CLASSIC THEORY OF CALCULUS OF VARIATION

LONG CHEN

ABSTRACT. We present the first order condition: Euler-Lagrange equation, and various second order conditions: Legendre condition, Jacobi condition, Weierstrass condition etc.

CONTENTS

1. PROBLEM FORMULATION

The classical problem is

$$\inf_{x\in\mathcal{M}} \int_a^b L(t, x(t), x'(t))\,\mathrm{d}t, \tag{1}$$

with

- $\mathcal{M} = \{x \in C^1(a, b), x(a) = A, x(b) = B\}$
- $L(t, x, v) : \mathbb{R}^3 \to \mathbb{R}$ is sufficient smooth, e.g. C^2.

In $L(t, x, v)$, (t, x, v) are three independent variables known as: time, state, and velocity. But in $L(t, x(t), x'(t))$, the three variables are coupled together. A function $x_* \in \mathcal{M}$ is a global minimizer if $I(x_*) \leq I(x)$ for all $x \in \mathcal{M}$, and is a local minimizer if $I(x_*) \leq I(x)$ for all $x \in \mathcal{N}(x_*)$ where the neighborhood $\mathcal{N}(x_*)$ will be defined more precisely in Definition 5.1. If $\leq$ is changed to $<$, it is called a strict minimizer.

We introduce a variation of a function. Let ϕ be a test function in $\mathcal{M}_0$ satisfies certain boundary conditions so that if $x \in \mathcal{M}$, then $x + \epsilon\phi \in \mathcal{M}$. For the example considered,

Date: June 2, 2024.

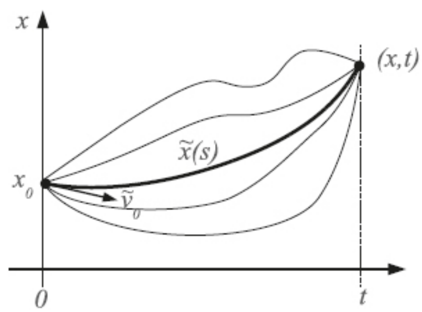

FIGURE 1. Variation of a curve.

$\mathcal{M}_0 = C_0^1(a,b) = \{\phi \in C^1(a,b), \phi(a) = \phi(b) = 0\}$. The term $\epsilon\phi$ is called an *variation* of x and the scaling ϵ indicates the variation is small.

Define

$$f(\epsilon) := I(x_* + \epsilon\phi).$$

We then back to a calculus problem. x_* is a local minimizer of $I(\cdot)$ if and only if 0 is a local minimum of $f(\epsilon)$. So the optimality condition of an extreme curve x of $I(x)$ is characterized as

$$f'(0) = \frac{\mathrm{d}}{\mathrm{d}\epsilon} I(x+\epsilon\phi)|_{\epsilon=0} = 0, \tag{2}$$

Furthermore, we have the following necessary and sufficient conditions for 0 being a local minimizer of $f(\epsilon)$ or equivalently x_* is a local minimizer of $I(x)$.

- *Necessary conditions.* If x_* is a local minimizer, then $f'(0) = 0$ and $f''(0) \geq 0$.
- *Sufficient conditions.* If $f'(0) = 0$ and $f''(0) > 0$, then x_* is a strict local minimizer.

2. THE EULER-LAGRANGE EQUATION

By the chain rule, we obtain the variational form of Euler-Lagrange equation, where $(\cdot,\cdot)$ is the stand L^2-inner product.

Variational form of Euler-Lagrange equation

$$(L_v(t,x,x'), \phi') + (L_x(t,x,x'), \phi) = 0 \quad \forall \phi \in H_0^1(a,b). \tag{3}$$

The space for the test function ϕ can be relaxed to $H_0^1(a,b) := \overline{C_0^1(a,b)}^{\|\cdot\|_1}$, where

$$(u,v)_1 := (u,v) + (u',v') = \int_a^b uv + u'v' \,\mathrm{d}t,$$

and $\|\cdot\|_1$ is the induced norm. In the variational form (3), the solution (if exists) $x \in H^1(a,b) = \overline{C^1(a,b)}^{\|\cdot\|_1}$. Functions in $H^1(a,b)$ possess weak derivatives, which are defined in terms of distributions rather than the classical notion of derivative as a pointwise limit of the difference quotient. More precisely, for a function $x \in L^1_{\rm loc}(a,b)$, if there exists a function $p \in L^1_{\rm loc}(a,b)$ satisfying

$$(p,\phi) = -(x,\phi') \quad \text{for all } \phi \in C_0^\infty(a,b),$$

then we say p is the weak derivative of x and still denoted by $p = x'$. Obviously when x is C^1, the weak derivative coincides with the classic derivative. For piecewise C^1 function and globally continuous functions, the weak derivative exists and equals to the piecewise derivative. It may not be differentiable in the classical sense, as demonstrated by examples such as the absolute value function and the ReLU function.

Using integration by parts, we get the strong form of E-L equation

Strong form of Euler-Lagrange equation

$$-\frac{\mathrm{d}}{\mathrm{d}t}L_v(t,x(t),x'(t)) + L_x(t,x(t),x'(t)) = 0. \tag{4}$$

The boundary term is invisible since $\phi \in H_0^1(a,b)$. And the test function ϕ is not present in (4) by using the fact: $\forall \phi \in L^2(a,b)$ holds $\to$ pointwise (a.e.) holds. If this were not the case, one could construct a test function ϕ near a point t that violates the weak form of the Euler-Lagrange equation (3), contradicting its validity for all functions in $H_0^1(a,b)$.

