HAMILTON JACOBI THEORY IN CALCULUS OF VARIATION

LONG CHEN

ABSTRACT. We introduce the Hamiltonian as the conjugate (Legendre transform) of the Lagrangian and derive the corresponding Hamiltonian system. By rewriting the action integral using the Poincaré–Cartan form, we obtain the Hamilton–Jacobi equation. As preparation, we give a concise introduction to convex analysis.

CONTENTS

1. Conjugate of Convex Functions 1
1.1. Convex Sets 1
1.2. Dual Representation 2
1.3. Convex Function and Epigraph 3
1.4. Conjugate Function 4
1.5. Smooth and Strongly Convex Functions 5
1.6. Regularity of Convex Functions 6
2. Hamiltonian and Hamiltonian Systems 7
2.1. Hamiltonian system 7
2.2. Equivalence with the Euler–Lagrange equation 8
2.3. Conservation of Hamiltonian 9
2.4. Conservation of volume: Liuville's theorem 9
3. Hamilton-Jacobi Equation 10
3.1. Poincaré–Cartan invariant form 11
3.2. Hamilton–Jacobi equation 12
3.3. An example 13
3.4. Directional field 14
3.5. Strong Weierstrass condition 15
3.6. Physics from the Hamilton–Jacobi equation 16
References 17

1. CONJUGATE OF CONVEX FUNCTIONS

In this section, we give a short introduction to convex analysis and focus on an *outside* characterization of convex sets and convex functions.

1.1. Convex Sets.

A subset S of a vector space V is convex if

$$x, y \in S,\ \alpha \in [0,1] \Longrightarrow (1-\alpha)x + \alpha y \in S. \tag{1}$$

A convex set need not be closed, for example an open ball. It need not be bounded. A basic example is a *convex cone* K, defined by the property that if $x \in K$, then $\lambda x \in K$ for

Date: May 4, 2026.

all $\lambda \geq 0$. A convex set may also have empty interior, for example a line segment in $\mathbb{R}^n$ for $n > 1$. Thus convexity is not a topological property.

For a linear subspace S, one requires $\alpha x + \beta y \in S$ for all $\alpha, \beta \in \mathbb{R}$. Convexity is weaker: it only allows coefficients with $\alpha + \beta = 1$ and $\alpha, \beta \geq 0$, which is called a convex combination. Geometrically, convexity means that the line segment joining x and y lies in the set, while for a linear subspace the whole plane spanned by x and y lies in the subspace. In particular, 0 must belong to a subspace, while a convex set need not contain the origin.

Convexity is translation invariant: if S is convex, then $S + z := \{x + z : x \in S\}$ is also convex for any $z \in V$. Hence many geometric properties of convex sets do not change under translation. Depending on the context, the condition $0 \in S$ can often be arranged by translation.

1.2. **Dual Representation.** The definition of convexity in (1) is an *inside* condition. We now present an outside characterization of convex sets.

Consider a convex polygon cut from a sheet of paper. One way is to trim along the boundary with scissors. If scissors are not available and we can only fold the paper, then we may fold along straight lines and keep the side that contains the polygon. Repeating this process gives an outside view of convexity through half-spaces.

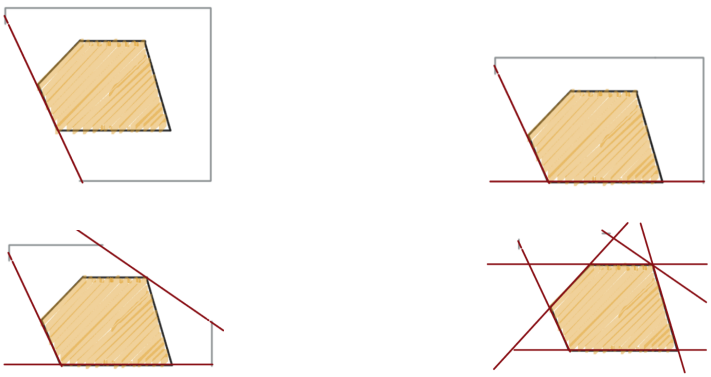


FIGURE 1. Dual representation of a convex set.

