October 10, 2006

Introduction to variational calculus: Lecture notes[1]

Edwin Langmann

Mathematical Physics, KTH Physics, AlbaNova, SE-106 91 Stockholm, Sweden

Abstract

I give an informal summary of variational calculus (complementary to the discussion in the course book).

Aims (what I hope you will get out of these notes):
(i) know a few important examples of variational problem and why it is useful to know how to solve then,
(ii) how to solve certain standard types of variational problems,
(iii) reformulate ODE and PDE problem as variational problems, and, in particular, know Hamilton's principle,
(iv) the chapter on variational calculus in the course book is rather brief, and thus the present notes should provide a more detailed presentation of the theory needed to solve the variational problems in the exercise classes.

WARNING: Beware of typos: I typed this in quickly. If you find mistakes please let me know by email.

1. Examples of variational problems

We now formulate a few standard such problems. We first will give an intuitive description and then a mathematical formulation. Further below we will discuss methods how to solve such problems.

Problem 1: *Find the shortest path in the plane connecting two different points A and B.*

Of course we intuitively know the answer to this problem: the straight line connecting these points. However, is instructive to give a precise mathematical formulation of this: Let (x_0, y_0) and (x_1, y_1), $x_0 < x_1$, be the Cartesian coordinates of the points A

[1]I thank Christian Adåker for pointing our several typos.

and B, and $y(x)$ a C^2 function,[2] $x_0 \leq x \leq x_1$, such that $y(x_i) = y_i$ for $i = 0, 1$. Then $y(x)$ describes a path connecting A and B, and the length of this path is

$$L = \int_A^B ds = \int_{x_0}^{x_1} \sqrt{dx^2 + dy^2}.$$

Using $dy = y'(x)dx$ we get

$$L[y] = \int_{x_0}^{x_1} \sqrt{1 + y'(x)^2} dx. \tag{1}$$

We now can rephrase Problem 1 in a mathematical language:

Problem 1': *Determine the real valued C^2 function y on $[x_0, x_1]$ (i.e. y is given by $y(x)$ with x in $[x_1, x_0]$), fixed at the end points: $y(x_i) = y_i$ for $i = 0, 1$, and such that the integral in (1) is minimal.*

The integral in (1) is an example of a *functional*: We have a set of functions F, i.e., the set of all C^2 function y on the interval $[x_0, x_1]$ such that $y(x_i) = y_i$ for $i = 0, 1$, and $L[y]$ is assigns to each y in F a real number. Moreover, Problem 1' is a *variational problem*: we look for the function in F minimizing the functional $L[y]$.

In general, **a functional is a mapping that assigns to each element in some function space a real number, and a variational problem amounts to searching for functions which are an extremum (minimum, maximum, or saddle points) of a given functional.**

In these notes we restrict ourselves to methods allowing to find extrema, and I will *not* discuss how one can determine if a known extremum is a minimum, a maximum, or a saddle points. One reason for that is that there is no general, simple method available for that, another that in many variational problems it is not so interesting to know this. Fortunately, in some important variational problems it is obvious (by some special reasons) what kind of extremum one has: for example, in Problem 1 any extremum must be a local minimum (since one always can always increase the length of a given the path by a further deformation), and a similar remark applies to Problem 2 below.

While in the case above it is possible to "guess" the solution, it is easy to formulate similar problems where the solution no longer is obvious:

[2] A C^2 function is one that is can be differentiated twice. We will not be very careful about such technicalities and assume that all functions are sufficiently "nice" that all derivatives we write are are well-defined: We only write "C^2" at some places to remind us that there is some condition on the functions concerning differentiability. A more precise mathematical discussion would discuss this in more depth.

Problem 2: *Find the shortest/extremal path on a given surface embedded in three dimensions which connects two given points A and B on this surface.*

To formulate this problem in a mathematical language I assume the surface is given by a C^2 function $z = z(x, y)$ in Cartesian coordinates, and that $x_i, y_i, z_i = z(x_i, y_i)$ for $i = 0$ and $i = 1$ are the Cartesian coordinates of the two points A and B, respectively. Then the path connecting A and B can be parametrized by a vector valued function $\mathbf{x}(t) = (x(t), y(t))$, $0 \leq t \leq 1$, and such that $\mathbf{x}(0) = \mathbf{x}_0 = (x_0, y_0)$ and $\mathbf{x}(1) = \mathbf{x}_1 = (x_1, y_1)$. The length of such a path is

$$L = \int_A^B ds = \int_A^B \sqrt{dx^2 + dy^2 + dz^2},$$

and using $dx = x'(t)dt$, $dy = y'(t)dt$, and $dz = [z_x(x(t), y(t))x'(t) + z_y(x(t), y(t))x'(t)]dt$ we get

$$L[\mathbf{x}] = \int_0^1 \sqrt{x'(t)^2 + y'(t)^2 + [z_x(x(t), y(t))x'(t) + z_y(x(t), y(t))x'(t)]^2} dt. \qquad (2)$$

$L[\mathbf{x}]$ again is a functional (now on a different space of functions), and mathematically we again have a variational problem, but now of a function with values in $\mathbb{R}^2$:

Problem 2': *Find the C^2 function $\mathbf{x}$ on the interval $[0, 1]$ with values in (a subset of) $\mathbb{R}^2$ (i.e. $\mathbf{x}$ is given by $\mathbf{x}(t) = (x(t), y(t))$ with t in $[0, 1]$), fixed at the end points: $\mathbf{x}(0) = \mathbf{x}_0$ and $\mathbf{x}(1) = \mathbf{x}_1$, and which minimizes the functional in (2).*

The following is an example of a variational problem for a function in two variables:

Problem 3: *A homogeneous membrane is fixed in a given frame (which is not even). Find the shape of the membrane.*

A possible shape of the membrane can be described by a function $u(\mathbf{x})$ where $\mathbf{x} = (x, y)$ is a point in a two dimensional domain Ω,[3] where the boundary $\partial\Omega$ parametrizes the frame where $u(\mathbf{x})$ is fixed: $u(\mathbf{x})$ on $\partial\Omega$ equals some given functions $\alpha(\mathbf{x})$. [To have a specific example: assume that the membrane is a circular disc of radius R. The $u(\mathbf{x})$ is a function defined for $r < R$ such that $u(\mathbf{r})|_{r=R} = \alpha(\theta)$ where $(x, y) = r(\cos(\theta), \sin(\theta))$ are polar coordinates.]

The physical principle determining the shape of the membrane is: "the potential energy is a minimum". The potential energy of a homogeneous membrane can be assumed to the proportional to is area,

$$A[u] = \int_\Omega dS = \int_\Omega \sqrt{1 + u_x(x, y)^2 + u_y(x, y)^2} dxdy \qquad (3)$$

[3] To be mathematically precise we should specify further what kinds of Ω we allow: the boundary should be a nice curve (which is differentiable) etc., but for simplicity we will ignore such technicalities here and in the following.

(since, to increase the area, one needs to stretch the membrane, which increases its potential energy; see our course book for a more detailed motivation); here and in the following, subscriped variables indicate differentiation, e.g. $u_x = \partial u/\partial x$ etc. For membrane shapes which do not deviate much from the flat membrane $u = 0$ one can approximate this functional by expanding the square root in a power series and only taking the leading non-trivial terms, $A \approx \int [1+\frac{1}{2}(u_x^2+u_y^2)]dxdy$. Dropping the irrelevant constant term $\int dxdy$ we get the following (approximate) energy functional,

$$J[u] = \int_\Omega \frac{1}{2}\nabla u(\mathbf{x})^2 d^2x = \int_\Omega \frac{1}{2}[u_x(x,y)^2 + u_y(x,y)^2]dxdy. \tag{4}$$

We thus have the following mathematical problem:

Problem 3': *Find the real valued functions u on the domain Ω in $\mathbb{R}^2$ which equals to the function α on $\partial\Omega$ and which minimizes the functional in (4).*

The following is a famous variational problem with a constraint given by a functional:

Problem 4: *Find shape of a homogeneous chain fixed in the two end points in the earth gravitational field.*

We can model the chain by a function $y(x)$, $x_0 \leq x \leq x_1$, where $(x_i, y_i = y(x_i))$ for $i = 0, 1$ the the two fixed end points. The physical principle determining the shape is again: "the potential energy is a minimum". We thus need to find a mathematical expression for the potential energy: we can think the the chain as a collection of small parts labeled by x: each part has the length $ds = \sqrt{1 + y'(x)^2}dx$ and mass $dm = \rho ds$ where ρ is the mass density (in kg/m) of the chain. The potential energy of such part is $gy(x)dm = g\rho y(x)ds$, and thus the total energy of the chain is

$$E_{\text{pot}} = \rho g \int_{x_0}^{x_1} y(x)ds.$$

It is important to note that the length of the chain is

$$L = \int_{x_0}^{x_1} ds,$$

and when searching for the function minimizing the potential energy we only can allow those where L equals the given length L_0 of the chain:

Problem 4': *Find the real-valued function y on $[x_0, x_1]$ fixed at the end points, $y(x_i) = y_i$ for $i = 0, 1$, and minimizing the functional*

$$J[y] = \int_{x_0}^{x_1} y(x)\sqrt{1 + y'(x)^2}dx \tag{5}$$

under the constraint

$$L[y] = \int_{x_0}^{x_1} \sqrt{1 + y'(x)^2} dx = L_0 \tag{6}$$

for a given L_0.

There are many more examples, of course...

2. Method of solution

As discussed, a variational problem amounts to extremizing some functional. The standard method to solve such a problem is to convert it into a ODE or PDE problem. It also can be useful to reformulate a given ODE or PDE problem as variational problem: this can provide a starting point for an approximate solution.

2.1 Functionals depending on one function of one variable.

In this section I discuss in detail the simples kind of variational problems. All essential ideas behind variational calculus appear already here, and the generalization of these ideas to more complicated cases in the following sections is quite straightforward (I will be rather brief there).