If the test space is changed to $H^1(a,b)$ without the zero boundary condition, then the variational form (3) still holds by first restricting the test function in the subspace $H_0^1(a,b)$ and again integration by parts to get (4). Next chose $\phi \in H^1(a,b)$ and use (4) to eliminate the volume contribution and obtain the Neumann boundary condition

$$L_v(t,x,x')\,\phi|_a^b = 0,\ \forall \phi \quad \Longrightarrow \quad L_v(a,x(a),x'(a)) = L_v(b,x(b),x'(b)) = 0.$$

In general (4) is a nonlinear second order elliptic ODE. When L is independent of t, which is called *autonomous*, we have the conservation of Hamiltonian which is known as the second Erdmann condition.

Erdmann condition for autonomous Lagrangian

$$x'(t)L_v(x(t),x'(t)) - L(x(t),x'(t)) = \text{const.} \tag{5}$$

In contrast to the strong form of the Euler-Lagrange equation given by (4), the equation (5) is a first order nonlinear ODE and can be thought of as a first integral of (4). The validity of (5) can be easily verified by taking its derivative and using the strong form of the Euler-Lagrange equation.

3. Second Order Conditions

In calculus, a critical point of the optimization problem $\min f(x)$ is a solution to $f'(x) = 0$. Similarly, a solution to the Euler-Lagrange equation is called an extremal of the functional. Showing an extremal is a local minimum of the functional requires second order conditions. Specifically, we need to examine the behavior of the second variation of the functional around the extremal. Throughout this section, we will assume that the Lagrangian L is twice continuously differentiable, which ensures that the second variation exists and is well-defined.

3.1. **Second order variation.** We first compute the second order variation:

$$f''(0) = \delta^2 I(x,\phi) := \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} I(x+\epsilon\phi)|_{\epsilon=0} = (A\phi',\phi') + 2(B\phi',\phi) + (C\phi,\phi),$$

where

$$A = L_{vv}(t,x(t),x'(t)), \quad B = L_{xv}(t,x(t),x'(t)), \quad C = L_{xx}(t,x(t),x'(t)).$$

In terms of the second order variation, we can write out conditions for 0 being a local minimizer of the single variable function $f(\epsilon)$.

- *Necessary conditions.* If $x_* \in \mathcal{M}$ is a local minimizer of $I(\cdot)$, then
  (1) $\delta I(x_*,\phi) = 0$, and
  (2) $\delta^2 I(x_*,\phi) \geq 0$ for all $\phi \in H_0^1(\Omega)$.

- *Sufficient conditions.* If
  (1) $\delta I(x_*,\phi) = 0$, and
  (2) there exists $\lambda > 0$ such that $\delta^2 I(x_*,\phi) \geq \lambda \|\phi\|_1^2$ for all $\phi \in H_0^1(\Omega)$,

then x_* is a strict local minimizer of $I(\cdot)$.

We then derive conditions without the test function ϕ.

3.2. **Hessian matrix.** The Hessian matrix of $L(t,x,v)$ with respect to variable (v,x) at x_* is denoted by

$$\mathcal{H}^*(t) = \begin{pmatrix} A^* & B^* \\ B^* & C^* \end{pmatrix}. \tag{6}$$

Here * is used to emphasize it is evaluated at a particular function x_* solving the E-L equation. Then the second order variation can be rewritten as

$$\delta^2 I(x_*,\phi) = \left(\mathcal{H}^* \begin{pmatrix} \phi' \\ \phi \end{pmatrix}, \begin{pmatrix} \phi' \\ \phi \end{pmatrix}\right).$$

Based on this formulation, we have the following sufficient condition.

A sufficient condition: Hessian is SPD

Suppose
  (1) x_* satisfies the E-L equation of $I(\cdot)$;
  (2) $\mathcal{H}^*(t) > 0$ for all $t \in [a,b]$;

then x_* is a strict local minimum.

Here, a symmetric $n \times n$ matrix $M > 0$ means $(Mv,v) \geq 0$ for all $v \in \mathbb{R}^n$, and $(Mv,v) = 0$ if and only if $v = 0$, or equivalently, $\lambda_{\min}(M) > 0$, where $\lambda_{\min}(M)$ is the minimum eigenvalue of M.

Use the continuity of $\lambda_{\min}(\mathcal{H}^*(t))$ on the compact interval $[a,b]$, which follows from the assumption that L is twice continuously differentiable, we conclude that there exists a minimum value $\lambda_0 > 0$ s.t. $\min_{t\in[a,b]} \lambda_{\min}(\mathcal{H}^*(t)) > \lambda_0$.

The condition $\mathcal{H}^* \geq 0$ is, however, not necessary. Here is an example.

Example 3.1. *Consider* $L(t,x,v) = v^2 - x^2$. *Then* $\mathcal{H}^* = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. *Consider zero boundary condition* $x(a) = x(b) = 0$. *Then* $x_* = 0$ *solves E-L equation. By the Poincaré inequality,*

$$\delta^2 I(x_*, \phi) = \|\phi'\|^2 - \|\phi\|^2 \geq \left(1 - \frac{(b-a)^2}{2}\right)\|\phi'\|^2, \quad \forall \phi \in H_0^1(a,b).$$

So for $(b-a)^2$ *is smaller than* 2*, we conclude* $x_* = 0$ *is a local minimizer.*

Exercise 3.2 (Poincaré inequality)**.** *For* $\phi \in H_0^1(a,b)$*, we have*

$$\|\phi\|^2 \leq \frac{(b-a)^2}{2}\|\phi'\|^2.$$

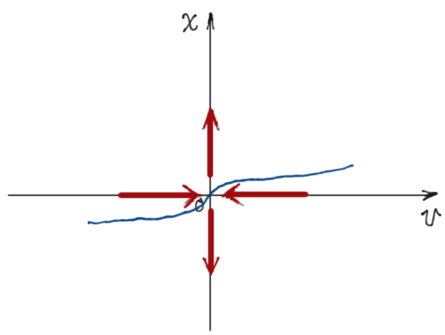

FIGURE 2. $\mathcal{H} \geq 0$ is not necessary.

