Chapter 6

Calculus of Variations

Real mathematics must be justified as art if it can be justified at all.

– G.H. Hardy

CALCULUS of variations is a branch of mathematical analysis that focuses on finding a function that maximizes or minimizes a given quantity. Instead of optimizing numbers or finite-dimensional vectors, the calculus of variations optimizes over functions. This field is closely related to physics, engineering, and modern optimization, primarily through the principle of least action in mechanics and variational methods in differential equations and machine learning.

6.1 Examples of Variational Problems

We first present several concrete examples of variational problems.

6.1.1 Brachistochrone

The problem was first proposed by Johann Bernoulli in 1696 and solved by Issac Newton shortly after. The Greek word *brachistochrone* means ”minimal time.”

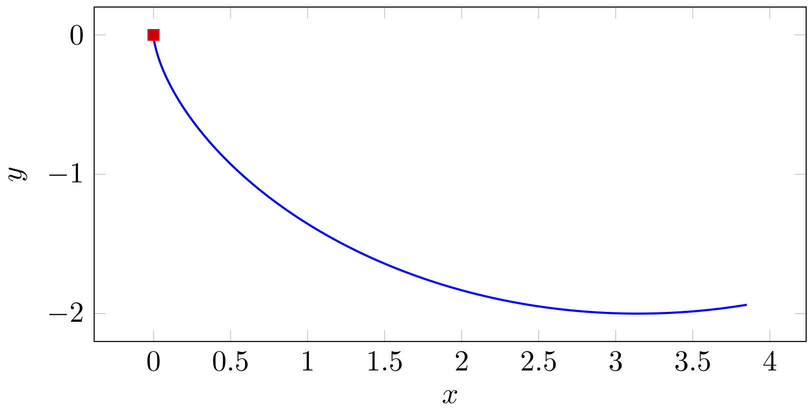

Figure 6.1: The brachistochrone curve.

Let a small bead stay at the origin $x = (0, 0)$ at time $t = 0$, and the ball begins to move downwards along a wire. The shape of the wire is described by the following function.

$$y = u(x), \quad y \leq 0.$$

The terminal point is assumed to be $y = (a, b)$, such that $a > 0$ and $b < 0$. Let $v(x)$ denote the speed of the bead when it reaches the point $(x, u(x))$. The total travel time of the bead is

$$T[u] = \int_0^L \frac{ds}{v} = \int_0^a \frac{\sqrt{1 + |u'(x)|^2}}{v(x)} dx$$

The speed $v(x) = \sqrt{-2gu(x)}$ is from the kinetic energy, then we need to find u such that $T[u]$ is minimized:

$$T[u] = \int_0^a \frac{\sqrt{1 + |u'(x)|^2}}{\sqrt{-2gu(x)}} dx, \quad u(0) = 0, \quad u(a) = b.$$

The variational problem is to find a function u that minimizes $T[u]$ among all reasonable functions. Note that we are not seeking an arbitrary function with the boundary constraints. A reasonable function should be at least C^1 or piecewise C^1 such that the arc length is well defined.

6.1.2 Shortest Distance

The shortest-distance problem arises in geometric optics, where a light ray travels through a medium along a path that minimizes travel time, a principle known as Fermat's principle. In an inhomogeneous medium in $\mathbb{R}^2$, the light speed $c(x, y)$ depends on the location $(x, y) \in \mathbb{R}^2$. If the light travels along a curve $y = u(x)$, then

$$c(x, u(x)) = \frac{ds}{dt} = \sqrt{1 + |u'(x)|^2}\frac{dx}{dt},$$

the travel time between two points (a, A) and (b, B) is

$$T[u] = \int_0^T dt = \int_a^b \frac{dt}{dx}dx = \int_a^b \frac{\sqrt{1 + |u'(x)|^2}}{c(x, u(x))}dx \quad u(a) = A, \quad u(b) = B.$$

6.1.3 Minimal Surface

The minimal surface problem generalizes the shortest-distance problem. Let $\Gamma \subset \mathbb{R}^3$ be a simple closed curve. The minimal surface is the surface of least total area among all surfaces whose boundary is Γ.

For simplicity, we allow Γ's projection onto the xy plane as a simple closed loop. Let the surface be $z = u(x, y)$, and the curve Γ is denoted by $z = h(x, y)$, then the surface area is

$$J[u] = \iint_\Omega \sqrt{1 + u_x^2 + u_y^2}\, dxdy,$$

where $u(x, y) = g(x, y)$ for $(x, y) \in \partial\Omega$.

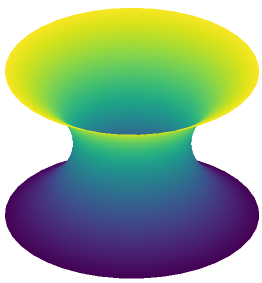

Figure 6.2: Minimal surface with revolutionary symmetry.

In particular, if the surface is rotationally symmetric, the problem can be simplified. The surface is determined by a curve $y = u(x)$, connecting two points

(a, A) and (b, B), and the area becomes

$$J[u] = 2\pi \int_a^b |u|\sqrt{1 + |u'(x)|^2}dx, \quad u(a) = A, \quad u(b) = B.$$

6.1.4 Isoperimetric Problem

We consider a simple version of the isoperimetric problem. Let Γ be a plane curve of a fixed length l connecting points $(0, 0)$ and another point on the x-axis. The objective is to find the domain Ω under the curve that maximizes the area. Let the curve be parameterized by its arc length $(x(s), y(s))$, $0 \le s \le l$. The area

$$J[y] = \int_\Gamma y dx = \int_0^l y(s)x'(s)ds = \int_0^l y(s)\sqrt{1 - |y'(s)|^2}ds,$$

where $y(0) = 0$ and $y(l) = 0$.

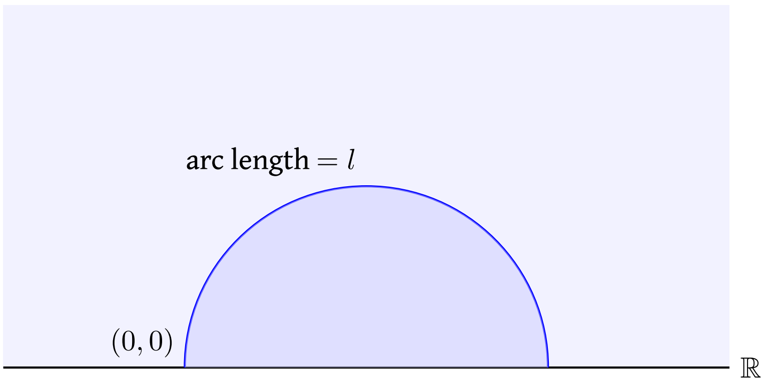

Figure 6.3: Isoperimetric curve.

