14 Generalizations of the Simplest Problem

You have to be odd to be number one
—Dr. Seuss

Clearly the simplest problem is not the only type of variational problem that we can consider. There are several variations that are of fundamental interest for a variety of circumstances. We will not consider all of the possible variations, but will go through in great detail several different possibilities. The goal of this Section is not to give a comprehensive set of Euler equations with corresponding boundary conditions for the most general setting or for every potentially interesting possibility. Instead the goal of the following sections is to go through the derivation for a few different cases so that you (the illustrious reader) will be comfortable confronting other variational settings that are not covered here, and you will be well equipped to face those challenging boundary conditions etc. head on, deriving the corresponding needed equations and conditions.

14.1 Variable endpoints and natural boundary conditions

14.1.1 Variable endpoints

To motivate this first topic, consider Tom Sawyer swimming across the Mississippi river (for those of you that know very much about the Mississippi river, you will recognize that this is a horrible idea as the undercurrents in the Mississippi are notoriously dangerous) while running away from Aunt Polly one warm afternoon. Tom enters the water in a patch of reeds (the boundary condition on one side of the river is fixed), but he doesn't care where he lands on the other side, as long as he gets to the other side of the river (the boundary condition on the other side of the river is free). If we suppose that the riverbank is perfectly straight (a bit of a stretch, but let's start with it) then we have the following type of optimization problem to determine Tom's path if he is trying to optimize his swimming, and we ignore the strength of the current for now. Tom wants to minimize the path across the river given his specific entry point, but allowing himself to land anywhere on the opposite bank, i.e. he may start from $x = a,\ y = c$ and he wants to end at $x = b$ but he really doesn't care what value of y he ends at.

For the purposes of deriving something realistic, we return to something more abstract (because why not?), and consider optimization of the functional

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

over all $y(x)$ with $y(a)$ and $y(b)$ free (this would be like Tom not caring where he enters or exits the river as long as he gets all the way across). As before, we compute the first variation (Gateaux differential) $\delta J(y; h)$, but now the increments $h(x)$ need not satisfy homogeneous BC's because $y(a) + \varepsilon h(a)$ and $y(b) + \varepsilon h(b)$ are not fixed, although $x = a$ and $x = b$ are. Thus

$$\begin{aligned}
\delta J(y; h) &= \frac{\partial}{\partial \varepsilon} \int_a^b L(x, y + \varepsilon h, y' + \varepsilon h')dx\Big|_{\varepsilon=0} \\
&= \int_a^b \left[L_y(x, y, y')h(x) + L_{y'}(x, y, y')h'(x)\right] dx \\
&= \int_a^b \left[L_y(x, y, y') - \frac{d}{dx}L_{y'}(x, y, y')\right] h(x)dx + L_{y'}(x, y, y')h(x)\big|_{x=a}^{x=b},
\end{aligned}$$

where we integrated by parts to reach the last line (as before there is a way to justify everything that follows without resorting to integration by parts for strong solutions, but we will just march on as though integration by parts were the most wonderful tool available to us...HINT it kind a sorta is). For $y(x)$ to be an extremal (optimizer) of $J[y]$, $\delta J(y; h) = 0$ for all $h(x) \in C^1[a, b]$ where now $h(a)$ and $h(b)$ are *NOT* fixed.

- Since this also means that $\delta J(y; h) = 0$ for all $h(x) \in C^1[a, b]$ where $h(x)$ vanishes at a and b (such functions form a subset of the full space $C^1[a, b]$), the (EL) are satisfied:
$$L_y(x, y, y') - \frac{d}{dx}L_{y'}(x, y, y') = 0. \tag{14.1}$$
- In general, if $y(x)$ is an extremal, then $\delta J(y; h) = 0$ for all $h(x) \in C^1[a, b]$ even those for which the endpoints do not vanish. We have just verified however that (EL) hold, so from the derivation above we see that it follows that
$$L_{y'}(x, y, y')h(x)|_{x=a}^{x=b} = 0 \tag{14.2}$$
as well for all $h(x) \in C^1[a, b]$ (regardless of their boundary values).
- This means that
$$L_{y'}|_{x=b}\, h(b) - L_{y'}|_{x=a}\, h(a) = 0,$$
which must be satisfied for all potential values of $h(a)$ and $h(b)$.
- This is only possible if
$$L_{y'}|_{x=b} = L_{y'}|_{x=a} = 0. \tag{14.3}$$

These conditions are called the natural boundary conditions for the problem.