The Hessian matrix $\mathcal{H}$ is defined for variables (v, x) where v and x are independent. However, in the second order variation, these two variables are dependent through $v = x'$ which defines a lower dimensional manifold Γ. The functional $I(x)$ achieves a local minimum on Γ rather than in the full space of (v, x). Furthermore, there is an integration over the interval $[a, b]$ to average out the point-wise information.

As we shall show in a moment, the derivative part dominates the scaling and can overwhelm the Hessian term in the second order variation. Therefore, in Example 3.1, although $\mathcal{H}$ is indefinite and is only SPD along the v-direction, it is still possible for x_* to be a local minimum.

Let us rewrite the second order variation in another operator form. Assume $L \in C^3$. Using integration by parts, the cross term

$$\begin{aligned} (B\phi, \phi') &= -\left((B\phi)', \phi\right) = -(B'\phi, \phi) - (B\phi', \phi) \\ \implies \quad 2(B\phi, \phi') &= -\left(B'\phi, \phi\right). \end{aligned}$$

Therefore

$$\delta^2 I(x, \phi) = (S\phi, \phi) := (P(t)\phi', \phi') + (Q(t)\phi, \phi),$$

where

$$P(t) = A(t) = L_{vv}(t, x(t), x'(t)), \quad Q(t) = C - B' = L_{xx} - L'_{xv}.$$

The operator $S : H_0^1(a,b) \to H_0^1(a,b)$ is self-adjoint. The sufficient condition

$$\delta^2 I(x_*, \phi) > 0 \quad \forall \phi \in H_0^1(a,b) \qquad \iff \qquad (S^*\phi, \phi) > 0 \quad \forall \phi \in H_0^1(a,b).$$

That is S^* is an SPD on $H_0^1(a,b)$.

Through integration by parts, we change the Hessian matrix $\mathcal{H}$ to a diagonal matrix S. The action of these two on $\Gamma = \{(\phi', \phi)\}$ are the same. As $\delta^2 I(x_*, \phi) = (S^*\phi, \phi)$, the operator S^* plays the role of the Hessian matrix for the calculus problem.

3.3. **Legendre condition.** In the Hessian matrix $\mathcal{H}$, $A = L_{vv}$ dominates in the sense that it acts on derivatives: $(A\phi', \phi')$. Using a scaling argument, one can easily construct a test function ϕ supported near a particular point t with length h and height 1 (e.g. a hat function which is in $H_0^1(a,b)$). Then $(\phi', \phi') = \mathcal{O}(h^{-1})$, $(\phi', \phi) = \mathcal{O}(h)$, and $(\phi, \phi) = \mathcal{O}(h^2)$. If we work with C^1 test functions instead of H^1, we can apply a mollifier to obtain the desired test function.

Using this scaling argument, we can obtain a necessary condition for a local minimum.

Legendre Condition

If $x_* \in \mathcal{M}$ is a local minimizer of $I(\cdot)$. Then the following Legendre condition holds

$$P^*(t) = A^*(t) := L_{vv}(t, x_*(t), x_*'(t)) \geq 0, \quad \forall t \in [a,b]. \tag{7}$$

Proof. Construct a hat function $\phi \in H_0^1(a,b)$ with support $[t-h, t+h]$ and height h. Then $|\phi'|^2 = 1$ only in $[t-h, t+h]$ and zero otherwise. From the second variation condition $\delta^2 I(x, \phi) \geq 0$, we conclude

$$\int_{t-h}^{t+h} \left[P^*(t) + |Q^*(t)| h^2\right] \, \mathrm{d}t \geq 0.$$

As h and t are arbitrary, we conclude the Legendre condition. □

The strict (or called strengthened) Legendre condition $A^* > 0$ is not sufficient as it is only part, although the main part, of the Hessian $\mathcal{H}^*$. Instead we should look at S^* as $\delta^2 I(x_*, \phi) = (S^*\phi, \phi)$. The operator S^* plays the role of the Hessian matrix for the calculus problem. Conditions involving $\mathcal{H}^*$, A^*, and S^* are summarized as follows:

$$\mathcal{H}^* > 0 \Longrightarrow x_* \text{ is a local minimizer } \not\Longrightarrow \mathcal{H}^* \geq 0.$$
$$A^* > 0 \not\Longrightarrow x_* \text{ is a local minimizer } \Longrightarrow A^* \geq 0.$$
$$S^* > 0 \Longrightarrow x_* \text{ is a local minimizer } \Longrightarrow S^* \geq 0.$$

3.4. **Conjugate points.** Assuming $P^* > 0$ and $Q^* \geq 0$, which is roughly equivalent to the strong sufficient condition $\mathcal{H} > 0$, we can conclude that $S^* > 0$. However, we cannot be certain of the sign of Q^* over the interval $[a,b]$ and it is possible $S^* > 0$ even Q^* is negative.

To study the condition $S^* > 0$, we introduce the concept of conjugate points. We will start with an example where the conjugate point has a clear geometric interpretation.

Exercise 3.3. Consider two points a, b on the unit sphere S. The geodesic connecting a, b includes two great circles. Which one is the shortest one?

