3 Calculus of Variation

Variational calculus is the study of linear response of a functional (a map which maps a function to another object, say, a number). We discuss the classical Euler-Lagrange theory, and then reconsider the theory from the functional differentiation point of view. A necessary and sufficient condition for a functional to have a minimum and direct methods to variational calculus are briefly discussed.

Keywoods: functional. Euler-Lagrange equation. Lagrange multiplier. variable end point case. functional derivative. delta function. second variation. Legendre's condition. Noether's theorem. Vainberg's theorem. direct method.

Summary
(1) Calculus of variation is the calculus on a function space (**3.1**, **3.7**).
(2) Euler-Lagrange equation is a necessary condition for extremity (**3.2**). A sufficient condition for extremity is more involved than the ordinary calculus case (**3.14**).
(3) In terms of functional derivative (**3.7-9**). the parallelism between the ordinary calculus and variational calculus becomes explicit (**3.10-13**).
(4) Variational principle is practically useful (**3.18.3.19**). so remember that there is a way to costruct a variational functional (if any) for a given equation (**3.17**).

as an example Dirick=hle problem 入p132

3.1 Variational calculus The study of the linear response (→**2A.1**. **2A.4**) of a functional (a map which maps a function to another function or to a number is called a *functional*) is called *variational calculus*. The essence of calculus of variation is the differential calculus of functionals. This point will be made more explicit later through the introduction of functional derivatives (→**3.7-9**). A typical problem of variational calculus is to extremize a given functional.

Green 1835
Dirichlet prinbciple

The best introductory book of calculus of variation is: I. M. Gel'fand and S. V. Fomin. *Calculus of Variation* (Englewood Cliffs. 1963).

Initially. it was believed that any extremum problem had a solution. Riemann relied on a variational formulation of the Laplace equation to demonstrate the fundamental theorem of conformal transformation (→**10.11**). However. soon later Weierstrass pointed out that it is not always the case. Consider, for example. the following problem:

入p137

Gauss also assumed the existence; what is `rigorous'? 人6 p82

Weierstrass example 人5 p130; Dedekind \infty

minimize

$$\int_0^1 v(x)^2 dx \tag{3.1}$$

under the condition $v(0) = 0$ and $v(1) = 1$.

3.2 Theorem [Euler]. Let $S[f]$ be a functional on the set of C^1-functions on $[a, b]$ such that $f(a) = A$, $f(b) = B$ (fixed) defined as

$$S[f] = \int_a^b dx L(f, f', x), \tag{3.2}$$

where L is a C^2-function of its variables. A necessary condition for g to give an extremal value of S is that g satisfies *Euler's equation* (or the Euler-Lagrange equation)(→**3.5**: **4.4** for some history; for a sufficient condition see **3.14**):

$$\frac{\partial L}{\partial f} - \frac{d}{dx}\left(\frac{\partial L}{\partial f'}\right) = 0. \tag{3.3}$$

□

Renark on differentiability du Bois Raymond's theorem, 人5 p139, p180 Th1

3.3 Proof of Euler-Lagrange equation. A necessary condition for f to give an extremal value with respect to the small change[92] of f is

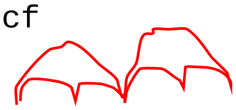
$$\delta S[f] \equiv S[f + \delta f] - S[f] = \int_a^b dx \left[\frac{\partial L}{\partial f}\delta f + \frac{\partial L}{\partial f'}\delta f'\right] = 0 \tag{3.4}$$

for any small δf (see the above footnote). Integrating this by parts, we get (note that $\delta f = 0$ at the boundaries)

$$\int_a^b dx \left[\frac{\partial L}{\partial f} - \frac{d}{dx}\frac{\partial L}{\partial f'}\right]\delta f = 0. \tag{3.5}$$

Thus, the quantity in the square brackets must vanish.[93]

cf



[92] This statement is actually tricky. We must fix a method to evaluate the size of a function, or we must be able to tell whether a function f and g are close or not. We measure the size of f in the present context in terms of the *norm* $\|f\|$ (→**20.3**) which is the sum of the largest value of $|f|$ in the relevant domain and that of $|f'|$. This is called the C^1-norm. $\delta S = 0$ in (3.4) means, more precisely, that $\|\delta S\|$ is much smaller than $\|\delta f\|$, or $\|\delta S\| = o\|\delta f\|$.

[93] More precisely: if continuous functions α and β satisfy

$$\int_a^b [\alpha(x)h(x) + \beta(x)h'(x)]dx = 0$$

for any C^1-function on $[a, b]$ such that $h(a) = h(b) = 0$, β is differentiable and $\alpha - \beta' = 0$. See Gel'fand and Fomin, Lemma 3 in Section 3.