6.2 The Euler–Lagrange Equation

In the examples above, the problems are finally reduced to a minimization in the following form:

$$J[u] = \int_a^b L(x, u, u')dx \tag{6.1}$$

with boundary constraits $u(a) = A$ and $u(b) = B$. The integrand L is known as the *Lagrangian*, and suppose the Lagrangian $L(x, u, p)$ is sufficiently smooth in all variables.

6.2.1 First Variation

The local minimizers of a sufficiently regular objective function must be critical points. Analogously, the local minimizers of $J[u]$ should satisfy a similar condition. Thus, we need to define the "gradient" of the functional J.

$$\langle \nabla J[u], \phi \rangle = \frac{d}{d\varepsilon} J[u + \varepsilon\phi] \bigg|_{\varepsilon=0}.$$

The function ϕ can be viewed as the "direction" for the directional derivative, but in infinite dimensions. Usually, we denote $\phi := \delta u$, called a *variation* of the function u. The derivative, denoted by δJ, is known as the *variational derivative*.

Next, we compute the variational derivative for (6.1).

$$\begin{aligned}
\delta J &= \frac{d}{d\varepsilon} \int_a^b L(x, u + \varepsilon \delta u, u' + \varepsilon \delta u') dx \bigg|_{\varepsilon=0} \\
&= \int_a^b \frac{d}{d\varepsilon} L(x, u + \varepsilon \delta u, u' + \varepsilon \delta u') dx \bigg|_{\varepsilon=0} \\
&= \int_a^b \left[\delta u \cdot L_u(x, u + \varepsilon \delta u, u' + \varepsilon \delta u') + \delta u' \cdot L_p(x, u + \varepsilon \delta u, u' + \varepsilon \delta u') \right] dx \bigg|_{\varepsilon=0} \\
&= \int_a^b \left[\delta u \cdot L_u(x, u, u') + \delta u' \cdot L_p(x, u, u') \right] dx.
\end{aligned}$$

The above is known as the *first variation* of $J[u]$. The derivative $\delta u'$ should be understood in the weak form since we do not require δu to be differentiable. The constraint on u implies that δu vanishes at the two boundary points. Therefore,

$$\int_a^b \delta u' \cdot L_p(x, u, u') dx = \delta u L_p(x, u, u') \bigg|_a^b - \int_a^b \delta u L_{x,p}(x, u, u') dx.$$

Thus,

$$\delta J = \int_a^b \delta u \left[L_u(x, u, u') - \frac{d}{dx} L_p(x, u, u') \right] dx = 0.$$

for all possible δu.

Lemma 6.2.1 (Fundamental Lemma). *If $f \in C[a, b]$ and*

$$\int_a^b f(x) v(x) dx = 0,$$

Continue on next page ⇓

Continue from last page ⇑

for every $v(x) \in C_c^\infty(a,b)$ *then* $f(x) \equiv 0$ *on* $[a,b]$.

Proof. If we only require $v \in C[a,b]$, then taking $v = f$ suffices. If $v \in C_c^\infty$, then if $u \neq 0$ at any interior point, we can find a local region where $u > 0$ (or < 0) and construct a compactly supported bump function for v. The boundary values are enforced through continuity. □

Therefore, we must have the *functional derivative*:

$$\frac{\delta J}{\delta u} = L_u(x,u,u') - \frac{d}{dx} L_p(x,u,u') = 0.$$

The "reasonable" solutions $u(x)$ of the above equation with prescribed boundary conditions are the critical functions. The equation is known as the *Euler-Lagrange* equation. Note that the local minimizers must be critical functions, but the converse is not valid. The above analysis is summarized in the following theorem.

Theorem 6.2.2. *Suppose* $L(x,u,p) \in C^2$*, then any* C^2 *local minimizer* $u(x)$ *to the functional* $J[u]$ *in* (6.1) *satisfies the Euler-Lagrange equation.*

However, the minimizer of $J[u]$ may not be a C^2 function (or even C^1). In addition, we usually require $L(x,u,p)$ to be convex in (u,p). Otherwise, the local critical function may be a local maximizer.

Theorem 6.2.3. *Suppose* $u(x) \in C^1$ *is a solution to the Euler-Lagrange equation and* $L(x,u,p)$ *is (strictly) convex in* (u,p)*, then* u *is a (unique) minimizer of* $J[u]$.

Proof. The convexity implies

$$L(x,\phi,\phi') \geq L(x,u,u') + L_u(x,u,u')(\phi - u) + L_p(x,u,u')(\phi' - u').$$

Therefore,

$$J[\phi] \geq J[u] + \int_a^b \left[L_u(x,u,u')(\phi - u) + L_p(x,u,u')(\phi' - u')\right] dx.$$

Using integration by parts and the boundary constraints,

$$J[\phi] \geq J[u] + \int_a^b \left[L_u(x,u,u') - \frac{d}{dx} L_p(x,u,u')\right] (\phi - u) dx.$$

Thus, u is a minimizer. Suppose L is strictly convex in (u,p), and permits two distinct minimizers $u \neq v$, then

$$\frac{1}{2}\left(L(x,u,u') + L(x,v,v')\right) > L(x, \frac{u+v}{2}, \frac{u'+v'}{2}).$$

Contradiction. □

Remark 6.2.4. *If the regularity assumption is weakened, local minimizers may not satisfy the Euler-Lagrange equation or its weak form; see Ball and Mizel (1987) for examples.*

As a special case of Theorem 6.2.2, if the Lagrangian does not depend on x, then the Euler-Lagrange equation can be solved easily (without writing down the equation).

Theorem 6.2.5. *If $L(x,u,p) = L(u,p)$, then*

$$H(u,p) = L(u,p) - pL_p(u,p)$$

is a first integral of the equation, that is, $H(u,u') \equiv C$ is a solution.

Proof. Simply differentiate $H(u,u')$,

$$\frac{d}{dx}H(u,u') = \frac{d}{dx}(L(u,u') - u'L_p(u,u')) = u'(L_u - \frac{d}{dx}L_p) = 0.$$

□

If $L(x,u,p) = L(p)$, then the Euler-Lagrange equation becomes $\frac{d}{dx}L_p = 0$, that is, $L(p) \equiv C$ is a constant. Note that if $u(x)$ is a linear function with the prescribed boundary condition, $p = u'$ is a constant, and satisfies the Euler-Lagrange equation. When $L(p)$ is convex in p, using the previous Theorem 6.2.3, we can show that the linear function is a minimizer (if strictly convex, it is the unique solution). However, if $L(x,u,p) = L(p)$ is non-convex, then the reasonable minimizer does not exist in general, as shown in the next example.