Let $\hat{a} = -a$ be called the conjugate point a. The shortest curve is the one which does not contains the conjugate point $\hat{a}$. The calculation for the sphere geodesic can be simplified by choosing the plane containing $(a, 0, b)$ as the the $x - y$ plane and $a = (1, 0, 0)$. □

In the general case, the concept of a conjugate point is not so straightforward. Given a number $\sigma \in (a, b]$, we can embed the subspace $H_0^1(a,\sigma) \hookrightarrow H_0^1(a,b)$ by zero extension, and naturally restrict the operator S to $H_0^1(a,\sigma)$. More precisely, for $\phi, \psi \in H_0^1(a,\sigma)$

$$(S_\sigma \phi, \psi) := (P(t)\phi', \psi') + (Q(t)\phi, \psi).$$

The notation S_σ is used to indicate the dependence of S on the parameter σ.

Definition 3.4 (Conjugate points)**.** *Let x_* be a solution to the Euler-Lagrange equation. A conjugate point of a (along x_*) is a number $\sigma \in (a, b]$ such that S^*_σ has a zero eigenvalue on $H_0^1(a,\sigma)$. In other words, there exists a nonzero function $u \in H_0^1(a,\sigma)$ satisfying $S^*_\sigma u = 0$, with the following strong form:*

$$-(P^*(t)u'(t))' + Q^*(t)u(t) = 0, \quad t \in (a,b), \quad u(a) = u(\sigma) = 0. \tag{8}$$

If a conjugate point exists, we can utilize the eigenfunction u from (8) as the variation. This leads to $\delta^2 I(x_*, u) = 0$, which does not allow us to draw any conclusion for x_* being a local minimizer.

When $P^* > 0$, S^* is a Sturm-Liouville (compact) operator and thus has countable real eigenvalues. When σ is sufficiently close to a, approximate the coefficient $P(t), Q(t)$ by constants $p = P(a) > 0, q = Q(a)$ and solve a linear eigenvalue equation

$$-pu'' + qu = \lambda u, \quad u(a) = u(\sigma) = 0,$$

to get $u = \sin(k\pi(x-a)/(\sigma - a)), k \in \mathbb{N}$, and $\lambda = q + pk^2\pi^2/(\sigma-a)^2$. So, even $q < 0$, we have

$$\lambda_{\min}(S_\sigma) > 0, \text{ for } \sigma \text{ is sufficiently close to } a.$$

Furthermore, we have the following dependence on σ.

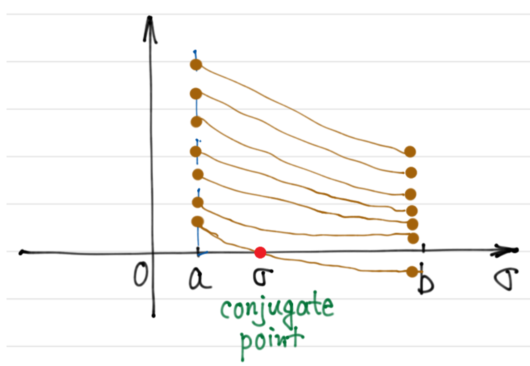

FIGURE 3. Eigenvalues of S_σ are decreasing with respect to σ.

Proposition 3.5. *Assume $P(t) > 0$ for all $t \in [a,b]$. Then $\lambda_{\min}(S_\sigma)$ is a decreasing function of $\sigma \in (a,b]$.*

Proof. For the self-adjoint operator S_σ,

$$\lambda_{\min}(S_\sigma) = \inf_{v \in H_0^1(a,\sigma)} \frac{(S_\sigma v, v)}{(v,v)}. \tag{9}$$

Take $\sigma_1 < \sigma_2$. According to (9), we have $\lambda_{\min}(S_{\sigma_1}) \geq \lambda_{\min}(S_{\sigma_2})$. Next we show $>$ holds by construction. Take v as the eigen-function of $\lambda_{\min}(S_{\sigma_1})$ with normalization $\|v\| = 1$.

If $v'(\sigma_1) = 0$, together with $v(\sigma_1) = 0$, we know locally $v = 0$ and can shift σ_1 to the left s.t. $v'(\sigma_1) \neq 0$. Without loss of generality, assume $v'(\sigma_1) > 0$. We then modify to a function $\tilde{v} \in H_0^1(a, \sigma_1 + h)$ and show $\frac{(S_{\sigma_2}\tilde{v}, \tilde{v})}{(\tilde{v}, \tilde{v})} < \frac{(S_{\sigma_1}v, v)}{(v, v)}$ when h is small enough.

include a figure here. □

Proposition 3.5 can be generalized to every eigenvalue of S_σ and the full proof can be found in [3].

3.5. **Jacobi condition.** The operator S^* plays the role of Hessian in the calculus problem.

Necessary condition to be a local minimum: no interior conjugate point

Suppose $P^* := L_{vv}(t, x_*(t), x'_*(t)) > 0$ for all $t \in [a, b]$. If x_* is a local minimum of $I(\cdot)$, then

(1) x_* satisfies the E-L equation;
(2) there is no interior conjugate point in (a, b).

Note that as $\delta^2 I(x_*, \phi) \geq 0$ for a local minimum, the right end point b could be a conjugate point. The above necessary condition says no interior conjugate point. Otherwise by Proposition 3.5, after the conjugate point σ, there exists a negative eigenvalue of $S^*_{\sigma_2}$, for $\sigma_2 \in (\sigma_1, b]$, and $\delta^2 I(x_*, u) < 0$ for the corresponding eigen-function u, which violates the local minimizer assumption. An elementary proof using the regularity result will be included in Exercise 4.4.

We present the following sufficient conditions for a local minimizer.

Sufficient conditions: E-L + strict Legendre condition + Jacobi condition

Suppose

(1) x_* satisfies the E-L equation of $I(\cdot)$;
(2) $P^*(t) := L_{vv}(t, x_*(t), x'_*(t)) > 0$ for all $t \in [a, b]$;
(3) no conjugate point in $(a, b]$,

then x_* is a strict local minimum.