Problem 1 above is a (rather trivial) example of the following kind of problem:

Problem A: *Given a functional*

$$J[u] = \int_{x_0}^{x_1} F(u(x), u'(x), x) dx \tag{7}$$

where $F = F(u, u', x)$ *is a* C^2 *function of three real variables. Find the function(s)* u *on the interval* $[x_0, x_1]$ *fixed at the end points,* $u(x_i) = y_i$ *for* $i = 0, 1$*, and extremizing this functional.*

To avoid confusion I should make clear here that I slightly abuse notation and use the symbols u and u' in two different meanings: firstly, in $F(u, u', x)$ as a symbols for the first and second variable of the function F, and secondly, as symbol for the function u on $[x_0, x_1]$ and its derivative. This is not the same, and a pedant would have written $F = F(r, p, x)$ and used different symbols. However, once one has thought this through carefully there is no more need to make such distinctions: as we will see, it is quite useful to be somewhat vague about what one means by u and u' since this avoids lengthy notation. However, I know that this "identifying u with $u(x)$ and u' with $u'(x)$" which is used in many physics texts (and here) can be quite confusing for beginners, and I hope my remarks here and below will help to avoid this confusion. (Similar remarks apply to my discussions of other variational problems below: in these cases I will also abuse notation in a similar manner without further comments.)

Below I derive and (try to) explain the following important

Fact A: *All solutions of the variational Problem A above satisfy the following so-called* **Euler-Lagrange equations***,*

$$\frac{\partial F}{\partial u(x)} - \frac{d}{dx}\frac{\partial F}{\partial u'(x)} = 0 \tag{8}$$

with the boundary conditions $y(x_i) = y_i$ *for* $i = 0, 1$.

Remarks on notation: A few important remarks are in order here: as discussed, the function F depends on three variables: $F = F(u, u', x)$, and in the integrand in (7) the first variable u is replaced by $u(x)$ and the second variable u' by $u'(x)$.[4] Then $\frac{\partial F}{\partial u(x)}$ means $F_u(u(x), u'(x), x)$ (partial derivative of $F(u, u', x)$ with respect to the first variable, and replacing u by $u(x)$ and u' by $u'(x)$), and similarly $\frac{\partial F}{\partial u'(x)}$ means $F_{u'}(u(x), u'(x), x)$. Thus if, e.g.,

$$F = F(u, u', x) = (u')^2 + xu^3$$

then

$$\frac{\partial F}{\partial u(x)} = 3xu(x)^2, \quad \frac{\partial F}{\partial u'(x)} = 2u'(x).$$

It is also important to distinguish partial and total derivative here: we have

$$\frac{d}{dx}\frac{\partial F}{\partial u'(x)} = \frac{d}{dx}\frac{\partial F(u(x), u'(x), x)}{\partial u'(x)} = \frac{\partial^2 F}{\partial u(x)\partial u'(x)}u'(x) + \frac{\partial F}{\partial u'(x)^2}u''(x) + \frac{\partial^2 F}{\partial x \partial u'(x)}$$

etc. If you find this confusing I recommend that you try to use the more careful notation when trying to understand the derivation below — I trust that you will soon be convinced that the (common) abuse of notation I use is much more convenient.

Derivation of the Fact above. Step I: The idea is to consider a small variation the the function, change $y(x)$ to a functions $y(x) + \varepsilon\eta(x)$ where ε is a "small" parameter and $\eta(x)$ is a "small variation". Since the end points are fixed we only consider functions $\eta(x)$ such that

$$\eta(x_i) = 0 \quad \text{for } i = 0, 1. \tag{9}$$

In general, the change of the functional by this variation:

$$\delta J = J[y + \varepsilon\eta] - J[y],$$

will be proportional to ε. However, if y is an extremum of $J[y]$ then the variation will vanish faster. *Thus, the function y on $[x_0, x_1]$ with fixed end points is an extremum of the functional in (7) if*

$$\left.\frac{d}{d\varepsilon}J[y + \varepsilon\eta]\right|_{\varepsilon=0} = 0 \tag{10}$$

[4] A pedant would have written: "... $F = F(r, p, x)$ ... r by $u(x)$ ... p by $u'(x)$."

for all η *on* $[x_0, x_1]$ *such that* $\eta(x_i) = 0$ *for* $i = 0, 1$.

Analogy providing additional motivation: It might be useful at this point to recall a similar problem which you know very well: *Find the extremal points of a* C^2 *function* $f(\mathbf{x})$ *in* N *variables,* $\mathbf{x} = (x_1, x_2, \ldots, x_N)$. To check if some point $\mathbf{x}$ is an extremum we can "go a little away from this point" and check how the function changes: $\mathbf{x}$ is extremum if, for an arbitrary vector $\boldsymbol{\eta}$, $[f(\mathbf{x} + \varepsilon\boldsymbol{\eta}) - f(\mathbf{x})]/\varepsilon$ vanishes as $\varepsilon \to 0$ or, equivalently, if

$$\frac{d}{d\varepsilon} f(\mathbf{x} + \varepsilon\boldsymbol{\eta})\bigg|_{\varepsilon=0} = 0 \quad \text{for all } \boldsymbol{\eta} \in \mathbb{R}^N.$$

Note that $\boldsymbol{\eta}$ specifies the direction in which "we go away from the point $\mathbf{x}$", and to know if we have an extremum we need to check all possible directions.

By the chain rule this is equivalent to the following well-known necessary condition for extremal points, of course: $\nabla f(\mathbf{x}) = 0$ (i.e., $\partial f(\mathbf{x})/\partial x_j = 0$ for $j = 1, 2, \ldots, N$).