Now we give a mathematical description. Let V be a Hilbert space with inner product $(\cdot,\cdot)$. For $n \in V$ and $b \in \mathbb{R}$, define the closed half-spaces

$$H^+_{n,b} := \{u \in V \mid (n,u) - b \geq 0\}, \qquad H^-_{n,b} := \{u \in V \mid (n,u) - b \leq 0\},$$

and the hyperplane

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

Theorem 1.1 (Dual representation of a convex set)**.** *A nonempty closed convex set S is the intersection of all closed half-spaces that contain it:*

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

Proof. This follows from the Hahn–Banach separation theorem for convex sets. □

If $S \subseteq H^+_{n,b_2}$, then $S \subseteq H^+_{n,b_1}$ for any $b_1 < b_2$. In the folding picture, we want the fold line to be as close to the boundary as possible. So we define

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

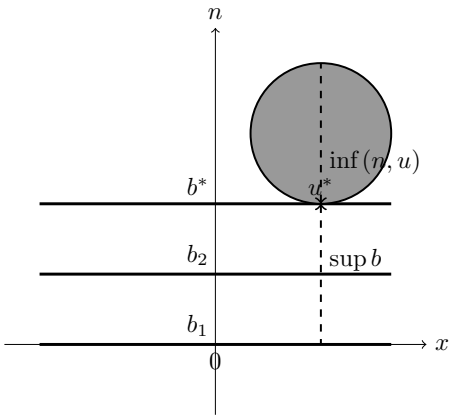


FIGURE 2. Duality for convex sets: $\inf_{u\in S}(n,u) = \sup_{S\subseteq H^+_{n,b}} b$. In the figure, we may take $n = (0,1)$, so $(n,u) = u_d$ is the height. The relation $S \subseteq H^+_{n,b}$ is simply $u_d \geq b$.

Lemma 1.2. *For any nonempty set $S \subset V$ and any fixed direction $n \in V$,*

$$\inf_{u\in S}(n,u) = \sup_{S\subseteq H^+_{n,b}} b. \tag{2}$$

Proof. This is immediate from the definition. If the infimum is attained at some $u^* \in S$, then $b^* = (n,u^*)$, as illustrated in Figure 2. □

The identity (2) already shows a basic duality: the variable changes from the inner point $u \in S$ to the outer parameter b, and the operation changes from $\inf$ to $\sup$.

The hyperplane H_{n,b^*} is called a supporting hyperplane of S. When ∂S is smooth, this supporting hyperplane agrees with the tangent hyperplane at the touching point. When ∂S is not smooth, for example at a corner, there may be more than one supporting hyperplane; see Fig. 4.

We can therefore refine the outside description of a convex set as

$$S = \bigcap_{n\in V} H^+_{n,b^*(n)}. \tag{3}$$

The function $b^*(n)$, viewed as a function of the direction n, determines the convex set.

1.3. **Convex Function and Epigraph.** We now connect convex sets with convex functions and use the geometric view of convex sets to study convex functions.

The *epigraph* of a function $f : X \to \mathbb{R} \cup \{+\infty\}$ is

$$\mathrm{epi}(f) = \{(x,t) \in X \times \mathbb{R} : t \geq f(x)\}.$$

For each $x \in \mathrm{dom}(f) := \{x \in X : f(x) < +\infty\}$, the set $\{x\} \times [f(x), \infty)$ is a vertical ray starting at $(x, f(x))$ and extending upward in $X \times \mathbb{R}$.

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

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) \qquad \forall\, x, y \in X,\ \forall\, \alpha \in [0,1]. \tag{4}$$

It is called *strictly convex* if the inequality is strict for $x \neq y$ and $\alpha \in (0,1)$.

The following result can be proved by definition.

Theorem 1.3. *A function f is convex if and only if* $\operatorname{epi}(f)$ *is convex.*

1.4. **Conjugate Function.** For a convex function f, we introduce the conjugate function, also called the *Legendre transform* or *Fenchel dual*, through the geometry of $\operatorname{epi}(f)$.

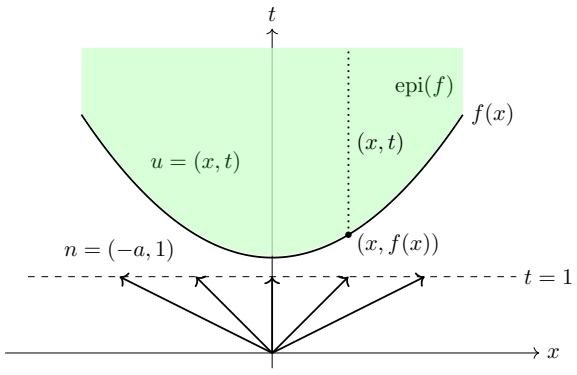


FIGURE 3. Outer characterization of $\operatorname{epi}(f)$.