Example 14.1.1. Consider finding the minimum of

$$J[y] = \int_0^1 ((y')^2 + y^2)dx,$$

where $y(0) = 1$ and $y(1)$ is free. The (EL) are

$$2y - 2\frac{d}{dx}y' = 0.$$

- This implies that $y'' = y$ so $y(x) = c_0 \sinh(x) + c_1 \cosh(x)$.
- The condition at $x = 0$ leads to $y(0) = c_1 = 1$ giving $y = c_0 \sinh(x) + \cosh(x)$.
- The natural boundary conditions give $L_{y'}(x = 1) = 0$ so $2y'(1) = 0$. This leads to $y'(1) = c_0 \cosh(1) + \sinh(1) = 0$ or $c_0 = -\tanh(1)$.

Thus
$$y = -\tanh(1) \sinh(x) + \cosh(x).$$

Just to beat a dead problem more than we should, let's consider how this affects the solution to Tom's problem of swimming across the Mississippi.

Example 14.1.2. If Tom is only interested in getting to the other side of the river then his cost function will look like
$$J[y] = \int_a^b \sqrt{1 + (y')^2} dx,$$
where $y(a) = c$ and $y(b)$ is free. This leads to the Euler-Lagrange equations
$$-\frac{d}{dx} \frac{y'}{\sqrt{1 + (y')^2}} = 0,$$
with $y(a) = c$ and natural boundary condition $\frac{y'(b)}{\sqrt{1+y'(b)^2}} = 0$.

Boringly enough, after some manipulation we find that this leads to the constant solution $y = c$ which shouldn't come as any surprise, i.e. the best route for Tom to swim across the river is to go straight across. Although this isn't surprising, the final result is important because it demonstrates a principle that we will see later on. The principle is that the optimal route to cross the river is to reach the other bank in a path that is perpendicular to the opposite river bank.

14.1.2 Endpoint costs

Consider
$$J[y] = \int_a^b L(x, y, y')dx + G(y(b)),$$
where $y \in C^1[a, b]$ and $y(a) = y_0$ is fixed but $y(a)$ is free. This may be of interest to Tom Sawyer if he knows where he enters the river and is primarily interested in getting across, but also wants to avoid the swamp on the other side meaning that certain values of $y(b)$ will 'cost' him more. In the mathematical setting we also need G to be continuously differentiable. Then if we wish to optimize $J[y]$ we compute it's Gateaux differential using increments $h(x) \in C^1[a, b]$ where $h(a) = 0$ to keep $y(a) + \varepsilon h(a) = y_0$ fixed, but $h(b)$ is free. This follows in the steps below:

-
$$\begin{aligned}
\delta J(y; h) &= \int_a^b \left[L_y - \frac{d}{dx} L_{y'} \right] h(x)dx + L_{y'}(x, y, y')h(x)\Big|_{x=a}^{x=b} + \frac{\partial}{\partial \varepsilon} G(y(b) + \varepsilon h(b))\Big|_{\varepsilon=0} \\
&= \int_a^b \left[L_y - \frac{d}{dx} L_{y'} \right] h(x)dx + \left\{ L_{y'}(x, y, y')\big|_{x=b} + G'(y(b)) \right\} h(b) = 0,
\end{aligned}$$
for all $h \in C^1[a, b]$ and $h(a) = 0$ (this condition was used to eliminate one of the boundary terms from the integration by parts).

- Since this must be true for all $h \in C^1[a, b]$ with the left hand boundary condition, then it is true for all $h(x)$ such that $h(b) = 0$, in which case, the boundary terms vanish, and hence (EL) are satisfied, i.e.

$$L_y - \frac{d}{dx} L_{y'} = 0.$$

- Returning to the more general case wherein $\delta J(y; h)$ vanishes for all $h(x)$ that vanish at $x = a$, but not necessarily $x = b$, we see that

$$\left\{ L_{y'}(x = b, y(b), y'(b)) + G'(y(b)) \right\} h(b) = 0,$$

for all possible $h(b)$.

- This implies that

$$L_{y'}(x = b, y(b), y'(b)) + G'(y(b)) = 0. \tag{14.4}$$

This last line is a variation of the natural boundary condition.

Further generalizations, particularly where both $y(a)$ and $y(b)$ are free is immediate. A more complicated setup occurs when we acknowledge that the opposite river bank that Tom is swimming to may not be straight, i.e. the boundary must satisfy a condition like $y(b) = \phi(b)$ where $\phi(x)$ is some function describing the opposite river bank. Such a situation is sufficiently complicated that it deserves its own Section to derive the corresponding boundary conditions (giving you something to look forward to).