Discussion [Canonical distribution].
The canonical distribution function

$$\rho = \frac{1}{Z} e^{-\beta H} \tag{3.6}$$

of statistical mechanics can be obtained as the solution to the following conditional maximization problem: Maximize entropy

$$S = -\int \rho \ln \rho \, d\Gamma \tag{3.7}$$

under the condition

$$E = \int H \rho \, d\Gamma, \quad 1 = \int \rho d\Gamma. \tag{3.8}$$

Here, H is the system Hamiltonian, and $d\Gamma$ is the phase volume element (the Liouville measure). Z and β are introduced as the Lagrange multipliers.

The above formulation is for classical cases, but with the replacement of $\int d\Gamma$ with Tr, we can easily obtain the quantum counterpart. The formula for the entropy was first given by Gibbs. Later, the same formula was used to define information by Shannon.

The reader might be tempted to conclude that in this way we can found statistical mechanics on the Baysian statistics, and can dispense with the principle of equal probability. However, the principle is already implicit in (3.7) in the choice of the volume.

add geodesic 入5 p144 Lobachevski geometry p 147

3.4 Conditional extremum, Lagrange multiplier. Let $S[f]$ be a functional. We wish to extremize this under the condition that $G[f] = 0$, where G is another functional. A necessary condition for $S[f]$ to be extremal is as follows. Define $I[f, \lambda] = S[f] + \lambda G[f]$. Extremize I w.r.t. f and λ. This condition would give f as a function(al) of λ. Insert this into S, and fix λ with the auxiliary condition. The result gives the extremal value of S under $G = 0$.

Geometric explanation 入3 p28-

$$\frac{\delta I}{\delta f} = 0, \quad \frac{\partial I}{\partial \lambda} = 0. \tag{3.9}$$

The parameter λ is called the *Lagrange multiplier.*

Dynamic proof
A P Knoerr: A dynamical proof of the method of Lagrange
SIAM Rev 40 941 (1998)

3.5 Who was Lagrange ?[94] Lagrange was born in Turin in 1736, where he stayed until 1766. In the mid-1750s he began to establish his reputation, and began his correspondence with Euler (→**4.4**), who became his tutor praising his work on variational calculus, and with d'Alembert, who became his political counselor. As a poorly paid professor of the Royal Artillery School at Turin from 1755-66, he worked

[94] Mainly based on R Calinger, *Classics of Mathematics* (Prentice Hall, 1982, 1995).

Exanmple: Maxwell's fisheye

relentlessly to the extent to harm his health, sustained by his association with Euler and d'Alembert (→**2B.7**). He brought the calculus of variation to maturity and applied it to mechanics.

In 1766, Lagrange succeeded Euler as director of the mathematical section of the Berlin Academy. The years in Berlin were extremely productive for Lagrange. He contributed to the three-body problem, various number theoretical problems, and his 1770 memoir opened a new era in algebra (group theory).

When Friedrich the Great died in 1787, he accepted an invitation of Louis XVI to join the Paris Academy of Science. A year later he published his classic, *Mécanique analytique*. This was the first book of mechanics without any geometrical argument.

Shy, diplomatic, and amenable, Lagrange not only survived the Revolution but was treated throughout with honor and respect. In 1790 he served on the committee which proposed the metric system. In 1794 he helped to establish Ecole Polytechnique. He taught elementary mathematics at Ecole Normale (with Laplace (→**33.3**) as his assistant).

He was the last great mathematician of the 18th century. He opened the abstract mathematics of the 19th century. He tried to give a sound foundation to calculus, which was to be given by Cauchy (→**6.11**), Weierstrass (→**13.3b**), and others. To denote derivatives with $'$ was due to Lagrange.

3.6 Variable end points, transversality. Consider

$$S[f] = \int_a^b dx L(f, f', x)dx, \tag{3.10}$$

but now with the unspecified end point values of f. An elementary calculation gives

$$\delta S = \int_a^b dt \left(L_f - \frac{d}{dt} L_{f'} \right) \delta f(t) + L_{f'} \delta f(t)|_a^b + (L - L_{f'} f')_{t=b} \delta b - (L - L_{f'} f')_{t=a} \delta a. \tag{3.11}$$

The first order variations must be killed to be extremal, so f must obey the Euler-Lagrange equation (→**3.3**).

