neq \varnothing$. If y is in this intersection, then $y = g^n x$, with $x \in U(n = k - l)$. Then $x \in U$ and $g^n x \in U(n = k - l)$. ☐

3. Hamilton-Jacobi Equation

The key idea is: *change of variable*.

3.1. **Poincaré-Cartan invariant form.** We first rewrite (18) as a line integral

$$J(\gamma, p) = \int_\gamma p(t, x(t))\,\mathrm{d}x - H(t, x(t), p(t, x(t)))\,\mathrm{d}t \tag{28}$$

where γ is a curve on the (t, x) plane connecting two points (a, A) and (b, B), and $\dot{q}\,\mathrm{d}t = \dot{x}\,\mathrm{d}t = \mathrm{d}x$. On the curve $\gamma = (t, x(t))$, $x(t)$ is a function of t. The vector field $p(t, x)$ is considered a function of (t, x) and $H(t, x, p)$ is the known Hamiltonian. The 1-form

$$\omega = p(t, x)\,\mathrm{d}x - H(t, x, p(t, x))\,\mathrm{d}t$$

is known as Poincaré-Cartan invariant form. When p and x are vector functions, $p\,\mathrm{d}x = \sum_i p_i\,\mathrm{d}x_i$. But we will stick to the simple notation $p\,\mathrm{d}x$.

When restricted to an extremal curve $\gamma_* = (t, x_*(t))$ where x_* solves the Euler-Lagrange equation which equivalently solves Hamilton system $\dot{x}_* = H_p(t, x_*, p_*)$ with the vector field $p_* = L_v(t, x_*(t), \dot{x}_*(t))$, we have

$$J(\gamma_*, p_*) = \int_a^b [(p_*, \dot{x}_*) - H(t, x_*, p_*)]\,\mathrm{d}t = \int_a^b L(t, x_*(t), \dot{x}_*(t))\,\mathrm{d}t = I(x_*).$$

For general curve γ and p, we may not have $J(\gamma, p) = I(x)$.

3.2. **Hamilton-Jacobi equation.** If the line integral is independent of path, i.e., there exists a potential function $u(t, x) : \mathbb{R} \times V \to \mathbb{R}$, which is called an eikonal, s.t.

$$J(\gamma, p) = u(b, B) - u(a, A)$$

for any simple curve from (a, A) to (b, B) on the (t, x) plane, then

$$\mathrm{grad}\,_{(t,x)}u = p(t, x)\,\mathrm{d}x - H(t, x(t), p(t, x))\,\mathrm{d}t,$$

which is equivalent to

$$\partial_t u = -H(t, x, p(t, x)), \quad \nabla u = p. \tag{29}$$

Can we find such potential function u?

We can eliminate the unknown $p(t, x)$ in (29) and obtain a PDE for the potential function u which is known as Hamilton-Jacobi equation.

Hamiton-Jacobi Equation

$$\partial_t u(t,x) + H(t,x,\nabla u(t,x)) = 0. \tag{30}$$

If we can solve (30), then we get such u and we can compute the action on the extreme curve without finding the path x_*, i.e.,

$$I(x_*) = u(b,B) - u(a,A).$$

On the other hand, to be closed, the form $\omega = p(t,x)\,\mathrm{d}x - H(t,x,p(t,x))\,\mathrm{d}t$ should satisfy the condition $\mathrm{d}\omega = 0$ which may not hold. Namely it is possible no solution exists to (30).

3.3. **Directional field.** Assume we have found a solution u to the H-J equation (30). Can we find the extreme curve x_* more easily? Yes. We can solve the ODE

$$\dot{x} = H_p(t,x,\nabla u(t,x(t))).$$

As u is known, the right hand side is a known vector field.

Let Ω be a simply connected domain in (t,x) domain containing the ending points (a,A) and (b,B). Consider a field $\psi(t,x)$ defined on Ω and assume the solution of ODE system

$$\dot{x} = \psi(t,x)$$

passing through a point (t_0,x_0) is locally unique and satisfies the Euler-Lagrange equation. Such $\psi(t,x)$ is called a directional field and the domain Ω is called a field of extremals. With such ψ, instead of solving a 2nd order we can solve a first order ODE with initial condition $x(a) = A$ to get the extreme curves. We just mentioned $\psi(t,x) = H_p(t,x,\nabla u(t,x))$ is a directional field.