Remark 14.1.3. The type of cost functional that is dependent on both the integral of the Lagrangian, and is a function of one of the endpoints appears in many applications, not just when Tom Sawyer is trying to escape his doting Aunt Polly. In particular, in financial applications the maximum utility is desired, but there may be some end state that should be maximized simultaneously (you want to enjoy retirement but not run out of money after 10 years). Another example we will see later is related to the use of chemotherapy drugs. The negative effect of the drug should be minimized, but the final number of cancer cells when the treatment ends should be minimized as well.

Example 14.1.4 (Retirement Planning with an endpoint). Suppose that an individual is planning out their retirement just as in example 13.2.3, however they want to make sure that they don't run out of money by the end time $t = t_f$. Hence, they modify the cost functional to maximize

$$J[r(t)] = \int_0^{t_f} e^{-\beta t} u(r(t)) dt + x(t_f),$$

which indicates that they would like as much remaining capital $x(t)$ at time t_f as possible. Following Example 13.2.3, this leads to the cost functional in terms of $x(t)$ alone:

$$J[x(t)] = 2 \int_0^{t_f} e^{-\beta t} \sqrt{\alpha x(t) - x'(t)} dt + x(t_f),$$

with $x(0) = S$ some initial amount of funds when retirement starts. Following the same derivation as was done previously we can find that because the same (EL) are present that

$$x'(t) = \alpha x(t) - (\alpha S - x'(0)) e^{2(\alpha - \beta)t},$$

where now we also have the endpoint condition

$$-\frac{e^{-\beta t_f}}{\sqrt{\alpha x(t_f) - x'(t_f)}} + 1 = 0,$$

which can be simplified to

$$e^{-2\beta t_f} = \alpha x(t_f) - x'(t_f).$$

Combined with the relationship found above evaluated at $t = t_f$ we find that

$$(\alpha S - x'(0))e^{2(\alpha-\beta)t_f} = e^{-2\beta t_f},$$

or after further manipulation,

$$\begin{aligned} \alpha S - x'(0) &= e^{-2\alpha t_f}, \\ \Rightarrow x'(0) &= \alpha S - e^{-2\alpha t_f}, \end{aligned}$$

which gives the initial rate at which we are spending the hard-earned savings.

Example 14.1.5. Once again we return to Tom trying to cross the Mississippi. Once again Tom is starting from $y(a) = c$ and trying to reach the other side at $x = b$. In this case however, Tom is trying to avoid a marsh at $y = d$ on the opposite shore (Tom is bright enough to know that alligators prefer the marshy part of the river). When computing his planned swimming trajectory (something that all runaways in the 19th century did before swimming across a drastically turbulent and deep river), Tom proposes the following cost functional to help avoid landing near the center of the marsh at $y = d$:

$$J[y] = \int_a^b \sqrt{1 + (y')^2}dx + \frac{\alpha}{(y(b) - d)^2},$$

where α is a tunable parameter that can be used to denote how wide Tom expects the marsh to be.

Following the derivation provided previously we see that Tom's optimal swimming path will be given by the solution of

$$\frac{d}{dx}\sqrt{1 + (y')^2} = 0,$$

where $y(a) = c$ and $\frac{y'(b)}{\sqrt{1+y'(b)^2}} - \frac{2\alpha}{(y(b)-d)^3} = 0$.

- To find the full solution, we recall that the solution to this differential equation is a linear function $y(x) = Ax + B$, where A and B are unknown constants dependent on the boundary conditions given above.
- Using the left hand boundary condition we find that $Aa + B = c$ so that $y(x) = A(x - a) + c$.
- Inserting this into the natural boundary condition, we arrive at

$$\frac{A}{\sqrt{1 + A^2}} - \frac{2\alpha}{(A(b - a) + c - d)^3} = 0,$$

which we can simplify to

$$\frac{A^2}{1+A^2} = \frac{4\alpha^2}{(A(b-a)+c-d)^6},$$

which is a rather messy condition on the slope of the route A that Tom swims. Gratefully, Tom also spent quite a bit of time practicing Newton's method, so he is fully aware of how to find an approximate solution for specific values of a, b, c and d.