Example 6.2.6. *Let $L(p) = e^{-p^2}$, and the boundary constraints are $u(0) = u(1) = 0$. By choosing*

$$u_n(x) = n(x - \frac{1}{2})^2 - \frac{n}{4}$$

The variation

$$\int_0^1 e^{-u_n'(x)^2}dx = \frac{1}{2n}\int_{-n}^{n} e^{-t^2}dt \to 0, \quad n \to \infty.$$

However, it is impossible to find a function $u \in C^1$ or even a piecewise C^1 function to make the variation equal to zero.

Example 6.2.7. *Let $L(p) = (p^2 - 1)^2$, and the boundary constraints are $u(0) = u(1) = 0$. It is not hard to see*

$$u_{\textit{piece}}(x) = \begin{cases} x, & x \in [0, \frac{1}{2}] \\ 1 - x, & x \in [\frac{1}{2}, 1] \end{cases}$$

is a minimizer, the minimum is zero. However, by taking a smooth version of $u_{\textit{piece}}$ near $x = \frac{1}{2}$, we can see the infimum over C^1 functions is also zero. However, it is impossible to find a C^1 function to make that happen.

Therefore, it is common to consider the variational problem among a slightly larger class of functions that are piecewise C^1 but globally continuous. On each piece, the Euler-Lagrange equation still holds, but we also need (why?) the continuity of both L_p and $L - pL_p$.

The above situation is related to the *Lavrentiev phenomenon*: the minimizer exists among *absolutely continuous* functions (denoted by **AC**) but not among Lipschitz continuous functions (denoted by **Lip**), see Dacorogna (2008).

Example 6.2.8. *Let $L(x, u, p) = (u^3 - x)^2 p^6$ with boundary constraints $u(0) = 0$ and $u(1) = 1$. It is shown in Mania (1934) that*

$$\inf_{u \in \textbf{\textit{AC}}^*[0,1]} \int_0^1 L(x, u, u')dx = 0 < \inf_{u \in \textbf{\textit{Lip}}^*[0,1]} \int_0^1 L(x, u, u')dx.$$

*The $*$ in the notations **AC**$^*[0, 1]$ and **Lip**$^*[0, 1]$ means boundary constraints are applied. The left side has a minimizer $u^*(x) = x^{1/3}$. On the right side, let $f(x) = \frac{x^{1/3}}{2}$, then there exists a real number $a \in (0, 1)$ such that $f(x) \geq u(x)$ on $[0, a]$ and $u(a) = f(a)$. Note that the number $a > 0$ exists because $u \in$ **Lip**$^*[0, 1]$. Therefore, on $[0, a]$ we have*

$$(u^3 - x)^2 p^6 \geq (f(x)^3 - x)^2 p^6 = \frac{7^2}{8^2} x^2 p^6.$$

By Hölder inequality,

$$\left(\int_0^a x^2 |u'(x)|^6 dx\right)^{1/6} \left(\int_0^a x^{-2/5}\right)^{5/6} \geq \int_0^a |u'(x)| dx \geq \frac{a^{1/3}}{2}.$$

Thus,

$$\begin{aligned}\int_0^1 L(x,u,u')dx &\geq \frac{7^2}{8^2}\int_0^a x^2|u'(x)|^6dx \int_0^a x^2|u'(x)|^6dx \\ &\geq \frac{7^2}{8^2}\left(\frac{a^{1/3}}{2}\right)^6\left(\frac{5}{3}a^{\frac{3}{5}}\right)^{-5} = \frac{1}{a}\cdot\frac{7^2 3^5}{8^2 2^6 5^5} \geq \frac{7^2 3^5}{8^2 2^6 5^5}.\end{aligned}$$

*In other words, it is impossible to construct a sequence of $u \in$ **Lip**$^*[0,1]$ to approximate $u^*(x) = x^{1/3}$ while minimizing the variational problem. This phenomenon is essentially different from the previous examples.*

*In fact, if we construct a sequence $u_n \in$ **Lip**$^*[0,1]$ to approximate u^*, then we define a_n be the real number that $f(a) = u_n(a)$ and $f(x) \geq u_n(x)$ on $[0,a_n]$, then it is necessary to have $a_n \to 0$ (at least subsequence), while the integral*

$$\int_0^1 L(x,u_n,u_n')dx \geq \frac{1}{a_n}\frac{7^2 3^5}{8^2 2^6 5^5} \to \infty.$$

6.2.2 Revisiting Examples

Using the Euler-Lagrange equation, we revisit the previously mentioned variational problems.

Brachistochrone

The Lagrangian does not depend on x, then the solution satisfies (Theorem 6.2.5):

$$u(x)(1+|u'(x)|^2) = C < 0, \quad u(0) = 0.$$

Solving the equation $u'(x) = -\sqrt{\frac{C-u}{u}}$, we obtain the parameterized solution

$$\begin{aligned} x &= -\frac{1}{2}C(\theta - \sin\theta) \\ u(x) &= \frac{1}{2}C(1-\cos\theta).\end{aligned}$$

The constraint $u(a) = b$ determines the unique constant C.

Fermat Principle

The Lagrangian is $L(x,u,p) = \sqrt{1+u'^2}/c(x,u)$. The Euler-Lagrange equation is

$$\frac{d}{dx}\left(\frac{1}{c(x,u)}\frac{u'}{\sqrt{1+u'^2}}\right) = \partial_u\left(\frac{1}{c(x,u)}\right)\sqrt{1+u'^2}.$$

That is (why?),

$$\left[\partial_x\left(\frac{1}{c(x,u)}\right)u' - \partial_u\left(\frac{1}{c(x,u)}\right)\right](1+u'^2) + \frac{1}{c(x,u)}u'' = 0.$$

Minimal Surface

Using Theorem 6.2.5, we have the solution u satisfies

$$\frac{u}{\sqrt{1+u'^2}} = C \geq 0.$$

Therefore, $Cu'(x) = \sqrt{u^2 - C^2}$. The solution is

$$u(x) = C\cosh(\frac{x+C'}{C}).$$

This shape is known as a catenary, which is quite similar to a parabola. This prob-

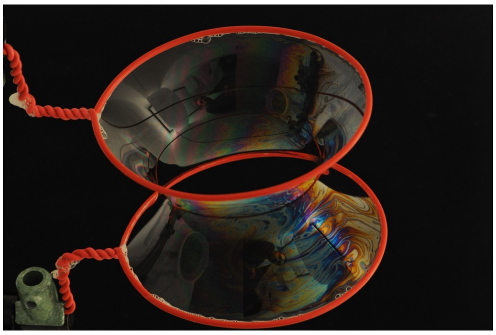

Figure 6.4: Soap film between two circles.

lem is closely related to the shape of a soap film (without gravity). The film's shape will minimize its potential energy due to surface tension. The energy is proportional to the surface area. See Figure 6.4 for an example of such a surface.