To use the dual description above, write

$$u = (x,t), \qquad n = (-a, 1), \qquad S = \operatorname{epi}(f).$$

Then the condition

$$\operatorname{epi}(f) \subseteq H^+_{n,b}$$

is equivalent to

$$t \geq \langle a, x\rangle + b, \qquad \forall\, (x,t) \in \operatorname{epi}(f). \tag{5}$$

For fixed x, the smallest such t is $f(x)$. Hence (5) is equivalent to

$$f(x) \geq \langle a, x\rangle + b, \qquad \forall\, x \in \operatorname{dom}(f). \tag{6}$$

Now consider the outside characterization

$$b^*(n) = \sup_{S \subseteq H^+_{n,b}} b = \inf_{u \in S} (n, u) = \inf_{x \in \operatorname{dom}(f)} \big[f(x) - \langle a, x\rangle \big].$$

This leads to the definition of the *conjugate function* $f^* : V^* \to \mathbb{R} \cup \{+\infty\}$:

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

The conjugate $f^*(a)$ may be $+\infty$ for general convex functions. For example, if $f(x) = e^{-x}$ on $\mathbb{R}$, then $f^*(a) = +\infty$ for every $a > 0$.

By definition, we have the Fenchel inequality.

Lemma 1.4 (Fenchel inequality)**.** *For a convex function f and its conjugate f^* defined by (7), we have*

$$f(x) + f^*(a) \geq \langle a, x\rangle \qquad \forall\, x \in \mathrm{dom}(f),\ \forall\, a \in V^*. \tag{8}$$

If f is differentiable at x, then equality holds in (8) when $a = \nabla f(x)$.

Since the convex set $\mathrm{epi}(f)$ is characterized by its supporting half-spaces, the convex function f is characterized by its conjugate f^*. The conjugate transform, like the Fourier transform, rewrites a function in dual variables.

The conjugate can be applied once more. We then obtain the biconjugate

$$f(x) = f^{**}(x) := \sup_{a \in V^*} \left[\langle a, x\rangle - f^*(a)\right] \tag{9}$$

provided that $\mathrm{epi}(f)$ is closed in $V \times \mathbb{R}$, equivalently, provided that f is proper, convex, and lower semicontinuous.

To keep the discussion simple, we now restrict attention to a regular class of convex functions for which the conjugate is easy to compute.

1.5. **Smooth and Strongly Convex Functions.** We switch the notation $x \to v$ and $a \to p$. A convex function is called *μ-strongly convex* with parameter $\mu > 0$ if

$$f(\cdot) - \frac{\mu}{2}\|\cdot\|^2$$

is convex. The gradient ∇f is Lipschitz continuous with constant $0 < L < \infty$ if

$$\|\nabla f(v) - \nabla f(w)\|_* \leqslant L\|v - w\| \qquad \forall\, v, w \in \mathrm{dom}(f).$$

For constants $0 \leqslant \mu \leqslant L < \infty$, we introduce the class

$$\mathcal{S}^1_{\mu,L} := \{f \text{ is } \mu\text{-strongly convex and } \nabla f \text{ is } L\text{-Lipschitz}\}.$$

The ratio $\kappa = L/\mu$ is called the condition number of f. It is known that

$$f \in \mathcal{S}^1_{\mu,L} \iff \mu\|v - w\|^2 \leqslant \langle \nabla f(v) - \nabla f(w), v - w\rangle \leqslant L\|v - w\|^2 \qquad \forall\, v, w \in V.$$

For any $p \in V^*$, the conjugate $f^*(p)$ is finite, and the supremum in

$$f^*(p) := \sup_{v \in V} \left[\langle p, v\rangle - f(v)\right] \tag{10}$$

is attained at a unique point $v(p) \in V$ satisfying $\nabla f(v(p)) = p$. Moreover,

$$f^*(p) = \langle p, v(p)\rangle - f(v(p)). \tag{11}$$

Thus the map $v(p) : V^* \to V$ is the inverse of the gradient map $\nabla f : V \to V^*$. Equivalently,

$$v = \nabla f^*(p) \qquad \Longleftrightarrow \qquad p = \nabla f(v). \tag{12}$$

Lemma 1.5. *Let $f \in \mathcal{S}^1_{\mu,L}$ and let f^* be its conjugate. Define*

$$v(p) := \arg\sup_{v \in V} \left[\langle p, v\rangle - f(v)\right].$$

Then:

1. *f^* is continuously differentiable, with gradient $\nabla f^*(p) = v(p)$, and $\nabla f(v(p)) = p$.*
2. *f^* is $1/L$-strongly convex.*
3. *∇f^* is Lipschitz continuous with constant $1/\mu$.*

Consequently, $f^* \in \mathcal{S}^1_{1/L,1/\mu}$.