Derivation of the Fact above. Step II: We now show how to derive (8) from (10): We compute (to come from the first to the second line below we interchange differentiation with respect to ε and integration, and we use the chain rule of differentiation)

$$\begin{aligned} \frac{d}{d\varepsilon} J[y + \varepsilon\eta]\bigg|_{\varepsilon=0} &= \frac{d}{d\varepsilon} \int_{t_0}^{t_1} F(u(x) + \varepsilon\eta(x), u'(x) + \varepsilon\eta'(x), x) dx|_{\varepsilon=0} \\ &= \int_{t_0}^{t_1} \Big[\frac{\partial F}{\partial u(x)} \eta(x) + \underbrace{\frac{\partial F}{\partial u'(x)} \eta'(x)}_{= \frac{d}{dx}(\frac{\partial F}{\partial u'(x)} \eta'(x)) - \eta(x) \frac{d}{dx} \frac{\partial F}{\partial u'(x)}} \Big] dx \\ = \frac{\partial F}{\partial u'(x)} \eta(x) \bigg|_{x=x_0}^{x_1} &+ \int_{x_0}^{x_1} \eta(x) \left(\frac{\partial F}{\partial u(x)} - \frac{d}{dx} \frac{\partial F}{\partial u'(x)} \right) dx \end{aligned} \tag{11}$$

which must hold true for every C^2 function η such that $\eta(x_i) = 0$ for $i = 0, 1$. The key step here was a partial integration so that the integrand is of the form $\eta(x) \times (\cdots)$, and by that we get a boundary term (the first term). Since $\eta(x_i) = 0$ for $i = 0, 1$ this boundary term vanishes for all "allowed" variational functions η, and we can conclude that (10) holds true if and only if (8) holds true. The completes our derivation of the Fact above.

Remark: In the last step we used the following fact: *The integral*

$$\int_{x_0}^{x_1} \eta(x) F(x) dx$$

vanishes for all C^2 *functions* η *on* $[x_0, x_1]$ *if and only if* $F(x) = 0$. This is very plausible: take a sequence of functions $\eta(x)$ converging to $\delta(x - x_0)$, and we get $F(x_0) \to 0$ for arbitrary x_0 (strictly speaking this fact requires a more careful proof which is beyond the scope of my notes).

Boundary conditions. Above we considered a variational problem where the allowed functions were fixed at the end points. There are other such problems where this is not the case, and then one also gets Euler-Lagrange equations but with different boundary conditions. The derivation of the Euler-Lagrange equations above is important since it also shows how the correct boundary conditions are. To show that I consider the derivation of the following variant of Fact A above:

Fact B: *All functions $u(x)$ extremizing the functional $J[u]$ in (7) (without any constraints at the end points!) are solutions of the Euler-Lagrange equations in (8) with the boundary conditions*

$$\left.\frac{\partial F}{\partial u'(x)}\right|_{x=x_0} = 0, \quad \left.\frac{\partial F}{\partial u'(x)}\right|_{x=x_1} = 0. \tag{12}$$

This is easy to understand from the computation in (11): the boundary term in the last line now no longer vanishes automatically (since $\eta(x_0)$ and $\eta(x_1)$ can now be arbitrary), and to have it zero we have to require (12). Of course, one could also have a variational problem where $\eta(x_0)$ is fixed but $\eta(x_1)$ is arbitrary - in this case we only get the second condition in (12), etc.

Remark on how to solve such Euler-Lagrange equations: In general, the Euler-Lagrange equation in (8) is a ODE of order 2 (since $\partial F/\partial u'(x)$ typically contains some power of $u'(x)$, which, when differentiated, gives a term $u''(x)$). The general solution of this equation thus contains two arbitrary constant which need to be fixed by two conditions: either the conditions in Fact A above, or otherwise the conditions in Fact B.

Anyway, solving a 2nd order ODE is not always easy, and there are two important special cases where the problem can be simplified to a 1st order ODE:

Fact C: *(a) For a function $F = F(u'(x), x)$ independent of $u(x)$, the Euler-Lagrange equations in (8) are equivalent to*

$$\frac{\partial F}{\partial u'(x)} = C \tag{13}$$

for an arbitrary integration constant C.

(b) For a function $F = F(u(x), u(x))$ independent of x, the Euler-Lagrange equations in (8) are implied by

$$u'(x)\frac{\partial F}{\partial u'(x)} - F = C \tag{14}$$

for an arbitrary integration constant C.

In these two cases we only need to solve a 1st order ODE which is a lot simpler. The proof of this is simple: For (a) we note that F being independent of $u(x)$ means that $\partial F/\partial u(x) = 0$ and thus (8) becomes

$$\frac{d}{dx}\frac{\partial F}{\partial u'(x)} = 0.$$

equivalent to (13). To prove (b) we differentiate (14) with respect to x:

$$\frac{d}{dx}\left(u'(x)\frac{\partial F}{\partial u'(x)} - F\right) = 0,$$

and by a simple computation using the chain rule we see that this is equivalent to

$$u''(x)\frac{\partial F}{\partial u'(x)} + u'(x)\frac{d}{dx}\frac{\partial F}{\partial u'(x)} - \frac{\partial F}{\partial u(x)}u'(x) - \frac{\partial F}{\partial u'(x)}u''(x) = 0$$

where we used $\partial F/\partial x = 0$. The terms proportional to $u''(x)$ cancel, and pulling out the common factor $u'(x)$ in the two remaining terms we get the Euler-Lagrange equations in (8).