Remark 6.2.9. *As one inspiring example, when the boundary condition is $u(-L) = A$ and $u(L) = A$, then $C' = 0$ and we observe that*

$$\frac{A}{L} = \frac{C}{L}\cosh\frac{L}{C}.$$

The function $h(x) = \cosh(x)/x$ attains the minimum at $x \approx 1.2$, see Figure 6.5. So we cannot find a solution that satisfies the Euler-Lagrange equation in this special example if $L \gg A$. In this case, it means we cannot have a smooth surface that satisfies the prescribed boundary conditions. Physically, the surface will break into two disks if they are

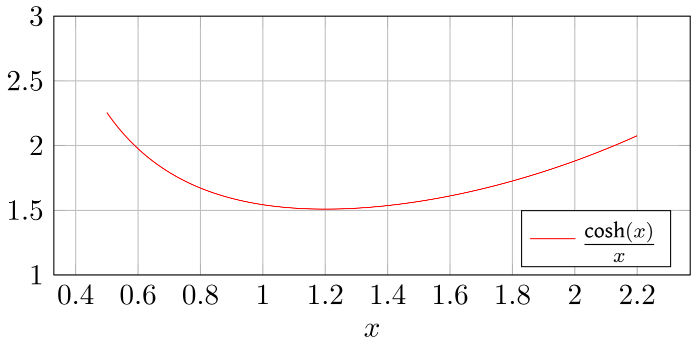

Figure 6.5: Plot of function $h(x) = \cosh(x)/x$ on $x > 0$.

far apart. Geometrically, the solution is not a smooth curve but rather two vertical lines at the endpoints and a horizontal line along the axis connecting them.

When $\frac{A}{L}$ is above the minimum value of h, there exist two solutions to the Euler-Lagrange equation with the given boundary conditions. The surface area is proportional to

$$\frac{1}{L^2}\int_{-L}^{L} u\sqrt{1+u'^2}dx = \frac{1}{L^2}\int_{-L}^{L} C\cos(\frac{x}{C})^2 dx = \frac{C^2}{L^2}\left(\frac{1}{2}\sinh(\frac{2L}{C}) + \frac{L}{C}\right).$$

The curve with a larger C has a smaller surface area (why?).

6.2.3 Multivariable Case

The Euler-Lagrange equation can be generalized to multi-dimensions. Consider the variational problem

$$\inf_{u\in X} J[u], \quad J[u] := \int_\Omega L(x, u, \nabla u)dx,$$

where $X = \{u \in C^1(\Omega) \mid u|_{\partial\Omega} = h(x)\}$. If u is a C^2 minimizer, then

$$\begin{aligned}\frac{\delta J}{\delta u}[\phi] &= \int_\Omega \left[\partial_u L(x, u, \nabla u) + \sum_i \partial_{p_i} L(x, u, \nabla u)\partial_{x_i}\phi\right] dx \\ &= \int_\Omega [\partial_u L(x, u, \nabla u) + \nabla_x \cdot [\nabla_p L(x, u, \nabla u)\phi] - \phi\nabla_x \cdot \nabla_p L(x, u, \nabla u)]\, dx \\ &= \int_\Omega [\partial_u L(x, u, \nabla u) - \nabla_x \cdot \nabla_p L(x, u, \nabla u)]\, \phi dx.\end{aligned}$$

The Euler-Lagrange equation becomes

$$\partial_u L(x, u, \nabla u) - \nabla_x \cdot \nabla_p L(x, u, \nabla u) = 0. \tag{6.2}$$

We can reformulate the same theory for the multi-dimensional case as Theorem 6.2.2 and Theorem 6.2.3. The proof is omitted here.

Theorem 6.2.10. *Let $L(x, u, p) \in C^2$ and consider the variational problem*

$$\inf_{u \in X} \int_\Omega L(x, u, \nabla u) dx,$$

where $X = \{u \in C^1 \mid u|_{\partial\Omega} = h(x)\}$.

1. *If $u \in C^2$ is a minimizer (or any critical function), then it satisfies the Euler-Lagrange equation (6.2).*
2. *If $u \in X$ satisfies (6.2) and L is convex in (u, p), then u is a minimier.*
3. *If $u \in X$ satisfies (6.2) and L is strictly convex in (u, p), then u is the unique minimier.*

Example 6.2.11. *For instance, the minimal surface problem*

$$J[u] = \int_\Omega \sqrt{1 + |\nabla_x u|^2} dx.$$

The Euler-Lagrange equation is

$$\nabla_x \cdot \frac{\nabla_x u}{\sqrt{1 + |\nabla_x u|^2}} = 0.$$

The left-hand side is known as the surface's mean curvature. The equation is equivalent to

$$(1 + u_x^2)u_{yy} - 2u_x u_y u_{xy} + (1 + u_y^2)u_{xx} = 0.$$

Example 6.2.12. *Consider the Dirichlet minimization problem*

$$J[u] = \int_\Omega \frac{1}{2} |\nabla u|^2 dx$$

The Euler-Lagrange equation is

$$\nabla_x \cdot \nabla_x u = \Delta u = 0.$$

As we can see later, the solution to the Laplace equation (if it exists) is the unique minimizer to the variational problem.

6.2.4 Second Variation

The first variation provides a *necessary* condition for a minimizer. As an analogue in finite-dimensional minimization problems, it is possible to find a *sufficient* condition for a minimizer using second-derivative information. Let the variational problem be

$$J[u] := \int_a^b L(x, u, u')dx.$$

Then, by Taylor expansion, if $L \in C^3$,

$$\begin{aligned} J[u + \varepsilon\phi] - J[u] = \varepsilon \int_a^b (L_u\phi + L_p\phi')\, dx \\ + \frac{\varepsilon^2}{2} \int_a^b \left(L_{uu}\phi^2 + 2L_{up}\phi\phi' + L_{pp}\phi'^2\right) dx + o(\varepsilon^2). \end{aligned}$$

The partial derivatives are evaluated at (x, u, u'). Let u be a critical function (Euler-Lagrange), we can rewrite

$$\int_a^b \left(L_{uu}\phi^2 + 2L_{up}\phi\phi' + L_{pp}\phi'^2\right) dx = \int_a^b \left(Q(x, u, u')\phi^2 + P(x, u, u')\phi'^2\right) dx,$$

where $Q(x, u, u') = L_{uu} - \frac{d}{dx}L_{up}$ and $P(x, u, u') = L_{pp}(x, u, u')$. A sufficient condition to make a critical function u a local minimizer is that

$$\int_a^b (Q\phi^2 + P\phi'^2)dx > 0$$

for all nonzero $\phi \in C^1[a, b]$ with $\phi(a) = \phi(b) = 0$.