We still have first order terms at the end points. A realistic situation is that the end points are constrained on prescribed curves c_a and c_b. Hence $\delta f|_{t=a} = c_a'(a)\delta a$ and $\delta f|_{t=b} = c_b'(b)\delta b$. Putting these conditions into (3.11), we get the following so-called *transversality conditions*:

$$[L + L_{f'}(c_a' - f')]_{t=a} = 0, \quad [L + L_{f'}(c_b' - f')]_{t=b} = 0. \tag{3.12}$$

These equations give the boundary conditions for the Euler-Lagrange equation to single out its solution.

3.7 Functional derivative. As we noted in **3.1**, calculus of variation is essentially the differential calculus on a function space.[95] As a preliminary step. let us review the differentiation of a scalar valued function of a vector $S(\boldsymbol{f})$. Its (strong) derivative (→**2A.4**) is the gradient of S and is a vector $gradS = (\partial S/\partial f_1, \cdots, \partial S/\partial f_n)$. We have

$$\delta S = \sum_{i=1}^{n} \frac{\partial S}{\partial f_i} \delta f_i. \tag{3.13}$$

Compare this with the formula (3.4). The parallelism becomes almost perfect. if we regard the value $f(a)$ of f at $x = a$ as the a-component of a vector f (see **20.21** for the idea). In this case a is a continuous parameter. so that the summation in (3.13) must be replaced by an integration over the parameter. and we have the form something like:

$$\delta S = \int dx \frac{\delta S}{\delta f(x)} \delta f(x). \tag{3.14}$$

Here the integration kernel $\delta S/\delta f(x)$ is called by physicists the *functional derivative* of S with respect to f. Its functional form can be read off by comparing this formula and the standard variational formula such as (3.4). Hence. the calculation in the proof of Euler's theorem tells us that

$$\frac{\delta S}{\delta f(x)} = \frac{\partial L}{\partial f(x)} - \frac{d}{dx}\frac{\partial L}{\partial f'(x)}. \tag{3.15}$$

for S given in **3.2**.

3.8 Delta function. We ought to be able to differentiate any (well-behaved) functional of f w.r.t. f. For example f itself is a functional of f just as the indentity map maps a vector $\boldsymbol{v}$ to itself. Because $\partial v_i/\partial v_j = \delta_{ij}$ (the identity matrix). we expect the functional derivative of f w.r.t. f itself should be an identity operator (or the integration kernel corresponding to the identity). We introduce δ as follows

$$\frac{\delta f(x)}{\delta f(y)} = \delta(x - y). \tag{3.16}$$

For any (integrable) variation δf. (3.4) in the present case reads[96]

$$\delta f(x) = \int dy \delta(x - y) \delta f(y). \tag{3.17}$$

[95] A set with a certain structure is often called a space. In the case of functional analysis. it is often a linear space. That is. linear combinations of the elements in the space are again in the space.

[96] The reader must remember that the definition of the delta 'function' is inseparable from the definition of the integral being used (→**20.25**).

$\delta(x-y)$ is called the *delta function*,[97] and later mathematically rationalized by Schwartz as a generalized function (→**14**). We will encounter this object later in many other contexts.

3.9 Formal rules of functional differentiation. With respect to the functional differentiation, the ordinary integration and differentiation just correspond to procedures to make linear combinations of the components of the vectors, so that we may freely change the order as

$$\frac{\delta f'(x)}{\delta f(y)} = \frac{d}{dx}\delta(x-y). \tag{3.18}$$

or

$$\frac{\delta}{\delta f(y)}\int_a^b dx f(x) = \int_a^b dx\delta(x-y). \tag{3.19}$$

Furthermore, the chain rule holds as

$$\frac{\delta F(f(x))}{\delta f(y)} = F'(f(x))\delta(x-y), \tag{3.20}$$

where F is a function. Hence, we can obtain Euler's equation **3.2** quite mechanically as follows:

$$\frac{\delta S[f]}{\delta f(y)} = \int dx\left[\frac{\partial L}{\partial f(x)}\delta(x-y) + \frac{\partial L}{\partial f'(x)}\frac{d}{dx}\delta(x-y)\right] = \frac{\partial L}{\partial f(y)} - \frac{d}{dy}\frac{\partial L}{\partial f'(y)}. \tag{3.21}$$

Here integration by parts has been used. examples needed

3.10 Intuitive introduction to minimization of functional I. Suppose we wish to minimize a well-behaved functional $S[f]$. **3.9** tells us that the essence of Euler's theorem **3.2** is that the necessary condition is

$$\frac{\delta S}{\delta f} = 0. \tag{3.22}$$

This is quite parallel to the ordinary calculus (→**A3.19**). Therefore, it is tempting to seek more parallelisms. To this end we need an analogue of the second derivative.