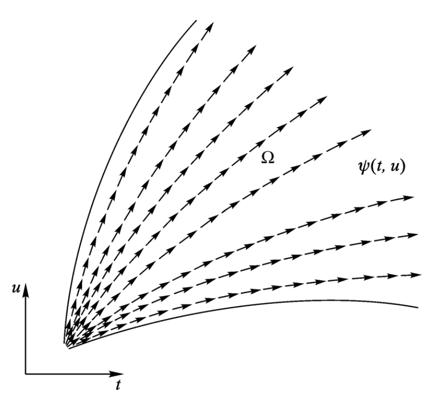

FIGURE 5. Direction field

Given a directional field ψ, consider $p_\psi(t,x) = L_v(t,x,\psi(t,x))$ and denote by

$$J(\gamma,\psi) = J(\gamma,p_\psi) = \int_\gamma L_v(t,x(t),\psi(t,x(t)))\,\mathrm{d}x - H(t,x(t),L_v(t,x(t),\psi(t,x(t)))\,\mathrm{d}t.$$

Again restricted the solution curve $\gamma_* : \dot{x}_* = \psi(t, x_*)$, here we use notation with $*$ to emphasize $x_*(t)$ solves the E-L equation, then using (26), we have

$$J(\gamma_*, \psi) = \int_\gamma [p_* \dot{x}_* - H(t, x_*, p_*)] \,\mathrm{d}t$$
$$= \int_a^b L(t, x_*(t), \dot{x}_*(t)) \,\mathrm{d}t = I(x_*).$$

The directional field ψ can be also derived from the no conjugate point condition. Assume $x(t)$ solves the E-L equation and along the extreme curve $(t, x(t))$ for $t \in [a, b]$, there is no conjugate point and $L_{vv}(t, x(t), \dot{x}(t)) > 0$. Then solve the E-L equation

$$\partial_t L_v(t, \varphi(t, \alpha), \dot{\varphi}(t, \alpha)) = L_x(t, \varphi(t, \alpha), \dot{\varphi}(t, \alpha))$$

but with initial condition

$$\varphi(a, \alpha) = x(a), \quad \partial_t \varphi(t, \alpha) = \dot{x}(a) + \alpha.$$

For $|\alpha|$ is small enough, we can guarantee the existence of the solution and the extremal curve $x(t) = \varphi(t, 0)$. Use the no conjugate point condition, we can conclude $\partial_\alpha \varphi(t, \alpha) > 0$ for $|\alpha|$ small enough. By the implicit function theorem, the equation

$$x = \varphi(t, \alpha)$$

has a unique continuously differentiable solution $\alpha = w(t, x)$ locally. Finally

$$\psi(t, x) = \partial_t \varphi(t, w(t, x))$$

is a directional field as

$$\dot{x} = \partial_t \varphi(t, w(t, x)) = \psi(t, x).$$

3.4. **Strong Weierstrass condition.** Next we verify γ_* is a local strong minimum if a stronger Weierstrass condition satisfies: there exists an $\epsilon > 0$ s.t.

$$\mathfrak{E}_L(t, x, v, \xi) := L(t, x, v + \xi) - L(t, x, v) - \xi \cdot L_v(t, x, v) \geq 0,$$

for all $\xi \in \mathbb{R}^N$, $\forall (x, v) \in B_\epsilon((x_*, \dot{x}_*), \| \cdot \|_{1,\infty})$. Recall that Weierstrass condition only requires the inequality holds for one point (x_*, x'_*).

a figure here

Assume we have the potential u solves the H-J equation. Set $\psi = H_p(t, x, \nabla u)$ and assume ψ is Lipschitz continuous. Now take a curve $\gamma = (t, x(t))$ with the same ending points of γ_* which may not solve the ODE, i.e. $\dot{x} \neq \psi(t, x)$. As the line integral is independent of path, i.e.,

$$I(x_*) = J(\gamma_*, p_*) = J(\gamma, p_\psi)$$
$$= \int_a^b L_v(t, x, \psi) \,\mathrm{d}x + [L(t, x, \psi) - \psi L_v(t, x, \psi)] \,\mathrm{d}t$$
$$= \int_a^b L(t, x, \psi) + (\dot{x} - \psi) L_v(t, x, \psi) \,\mathrm{d}t.$$

Then we get the difference

$$I(x) - I(x_*) = \int_a^b L(t, x, \dot{x}) - L(t, x, \psi) - (\dot{x} - \psi) L_v(t, x, \psi) \,\mathrm{d}t$$
$$= \int_a^b \mathfrak{E}_L(t, x, \psi, \dot{x} - \psi) \,\mathrm{d}t \geq 0,$$

for $x \in B_\epsilon(x_*, \|\cdot\|_\infty), t \in [a,b]$ and $\|\psi(t,x) - \psi(t,x_*)\| \le C\|x - x_*\| \le C\epsilon$. Therefore x_* is a local minimum.

3.5. **An example.** We provide an example in [2] to illustrate the procedure to solve the Hamilton-Jacobi equation first and then solve the Hamiltonian system afterwards.