14.2 Endpoint on a curve

To motivate this problem we recognize that clearly the Mississippi river does not have a perfectly straight bank, and thus when Tom is trying to reach the opposite shore he is not really aiming for any value of y with $x = b$, but instead he should be looking to land on a curve defined by $\phi(x)$ where $\phi \in C^1(\mathbb{R})$ (banks of the Mississippi are actually quite smooth). Thus Tom wants to start at $y(0) = 0$ (for simplicity we choose his starting point as $(0, 0)$) and is aiming for the point $(x_f, y(x_f))$ where $y(x_f) = \phi(x_f)$ and the function $\phi(x)$ describes the other side of the river bank.

In the general setting, this means that Tom is interested in the optimization problem:

$$J[y] = \int_0^{x_f} L(x, y, y')dx, \tag{14.5}$$

with the constraints on the potential optimizers $y(x)$ that $y(0) = 0$ and $y(x_f) = \phi(x_f)$. Looking at the Gateaux differential of this cost functional, we need to look at variations $y(x) \to y(x)+\varepsilon h(x)$ as well as $x_f \to g(\varepsilon) = x_f + g'(0)\varepsilon + O(\varepsilon^2)$, but guarantee at the same time that the auxiliary conditions are all maintained. Derivation of the Euler-Lagrange equations follows much like we did before, but now there is an additional effect that comes from the endpoint condition at x_f, and guaranteeing that variations in the curve itself still end on the curve $\phi(x)$.

Marching forward, we see that the Gateaux differential is given by:

$$\begin{aligned}
\delta J[y; h] &= \frac{\partial}{\partial \varepsilon} \int_0^{g(\varepsilon)} L(x, y+\varepsilon h, y'+\varepsilon h')dx \Bigg|_{\varepsilon=0} && (14.6)\\
&= \int_0^{x_f} \left[L_y h + L_{y'} h'\right] dx + L(x_f, y(x_f), y'(x_f))g'(0) && (14.7)\\
&= \int_0^{x_f} \left[L_y - \frac{d}{dx} L_{y'}\right] h(x)dx + L(x_f, y(x_f), y'(x_f))g'(0) && (14.8)\\
&+ L_{y'}(x_f, y(x_f), y'(x_f))h(x_f).
\end{aligned}$$

Now the real trick here is that $h(x_f)$ is not arbitrary, i.e. $y(g(\varepsilon)) + \varepsilon h(g(\varepsilon))$ must remain on the curve $\phi(x)$. Hence we can't just say that $\delta J = 0$ for all possible variations $h(x)$ for any potential $h(x_f)$ because x_f is actually varying with ε as well and we must find out how these are related.

- Note that in order for the endpoint condition to be satisfied for all variations $x_f + g'(0)\varepsilon+O(\varepsilon^2)$ and $y(x)+\varepsilon h(x)$ then we must satisfy the constraint that the perturbed endpoint is still on the bank of the river (in Tom's case) or more explicitly in the general setting:

$$y\left(x_f + g'(0)\varepsilon + O(\varepsilon^2)\right) + \varepsilon h\left(x_f + g'(0)\varepsilon + O(\varepsilon^2)\right) = \phi\left(x_f + g'(0)\varepsilon + O(\varepsilon^2)\right). \tag{14.9}$$

- Now we consider a Taylor series of all of these terms in ε, recalling that $y(x_f) = \phi(x_f)$ already, then we see that:

$$y(x_f) + \varepsilon g'(0)y'(x_f) + \frac{\varepsilon^2 g'(0)^2}{2}y''(x_f) + \varepsilon h(x_f) + \varepsilon^2 g'(0)h'(x_f) \tag{14.10}$$

$$= \phi(x_f) + \varepsilon g'(0)\phi'(x_f) + \frac{\varepsilon^2 g'(0)^2}{2}\phi''(x_f) + O\left(\varepsilon^3\right) \tag{14.11}$$

$$\Rightarrow g'(0)y'(x_f) + h(x_f) = g'(0)\phi'(x_f), \tag{14.12}$$

  where the final equation comes from the $O(\varepsilon)$ terms (the $O(1)$ terms are already satisfied).

- Hence we see that

$$h(x_f) = g'(0)(\phi'(x_f) - y'(x_f)). \tag{14.13}$$

- Inserting this back into δJ we see that

$$\delta J[y;h] = \int_0^{x_f} \left[L_y - \frac{d}{dx}L_{y'}\right] h(x)dx \tag{14.14}$$

$$+ \left[L(x_f, y(x_f), y'(x_f)) + (\phi'(x_f) - y'(x_f))L_{y'}(x_f, y(x_f), y'(x_f))\right] g'(0) = 0, \tag{14.15}$$

  for all variations $h(x)$ with $h(0) = 0$ and all potential linear variations in the endpoint $g'(0)$, i.e. at this point we have enforced the dependence of $g'(0)$ on the endpoint $h(x_f)$ so in essence $g'(0)$ is arbitrary.