3.11 Second variation. If the change of $S[f]$ can be written as

$$S[f+h] = S[f] + \varphi_1[h] + \varphi_2[h] + o[\|h\|], \tag{3.23}$$

[97] Physicists seem to believe that this was introduced by Dirac, but actually, this has been used for more than 100 years.

where φ_1 is a linear functional and φ_2 is a bilinear form, we say S has the *second variation*. In physicists' way, we can write

$$\varphi_2[h] = \frac{1}{2}\int dxdy \frac{\delta^2 S}{\delta f(x)\delta f(y)} h(x)h(y). \tag{3.24}$$

Actually, we can formally compute the second functional derivative as explained in **3.9**.

3.12 Legendre's condition. If we can write

$$S[f] = \int_a^b dx L(f(x), f'(x), x), \tag{3.25}$$

then the second variation can be written as (after integration by parts, taking into account $h(a) = h(b) = 0$)

$$\varphi_2[h] = \int_a^b (Qh^2 + Ph'^2)dx, \tag{3.26}$$

proof
\int \der{}{x}(h^2w)dx

only valid locally.

where

$$Q = \frac{1}{2}\left(L_{ff} - \frac{d}{dx}L_{ff'}\right), \quad P = \frac{1}{2}L_{f'f'}. \tag{3.27}$$

A necessary condition for φ_2 to be nonnegative is

$$P \geq 0. \tag{3.28}$$

3.13 Intuitive introduction to minimization of functional II. Legendre wished to establish a necessary and sufficient condition for the minimization of $S[f]$. Naturally, he guessed that the nonnegativity of the second variation as a sufficient condition. Therefore, he wished to claim that $P \geq 0$ in **3.12** was a sufficient condition, but failed to prove the assertion. Actually, the assertion is false, because the condition is only local. That is, if we change f only locally in space, indeed $P \geq 0$ implies the positivity of the second variation. However, a small change of f need not be spatially locally confined, and for such changes $P \geq 0$ does not guarantee the positivity of the second variation. We need a supplementary global condition. The final form of a sufficient condition reads:

3.14 Theorem [Sufficient condition for minimum w.r.t. C^1-norm][98] A sufficient condition for g to give a minimum of (3.2) is

[98] Gel'fand and Fomin, Section 24 Theorem.

(1)+(2)+(3) below:
(1) g satisfies Euler's equation (3.3).
(2) $\partial^2 L/\partial f'^2(g, g', x) > 0$.
(3) The interval $[a, b]$ does not contain the conjugate point[99] of a.□

no more max beyond

We need (3) to exclude the global pathology. (To understand the meaning of (3) consider the shortest distance between the points on a great circle of a 2-sphere.) The global condition cannot be derived easily by a formal consideration alone.

Discussion. Discuss the relation between the conjugate point and focus in geometrical optics.

Notice that we do not need eq of motion

3.15 Noether's theorem.[100] Let the functional S in **3.2** be invariant under the following one to one map $g_\alpha : (x, f) \to (x^*, f^*)$, where $x^* = \varphi(x, f, \alpha)$ and $f^* = \psi(x, f, \alpha)$ such that $x = x^*$ and $f = f^*$ for $\alpha = 0$.[101] That is, $S[f] = S[f^*]$. We assume the transformation is differentiable with respect to α. Then, along each stationary curve, the following quantity is constant:

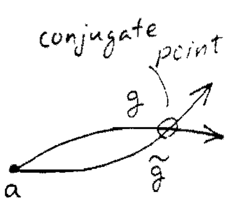
$$\hat{\psi} F_{f'} + (F - f' F_{f'})\hat{\phi} = \text{const.} \tag{3.29}$$

Here $\hat{}$ denotes the partial derivative w.r.t. α evaluated at $\alpha = 0$.□
This should be easily demonstrated, if we look at the calculation in **3.6**.

3.16 Usefulness of variational principle. As we see in **3.18**, if we could cast a (partial) differential equation in the variational principle form (i.e., if we know the variational functional whose Euler's equation (→**3.2**) is the desired equation), then there is a means to get its solution numerically, at least approximately. Hence, to construct a variational principle (if any) is of practical importance. The following Vainberg's theorem tells us when we can expect a variational principle.

3.17 Vainberg's theorem. Suppose
(1) N is an operator from a Hilbert space (→**20.3**) into its conjugate space.
(2) N has a linear Gâteaux[102] derivative $DN(u, h)$ at every point of the

This does not cover the cases with integral



---

[99] Let g and $\tilde{g}$ be two solutions of (3.3) starting from point a. The *conjugate point* of a is the crossing point of g and $\tilde{g}$ in the limit of $\tilde{g} \to g$ in the C^1-norm (→**3.3** footnote).