Consider the harmonic oscillator example. Given m and k two positive constants, the Lagrangian and Hamiltonian is

$$L(t,x,v) = \frac{1}{2}\left(mv^2 - kx^2\right), \quad H(t,p,q) = \frac{1}{2}\left(\frac{p^2}{m} + kx^2\right),$$

where the dual variable $p = mv$ represents the momentum. H-J equation becomes

$$\partial_t u + \frac{1}{2}\left(\frac{|\nabla u|^2}{m} + kx^2\right) = 0.$$

To solve the H-J equation, try a special form (change of variable)

$$u(t,x) = \phi(x) - \alpha t, \tag{31}$$

where α is a parameter and ϕ is an unknown function to be determined.

Substitute back to H-J equation, we get

$$\frac{1}{2}\left(\frac{(\phi'(x))^2}{m} + kx^2\right) = \alpha$$

i.e.

$$\phi'(x) = \sqrt{m\left(2\alpha - kx^2\right)},$$

which can be solved by separation of variables:

$$u(t,x,\alpha) = \int_0^x \sqrt{m\left(2\alpha - ky^2\right)}\,\mathrm{d}y - \alpha t.$$

The integral on the right hand side is still not easy to compute. We further take partial derivative of α We now solve the equation

$$-\partial_\alpha u = t - \int_0^x \frac{m}{\sqrt{m\left(2\alpha - ky^2\right)}}\,\mathrm{d}y = t - \sqrt{\frac{m}{k}}\arcsin\left(\sqrt{\frac{k}{2\alpha}}x\right).$$

Solving $\beta = -\partial_\alpha u$, we get

$$x = \sqrt{\frac{2\alpha}{k}}\sin\left(\sqrt{\frac{k}{m}}(t-\beta)\right).$$

Substituting it back into u, it follows that

$$p = \nabla u = \phi'(x) = \sqrt{2\alpha m}\cos\left(\sqrt{\frac{k}{m}}(t-\beta)\right).$$

This gives the solution of the Hamiltonian system of a harmonic oscillator involving the two parameters α, β.

For this example, it seems easier to solve the Euler-Lagrange equation or Hamiltonian system. The reader is encouraged to try the Lagrangian

$$L(t,x,v) = e^{-x}\sqrt{1+v^2},$$

and compare the three approaches.

TABLE 1. Comparison of Euler-Lagrange equation, Hamiltonian system, and Hamilton-Jacobi equation

Function	$L(t,x,v)$	$H(t,q,p)$	$u(t,x)$
Unknown	$x(t)$	$(p(t),q(t))$	$u(t,x)$
Equation	$\frac{\mathrm{d}}{\mathrm{d}t}L_v = L_x$	$\begin{cases}\dot p = -H_q,\\ \dot q = H_p,\end{cases}$	$\partial_t u + H(t,x,\nabla u) = 0$
	2nd order ODE	1st order ODE system	1st order PDE

3.6. **Physics from H-J equation.** We use the light through a medium example to explain the physical property from H-J equation.

Denote the density of a medium by $\rho(t,x)$. Consider the Lagrangian

$$L(t,x,v) = \rho(t,x)\sqrt{1+v^2}.$$

By Fermat's principle, the path taken by a ray between two given points is the path that can be traveled in the least time which is equivalent the problem

$$\inf_x \int L(t,x(t),\dot x(t))\,\mathrm{d}t$$

By direct calculation, we get the Hamiltonian

$$H(t,p,q) = -\sqrt{\rho^2(t,q)-p^2}.$$

Then the Hamilton-Jacobi equation becomes

$$(\partial_t u)^2 + |\nabla u|^2 = \rho^2.$$

The level set

$$u(t,x) = \text{const}$$

is the wavefront of light.

The Jacobi field is

$$\psi(t,x) = H_p(t,x,\nabla u) = \frac{\nabla u}{\sqrt{\rho^2 - |\nabla u|^2}} = \frac{\nabla u}{\partial_t u}.$$

For the extremal curve, $\dot x = \psi(t,x)$, the tangential direction is

$$(\dot t, \dot x) = (1,\dot x) = c(\partial_t u, \nabla u) \quad \text{with } c = (\partial_t u)^{-1}.$$

Therefore a ray of light travels perpendicularly to the wavefronts. more from Arnold [1].

3.7. **Hopf-Lax formula.** We consider solving a special form of H-J equation

$$u_t + H(\nabla u) = 0, \tag{32}$$

$$u(0,x) = g(x). \tag{33}$$

That is the Hamiltonian $H(p)$ only depends on p and in turn the Lagrangian $L(v)$ depends on v only.