As this must apply for all possible $g'(0)$ we see that the standard (EL) are satisfied, i.e. $\delta J = 0$ even for $g'(0) = 0$:

$$L_y - \frac{d}{dx}L_{y'} = 0. \tag{14.16}$$

In addition to the endpoint condition $y(0) = 0$, and $y(x_f) = \phi(x_f)$, we also have the additional *transversality* condition

$$L(x_f, y(x_f), y'(x_f)) + (\phi'(x_f) - y'(x_f))L_{y'}(x_f, y(x_f), y'(x_f)) = 0. \tag{14.17}$$

Remark 14.2.1. Before making use of this remarkable boundary condition in an example we make a few comments about the derivation provided here. The key is to note what is allowed to vary in the calculation of the Gateaux differential and what must remain fixed. Once the Gateaux differential is calculated, we then used Taylor series expansions to relate the two unknown terms $h(x_f)$ and $g'(0)$ so that the remaining boundary terms could be written as a multiple of just one of these terms which yielded the transversality condition. These steps are critical, and illustrative of what could be done for an implicitly defined curve as well.

Remark 14.2.2. In the spirit of adding remarks because the author has all authority over what is written here we note that the derivation above has used two of the most powerful tools in an applied mathematician's ever-expanding arsenal: integration by parts, and Taylor series. Dissertations have been written, and careers dedicated to very careful usage of Taylor series and integration by parts. It is in the best interest of any budding applied mathematician to familiarize themselves with both of these tools...this isn't the last time you will see them.

Example 14.2.3. As an example, suppose that the Mississippi river's far bank follows $\phi(x) = x^2 - 5$ for $x > 0$. If Tom is such a strong swimmer that he can neglect the current of the river in his calculations (if this were true Tom would have given Michael Phelps a run for his money even in his prime), then his cost function is just $L(x, y, y') = \sqrt{1 + (y'(x))^2}$ just as we saw when finding the shortest path between two points.

Following the derivation above we see that the (EL) are satisfied:

$$\frac{d}{dx}\frac{y'(x)}{\sqrt{1+(y'(x))^2}} = 0,$$

with boundary conditions

$$y(0) = 0, \quad \text{and} \quad \sqrt{1 + y'(x_f)^2} + (2x_f - y'(x_f))\frac{y'(x_f)}{\sqrt{1+y'(x_f)^2}} = 0.$$

- Note that the only functions that satisfy these (EL) are linear functions, i.e. $y(x) = ax + b$.
- The boundary condition $y(0) = 0$ implies that $b = 0$ so that $y(x) = ax$.
- Thus, we have the following system of equations that we need to solve for x_f and a,

$$\begin{aligned} ax_f &= x_f^2 - 5 \\ \sqrt{1+a^2} + (2x_f - a)\frac{a}{\sqrt{1+a^2}} &= 0. \end{aligned}$$

Normally Tom would be quite happy to solve this numerically, but as with all good textbook examples, we will actually go through the nasty process of finding the solution. After some algebraic manipulation, this reduces to the system of equations given by:

$$\begin{aligned} x_f^2 - ax_f - 5 = 0 \\ a = -\frac{1}{2x_f}. \end{aligned}$$

This leads to $x_f^2 = \frac{9}{2}$ so that $x_f = \frac{3}{\sqrt{2}}$ and hence $a = -\frac{1}{3\sqrt{2}}$. In other words, Tom should swim along the path described by $y = -\frac{x}{3\sqrt{2}}$ and he will arrive at the far bank at the point $\left(\frac{3}{\sqrt{2}}, -\frac{1}{2}\right)$.

Remark 14.2.4. We have focused on a case that is quite tractable, when one endpoint is fixed by a nice differentiable curve. There are of course further generalizations of this where the endpoint is setup via an implicit relationship such as $f(x_f, y(x_f)) = 0$ where f is some function that defines the curve at the endpoint. This leads to a slightly more complicated setting, but overall it works out quite reasonably, and is certainly more general than the case considered here.

The next example is less motivated, and simultaneously less tractable but still of significant interest.