I stress that Fact C is important for solving many examples. To illustrate this I consider the following example: *Extremize the functional*

$$S[x] = \int_{t_0}^{t_1} \left(\frac{m}{2}x'(t)^2 - V(x(t), t)\right). \tag{15}$$

As I discuss below, this is the action functional for a particle with mass $m > 0$ moving in x-direction and in an external potential $V(x, t)$ depending on position and time t. The Euler-Lagrange equations in this example are nothing but Newton's equations,

$$mx''(t) = -\frac{\partial V(x(t), t)}{\partial x(t)} \tag{16}$$

(check that!). Moreover, in case $V(x, t) = V(x)$ is independent of t then Fact C (b) above implies that

$$\frac{m}{2}x'(t)^2 + V(x(t)) = C, \tag{17}$$

which is nothing but energy conservation. This is useful since the latter ODE can be solved (in principle) by separation:

$$\int \frac{dx}{\sqrt{\frac{2}{m}(C - V(x))}} = t + const.$$

This you knew before: the energy of a system is conserved only for potentials which do not explicitly depend on time.

For further examples see the course book.

2.2 Functionals depending on several functions of one variable.

Problem 2 above is a special case of the following generalizing Problem A above:

Fact D: *Consider the following functional*

$$J[\mathbf{x}] = \int_{t_0}^{t_1} F(\mathbf{x}(t), \mathbf{x}'(t), t)dt \tag{18}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ *is a collection of* N *functions (or equivalently, a function with values in* $\mathbb{R}^N$*) and* $F = F(\mathbf{x}, \mathbf{x}', t)$ *a* C^2 *function of* $2n+1$ *variables* $\mathbf{x}$*,* $\mathbf{x}'$*, and* t*. Then the function* $\mathbf{x}$ *on* $[t_0, t_1]$ *extremizes this functional if the following Euler-Lagrange equations are fulfilled:*

$$\frac{\partial F}{\partial x_j(t)} - \frac{d}{dt}\frac{\partial F}{\partial x_j'(t)} = 0 \quad \textit{for all } j = 1, 2, \ldots, N. \tag{19}$$

The derivation is very similar to the one above and we therefore are rather sketchy: The condition for $\mathbf{x}$ to extremize $J[\mathbf{x}]$ is

$$\frac{d}{d\varepsilon} J[\mathbf{x} + \varepsilon\boldsymbol{\eta}]\Big|_{\varepsilon=0} = 0 \tag{20}$$

for variational functions $\boldsymbol{\eta}$ on $[t_0, t_1]$ with values in R^N, and by a computation similar to the one in (11) we find

$$\begin{aligned} \frac{d}{d\varepsilon} J[\mathbf{x} + \varepsilon\boldsymbol{\eta}]\Big|_{\varepsilon=0} &= (\cdots) \\ = \int_{t_0}^{t_1} \Big(\sum_{j=1}^{N} \frac{\partial F}{\partial x_j(t)}\eta_j(t) &+ \sum_{j=1}^{N} \frac{\partial F}{\partial x_j'(t)}\eta_j'(t)\Big)dt \\ = BT + \int_{t_0}^{t_1} \sum_j \eta_j(t)\Big(\frac{\partial F}{\partial x_j(t)} &- \frac{d}{dt}\frac{\partial F}{\partial x_j'(t)}\Big) \end{aligned} \tag{21}$$

where

$$BT = \sum_{j=1}^{N} \frac{\partial F}{\partial x_j'(t)}\eta_j(t)\Big|_{t=t_0}^{t_1} \tag{22}$$

is a boundary term obtained by partial integration, as above. If the function $\mathbf{x}$ is fixed at the end points then this boundary term vanishes, otherwise it will provide some boundary conditions, as above (we will now spell out these different cases in more detail: I hope it will be clear in particular examples, and if not the safe method is to go back to the computation in (21) and treat the boundary term in more detail). Anyway, we see that for the expression in (21) to be zero for *every* function $\boldsymbol{\eta}$ we need the Euler-Lagrange equations in (19) to hold true.

In general this is a system of N ODEs of order 2 which are difficult to solve. Let me mention an

Important example in physics. I. For a mechanical system, the functional

$$S = \int_{t_0}^{t_1} \Big(\text{kinetic energy} - \text{potential energy}\Big) dt$$

is called **action** of this system, and the **Hamilton principle** states that **the time evolution of a system is such that its action is extremal**.

In many important examples the action is a functional of the form

$$S[\mathbf{q}] = \int_{t_1}^{t_1} \sum_{j=1}^{N} \Big(\frac{m_j}{2} q_j'(t)^2 - U(q_1(t), \ldots, q_N(t)) \Big) dt \tag{23}$$

where $\mathbf{q}$ is a function on $[t_0, t_1]$ with values in R^N (e.g. for a system with 2 particles moving in $\mathbb{R}^3$ we have $N = 6$, $m_j = M_1$ (mass of particle 1) for $j = 1, 2, 3$, $m_j = M_2$ (mass of particle 3) for $j = 4, 5, 6$, $\mathbf{q} = (x_1, y_1, z_1, x_2, y_2, z_2)$ the collection of all Cartesian coordinates of the particles, and U the sum of two-body and external potentials). In this case it is easy to see that Hamilton's principle implies the usual Newton's equations by Fact D above:

$$m_j q_j''(t) = -\frac{\partial U(\mathbf{q}(t), t)}{\partial q_j(t)} \quad \text{for } j = 1, 2, \ldots, N \tag{24}$$

(check that!). One thus can use the action functional to define a mechanical system (rather than Newton's equations). This has many advantages…

2.3 Functionals depending on functions of several variable.

Problem 3 above is a special case of the following:

Problem E: *Consider the following functional*

$$J[u] = \int_{\Omega} F(u(x,y), u_x(x,y), u_y(x,y), x, y) dx dy \tag{25}$$

where Ω is some domain in $\mathbb{R}^2$, and $F = F(u, u_x, u_y, x, y)$ a C^2 function in 5 variables. Find the C^2 function(s) u on Ω which are fixed on the boundary $\partial\Omega$ of Ω and which extremize this functional.