As $\lambda_{\min}(S^*_\sigma) > 0$ for σ sufficiently close to a and $\lambda_{\min}(S^*_\sigma)$ is a decreasing function of σ, "no conjugate point condition" implies $\lambda_{\min}(S^*) = \lambda_0 > 0$ and thus $\delta^2 I(x_*, \phi) = (S^*\phi, \phi) \geq \lambda_0 \|\phi\|^2$. Use the following exercise to change to the coercivity in H^1-norm.

Exercise 3.6. Assume $P^* > 0$ for all $t \in [a, b]$ and $\delta^2 I(x_*, \phi) \geq \mu \|\phi\|^2$ for all $\phi \in H_0^1(a, b)$. Then it is also coercive in H^1 norm, i.e., there exists $\lambda > 0$ such that

$$\delta^2 I(x_*, \phi) \geq \lambda \|\phi\|_1^2 \quad \forall \phi \in H_0^1(a, b).$$

3.6. **Jacobi equation.** We present another elementary approach given in the book [2]. A key trick is that $\int_a^b (w\phi^2)' \, dt = 0$, given that $\phi \in H_0^1(a, b)$. We can therefore add w' to the (ϕ, ϕ) term for any arbitrary function w. If w satisfies the nonlinear ODE

$$w' + Q^* = \frac{w^2}{P^* - \delta}, \tag{10}$$

where $\delta = \min_{t\in[a,b]} P^*(t)/2$ subject to $P^* - \delta > 0$, we can add $(w'\phi, \phi)$ to $\delta^2 I(x_*, \phi)$ and use integration by parts $(w'\phi, \phi) = -2(w\phi, \phi')$ to get

$$\delta^2 I(x_*, \phi) = \|\sqrt{P^* - \delta}\phi' + \frac{w}{\sqrt{P^* - \delta}}\phi\|^2 + \delta\|\phi'\|^2 \geq \delta\|\phi'\|^2.$$

This is exactly the original proof given by Legendre.

By the continuation argument, we can consider the case $\delta = 0$

$$w'(t) + Q^*(t) = \frac{w^2(t)}{P^*(t)},\ t \in (a, b), \quad w(a) = w(b) = 0. \tag{11}$$

If solution of (11) exists, then by ODE theory, for small enough δ, solution to (10) also exists. It is worth noting that the existence of a solution to (11) is not immediately apparent, as (11) is nonlinear and only local existence near the ending points can be guaranteed.

How to solve the nonlinear ODE (11)?

Change of variable! Jacobi introduced a change of variable with a positive function u

$$w(t) = -u'(t)P^*(t)/u(t).$$

The positivity $u(t) > 0$ for all $t \in [a, b]$ is required as u appears in the denominator. With this change of variable, w solves (11) if u is a solution to the Jacobi equation

$$-(P^*u')' + Q^*u = 0, \quad t \in (a, b), \quad u \neq 0 \text{ in } [a, b]. \tag{12}$$

To prove the existence of a positive solution to (12), we rely on the no conjugate point condition. Here is a rough outline of a proof. We introduce $v = P^*u'$ and write the Jacobi equation (12) as a first-order system. Then, by assuming that there are no conjugate points, we show the existence of a positive solution for $u(a) = 0$ and $v(a) = P^*(a) > 0$. Finally, we shift to $u(a) > 0$ through continuation.

How to check the existence of a conjugate point? Solve Jacobi equation (12) with boundary condition $u(a) = 0$ and $u'(a) = 1$. Then check if there exists a zero point of u inside the interval. Or solve the equation (11) of w directly.

4. Non-smooth Extremal

In this section, we will discuss the relaxation of the smoothness assumptions on the admissible function x. Previous treatments have typically focused on smooth solutions, where the Euler-Lagrange equation takes the form of a second-order ODE, and the solution belongs to the class $C^2(a, b)$. Similarly, the Legendre and Jacobi conditions involve the second derivative L_{vv} of the Lagrangian, which requires $L \in C^2$. However, these smoothness assumptions can be relaxed by considering weaker notions of differentiability.

4.1. **Relax smoothness.**

In the context of the first-order variation $\delta I(x, \phi)$, it is sufficient for x and ϕ to belong to the class C^1 without any need for further integration by parts. However, it should be noted that the requirement of C^1 is a rather strong one, as it implies that the derivative exists in the classical sense as the limit of the rate of change. A more general class of functions that is of great importance in various applications is that of continuous and piecewise smooth functions. Such functions may contain corners or kinks, which are not differentiable in the classical sense. This highlights the need for a more flexible notion of differentiability, such as the concept of weak derivatives.

Example 4.1. Consider the following example (Example 15.1 p307 in [2]):

$$\min I(x) := \int_{-1}^{1} x^2(x' - 1)^2 \,\mathrm{d}t.$$

Obviously $I(x) \geq 0$ and $x_* = \max\{t, 0\}$ (ReLU function) is a global minimum which is not in C^1. One can easily construct a sequence of C^1 curves x_n s.t. $I(x_n) \to 0$ and $x_n \to x_*$ pointwisely. That is we find a minimizing sequence but the limit is out of C^1. In other words, C^1 is not complete under the norm $\|\cdot\|_{1,\infty}$.

We introduce the Sobolev space $W^{1,\infty}(a,b) = \overline{C^1(a,b)}^{\|\cdot\|_{1,\infty}}$ which is a Banach space under the norm $\|\cdot\|_{1,\infty}$. As the functions are League measurable, the $\|\cdot\|_\infty$ is the essential sup. The classic derivative defined as pointwise limit will be extended to weak derivatives defined by integration by parts.