Example. Let $f(v) = \frac{1}{2}(Av, v)$, where $A : V \to V$ is a symmetric positive definite operator. Then $f \in \mathcal{S}^1_{\mu,L}$ with $\mu = \lambda_{\min}(A)$ and $L = \lambda_{\max}(A)$. The conjugate function is

$$f^*(p) = \frac{1}{2}(A^{-1}p, p),$$

and the gradient maps are

$$\nabla f(v) = Av, \qquad \nabla f^*(p) = A^{-1}p.$$

Thus the dual of a strongly convex and Lipschitz smooth function stays in the same class, with the parameters reversed.

1.6. **Regularity of Convex Functions.** Convexity gives a useful form of regularity, though in general it is weaker than differentiability. We summarize several basic properties below and illustrate them in Fig. 4.

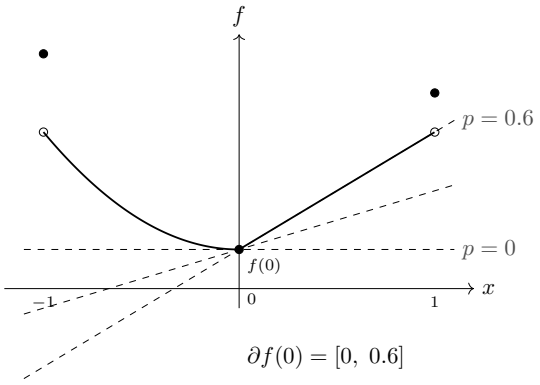


FIGURE 4. A convex and non-smooth function.

1. Continuity on the interior of the domain. While on the boundary of $\mathrm{dom}(f)$, the function may be discontinuous.
2. Existence of one-sided directional derivatives. For a convex function f, the directional derivative
$$f'(x; d) := \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$
exists, though it may be infinite, for all $x \in \mathrm{dom}(f)$ and all directions $d \in \mathbb{R}^n$.
3. Instead of the classical derivative, convex analysis uses the *subdifferential*. For $x \in \mathrm{dom}(f)$, define
$$\partial f(x) := \{v \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle v, y - x\rangle \quad \text{for all } y \in \mathbb{R}^n\}.$$
If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. If f is not differentiable at x, then $\partial f(x)$ contains all slopes of supporting hyperplanes to the graph of f at x. In Fig. 4, $\partial f(0) = [0, 0.6]$ is a subset of slopes.
4. Almost everywhere differentiability. If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then f is differentiable almost everywhere. In other words, the set where f is not differentiable has Lebesgue measure zero.

2. Hamiltonian and Hamiltonian Systems

In this section, we apply the Legendre transform to the Lagrangian to define the Hamiltonian. The Euler–Lagrange equation is then rewritten as a Hamiltonian system.

2.1. **Hamiltonian system.** Assume that $L(t,q,\cdot)$ is strongly convex in v. We apply the Legendre transform to L with respect to v and define the Hamiltonian by

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

Equivalently, if p and v are related by

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

then

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

Since L is strongly convex in v, the relation $p = L_v(t,q,v)$ can be inverted, so v can be written as a function of (t,q,p).

Example 2.1. An important example is

$$L(q,v) = T(v) - U(q).$$

If

$$T(v) = \frac{1}{2}m|v|^2,$$

then

$$p = L_v(q,v) = mv$$

is the momentum. Hence

$$H(q,p) = (p,v) - L(q,v) = (p,v) - T(v) + U(q) = T(v) + U(q),$$

which is the total energy.

Since $f^{**} = f$ for a proper closed convex function, we can recover the Lagrangian from the Hamiltonian by

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

Therefore,

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

Thus the variational problem can be rewritten as

$$\inf_q I(q) = \inf_q \sup_p F(p,q),$$

where the functional of $(p(t),q(t))$ is

$$F(p,q) = \int_a^b (p(t),\dot{q}(t)) - H(t,q(t),p(t))\,\mathrm{d}t. \tag{17}$$

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

Hamiltonian system

$$\begin{cases} \dot{p} = -H_q(t,q,p), \\ \dot{q} = H_p(t,q,p). \end{cases} \tag{18}$$

The Hamiltonian system can also be written as the skew-gradient flow

$$\frac{\mathrm{d}y}{\mathrm{d}t} = J\nabla H(y) = \nabla^{\perp} H, \tag{19}$$

where

$$y = \begin{pmatrix} q \\ p \end{pmatrix}, \qquad J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

The matrix J is skew-symmetric and orthogonal:

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

Geometrically, in two dimensions, J is the rotation by 90° counterclockwise. One also writes $\nabla^{\perp} := J\nabla$.