As above one can make precise this condition on u by

$$\frac{d}{d\varepsilon}J[u+\varepsilon\eta]\Big|_{\varepsilon=0}=0 \tag{26}$$

for all functions $\eta=\eta(x,y)$ on Ω. The following computation similar to the one in (11),

$$\begin{aligned}
&\frac{d}{d\varepsilon}J[u+\varepsilon\eta]\Big|_{\varepsilon=0}\\
&=\frac{d}{d\varepsilon}\int_\Omega F(u(x,y)+\varepsilon\eta(x,y),u_x(x,y)+\varepsilon\eta_x(x,y),u_y(x,y)+\varepsilon\eta_y(x,y),x,y)dxdy\\
&=\int_\Omega\Big(\frac{\partial F}{\partial u(x,y)}\eta(x,y)+\underbrace{\frac{\partial F}{\partial u_x(x,y)}\eta_x(x,y)}_{\frac{d}{dx}(\frac{\partial F}{\partial u_x(x,y)}\eta(x,y))-\eta(x,y)\frac{d}{dx}\frac{\partial F}{\partial u_x(x,y)}}+\underbrace{\frac{\partial F}{\partial u_y(x,y)}\eta_y(x,y)}_{\cdots}\Big)dxdy\\
&=BT+\int_\Omega\eta(x,y)\Big(\frac{\partial F}{\partial u(x,y)}-\frac{d}{dx}\frac{\partial F}{\partial u_x(x,y)}-\frac{d}{dy}\frac{\partial F}{\partial u_y(x,y)}\Big)dxdy
\end{aligned} \tag{27}$$

where

$$\begin{aligned}
BT&=\int_\Omega\Big(\frac{d}{dx}(\frac{\partial F}{\partial u_x(x,y)}\eta(x,y))+\frac{d}{dy}(\frac{\partial F}{\partial u_y(x,y)}\eta(x,y))\Big)dxdy\\
&=\oint_{\partial\Omega}\Big(\frac{\partial F}{\partial u_x(x,y)}\eta(x,y)dy-\frac{\partial F}{\partial u_y(x,y)}\eta(x,y)dx\Big)
\end{aligned} \tag{28}$$

is a boundary term (we used Green's theorem) which vanishes since u is fixed on $\partial\Omega$, and thus the allowed variation functions η vanish on $\partial\Omega$. We thus can conclude, as above:

Fact E: *All solutions of the Problem E above satisfy the following Euler-Lagrange equations*

$$\frac{\partial F}{\partial u(x,y)}-\frac{d}{dx}\frac{\partial F}{\partial u_x(x,y)}-\frac{d}{dy}\frac{\partial F}{\partial u_y(x,y)}=0 \tag{29}$$

where $u|_{\partial\Omega}$ is fixed.

Again, it is imporant to distinguish partial and total derivatives, e.g.,

$$\frac{d}{dx}\frac{\partial F}{\partial u_x}=\frac{\partial^2 F}{\partial u\partial u_x}u_x+\frac{\partial^2 F}{\partial u_x^2}u_{xx}+\frac{\partial^2 F}{\partial u_y\partial u_x}u_{xy}+\frac{\partial^2 F}{\partial x\partial u_x}$$

(we suppress arguments x,y of u and its derivatives) etc.

Similarly as above there are also modified variational problems where u is allowed to vary on all of $\partial\Omega$ or on parts of it. In this case one also gets the Euler-Lagrange equations but with different boundary conditions: if $u(x,y)$ is allowed to vary on $\partial\Omega$ then one gets the following boundary conditions,

$$\left(n_x \frac{\partial F}{\partial u_x} + n_y \frac{\partial F}{\partial u_y}\right)\Big|_{\partial\Omega} = 0. \tag{30}$$

where $\mathbf{n} = (n_x, n_y)$ is the unit vector normal to the curve $\partial\Omega$. In various important examples one has $F = const.\nabla u^2 + \ldots$ where the dots indicate terms independent of ∇u, and in these cases (30) corresponds to Neumann boundary condtions.

To obtain (30), we note that if $\partial\Omega$ is parametrized by a curve $\mathbf{x}(t) = (x(t), y(t))$, $0 \leq t \leq 1$, then we can write the boundary term above as

$$BT = \int_0^1 \left(\frac{\partial F}{\partial u_x} n_x + \frac{\partial F}{\partial u_y} n_y\right)\Big|_{x=x(t),y=y(t)} \eta(x(t), y(t)) ds \tag{31}$$

where we suppress the arguments and used

$$y'(t)dt = n_x(t)ds, \quad x'(t)dt = -n_y(t)ds, \quad ds = \sqrt{x'(t)^2 + y'(x)^2}dt.$$

If $\eta(x(t), y(t))$ is allowed to be non-zero on (parts of) $\partial\Omega$, we must require the condition in (30) (on these very parts).