Theorem 6.2.13. *A necessary condition for u to be a local minimizer is*

$$P(x, u, u') \geq 0$$

for each $x \in [a, b]$.

Proof. Prove by construction. For each point $c \in (a, b)$, we take a bump function

$\phi = \varepsilon^{1/2}\psi(\frac{x-c}{\varepsilon})$, where $\psi \in C_c^\infty[-1,1]$.

$$\begin{aligned}\int_a^b Q\psi^2 + P\phi'^2 dx &= \int_{c-\varepsilon}^{c+\varepsilon} \varepsilon Q\psi\left(\frac{x-c}{\varepsilon}\right)^2 + \frac{P}{\varepsilon}\psi'\left(\frac{x-c}{\varepsilon}\right)^2 dx \\ &= \int_{-1}^{1} \varepsilon^2 Q(c+\varepsilon r)\psi(r)^2 + P(c+\varepsilon r)\psi'(r)^2 dr \\ &\to P(c)\int_{-1}^{1}\psi'(r)^2 dr, \quad \varepsilon \to 0.\end{aligned}$$

□

Definition 6.2.14. *Define $c \in (a,b)$ as a conjugate point of a if the differential equation*

$$-\frac{d}{dx}(Ph') + Qh = 0 \tag{6.3}$$

permits a nontrivial solution h that $h(a) = h(c) = 0$.

If the point a does not have any conjugate points (e.g., when $P > 0$ and $Q \geq 0$), then it means the equation (6.3) with boundary values $h(a) = h(c) = 0$ for any $c \in (a,b)$ forces $h \equiv 0$.

Theorem 6.2.15. *If $P > 0$ and a does not have any conjugate points on (a,b), then*

$$\int_a^b (Q\phi^2 + P\phi'^2)dx > 0$$

for all nonzero $\phi \in C^1[a,b]$ with $\phi(a) = \phi(b) = 0$.

Proof. Since $P > 0$, we know that if ϕ is not trivial, then

$$\int_a^b P(\phi' + \frac{w}{P}\phi)^2 dx > 0 \tag{6.4}$$

for any function w. Otherwise $\phi' + \frac{w}{P}\phi = 0$ forces $\phi \equiv 0$ from the boundary condition. The difference between this formulation and the desired one is

$$(\frac{w^2}{P} - Q)\phi^2 + w\frac{d}{dx}\phi^2.$$

The integral of this part equals zero if

$$(\frac{w^2}{P} - Q) = w' \Rightarrow w' - \frac{w^2}{P} = -Q.$$

This is the Riccati equation. If we take a change of variable

$$\frac{\psi'}{\psi} = -\frac{w}{P},$$

it becomes $-(P\psi')' + Q\psi = 0$. The non-existence of conjugate points implies that we can find a solution $\psi > 0$ on $[a, b]$ through the continuous dependence of the solution on the initial condition (why?). Then w is well-defined, and leads to (6.4). □

Remark 6.2.16. *The theorem slightly generalizes the convexity requirement on L. The essential requirement is $P = L_{pp} \geq 0$, but this condition alone is not sufficient. In fact, Theorem 6.2.15's argument should be if and only if.*

6.3 Legendre Transform

Recall the variational problem that minimizes

$$J[u] := \int_a^b L(x, u, u')dx.$$

It can be reduced to the Euler-Lagrange equation if the solution is sufficiently regular. In this section, we introduce the Legendre transform to prepare the reformulation of the variational problem in the next section.

Definition 6.3.1. *Let $f : I \to \mathbb{R}$ and we define the **Legendre transform** of f by*

$$f^*(p) := \sup_{x \in I}(xp - f(x))$$

For each $p \in I^*$, the supremum is taken over all $x \in I$. The resulting function f^* (if exists) is defined on $I^* \to \mathbb{R}$, where

$$I^* = \{p \in \mathbb{R} \mid f^*(p) < \infty\}.$$

The Legendre transform of f is called the convex conjugate of f. The definition of the Legendre transform can be easily extended to higher dimensions.

Definition 6.3.2. *In higher dimensions, let $\Omega \subset \mathbb{R}^n$, and $f : \Omega \to \mathbb{R}$. We define the* ***Legendre transform*** *of f by*

$$f^*(p) := \sup_{x\in\Omega}(\langle x, p\rangle - f(x))$$

Example 6.3.3. *If f is a convex function on $x \in I$, then*

$$f^*(p) = \sup_{x\in I}(xp - f(x)) = (xp - f(x))\Big|_{x=(f')^{-1}(p)}.$$

Note that f' is monotone (since $f'' > 0$), the set $\{(f')^{-1}(p)\}$ is a singleton. Geometrically, we can view f^* as the negative of the y-intercept of the tangent line to the graph of f with slope p.

Theorem 6.3.4. *If $f \geq g$, then $f^* \leq g^*$.*

Proof. $f^*(p) = \sup_{x\in I} xp - f(x) \leq \sup_{x\in I} xp - g(x) = g^*(p)$. □

Theorem 6.3.5. *If $f \in C^2(I)$ is convex, then f^* is also convex.*

Proof. Let $p = f'(x_0)$, we define its inverse $x_0 = g(p)$. Then,

$$f^*(p) = g(p)p - f(g(p)) \Rightarrow \frac{d}{dp}f^*(p) = g'(p)p + g(p) + f'(g(p))g'(p) = g(p).$$

The second derivative $\frac{d^2}{dp^2}f^*(p) = g'(p) = \frac{1}{f''(x_0)} \geq 0$. □

Remark 6.3.6. *If $f \in C(I)$ is convex but not smooth, the convolution product of a non-negative bump function $\frac{1}{\varepsilon}K(\frac{x}{\varepsilon})$ with f makes the function back to $C^2(I)$ and the convexity is preserved (why?), and then take $\varepsilon \to 0$. In the following, we assume convex functions are smooth.*

Remark 6.3.7. *In fact, if f is not convex, f^* is still convex, thus f^* is also called Fenchel's* *convex conjugate. Because*

$$\begin{aligned}\alpha f^*(p_1) + (1-\alpha) f^*(p_2) &= \alpha \sup_{x\in\Omega}(\langle p_1, x\rangle - f(x)) + (1-\alpha)\sup_{y\in\Omega}(\langle p_2, y\rangle - f(y)) \\ &\geq \sup_{x\in\Omega}(\alpha(\langle p_1, x\rangle - f(x)) + (1-\alpha)(\langle p_2, x\rangle - f(x))) \\ &= f^*(\alpha p_1 + (1-\alpha)p_2).\end{aligned}$$

Therefore, f^{**} is always a convex function in the x variable. It is the closed convex hull of the epi-graph of f, that is, the largest convex function below f.