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

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\nabla E(y), \tag{20}$$

where E is called energy. The flow (19) and (20) will define different dynamic systems.

2.2. **Equivalence with 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

$$H(t,q,p) = (p,\dot{q}) - L(t,q,\dot{q}), \qquad \text{when } p = L_v(t,q,\dot{q}) \tag{21}$$

$$L(t,q,\dot{q}) = (p,\dot{q}) - H(t,q(t),p(t)), \qquad \text{when } \dot{q} = H_p(t,q,p), \tag{22}$$

without writing the $\sup$ operators.

While taking differentials we momentarily regard (t,q,p,v) as independent variables:

$$\begin{aligned} \mathrm{d}H &= H_t\,\mathrm{d}t + H_q\,\mathrm{d}q + H_p\,\mathrm{d}p, \\ \mathrm{d}\big(\langle p,v\rangle - L(t,q,v)\big) &= v\,\mathrm{d}p + p\,\mathrm{d}v - L_t\,\mathrm{d}t - L_q\,\mathrm{d}q - L_v\,\mathrm{d}v. \end{aligned}$$

Equating the two expressions and collecting coefficients of the independent differentials $\mathrm{d}v$, $\mathrm{d}t$, $\mathrm{d}q$, $\mathrm{d}p$ yields

$$p = L_v, \tag{23}$$

$$v = H_p, \tag{24}$$

$$H_q = -L_q, \tag{25}$$

$$H_t = -L_t. \tag{26}$$

Relation (23) and (24) are exactly the Legendre correspondence so that H and L can be related without sup operator. So from now on p, q, v, t are linked by $p = L_v(t,q,v)$.

From Euler–Lagrange to Hamilton. Assume that $q(\,\cdot\,)$ satisfies the Euler–Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\, L_v\big(t,q,\dot q\big) = L_q\big(t,q,\dot q\big).$$

As $p(t) := L_v(t,q(t),\dot q(t))$ and $v=\dot q$, using (24)–(25), we obtain the Hamiltonian system

$$\dot q = v = H_p(t,q,p), \qquad \dot p = L_q = -H_q(t,q,p). \tag{27}$$

Thus every Euler–Lagrange trajectory $(q,\dot q)$ produces a Hamiltonian trajectory (q,p).

From Hamilton to Euler–Lagrange. Conversely, let $(q(\,\cdot\,),p(\,\cdot\,))$ solve (27). Set $v(t) := H_p(t,q(t),p(t))$; then $p = L_v(t,q,v)$ by (23), so $v=\dot q$. Differentiating $p = L_v(t,q,\dot q)$ and using (27) gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\, L_v(t,q,\dot q) = \dot p = -H_q(t,q,p) = L_q(t,q,\dot q),$$

which is exactly the Euler–Lagrange equation. Therefore the two formalisms are equivalent.

In summary, under the regularity conditions that make the Legendre transform well defined and invertible, a curve $q(t)$ satisfies the Euler–Lagrange equation if and only if the pair $\big(q(t),p(t)\big)$ with $p = L_v(t,q(t),\dot q(t))$ satisfies the Hamiltonian system (27).

We emphasize that for Euler-Lagrange equation, two boundary conditions are provided while for Hamiltonian system (27), two initial conditions are given.

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 (19), the calculation is

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

Or even simpler: $\operatorname{div}\operatorname{curl} = 0$.

As a comparison, for the gradient flow (20), we have the energy dispersion

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

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

2.4. **Conservation of volume: 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 2.2 (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: Let $g^t(x) = x + f(x)t + O(t^2)$ be the flow map associated with the ODE

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

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

Lemma 2.3. *For any fixed $n\times n$ matrix A,*

$$\det(I+tA) \;=\; 1 + t\,\operatorname{tr}(A) + O(t^2) \qquad (t\to 0).$$

Proof. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A (in $\mathbb{C}$), then

$$\det(I + tA) \;=\; \prod_{i=1}^{n}\big(1 + t\lambda_i\big) \;=\; 1 + t\sum_{i=1}^{n}\lambda_i + O(t^2) \;=\; 1 + t\ \mathrm{tr}(A) + O(t^2),$$

because $\sum_{i=1}^{n}\lambda_i = \mathrm{tr}(A)$. □

Lemma 2.4.

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

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 2.3, we have

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

□

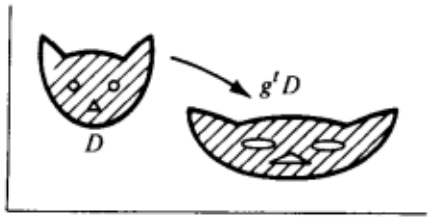


FIGURE 5. Arnold's cat: Conservation of volume of the Hamiltonian flow.