It is easy to generalize this to functional of functions u in N variables which, obviously, also are important in physics:

Fact F: *Functions u on the domain Ω in $\mathbb{R}^N$ extremizes the functional*

$$J[u] = \int_\Omega F(u(\mathbf{x}), \nabla u(\mathbf{x}), \mathbf{x}) d^N x, \tag{32}$$

$\nabla u = (u_{x_1}, \ldots, u_{x_N})$, F a C^2 function of $2N+1$ variables u, ∇u, $\mathbf{x}$, if the following Euler-Lagrange equation holds true:

$$\frac{\partial F}{\partial u(\mathbf{x})} - \sum_{j=1}^{N} \frac{d}{dx_j} \frac{\partial F}{\partial u_{x_j}(\mathbf{x})} = 0. \tag{33}$$

The derivation is as above and recommended as exercise to the reader.

The most general case are functionals of functions $\mathbf{u}$ on a domain Ω in $\mathbb{R}^N$ with values in $\mathbb{R}^M$: $\mathbf{u} = (u_1, \ldots, u_M)$ is a function of the variable $\mathbf{x} = (x_1, \ldots, x_N)$. Again, it is straightforward derive the Euler-Lagrange equations for such cases. I only write them down and leave it to the reader to fill in the details:

$$\frac{\partial F}{\partial u_k(\mathbf{x})} - \sum_{j=1}^{N} \frac{d}{dx_j} \frac{\partial F}{\partial (u_k)_{x_j}(\mathbf{x})} = 0 \ \text{ for } k = 1, 2, \ldots, M. \tag{34}$$

This is a coupled system of PDEs and, in general, difficult to solve. I only mention that action functionals used to define important theories in physics like electrodynamics or general relativity are of this type.

Important example in physics. II. Consider the oscillations of a homogeneous membrane described by a function $u = u(\mathbf{x}, t)$ where $\mathbf{x}$ in Ω are spatial coordinates and $t > 0$ time. As discussed in more detail in our course book, the kinetic energy of the membrane is

$$E_{\text{kin}} = \int_\Omega \frac{\rho}{2} u_t(\mathbf{x}, t)^2 dxdy$$

with $\rho > 0$ the mass density of the membrane (in kg/m^2), and the potential energy is

$$E_{\text{pot}} = \int_\Omega \frac{S}{2} [u_x(\mathbf{x}, t)^2 + u_y(\mathbf{x}, t)^2] dxdy$$

with a constant $S > 0$ characterizing material properties of the membrane.

Thus the action functional for the membrane is

$$S[u] = \int_{t_0}^{t_1} \Big(\frac{\rho}{2} u_t^2 - \frac{S}{2} [u_x^2 + u_y^2] \Big) dxdy, \tag{35}$$

(we suppress the common arguments t, x, y of the derivatives of the function u). Hamilton's principle can also be applied to this membrane, and from Fact F we obtain the following equations of motion for the membrane:

$$\rho u_{tt} - S[u_{xx} + u_{yy}] = 0.$$

This is the wave equation which we studied repeatedly in this course, of course. However, using variational calculus together with simple physical reasoning we now *derived* this equation, which otherwise would have been rather difficult.

In a similar manner one can derive our string model derived and discussed extensively in this course in a much simpler way using variational calculus.

2.4 Variational problems with constraints given by functionals.

Problem 4 above is a special case of the following important class of problems:

Problem G: *Given two functionals*

$$J[u] = \int_{x_0}^{x_1} F(u(x), u'(x), x) dx \tag{36}$$

and

$$K[u] = \int_{x_0}^{x_1} G(u(x), u'(x), x) dx \tag{37}$$

where $F = F(u, u', x)$ and $G = G(u, u', x)$ are two C^2 functions of three real variables. Find the function(s) u on the interval $[x_0, x_1]$ fixed at the end points, $u(x_i) = y_i$ for $i = 0, 1$, and extremizing the functional $J[u]$ under the constraint $K[u] = K_0$ where K_0 is a given constant.

This kind of problem can be reduced to a Problem A above by Lagrange's multiplicator method; see Fact G below. I will now give a derivation, but only motivate it by analogy with functions on $\mathbb{R}^N$:

Lagrange's multiplicator method for functions on $\mathbb{R}^N$:[5] Suppose we want to solve the following problem: *Find the minimum of a function f on $\mathbb{R}^N$ with the constraint that another function g on $\mathbb{R}^N$ has a fixed given value g_0.*

Geometrically, all points $\mathbf{x}$ in $\mathbb{R}^N$ satisfying $g(\mathbf{x}) = g_0$ define a hypersurface of dimension $N - 1$ in $\mathbb{R}^N$, and our problem is to extremize f on this hypersurface (for $N = 2$ this hypersurface is a curve, for $N = 3$ a surface, etc.). To check if a point $\mathbf{x}$ fulfills these requirements we check what happens if we go away from this point, remaining on this hypersurface: if $\mathbf{x}$ is on this hypersurface and $\mathbf{x} + \varepsilon\boldsymbol{\eta}$ is on this hypersurface as well then $g(\mathbf{x} + \varepsilon\boldsymbol{\eta}) - g(\mathbf{x}) = g_0 - g_0 = 0$, and taking the derivative of this with respect to ε gives

$$\frac{d}{d\varepsilon} g(\mathbf{x} + \varepsilon\boldsymbol{\eta})\bigg|_{\varepsilon=0} = \boldsymbol{\eta} \cdot \nabla g(\mathbf{x}) = 0. \tag{38}$$