We present a motivation of Hopf-Lax formula from [4]. For any fixed y,z, the function

$$v(t,x) = g(y) + (x-y)\cdot z - tH(z)$$

solves (32) which can be verified by direct calculation: $\nabla_x v = z$ and $\partial_t v = -H(z) = -H(\nabla_x v)$. But $v(0,x) = g(y) + (x-y)\cdot z \neq g(x)$.

The Hopf-Lax formula is obtained as a two-parameter envelope solution:

$$u(x,t) = \inf_{y\in\mathbb{R}^n} \sup_{z\in\mathbb{R}^n} \{g(y) + (x-y)\cdot z - tH(z)\}$$

Interpreting the inner sup as a Legendre transform, we arrive the Hopf-Lax formula

$$u(x,t) = \inf_{y\in\mathbb{R}^n} \left\{ g(y) + tL\left(\frac{x-y}{t}\right)\right\}. \tag{34}$$

When g is convex, we can obtain another Hopf-Lax formula by switching inf and sup:

$$u(x,t) = \sup_{z\in\mathbb{R}^n} \inf_{y\in\mathbb{R}^n} \{g(y) + (x-y)\cdot z - tH(z)\}$$

The inner inf can be interpreted as a Legendre transform to obtain

$$u(x,t) = \sup_{z\in\mathbb{R}^n} \{x\cdot z - g^*(z) - tH(z)\}$$

The verification

$$u(x,0) = g(x)$$

is given by Lemma 2 in Evans [3, Chapter 3] and $u(t,x)$ defined by (34) solves the H-J equation (32) is Theorem 5 [3, Chapter 3].

We will verify a simple but important case

$$L(v) = \frac{1}{2}\|v\|^2, \quad H(p) = \frac{1}{2}\|p\|^2.$$

The Hopf-Lax formulae becomes

$$u(x,t) = \inf_{y\in\mathbb{R}^n} \left[g(y) + \frac{1}{2t}\|y-x\|^2\right], \tag{35}$$

which is known as the Moreau envelope of g. Assume g is bounded below so that the inf in (35) is achievable.

To simplify the further discussion, we also assume $g \in C^1$. Then we conclude the inf is achieved at a point $\hat{x}$ s.t.

$$-\nabla g(\hat{x}) = \frac{\hat{x}-x}{t}$$

and

$$u(x,t) = g(\hat{x}) + \frac{1}{2t}\|\hat{x}-x\|^2.$$

The $\hat{x}$ is a function of x and is known as the proximal operator $\hat{x} = \operatorname{prox}_{tg}(x)$. We can write as

$$\hat{x} = x - t\nabla g(\hat{x}). \tag{36}$$

That is $\hat{x}$ is the implicit Euler method for the gradient flow $\dot{x} = -\nabla g(x)$ with step size t.

From formulae (36), we immediately conclude $u(x,0) = g(x)$ as $\hat{x} = x$. On the other extreme, for $t = +\infty$, $u(x,+\infty) = \min_y g(y)$ is a constant function. For a fixed x, $u(x,\cdot)$ will decrease from $g(x)$ to $\min_y g(y)$ as t increases from 0 to $+\infty$.

Computing derivatives is a little bit complicated. Let $\hat{x} = \operatorname{prox}_{tg}(x)$. The function $u(x,t)$ can be evaluated at $\hat{x}$ and by definition $u(x+\epsilon h,t) \leq g(\hat{x}) + \frac{1}{2t}\|\hat{x} - (x+\epsilon h)\|^2$. Then the difference

$$u(x+\epsilon h,t) - u(x,t) \leq \frac{1}{2t}\left[\|\hat{x} - (x+\epsilon h)\|^2 - \|\hat{x}-x\|^2\right].$$

From that we can conclude

$$\nabla u(x,t) = \frac{\hat{x}-x}{t}.$$

Similarly

$$u(x, t+\epsilon) - u(x,t) \leq \frac{1}{2(t+\epsilon)}\|\hat{x} - x\|^2 - \frac{1}{2t}\|\hat{x} - x\|^2.$$

From which we can conclude

$$\partial_t u = -\frac{1}{2t^2}\|\hat{x} - x\|^2 = -\frac{1}{2}\|\nabla u\|^2.$$

References

[1] V. I. Arnol'd. *Mathematical methods of classical mechanics*, volume 60. Springer Science & Business Media, 2013. 7, 8, 13

[2] K.-C. Chang. *Lecture Notes on Calculus of Variations*, volume 6. World Scientific, 2016. 11

[3] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, 1997. 14

[4] Jeff (https://math.stackexchange.com/users/313346/jeff), Hopf-Lax formula motivation, URL (version: 2017-03-25): https://math.stackexchange.com/q/2202914 13