For a function $x \in L^1_{\rm loc}(a,b)$, if there exists a function $p \in L^1_{\rm loc}(a,b)$ satisfying

$$(p, \phi) = -(x, \phi') \quad \text{for all } \phi \in C_0^\infty(a,b),$$

then we say p is the weak derivative of x and still denoted by $p = x'$. Obviously when x is C^1, the weak derivative coincides with the classic derivative. We can define, for $1 \leq p \leq \infty$,

$$W^{1,p}(a,b) = \{x \in L^p(a,b) : x' \in L^p(a,b)\}, \quad \|x\|_{1,p} = (\|x\|_{L^p}^p + \|x'\|_{L^p}^p)^{1/p}.$$

Exercise 4.2. Verify $W^{1,p}(a,b)$ endowed with norm $\|\cdot\|_{1,p}$ is a Banach space.

By definition, for a continuous and piecewise smooth function, the weak derivative is equal to the piecewise derivative. It is important to note that continuity is still required to avoid discontinuities in the integration by parts. In one dimension, if $x \in W^{1,1}(a,b)$, then we have the representation formula

$$x(t) = x(a) + \int_a^t x'(s)\,\mathrm{d}s. \tag{13}$$

Furthermore, the fact that x' exists almost everywhere implies that x is absolutely continuous. In particular, the space $W^{1,\infty}(a,b)$ can be identified with the space of Lipschitz continuous functions $\mathrm{Lip}(a,b)$ which is also denoted by $C^{0,1}(a,b)$.

The strong form of Euler-Lagrange equation is present for $x \in C^2(a,b)$ and (4) is a nonlinear second order elliptic ODE. When the function is not smooth enough, we can write it in the integral form.

Integral form of the Euler-Lagrange equation

$$-L_v(t, x(t), x'(t)) + \int_a^t L_x(s, x, x')\,\mathrm{d}s = \text{const.} \quad t \in [a,b] \quad \text{a.e.} \tag{14}$$

Proof. Start with the weak form (3) of the Euler-Lagrange equation and use integration by part but for the low order term

$$(L_x, \phi) = ((c + \int_a^t L_x)', \phi) = -(c + \int_a^t L_x, \phi').$$

Then for any $c \in \mathbb{R}^n$ and any $\phi \in H_0^1(a,b)$, we have

$$(L_v - \int_a^t L_x - c, \phi') = 0. \tag{15}$$

Choose a special test function

$$\phi(t) = \int_a^t \left[L_v - \int_a^s L_x - c \right] \mathrm{d}s$$

with property $\phi' = L_v - c - \int_a^t L_x$ and chose $c \in \mathbb{R}^n$ to satisfy the boundary condition $\phi(a) = \phi(b) = 0$. Then (15) becomes $\|\phi'\| = 0$ and thus $\phi = 0$. □

Introduce $p(t) = L_v(t, x(t), x'(t))$ which is called momentum or adjoint variable (in physics) or co-state (in control theory). Then the integral form of E-L equation can be written as

Euler-Lagrange equation as a first order system

(16) $$p = L_v, \quad p' = L_x \quad t \in [a, b] \quad \text{a.e.}$$

(16) is the standard way to write a 2nd order ODE (4) into a 1st order ODE system.

4.2. **Regularity.** We relax $C^1[a, b]$ to $\mathrm{Lip}[a, b]$ to seek for a minimum in a larger admissible function set. With certain conditions, we can prove the founded minimum is indeed in the smaller space $C^1[a, b]$.

Theorem 4.3. *Let $x_* \in \mathrm{Lip}[a, b]$ satisfy the integral Euler equation and assume for almost every $t \in [a, b]$, the Lagrangian $L(t, x_*(t), v)$ is strictly convex in v. Then $x_* \in C^1[a, b]$.*

Proof. In the integral form

$$x(t) = x(a) + \int_a^t x'(s) \,\mathrm{d}s,$$

the weak derivative x' exists a.e. If we can find a continuous representative of x', then the smoothness is improved from $\mathrm{Lip}[a, b]$ to $C^1[a, b]$.

For function x_* in $\mathrm{Lip}[a, b]$, x'_* could be discontinuous at a point τ. Then $L_v(\tau, x_*(\tau), \ell_-)$ may not be $L_v(\tau, x_*(\tau), \ell_+)$, where $\ell_\pm$ is the right and left limit of x'_* at τ.

Use the integral form of E-L equation (14), we have $L_v(\tau, x_*(\tau), \ell_-) = L_v(\tau, x_*(\tau), \ell_+)$ by taking two sequences convergent to τ from the left and right. As L is strictly convex in v, i.e., L_v is strictly increasing as a function of the third variable, then $\ell_- = \ell_+$. So x'_* is continuous at τ.

□

In Example 4.1, the condition 'L is strictly convex in v' is not valid as $x_*(t) = 0$ hold on a non-zero measure set. When L is C^2 and strictly convex, then $L_{vv} > 0$ and E-L equation is an elliptic equation and a stronger regularity result $x \in C^2(a, b)$ can be established.

Exercise 4.4. Let x_* be an extremal which satisfies the strengthened Legendre condition. We shall prove the necessary condition "no interior conjugate point":

If x_* is a weak local minimizer, then there is no conjugate point to a in the interval (a, b).

Suppose a conjugate point $\sigma \in (a, b)$ exists. By definition of the conjugate point, there exists a non-trivial solution u of Jacobi's equation

$$-(P^* u')' + Q^* u = 0, \; t \in (a, \sigma), \quad u(a) = u(\sigma) = 0.$$

(1) Prove that $u'(\sigma) \neq 0$ and that

$$\int_a^\sigma \left[P^*(t)u'(t)^2 + Q^*(t)u(t)^2\right] \, \mathrm{d}t = 0.$$

(2) Extend u to $[a, b]$ by zero. Consider the minimization problem

$$\min_{y \in \mathrm{Lip}_0[a,b]} I(y) := \int_a^\sigma \left[P^*(t)y'(t)^2 + Q^*(t)y(t)^2\right] \, \mathrm{d}t.$$

Then u is a minimizer. Use the regularity result to obtain a contradiction.