For f to be extremal on this hypersurface we need that

$$\frac{d}{d\varepsilon} f(\mathbf{x} + \varepsilon\boldsymbol{\eta})\bigg|_{\varepsilon=0} = \boldsymbol{\eta} \cdot \nabla f(\mathbf{x}) = 0. \tag{39}$$

for all vectors $\boldsymbol{\eta}$ satisfying the condition in (38) (this means that f only changes faster then linear in ε if we go away from $\mathbf{x}$ in the direction $\boldsymbol{\eta}/|\boldsymbol{\eta}|$ tangent to the hypersurface by a distance $\varepsilon|\boldsymbol{\eta}|$, similarly as discussed after Fact A above). Now $\nabla g(\mathbf{x})$ (for fixed $\mathbf{x}$) is a vector in $\mathbb{R}^N$, and thus (38) defines a hyperplane of dimension $N - 1$ (this is just the tangent plane to the above-mentioned hypersurface, of course). The condition in (39) means that $\nabla f(\mathbf{x})$ is also orthogonal to this very hyperplane, and this obviously can be true if and only if $\nabla f(\mathbf{x})$ and $\nabla g(\mathbf{x})$ are parallel, i.e., there must exist a constant λ such that

$$\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x}). \tag{40}$$

The latter is obviously equivalent to the function

$$f_\lambda(\mathbf{x}) = f(\mathbf{x}) - \lambda g(\mathbf{x}) \tag{41}$$

[5] To understand the following I suggest you write out everything for $N = 2$ first: in this case you can draw pictures and thus nicely visualize the argument geometrically, which I something I recommend.

being extremal: the extremal points $\mathbf{x} = \mathbf{x}_\lambda$ will depend on λ, and the value of λ is then to be fixed so that $g(\mathbf{x}_\lambda) = g_0$. The paramter λ introduced here is often called **Lagrange multiplier**.

To summarize: *The problem "extremize the function f on $\mathbb{R}^N$ with the constraint that another function g on $\mathbb{R}^N$ has a fixed given value g_0" is equivalent to "extremize the function $f - \lambda g$ on $\mathbb{R}^N$ and determine the parameter λ so that the extremal point fulfills the constraint."*

This argument can be generalized to functionals and used to derive the following

Fact G: *Problem G above is equivalent to extremizing the functional*

$$J[u] - \lambda K[u] \tag{42}$$

and then determining the parameter λ so that the constraint $K[u] = K_0$ is fulfilled: To solve Problem G one solves the Euler-Lagrange equations

$$\frac{\partial(F - \lambda G)}{\partial u(x)} - \frac{d}{dx}\frac{\partial(F - \lambda G)}{\partial u'(x)} = 0 \tag{43}$$

(which has solutions u depending on the parameter λ, of course) and then fixing λ by the constraint.

It is easy to generalize this method to other kinds of functions and/or to variational problems with several constraints given by functional: *Minimizing any kind of functional $J[u]$ with constraints $K_J[u] = k_J$, $J = 1, 2, \ldots, L$, is equivalent to minimizing*

$$J[u] - \sum_{J=1}^{L} \lambda_J K_J[u] \tag{44}$$

and then fixing the parameters λ_J so that all constraints are fulfilled: The solutions of the Euler-Lagrange equations which amount to minimizing the functional in (44) will depend on the parameters $\lambda_1, \lambda_2, \ldots, \lambda_L$, and this is exactly the number of free parameters needed to fulfill the L constraints. Again I leave details to the interested reader.

2.5 Variational problems with higher derivatives.

I finally shortly discuss variational problems involving higher derivatives and leading to higher order differential equations. For simplicity I only consider such problems for real-valued functions u of one real variable — the generalization to other cases is straightforward: We consider a functional

$$J[u] = \int_{x_0}^{x_1} F(u(x), u'(x), u''(x), x)dx \tag{45}$$

with $F = F(u, u', u'', x)$ a C^2 function depending now on four variables: the integrand of this functional depends not only on the function u and its derivative u' but also on the the second derivative u''. The Euler-Lagrange equations for such functional are

$$\frac{\partial F}{\partial u(x)} - \frac{d}{dx}\frac{\partial F}{\partial u'(x)} + \frac{d^2}{dx^2}\frac{\partial F}{\partial u''(x)} = 0. \tag{46}$$

It is straightforward to prove this by generalizing our derivation of Fact A in Section 2.1 above — I only note that the last term comes from two partial integrations, and it has a plus sign since $(-)^2 = +$: one minus sign for each partial integration. Note that the equation in (46) is in general a ODE of order 4. We also note that in such problems it is usually worth going though the derivation of the Euler-Lagrange equations from (10) since this allows one to obtain the correct boundary conditions. For example, one may fix u and u' in $x = x_0$ but leave u and u' in $x = x_1$ arbitrary. One then obtains Euler-Lagrange equations with particular boundary conditions which one derives by computing (10) allowing variational functions such that $\eta(x_0) = \eta'(x_0) = 0$ but with $\eta(x_1)$ and $\eta'(x_1)$ arbitrary: keeping track of the boundary terms obtained by the partial integrations one obtains the correct boundary conditions.