When applied to the Hamiltonian flow, the setting is $x = (p, q)^\intercal$ and $f(x) = \nabla^\perp H$ so that

$$\mathrm{div}\, f = \nabla^\perp H = \partial_p(H_q) - \partial_q H_p = 0.$$

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

3. HAMILTON-JACOBI EQUATION

In this section, we derive the Hamilton–Jacobi equation by rewriting the action integral as a line integral involving the Poincaré–Cartan form. The key idea is: *change of variables*.

We list comparison of Euler-Lagrange equation, Hamiltonian system, and Hamilton-Jacobi equation below.

	Euler-Lagrange	Hamilton	Hamilton-Jacobi
Function	$L(t,x,v)$	$H(t,q,p)$	$H(t,x,\nabla u)$
Unknown	$x(t)\in\mathbb{R}^n$	$(p(t),q(t))\in\mathbb{R}^n\times\mathbb{R}^n$	$u(t,x)\in\mathbb{R}$
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

3.1. **Poincaré–Cartan invariant form.** We rewrite the functional

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

as the line integral

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

where γ is a piecewise smooth curve in the (t,x)-plane joining (a,A) to (b,B). If $\gamma=\{(t,x(t))\mid a\le t\le b\}$, then $\mathrm{d}x=\dot x\,\mathrm{d}t$, and (28) becomes

$$J(\gamma,p)=\int_a^b \Big[(p(t,x(t)),\dot x(t)) - H\big(t,x(t),p(t,x(t))\big)\Big]\,\mathrm{d}t.$$

Here x depends on t along the curve, while $p(t,x)$ is viewed as a field in the variables (t,x). Strictly speaking, the momentum is first defined only along an extremal curve. Writing $p(t,x)$ means that we choose, when possible, a local extension of this momentum to a field in the (t,x)-plane. Such an extension is not automatic globally and usually requires a smooth local family of extremals.

The 1-form

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

is called the Poincaré–Cartan invariant form. If p and x are vectors, then

$$p\;\mathrm{d}x=\sum_i p_i\;\mathrm{d}x_i,$$

and we keep the compact notation $p\;\mathrm{d}x$.

Now restrict ω to an extremal curve $\gamma_*=(t,x_*(t))$, where x_* solves the Euler–Lagrange equation, and define

$$p_*(t):=L_v\big(t,x_*(t),\dot x_*(t)\big).$$

Then

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

For a general curve $\gamma=(t,x(t))$ and a general field p, the line integral $J(\gamma,p)$ is in general different from the action $I(x)$. Along the curve one does not necessarily have

$$p(t,x(t))\overset{?}{=}L_v\big(t,x(t),\dot x(t)\big).$$

This also explains why, in general, $F(p,q) \neq I(q)$. The functional $F(p,q)$ is defined for two independent variables (p,q) and agrees with the original functional only when

$$p = L_v(t,q,\dot{q}), \qquad \text{i.e. } I(q) = \sup_p F(p,q),$$

with equality attained at the momentum determined by the Legendre relation.

3.2. **Hamilton–Jacobi equation.** The goal is to replace the line integral by a difference of potential values. Suppose the line integral is path-independent on a simply connected region of the (t,x)-plane. Then there exists a potential function

$$u(t,x) : \mathbb{R} \times V \to \mathbb{R}$$

such that

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

for any piecewise smooth curve γ from (a,A) to (b,B). Equivalently,

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

which gives

$$\partial_t u = -H\big(t,x,p(t,x)\big), \qquad \nabla_x u = p. \tag{29}$$

Can we find such a potential u?

Eliminating the variable $p(t,x)$ from (29), we obtain a partial differential equation for u alone, called the Hamilton–Jacobi equation.

Hamilton–Jacobi Equation

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

The natural data are usually prescribed in time, either as an initial condition

$$u(0,x) = u_0(x),$$

or as a terminal condition

$$u(T,x) = g(x).$$

If the spatial domain is bounded, one may also prescribe boundary data on the inflow part of the boundary, determined by the characteristic flow. In variational problems, $u(t,x)$ is often a value function, and the initial or terminal condition is set by the endpoint cost.