5. Strong Minima

We discuss the perturbation in different norms. For the function space $\phi \in C_0^1[a, b]$, the default norm is

$$\|\phi\|_{1,\infty,[a,b]} = \max\{\|\phi\|_{\infty,[a,b]}, \|\phi'\|_{\infty,[a,b]}\}.$$

As a subspace of $C_0[a, b]$, a weaker norm $\|\cdot\|_{\infty,[a,b]}$ can be used. To simplify notation, we shall skip the interval in the norm. Using different norms, we can define different open balls near a point

$$B_\epsilon(x_*, \|\cdot\|_\infty) = \{x \in C_0[a, b] : \|x - x_*\|_\infty < \epsilon\}$$
$$B_\epsilon(x_*, \|\cdot\|_{1,\infty}) = \{x \in C_0^1[a, b] : \|x - x_*\|_{1,\infty} < \epsilon\}.$$

Definition 5.1. *A function $x_* \in \mathcal{M}$ is a* strong local minimum *of $I(\cdot)$ if there exists $\epsilon > 0$ such that for all $\phi \in C_0^1(a, b) \cap B_\epsilon(x_*, \|\cdot\|_\infty)$, we then have*

$$I(x_* + \phi) \geq I(x_*). \tag{17}$$

If the condition is changed to $\phi \in B_\epsilon(x_*, \|\cdot\|_{1,\infty})$, then it is called a weak local minimum.

We use Fig. 4 to illustrate different perturbations. The small square represents the $\|\cdot\|_\infty$ ball, which is centered at an extremum (x_*, x_*') with a radius of ϵ in the (x, v)-plane. However, it should be noted that the variables (x, v) are not independent for the variation. Specifically, the relation $v = x'$ defines a curve passing through (x_*, x_*').

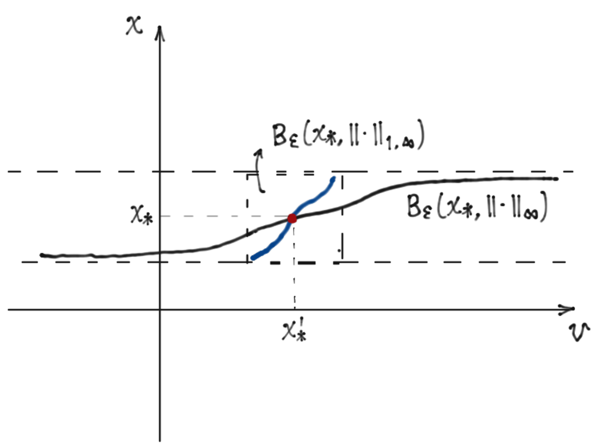

FIGURE 4. Perturbation in the weak minimum v.s. the strong minimum.

For a weak minimum, the perturbation is restricted to the curve inside the ball such that both the $\|\phi\|_\infty$ and $\|\phi'\|\infty$ are bounded by ϵ.

For a strong minimum, the norm of $\|\phi\|_\infty$ is still bounded by ϵ, but the norm of $\|\phi'\|_\infty$ is not necessarily bounded. Thus, the curve is only restricted to the strip $B_\epsilon(x_*, \|\cdot\|_\infty)$ and may go outside of the square. In other words, the curve can potentially move outside of the square, while the norm of $\|\phi\|_\infty$ remains controlled. This is shown in Figure 4, where the derivative of the variation may be unbounded outside of the square.

Notice that the C^1-norm is stronger in the sense that

$$\|\phi\|_\infty \leq \|\phi\|_{1,\infty} \Longrightarrow B_\epsilon(x_*, \|\cdot\|_{1,\infty}) \subseteq B_\epsilon(x_*, \|\cdot\|_\infty)$$

A topology induced by a stronger norm will have more open sets than a topology induced by a weaker norm. For instance, let us consider the $\|\cdot\|_\infty$ and $\|\cdot\|_{1,\infty}$ norms. Both norms induce topologies where the strip is open, but only the $\|\cdot\|_{1,\infty}$ topology has the square as an open set. Therefore, we say that $\|\cdot\|_{1,\infty}$ is stronger than $\|\cdot\|_\infty$, as it has more open sets, such as small squares.

However, note that a stronger norm assigns a smaller size to its unit ball compared to a weaker norm. This means that the set of points with norm at most 1 is smaller for a stronger norm than a weaker one.

- *Strong* minimum: the perturbation is in a *larger* ϵ-ball of a *weaker* norm.
- *Weak* minimum: the perturbation is in a *smaller* ϵ-ball of a *stronger* norm.

Obviously, a strong minimum is also a weak minimum but not vice verse. An example can be easily constructed by using a Lagrangian of x' only. The variation ϕ is small in L^2 but not in H^1. For example, a high frequency perturbation $\phi = \epsilon \sin(k\pi t)$.

Example 5.2 (Example 14.14 (page 301 and 318 in [2])). Consider

$$\min \int_0^1 (x'(t))^3 \, \mathrm{d}t : x(0) = 0, x(1) = 1.$$

A weak local minimum is $x_*(t) = t$ satisfying E-L condition, strengthened Lendre condition, no conjugate condition, but is not a strong local minimum.