Example 14.2.5. Consider minimizing the cost functional

$$J[y] = \int_0^{x_f} \left[(y')^2 + y^2\right] dx,$$

where $y(0) = 0$ and $y(x_f) = 100(x_f + 5)$

The Euler-Lagrange equations in this case are

$$2y - 2y'' = 0 \Rightarrow y'' = y.$$

Applying the boundary condition at the left endpoint (with skipping quite a few steps in between) we end up with the solution $y(x) = c\sinh(x)$. The endpoint condition is

$$c\sinh(x_f) = 100(x_f + 5),$$

and finally the transversality condition is

$$c^2(\sinh^2(x_f) + \cosh^2(x_f)) + 2(100 - c\cosh(x_f))c\cosh(x_f) = -c^2 + 200c\cosh(x_f) = 0.$$

Solving these final two equations simultaneously yields the constant c and final end point x_f, which are approximately $x_f \approx 1.2675$ and $c \approx 383.283$.

14.3 Functionals of several variables

Now we consider the higher dimensional setting where there are several independent functions that go into the cost functional. For instance, when Tom and Huck took off in a hot air balloon, their position is given by both horizontal and vertical coordinates (Sam Clemens should be very proud that his literature inspired not just American fictional literature but is now being used to inspire a new generation of applied mathematicians).

Consider a cost functional of the form:

$$J[y_1, y_2, ..., y_n] = \int_a^b L(x, y_1, y_2, ..., y_n, y_1', y_2', ..., y_n')dx. \tag{14.18}$$

A necessary condition for an extremal of J is found the same way as before, i.e.

$$\delta J[y_1, y_2, ..., y_n; h_1, h_2, ..., h_n] = 0$$

is satisfied where the $h_i(x)$ are chosen independently to satisfy the appropriate boundary conditions corresponding to each $y_i(x)$. Supposing that all the $y_i(x)$ are fixed at the endpoints meaning that each of the $h_i(x)$ must vanish at the endpoints. Then the Gateaux differential is computed as

$$\begin{aligned}
\delta J &= \frac{\partial}{\partial \varepsilon} J(y_1 + \varepsilon h_1, ..., y_n + \varepsilon h_n)\Big|_{\varepsilon=0} \\
&= \frac{\partial}{\partial \varepsilon} \int_a^b L(x, y_1 + \varepsilon h_1, ..., y_n + \varepsilon h_n, y_1' + \varepsilon h_1', ..., y_n' + \varepsilon h_n')dx\Big|_{\varepsilon=0}
\end{aligned}$$

It follows that for these boundary conditions the differential can be rewritten as:

$$\begin{aligned}\delta J &= \int_a^b [L_{y_1} h_1(x) + ... + L_{y_n} h_n(x) + L_{y_1'} h_1'(x) + ... + L_{y_n'} h_n'(x)]dx \\ &= \int_a^b \sum_{i=1}^n \left[L_{y_i} h_i(x) + L_{y_i'} h_i'(x)\right] dx \\ &= \int_a^b \sum_{i=1}^n \left[L_{y_i} - \frac{d}{dx} L_{y_i'}\right] h_i(x) dx = 0.\end{aligned}$$

Now since the $h_i(x)$ are chosen independently, this must be true even if we set all of them but one equal to zero, implying that for that particular index i,

$$L_{y_i} - \frac{d}{dx} L_{y_i'} = 0.$$

Now because we chose i arbitrarily, this is true for all i, i.e. the Euler-Lagrange equations are (EL)

$$L_{y_i} - \frac{d}{dx} L_{y_i'} = 0 \quad \text{for all } i = 1, ..., n. \tag{14.19}$$

Example 14.3.1. Consider the surface parameterized by

$$x(u, v) = \cos u, \quad y(u, v) = \sin u, \quad \text{and} \quad z(v) = v$$

for $0 \leq u \leq 2\pi$ and $-\infty < v < \infty$ which defines the infinite cylinder of radius 1 centered about the z-axis. We want to compute the geodesics on this surface, i.e. the shortest path between two points when the traveler between the two points is restricted to live on the surface of the cylinder.

To do so, consider first the two points defined by $(x(u_0, v_0), y(u_0, v_0), z(v_0))$ and $(x(u_1, v_1), y(u_1, v_1), z(v_1))$ and the curve in u-v space that connects (u_0, v_0) and (u_1, v_1) denoted by ξ. This curve ξ can itself be parameterized by $\xi = \xi(t) = (u(t), v(t))$ for $t \in [a, b]$ and $u(a) = u_0$, $u(b) = u_1$, and $v(a) = v_0$, $v(b) = v_1$, implying that

$$\begin{aligned}dx &= \frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv, \\ dy &= \frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv, \\ dz &= \frac{\partial z}{\partial u} du + \frac{\partial z}{\partial v} dv \\ &= \frac{\partial z}{\partial v} dv,\end{aligned}$$

where the final line is because z is independent of the radial independent variable u.