If a sufficiently smooth solution of (30) exists, then along an extremal path we have

$$p = \nabla_x u,$$

and therefore the action can be computed without first solving for x_*:

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

On the other hand, for the 1-form

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

to be exact, it must at least be closed, that is, $\mathrm{d}\omega = 0$. In one space dimension,

$$\mathrm{d}\omega = \big(p_t + H_x + H_p\, p_x\big)\,\mathrm{d}t \wedge \mathrm{d}x.$$

Hence the condition $\mathrm{d}\omega = 0$ becomes

$$p_t + H_x + H_p\, p_x = 0.$$

If $p = \nabla_x u$, this is exactly the x-derivative of the Hamilton–Jacobi equation. Thus $\mathrm{d}\omega = 0$ is the local compatibility condition for the existence of a potential u. This condition need not hold in general, so the field $p(t,x)$ may fail to come from a global smooth potential. We refer to [3, Chapter 10] for further discussion of the Hamilton–Jacobi equation, including viscosity solutions.

3.3. **An example.** We illustrate the method on a standard example from [2].

When $H(x,p)$ is independent of t, we look for a separated solution of the form

$$u(t,x) = \phi(x) + A(t). \tag{31}$$

Substituting this ansatz into the Hamilton–Jacobi equation gives

$$A'(t) + H(x, \nabla\phi(x)) = 0.$$

Hence

$$H(x, \nabla\phi(x)) = -A'(t) = \alpha$$

for some constant α. Therefore

$$A(t) = -\alpha t + C,$$

and the constant C can be absorbed into ϕ. We are thus led to solve

$$H(x, \nabla\phi(x)) = \alpha.$$

Consider the harmonic oscillator. Let $m, k > 0$. The Lagrangian and Hamiltonian are

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

where

$$p = L_v(t,x,v) = mv$$

is the momentum. The Hamilton–Jacobi equation becomes

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

Using the ansatz $u(t,x) = \phi(x) - \alpha t$, we obtain

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

Thus

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

To stay in the real setting, we assume

$$\alpha > 0, \qquad |x| \le \sqrt{\frac{2\alpha}{k}}.$$

Integrating, we get the complete integral

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

To recover the characteristics, we introduce the second parameter

$$\beta = -\partial_\alpha u.$$

Differentiating with respect to α, we find

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

Hence

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

Also,

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

These formulas recover the Hamiltonian trajectories of the harmonic oscillator, with parameters α and β.

In this example, it is simpler to solve the Euler–Lagrange equation or integrate the Hamiltonian system directly. The reader may compare these methods with the Hamilton–Jacobi approach for the Lagrangian

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

3.4. **Directional field.** Assume that we have found a sufficiently smooth solution u of the Hamilton–Jacobi equation (30). Define the vector field

$$\psi(t,x) := H_p\big(t,x,\nabla_x u(t,x)\big). \tag{32}$$

The extremal curve can then be recovered by solving the first-order ODE

$$\dot{x} = \psi(t,x), \qquad x(a) = A. \tag{33}$$

Thus the second-order Euler–Lagrange equation is replaced by a first-order equation.

Let $\Omega \subset \mathbb{R}\times V$ be a simply connected region containing the endpoints (a,A) and (b,B), and let $\psi(t,x)$ be a vector field on Ω. We call ψ a *directional field* if its integral curves are extremals, that is, they satisfy the Euler–Lagrange equation.

Given a directional field ψ, define

$$p^\psi(t,x) := L_v\big(t,x,\psi(t,x)\big).$$

For a general curve $\gamma = (t,x(t))$, the field p^ψ is independent of the velocity $\dot{x}(t)$ along γ. Hence the line integral built from p^ψ need not agree with the original action. Indeed,

$$\begin{aligned} J(\gamma,p^\psi) &= \int_\gamma p^\psi(t,x)\,\mathrm{d}x - H\big(t,x,p^\psi(t,x)\big)\,\mathrm{d}t \\ &= \int_a^b \Big[(p^\psi(t,x(t)),\dot{x}(t)) - H\big(t,x(t),p^\psi(t,x(t))\big)\Big]\,\mathrm{d}t, \end{aligned}$$

which is in general different from

$$I(x) = \int_a^b L\big(t,x(t),\dot{x}(t)\big)\,\mathrm{d}t,$$

because one does not usually have

$$p^\psi(t,x(t)) = L_v\big(t,x(t),\dot{x}(t)\big).$$

Now let $\gamma_* = (t,x_*(t))$ be an integral curve of the directional field, that is,

$$\dot{x}_* = \psi(t,x_*).$$

Along this curve,

$$p_*(t) := p^\psi(t,x_*(t)) = L_v\big(t,x_*(t),\psi(t,x_*(t))\big) = L_v\big(t,x_*(t),\dot{x}_*(t)\big).$$

So the Legendre relation holds along γ_*. Using (22), we obtain

$$\begin{aligned} J(\gamma_*, p^\psi) &= \int_a^b \Big[(p_*(t), \dot{x}_*(t)) - H\big(t, x_*(t), p_*(t)\big)\Big]\, \mathrm{d}t \\ &= \int_a^b L\big(t, x_*(t), \dot{x}_*(t)\big)\, \mathrm{d}t \\ &= I(x_*). \end{aligned}$$

When u solves the Hamilton–Jacobi equation, the field

$$\psi(t,x) = H_p\big(t, x, \nabla_x u(t,x)\big)$$

defines such a directional field. Therefore, once a solution of the Hamilton–Jacobi equation is known, its associated directional field generates extremals through a first-order ODE, and along those curves the Poincaré–Cartan integral agrees with the original action.