All minima discussed previously are weak local minima. The sufficient condition $\delta^2 I(x_*, \phi) \geq \lambda \|\phi\|_1^2$ for x_* being a local weak minimum is no longer sufficient for being a strong local minimum. Take the calculus problem $\min f(x)$ as an example. Assume 0 is a critical point. Then

$$f(\phi) - f(0) = \frac{1}{2} f''(0)\phi^2 + \frac{1}{6} f'''(\xi)\phi^3.$$

If $f''(0) \geq \lambda > 0$, then for $|\phi|$ sufficient small, the third order term $|f'''(\xi)||\phi|^3 \leq \lambda |\phi|^2$, then $f(\phi) - f(0) \geq \lambda |\phi|^2/3 \geq 0$ which means 0 is a local minimum.

Now the perturbation of the norm is different. We can still assume the norm $\|\phi\|_\infty < \epsilon$ but have no control of $\|\phi'\|_\infty$. In a similar expansion

$$I(x) - I(x_*) = \delta^2 I(x_*, \phi) + \text{Remainder},$$

the remainder will involve $\|\phi'\|_\infty$ which cannot be controlled by $\delta^2 I(x_*, \phi)$ when only $\|\phi\|_\infty$ is small.

We need more conditions to control the part out of the small ball. Such condition should be imposed for arbitrary perturbation in v-direction. Given a C^1 Lagrangian $L(t, x, v)$, define the Weierstrass excess function

$$\mathfrak{E}_L(t, x, v, q) = L(t, x, q) - L(t, x, v) - (q - v) \cdot L_v(t, x, v). \tag{18}$$

That is the difference to the linear expansion at point v. If $\mathfrak{E}_L(t,x,v,q) \geq 0$ for any (v,q), then $L(t,x,v)$ is convex w.r.t. v and $\mathfrak{E}_L$ is known as the Bregman divergence. To be a strong minimum, we require the positivity of $\mathfrak{E}_L$ at the minimum point only. The function $v \to L(t,x,v)$ is not necessarily convex; see Fig. 6.

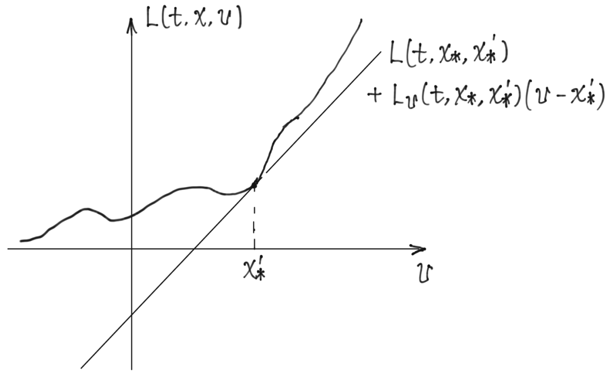

FIGURE 5. Weierstrass condition.

Weierstrass condition

If $x_* \in \mathcal{M}$ is a local strong minimizer of $I(\cdot)$. Then the following Weierstrass condition holds

$$\mathfrak{E}_L(t, x_*, x_*'(t), x_*'(t) + \xi) \geq 0, \quad \forall\, \xi \in \mathbb{R}^N, \forall\, t \in [a,b]. \tag{19}$$

Proof. We outline a proof of Weierstrass condition given in [1, page 45].

We start by considering the difference

$$I(x_* + \phi) - I(x_*) - \delta I(x_*, \phi) \geq 0. \tag{20}$$

We include the first variation $\delta I(x_*, \phi) = 0$ such that the integrand is $o(\|\phi\|_\infty + \|\phi'\|_\infty)$

$$\int_a^b \left[L(t, x_* + \phi, x_*' + \phi') - L(t, x_*, x_*') - L_x \cdot \phi - L_v \cdot \phi'\right] \mathrm{d}t$$

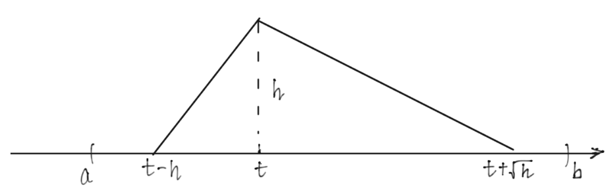

FIGURE 6. Variation in the proof of Weierstrass condition.

Given a $\xi \in \mathbb{R}^N$, construct ϕ_h as the Fig. 6 and use $\xi\phi_h$ as the variation in (20). Only the integral from $(t-h,t)$ and $(t, t+\sqrt{h})$ is nonzero. By the scaling argument, the integral from $(t, t+\sqrt{h})$ is $o(\xi h)$ while the integral from $(t-h, t)$ is $O(h)$. Thus only consider the limit

$$\begin{aligned}&\lim_{h\to 0}\frac{1}{h}\int_{t-h}^{t}\left[L(t, x_* + \phi, x_*' + \phi') - L(t, x_*, x_*') - L_x\cdot\phi - L_v\cdot\phi'\right]\,\mathrm{d}t\\&= L(t, x_*(t), x_*'(t)+\xi) - L(t, x_*(t), x_*'(t)) - \xi L_v(t, x_*(t), x_*'(t))\end{aligned}$$

which leads to the Weierstrass condition (19).

The slope of ϕ_h from the left is 1 and from the right is $-\sqrt{h}$ so that when $h \to 0$, $\xi\phi_h' \to \xi$ not $-\xi$. □

A sufficient condition for a strong minimum will use the extremal fields and will be discussed later in Hamilton-Jacobi theory.

References

[1] K.-C. Chang. *Lecture Notes on Calculus of Variations*, volume 6. World Scientific, 2016. 14

[2] F. Clarke. *Functional analysis, calculus of variations and optimal control*, volume 264. Springer Science & Business Media, 2013. 8, 9, 13

[3] R. S. Manning. Conjugate Points Revisited and Neumann–Neumann Problems. *SIAM Review*, 51(1):193–212, Feb. 2009. 8