[100] The theorem can be restated as: If a system is invariant under a continuous symmetry operation, then the corresponding generator of the symmetry operation is an integral of motion.

[101] That is, $\{g_\alpha\}$ is a one parameter transformation group, and $\alpha = 0$ corresponds to the unit element.

[102] A functional $\mathcal{N}[f]$ is said to be Gâteaux differentiable if there is a linear operator

ball $\|u - u_0\| < \epsilon$[103] for some positive ϵ.
(3) The scalar product $\langle h_1, DN(u, h_2)\rangle$ is continuous at every point of D.
Then, a necessary and sufficient condition for $N(u) = 0$ to be the Euler-Lagrange equation of a variational functional in the ball D is the symmetry

$$\langle h_1, DN(u, h_2)\rangle = \langle h_2, DN(u, h_1)\rangle. \tag{3.30}$$

A desired variational functional is given by

$$F(u) = -\int dt \int_0^1 d\lambda u N(\lambda u). \tag{3.31}$$

Here $-$ is only cosmetic. □[104]

3.18 Direct method. The Euler-Lagrange equation often becomes a complicated partial differential equation, so a method to use approximation sequence directly in the variational functional was conceived.[105] For a functional $S[f]$, let us assume that it has an infimum $\inf_f S[f] = \mu > -\infty$. Then, due to the definition of infimum, there is a sequence $\{f_n\}$ such that $S[f_n] \to \mu$. Such a sequence is called a *minimization sequence*.
Theorem.[106] If this sequence has a limit $\hat{f}$, and S is lower semicontinuous,[107] then

$$\lim_n S[f_n] = S[\hat{f}]. \tag{3.32}$$

□

3.19 Ritz's method. To construct a minimization sequence, Ritz used a complete function set (practically an orthonormal basis →**20.10**) $\{u_n\}$:

$$f_n = \sum_{j=1}^{n} c_j u_j. \tag{3.33}$$

Q such that for a function f, g and for sufficiently small λ $N[f+\lambda g] \simeq N[f]+\lambda Q[f]g$. This is a much weaker condition than the strong differentiability (→**2A.4**).

[103] $\|\ \|$ is the C^1 norm we discussed in the footnote of **3.3**.

[104] Actual applications can be seen in: R. W. Atherton and G. M. Homsy, "On the existence and formulation of variational principles for nonlinear differential equations", *Studies Appl. Math.* **LIV**, 31-60 (1975), and the references cited therein. For ODE see I. A. Anderson and G. Thompson, *The inverse problem of the calculus of variations for ordinary differential equations*, Memoirs of Am. Math. Soc. **98**, Number 473 (1991).

[105] Already Euler used it.

[106] Gel'fand and Fomin, Section 36.

[107] That is, for $\epsilon > 0$ there is $\delta > 0$ such that for any h such that $|h| < \delta$ $S[f + h] - S[f] > -\epsilon$.

Let μ_n be the minimum of $S[f_n]$ obtained by varying the coefficients in f_n. Then, obviously $\{\mu_n\}$ is a monotonically decreasing sequence.

Theorem. If $S[f]$ is continuous, and the function set $\{u_n\}$ is complete,[108] then μ_n converges to the desired minimum μ.

3.20 Why variational principle? The study of variational calculus was initiated to understand or to organize classical mechanics. The fundamental equation of motion is given as Newton's equation of motion. But why is this the form chosen by Creator? Under the strong influence of Christianity they thought the equation had to be a special one, for example, characterized by a sort of maximum or minimum principle. Thus a variational principle was pursued. Such a reasoning may sound irrational, but all the creative activities must have irrational components. We should not forget that Newton was a serious student of alchemy (his hair contains large amount of mercury, because he tasted reaction products) and the Bible chronology; his research was almost a religious activity to glorify God: he was a devout Unitarian. John Keynes wrote that Newton was the last magician.

3.21 Hamilton-Jacobi's equation, Jacobi's theorem, etc. These are best understood in the context of classical mechanics, so they will not be covered here. Although there is no balanced modern textbook of classical mechanics, read the first and the last chapters of Landau-Lifshitz, *Classical Mechanics* to start with. For a more serious student, V I Arnol'd, *Mathematical Methods of Classical Mechanics* (Springer, 1979) is recommended. Especially read all the appendices.

[108] Roughly speaking, this means that any function can be described as a linear combination of this set of functions (→**17.3**).