Let s be the arclength along the curve $(x(u(t), v(t)), y(u(t), v(t)), z(v(t)))$ in $\mathbb{R}^3$. Then

$$\begin{aligned}ds^2 &= dx^2 + dy^2 + dz^2 \\ &= (x_u du + x_v dv)^2 + (y_u du + y_v dv)^2 + z_v^2 dv^2 \\ &= x_u^2 du^2 + x_v^2 dv^2 + 2x_u x_v du dv + y_u^2 du^2 + y_v^2 dv^2 + 2y_u y_v du dv + z_v^2 dv^2.\end{aligned}$$

At this point we note that

$$x_u = \frac{\partial x}{\partial u} = -\sin u, \quad x_v = \frac{\partial x}{\partial v} = 0$$
$$y_u = \frac{\partial y}{\partial u} = \cos u, \quad y_v = \frac{\partial y}{\partial v} = 0$$
$$z_u = \frac{\partial z}{\partial u} = 0, \quad z_v = \frac{\partial z}{\partial v} = 1.$$

Inserting all of these into the previous calculation,

$$\begin{aligned} ds^2 &= \sin^2 u du^2 + \cos^2 u du^2 + dv^2 \\ &= du^2 + dv^2 \\ &= \left[\left(\frac{du}{dt}\right)^2 + \left(\frac{dv}{dt}\right)^2\right] dt^2. \end{aligned}$$

Thus

$$ds = \sqrt{\left(\frac{du}{dt}\right)^2 + \left(\frac{dv}{dt}\right)^2}\, dt.$$

It follows that the total length of the curve is

$$S[u,v] = \int_a^b ds = \int_a^b \sqrt{\left(\frac{du}{dt}\right)^2 + \left(\frac{dv}{dt}\right)^2}\, dt.$$

If we call

$$L[u,v] = \sqrt{\left(\frac{du}{dt}\right)^2 + \left(\frac{dv}{dt}\right)^2}$$

then S has an extremal if (EL) are satisfied, i.e.

$$-\frac{d}{dt} L_{u'} = 0 \quad \text{and} \quad -\frac{d}{dt} L_{v'} = 0.$$

Thus

$$\frac{u'}{\sqrt{(u')^2 + (v')^2}} = c_1 \quad \text{and} \quad \frac{v'}{\sqrt{(v')^2 + (u')^2}} = c_2.$$

Now a lot of the time we would be satisfied with stopping at this point and finding a numerical solution to this set of equations where the initial and end points are provided, but since this is an example in a set of notes that (we are told) will one day be turned into a textbook, we will take it a few additional steps.

If we consider

$$\frac{dv}{du} = \frac{\frac{dv}{dt}}{\frac{du}{dt}} = \frac{c_2}{c_1} = c_0,$$

we see that in the parameter space (u-v), extremals are given by $v = Au + b$ i.e. the shortest paths are straight lines in the parameter space. If $A \neq 0$ this gives a spiral: $x = \cos u$, $y = \sin u$, $z = Au + B$ where the constants A and B are determined by the beginning and end points of the problem. If $A = 0$, i.e. a horizontal line in u-v space, the geodesic will be an arc parallel to the x-y plane.

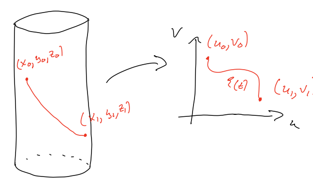

Figure 14.1: Geodesics on a cylinder.

This example is an important one even though it may seem a bit contrived. Finding the shortest path between two points on a prescribed surface is an important problem that arises in many contexts (believe it or not) including the theory of relativity where the geodesics will depend on space and time (not just space).

This example is also important because you can quickly see that the derivation of the actual solution rapidly gets messy for this type of problem. The more variables that we add into the variational setup, the more complicated the (EL) get, and the messier the answer is to reach. This is part of the reason that numerical solutions are so valuable as long as we set up the differential equation correctly and we know enough about the numerical methods to guarantee we are getting an accurate approximation of the 'correct' solution.