3.5. **Strong Weierstrass condition.** We use the directional field to prove a sufficient condition for x_* to be a local strong minimum.

Strong Weierstrass Condition

There exists an $\epsilon > 0$ such that

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

for all $\xi \in \mathbb{R}^N$ and all $(x,v) \in B_1((x_*(t), \dot{x}_*(t)), \epsilon)$.

Recall that the classical Weierstrass condition requires the inequality to hold only at a single point $(x_*(t), \dot{x}_*(t))$. The strong Weierstrass condition ensures that the map $v \mapsto L(t,x,v)$ is locally convex near $\dot{x}_*(t)$, which allows us to control perturbations in the velocity component.

[a figure here to illustrate the strong Weierstrass condition.]

Assume we have a potential function u that solves the Hamilton–Jacobi equation. Define the directional field $\psi(t,x) := H_p(t,x,\nabla u(t,x))$, and assume ψ is Lipschitz continuous. Let $\gamma = (t, x(t))$ be any curve with the same endpoints as γ_*, but not necessarily satisfying $\dot{x} = \psi(t,x)$. Since $\psi = H_p(t,x,\nabla u)$ and u solves the H–J equation, the line integral is path-independent. Also note:

$$H(t,x,p^\psi) = \langle \psi, p^\psi \rangle - L(t,x,\psi), \qquad \text{with } p^\psi = L_v(t,x,\psi). \tag{34}$$

Then:

$$\begin{aligned} I(x_*) = J(\gamma_*, p^\psi) &= J(\gamma, p^\psi) \qquad \text{(path independence)} \\ &= \int_a^b L_v(t,x,\psi)\, \mathrm{d}x + \left[L(t,x,\psi) - \psi \cdot L_v(t,x,\psi)\right] \mathrm{d}t \quad \text{by (34)} \\ &= \int_a^b \left[L(t,x,\psi) + (\dot{x} - \psi) \cdot L_v(t,x,\psi)\right] \mathrm{d}t. \end{aligned}$$

So the difference becomes

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

provided that $x(t) \in B(x_*(t),\epsilon)$ and $\|\psi(t,x) - \psi(t,x_*)\| \leq C\|x - x_*\| \leq C\epsilon$ due to Lipschitz continuity. This shows that x_* is a local strong minimum.

Proposition 3.1. *Suppose x_* is an extremal, i.e., a solution to the Euler–Lagrange equation. Assume the strong Weierstrass condition holds near (x_*, x_*'). Then x_* is a strong local minimum.*

3.6. **Physics from the Hamilton–Jacobi equation.** We use the example of light propagating through a medium to illustrate the physical meaning of the Hamilton–Jacobi equation.

Let $\rho(t,x)$ denote the density (or refractive index) of a medium. Consider the Lagrangian

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

According to Fermat's principle, the path taken by a light ray between two points minimizes the travel time, which leads to the variational problem

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

A direct computation yields the Hamiltonian:

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

Substituting into the Hamilton–Jacobi equation, we obtain

$$\partial_t u + H(t,\nabla u,x) = 0 \quad \Rightarrow \quad \partial_t u = \sqrt{\rho^2 - |\nabla u|^2}.$$

Squaring both sides, we get the eikonal-type form:

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

The level sets

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

represent the wavefronts of the light. These are the surfaces reached by light at time t.

The associated 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 $x_*(t)$ satisfying $\dot{x} = \psi(t,x)$, the tangent vector to the lifted path in (t,x)-space is

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

Thus, the direction of light propagation is orthogonal to the level sets of $u(t,x)$—that is, light rays are perpendicular to the wavefronts.

See also further discussion in Arnold [1].

REFERENCES

[1] V. I. Arnol'd. *Mathematical methods of classical mechanics*, volume 60. Springer Science & Business Media, 2013.
[2] K.-C. Chang. *Lecture Notes on Calculus of Variations*, volume 6. World Scientific, 2016.
[3] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, 1997.