Theorem 6.3.8. *If f is convex, then the Legendre transform is an involution, that is, $f^{**} = f$.*

Proof. Let $f^*(p) = pg(p) - f(g(p))$, and $\frac{d}{dp} f^*(p) = g(p)$. Similarly,

$$f^{**}(y) = p_0 y - f^*(p_0), \quad y = \frac{d}{dp} f^*(p_0) = g(p_0).$$

Therefore,

$$\begin{aligned}f^{**}(y) &= p_0 g(p_0) - f^*(p_0) \\ &= p_0 g(p_0) - (p_0 g(p_0) - f(g(p_0))) = f(g(p_0)) = f(y).\end{aligned}$$

□

Remark 6.3.9. *In fact, an involution on the class of lower semi-continuous convex functions with order-reversing must be the Legendre transform up to a linear transform. A proof found in Artstein-Avidan and Milman (2009).*

The following theorems are straightforward.

Theorem 6.3.10 (Young's Inequality). *$f^*(p) + f(x) \geq xp$.*

Proof. By definition. □

Theorem 6.3.11. *$f^{**}(x) \leq f(x)$.*

Proof. This is also by definition. From Young's inequality, $f(x) \geq xp - f^*(p)$, thus $f(x) \geq \sup_{p\in I^*}(xp - f^*(p)) = f^{**}(x)$. □

Example 6.3.12. *Let $f(x) = \frac{1}{q}x^q$ for $q \geq 1$, we compute $f^*(p) = \frac{1}{q'}p^{q'}$, where $\frac{1}{q} + \frac{1}{q'} = 1$. The previous theorem reduces to the standard Young's inequality*

$$\frac{1}{q}x^q + \frac{1}{q'}p^{q'} \geq xp, \quad x, p > 0.$$

Example 6.3.13. *Let $f(x) = e^x$, then $f^*(p) = p \ln p - p$. Then the Young's inequality becomes*

$$e^x + p \ln p - p \geq px.$$

Especially, if we take $x_i = \log q_i$ and $\sum_i q_i = \sum_i p_i = 1$ (discrete probability distribution), then $q_i + p_i \ln p_i - p_i \geq p_i \log q_i$. Taking the sum over i leads to the classical Gibbs' inequality (and proves the Kullback-Leibler (KL) divergence's nonnegativity):

$$\sum_i p_i \ln p_i \geq \sum_i p_i \ln q_i\,.$$

Theorem 6.3.14 (Fenchel's duality theorem)**.** *Let $f : I \to \mathbb{R}$ be a continuous **convex** function and $g : I \to \mathbb{R}$ be a continuous **concave** function. Then*

$$\mu := \inf_{x \in I}(f(x) - g(x)) = \sup_{p \in I^*}(g^*(p) - f^*(p)),$$

where g^* is defined as follows (known as concave conjugate).

$$g^*(p) = \inf_{x \in I}(xp - g(x)).$$

Proof. First, for all $x \in I$ and $p \in I^*$, the definition of Legendre transform implies

$$f(x) + f^*(p) \geq xp,$$
$$g(x) + g^*(p) \leq xp.$$

Thus, for all possible x, p,

$$f(x) - g(x) \geq g^*(p) - f^*(p).$$

Therefore,

$$\mu := \inf_{x \in I} f(x) - g(x) \geq \sup_{p \in I^*} g^*(p) - f^*(p).$$

Let $K_f :=$ epi$(f - \mu)$ and $K_g :=$ hypo(g). Both sets are closed and convex such that $K_f \cap$ int$(K_g) = \emptyset$, there exists a hyperplane with slope p_0 exactly separating

K_f and K_g. Hence,

$$\inf_{x\in I} xp_0 - g(x) = q_0 = g^*(p_0)$$
$$\sup_{x\in I} xp_0 - f(x) + \mu = q_0 = f^*(p_0) + \mu.$$

Thus, $\mu = g^*(p_0) - f^*(p_0)$. □

Example 6.3.15. *In empirical risk minimization problems, the function has the form*

$$F(w_1, \cdots, w_n) = \sum_{i=1}^{n} \phi_i(w_i) + \mathcal{R}(w).$$

If ϕ_i and $\mathcal{R}$ are convex, the minimization problem is equivalent to the maximization problem

$$G(\beta_1, \cdots, \beta_n) = -\sum_{i=1}^{n} \phi_i^*(\beta_i) - \mathcal{R}^*(-\beta).$$

For instance, if $\phi_i(w_i) = \max(0, 1 - w_i)$ is the hinge loss, then $\phi_i^*(\beta_i) = \beta_i$.

6.4 Hamiltonian Formulation

Definition 6.4.1. *Let $H(x, u, v)$ be defined as follows.*

$$H(x, u, v) := \sup_{p\in\mathbb{R}}\{vp - L(x, u, p)\}$$

where H is called ***Hamiltonian****, and the right-hand side is the Legendre transform of L.*

As we have seen previously, $L_{pp}(x, u, u') \geq 0$ is a necessary condition for the existence of a local minimizer u. When the Lagrangian L is convex in p, we can use the Legendre transform on the Hamiltonian H to recover L via the involution property.

The following lemma follows directly from the previous section. The superlinear growth guarantees that $\text{dom}(H) = \mathbb{R}$.

Lemma 6.4.2. *Let $f \in C^2$ such that $f'' > 0$ and*

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

Continue on next page ⇓

Continue from last page ⇑

Let $H(p) = \sup_{x\in\mathbb{R}}(px - f(x))$*. Then* $H \in C^2(\mathbb{R})$ *and*

$$H(p) = pH'(p) - f(H'(p)),$$

and $p = f'(x) \Leftrightarrow x = H'(p)$*.*

Remark 6.4.3. *Note that* f *is strictly convex does not imply* $f'' > 0$*. For instance* $f(x) = x^4$ *is strictly convex while* $f''(0) = 0$*.*

For a general Lagrangian, the previous lemma can be extended naturally.

Lemma 6.4.4. *If* $L(x, u, p) \in C^2(I \times \mathbb{R} \times \mathbb{R})$ *such that*

$$L_{pp}(x, u, p) > 0$$

and $L(x, u, p) \geq A(p) + B(x, u)$*, where* B *is continuous and* A *is non-negative such that*

$$\lim_{|p|\to\infty} \frac{A(p)}{|p|} = \infty.$$

Let $H(x, u, v)$ *be the Hamiltonian, then* $H \in C^2(I \times \mathbb{R} \times \mathbb{R})$ *and*

$$\begin{aligned} H_x(x, u, v) &= -L_x(x, u, H_v(x, u, v)), \\ H_u(x, u, v) &= -L_u(x, u, H_v(x, u, v)), \\ H(x, u, v) &= vH_v(x, u, v) - L(x, u, H_v(x, u, v)), \end{aligned}$$

and $v = L_p(x, u, p) \Leftrightarrow p = H_v(x, u, v)$*.*

Proof. We first show that $H(x, u, v)$ (by definition) is attained at a certain $p \neq \infty$. Otherwise, wecan find a maximizing sequence $|p_k| \to \infty$ for $H(x, u, v)$ such that $(\varepsilon_k \to 0)$

$$\begin{aligned} -L(x, u, 0) \leq H(x, u, v) &\leq \varepsilon_k + vp_k - L(x, u, p_k) \\ &\leq \varepsilon_k + vp_k - A(p_k) - B(x, u) \\ &\leq \varepsilon_k + |p_k| \left(v\frac{p_k}{|p_k|} - \frac{A(p_k)}{|p_k|} \right) - B(x, u). \end{aligned}$$

The right-hand side must diverge to $-\infty$ as $k \to \infty$, a contradiction with the left-hand side. There exists $p = p(x, u, v)$ such that

$$H(x, u, v) = vp - L(x, u, p), \quad v = L_p(x, u, p).$$

The equations for H_x, H_u, and H_v are straightforward. Meanwhile, $p(x, u, v)$ is bounded if (u, v) are bounded. Since $L \in C^2$, we have $p \in C^1$ from the inverse function theorem. Thus, $H \in C^2$. □

We now state the Hamiltonian formulation of the variational problem. Compared with the Lagrangian formulation, the Hamiltonian formulation is more symmetric and has a strong connection to symplectic geometry.

Theorem 6.4.5. *Let H and L be the Hamiltonian and Lagrangian in Lemma 6.4.4. Let $u, v \in C^2(I)$ such that*

$$\begin{aligned} u'(x) &= H_v(x, u(x), v(x)), \\ v'(x) &= -H_u(x, u(x), v(x)). \end{aligned} \tag{6.5}$$

This system is called the **Hamiltonian system**. *Then $u(x)$ satisfies the Euler-Lagrange equation*

$$L_u(x, u, u') - \frac{d}{dx} L_p(x, u, u') = 0.$$

Conversely, if $u \in C^2(I)$ solves the Euler-Lagrange equation, then (u, v) are the solutions to the system (6.5), where

$$v(x) = L_p(x, u(x), u'(x)).$$

Proof. If (u, v) are solutions to (6.5), then $u'(x) = p(x, u(x), v(x))$, and $v(x) = L_p(x, u(x), u'(x))$. We also have

$$v'(x) = -H_u(x, u(x), v(x)) = L_u(x, u(x), u'(x)).$$

Thus,

$$\frac{d}{dx} v(x) = \frac{d}{dx} L_p(x, u, u') = L_u(x, u, u').$$

The converse part is almost the same. □

Corollary 6.4.6. *Let H be the Hamiltonian in Lemma 6.4.4, and u, v be the solutions in (6.5), we have*

$$\frac{d}{dx} H(x, u(x), v(x)) = H_x(x, u(x), v(x)).$$

Proof. The computation is straightforward.

$$\begin{aligned}\frac{d}{dx}H(x,u,v) &= H_x(x,u,v) + H_u(x,u,v)u' + H_v(x,u,v)v' \\ &= H_x(x,u,v).\end{aligned}$$

□

It means that if the Lagrangian $L = L(u,p)$ is independent of x, then $H_x \equiv 0$ (from Lemma 6.4.4), and $H = H(u(x), v(x)) \equiv C$ is a constant along the trajectory $(u(x), v(x))$.

Example 6.4.7 (Hamiltonian Mechanics). *In mechanics, the Hamiltonian can be reviewed as the total energy. Let's start with Newton's Laws for a system of N particles, with masses and positions $(m_i, \mathbf{r}_i)$, $i = 1, 2, \cdots, N$. Denote $\mathbf{r} = (\mathbf{r}_1, \cdots, \mathbf{r}_N) \in \mathbb{R}^{3N}$. The kinetic energy is*

$$T(\mathbf{r}') = \frac{1}{2}\sum_{i=1}^{N} m_i |\mathbf{r}_i'(t)|^2.$$

The potential energy is denoted by a function $U(\mathbf{r})$, then the Hamiltonian (or the conserved total energy) is

$$H(\mathbf{r}, \mathbf{v}) = T(\mathbf{r}') + U(\mathbf{r}),$$

where $\mathbf{v} = \nabla_{\mathbf{p}} L(x, \mathbf{r}, \mathbf{r}')$. Therefore,

$$L(\mathbf{r}, \mathbf{r}') = \mathbf{v} \cdot \mathbf{r}' - H(\mathbf{r}, \mathbf{r}') = \mathbf{r}' \cdot \nabla_{\mathbf{p}} L(x, \mathbf{r}, \mathbf{r}') - T(\mathbf{r}') - U(\mathbf{r}).$$

Solving this equation, we obtain the Lagrangian

$$L(\mathbf{r}, \mathbf{r}') = T(\mathbf{r}') - U(\mathbf{r}) \Rightarrow H(\mathbf{r}, \mathbf{v}) = \frac{1}{2}\sum_{i=1}^{N} \frac{|\mathbf{v}_i|^2}{m_i} - U(\mathbf{r}).$$

The Euler-Lagrange equation is Newton's law.

$$-\nabla_{\mathbf{r}} U - \frac{d}{dt}\nabla_{\mathbf{p}} T(\mathbf{r}') = 0 \Rightarrow m_i \mathbf{r}_i''(t) = -\partial_{\mathbf{r}_i} U.$$

The Hamiltonian system is

$$\begin{aligned}\mathbf{r}_i' &= \frac{\mathbf{v}_i}{m_i}, \\ \mathbf{v}_i' &= -\partial_{\mathbf{r}_i} U.\end{aligned}$$

We can see the meaning of $\mathbf{v}_i$ is the momentum for the i-th particle.

We note that the Hamiltonian is independent of the Cartesian coordinates. Thus, (u, v) are also called generalized coordinates and momenta.

Example 6.4.8 (Pendulum)**.** *The simple pendulum's Lagrangian is (see below)*

$$L(u, u') = \frac{1}{2}m(Lu')^2 - (-mgL\cos(u)).$$

Therefore, the Euler-Lagrange equation is

$$mgL\sin(u) + mL^2u'' = 0 \Rightarrow u'' = -\frac{g}{L}\sin(u).$$

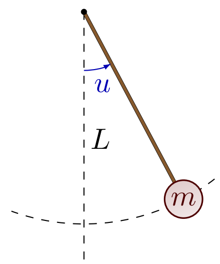

Figure 6.6: Pendulum

Remark 6.4.9. *The Hamiltonian system* (6.5) *is an important topic in dynamical systems. Lots of classical Hamiltonian systems exhibit chaotic behavior, for instance, three body system, double pendulum, etc.*

6.5 Hamilton-Jacobi Equation

The *Hamilton-Jacobi equation* provides an alternative to the Lagrangian and Hamiltonian formulations. Instead of treating the Hamiltonian system (and Lagrangian too) as trajectories of particles, the Hamilton-Jacobi equation treats the particles as *waves*.

Let the action or Hamilton's principal function:

$$S(t, u(t)) = \int_{t_0}^{t} L(x, u(x), u'(x))dx + \text{some constant},$$

and we try to derive the corresponding equation for S.

$$\frac{d}{dt}S(t, u(t)) = L(t, u(t), u'(t)) = \partial_t S(t, u) + \partial_u S(t, u) \cdot u'.$$

Taking partial derivative in u' (depending on the initial configuration, we assume $u(t)$ and $u'(t)$ can be made arbitrary), $L_p(t, u(t), u'(t)) = \partial_u S(t, u(t))$. Moreover,

$$H(x, u(x), v(x)) = v(x)u'(x) - L(x, u(x), u'(x)),$$

the generalized momentum $v(x) = L_p(x, u(x), u'(x)) = \partial_u S(x, u(x))$, and

$$\partial_t S(t, u) = L(t, u(t), u'(t)) - \partial_u S(t, u) \cdot u' = -H(t, u(t), \partial_u S(t, u(t)).$$

Therefore, it leads to the following Hamilton-Jacobi equation (we use x as the time variable).

Definition 6.5.1. *The Hamilton-Jacobi equation is*

$$\partial_x S(x, u) + H(x, u, \partial_u S) = 0.$$

From the above derivation, if $u(x)$ solves $u'(x) = H_v(x, u(x), S_u(x, u(x)))$, then let $v(x) = S_u(x, u(x))$, the functions $(u(x), v(x))$ will solve the Hamiltonian system.

Example 6.5.2. *Let the Hamiltonian $H(u, v) = \frac{1}{2}v^2 + V(u)$. Then the Hamilton-Jacobi equation is*

$$\partial_x S + \frac{1}{2}(S_u)^2 + V(u) = 0.$$

Since V does not depend on x, the above equation permits a solution $S(x, u) = S(u) - Cx$,

$$S_u = \sqrt{2(C - V(u))} \Rightarrow S(x, u) = \int_0^u \sqrt{2(C - V(s))} ds.$$

Then the dynamics satisfies

$$u'(x) = H_v(x, u(x), S_u(x, u(x))) = S_u(x, u(x)) = \sqrt{2(C - V(u(x)))}.$$

Example 6.5.3. *Let the Hamiltonian $H(x, \mathbf{u}, \mathbf{v}) = H(\mathbf{v}) = |\mathbf{v}|^2 - 1$, where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. This Hamiltonian describes the physical process in which the* ***magnitude*** *of momentum (or velocity) is a constant.*

The Hamiltonian-Jacobi equation is (we have reduced the x dependence, otherwise $S^*(x, \mathbf{u}) = S(\mathbf{u}) - Cx$ is a solution if $H(x, \nabla_{\mathbf{u}} S^*(x, \mathbf{u})) = C$)

$$H(\nabla_{\mathbf{u}} S(\mathbf{u})) = 0 \Rightarrow |\nabla_{\mathbf{u}} S(\mathbf{u})| = 1.$$

This equation is known as the eikonal equation. If we supply a boundary condition $S(\mathbf{u}) = 0$ on $\mathbf{u} \in \partial\Omega$, then

$$S(\mathbf{u}) = \text{dist}(\mathbf{u}, \partial\Omega)$$

is a solution. It describes the shortest distance from any point $\mathbf{u}$ to the boundary set $\partial\Omega$, or we call it the ***level set****.*

Remark 6.5.4. *The eikonal equation arises when solving the Helmholtz equation*

$$\Delta\phi + k^2\phi = 0.$$

Assuming the solution's form (WKB) $\phi(x) = e^{ikS(x)}$*, then*

$$k^2|\nabla S|^2 - ik\Delta S(x) = k^2$$

For $k \to \infty$*, it leads to* $|\nabla S|^2 = 1$*.*

6.6 Viscosity Solution

Theorem 6.6.1 (Hopf-Lax Formula)**.** *TODO*

6.7 Exercises

Problem 6.7.1. *Consider the variational problem*

$$\inf_{u\in X} \int_a^b L(x,u,p)dx$$

where $L \in C^2$ *and* $X = \{u \in C[a,b] \cap C^2[a,c] \cap C^2[c,b] \mid u(a) = A, u(b) = B\}$*,* $c \in (a,b)$ *is a given point. Find the necessary conditions* $u \in X$ *must satisfy to be a minimizer.*

Problem 6.7.2. *Consider the variational problem*

$$\inf_{u\in X} \int_0^1 x|u'(x)|^2 dx$$

where $X = \{u \in C^1[0,1] \mid u(0) = 1, u(0) = 0\}$*. Show that the minimizer does not exist in* X *and the infimum value of the variational problem is zero.*

Problem 6.7.3. *Show that the functional*

$$\mathcal{Q}[u] = \int_0^1 [u'^2 - u^2]dx$$

is non-negative if $u \in C^1[0,1]$ *with boundary condition* $u(0) = u(1) = 0$*.*

Problem 6.7.4. *Formulate a variational problem to find the shortest distance on a unit sphere between two points.*

Problem 6.7.5. *Derive the Euler-Lagrange equation for the following double pendulum example.*

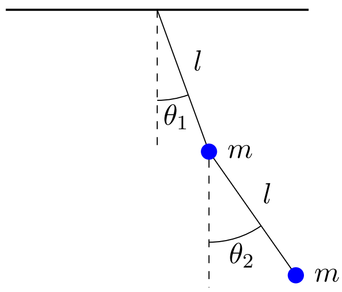

Figure 6.7: Double Pendulum.

Extended Reading

Artstein-Avidan, S. and Milman, V. (2009). The concept of duality in convex analysis, and the characterization of the legendre transform. *Annals of mathematics*, pages 661–674.

Ball, J. M. and Mizel, V. J. (1987). One-dimensional variational problems whose minimizers do not satisfy the euler-lagrange equation. In *Analysis and Thermomechanics: A Collection of Papers Dedicated to W. Noll on His Sixtieth Birthday*, pages 285–348. Springer.

Dacorogna, B. (2008). *Direct methods in the calculus of variations*. Springer.

Mania, B. (1934). Sopra un esempio di Lavrentieff. *Boll. Unione Mat. Ital.*, 13:147–153.