Example 14.3.2 (Upside down volcano). To overly emphasize the importance of geodesics on strange surfaces, we will investigate the shortest path between two points on on upside down volcano, i.e. a surface defined by $z = \sqrt{x^2 + y^2}$ which is parameterized $(x, y, z)^T = (u \cos v, u \sin v, u)^T$ where $v \in [0, 2\pi]$ and $u \in [0, 1]$. Following the previous example we see that the arc-length of the path is given by

$$\begin{aligned}
ds^2 &= dx^2 + dy^2 + dz^2 \\
&= (x_u du + x_v dv)^2 + (y_u du + y_v dv)^2 + z_u^2 du^2 \\
&= (\cos v du - u \sin v dv)^2 + (\sin v du + u \cos v dv)^2 + du^2 \\
&= \cos^2 v du^2 - 2u \sin v \cos v du dv + u^2 \sin^2 v dv + \sin^2 v du^2 \\
&+ 2u \sin v \cos v du dv + u^2 \cos^2 v dv^2 + du^2 \\
&= 2du^2 + u^2 dv^2 \\
&= \left[2 \left(\frac{du}{dt} \right)^2 + u(t)^2 \left(\frac{dv}{dt} \right)^2 \right] dt^2,
\end{aligned}$$

where t is the independent variable that parameterizes the desired curve $u(t)$, $v(t)$, implying that

$$ds = \sqrt{2 \left(\frac{du}{dt} \right)^2 + u(t)^2 \left(\frac{dv}{dt} \right)^2}\, dt.$$

Hence the shortest path between two points $(u(a) \cos(v(a)), u(a) \sin(v(a)), u(a))^T$ and $(u(b) \cos(v(b)), u(b) \sin(v(b)), u(b))^T$ is given by minimizing the functional

$$S[u, v] = \int_a^b \sqrt{2 \left(\frac{du}{dt} \right)^2 + u(t)^2 \left(\frac{dv}{dt} \right)^2}\, dt,$$

which leads to the Euler-Lagrange equations

$$\frac{u(v')^2}{\sqrt{2(u')^2+u^2(v')^2}} - \frac{d}{dt}\frac{u'}{\sqrt{2(u')^2+u^2(v')^2}} = 0,$$
$$-\frac{d}{dt}\frac{u^2v'}{\sqrt{2(u')^2+u^2(v')^2}} = 0.$$

These equations, it turns out, are not as simple to solve as what we had looked at in the last example. Nevertheless, numerical solutions of these equations are possible with the prescribed boundary conditions.

To make the previous example a bit more realistic, we would of course be interested in a surface defined by $z = A - \sqrt{x^2+y^2}$, i.e. a perfectly circular volcano of height A.

Example 14.3.3. Following a tradition of completing a section with a less motivated example, we consider minimizing

$$\int_0^1 \left[(y')^2 - 2yzw + (w')^2 + x(z')^2\right] dx,$$

subject to $y(0) = 1$, $y(1) = 2$, $z(0) = 0$, $z(1) = 2$, $w(0) = 2$, and $w(1) = 3$. The Euler-Lagrange equations in this case are given by:

$$\begin{aligned} y'' &= -zw, \\ w'' &= -yz, \\ xz'' + z' &= yw, \end{aligned}$$

with the corresponding boundary conditions given above. Solution of this system of second order differential equations will yield the optimal solution.

Just because the previous example had a trick to get an 'explicit' solution doesn't mean this one will, does it?

14.4 Higher dimensional objects

We saw in the previous section that multiple dependent variables made the solution of the variational problem far more complicated, but even then the application of the optimization principles were quite straightforward extensions of the theory we had developed. We will now consider what happens when we instead have multiple independent variables in the problem, and find that although theoretically the extension to this setting is straightforward, it is anything but that in practice. Before diving straight in, we will take some time to review some key concepts from multivariable calculus that will be necessary in what follows.

14.4.1 Multivariable Calculus: a 15 minute refresher

Notationally, we will define $\mathbf{x} \in \mathbb{R}^n$ to be a vector valued independent variable, and for now consider scalar valued functions of $\mathbf{x}$, i.e.

$$y(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}.$$

Definition 14.4.1. *The gradient of $y(\mathbf{x}) \in C^1$ is defined as the vector valued function:*

$$\nabla y = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{pmatrix},$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$.

For scalar valued functions the second derivatives are all included as a single tensor valued operator.

Definition 14.4.2. *The Hessian of $y(\mathbf{x}) \in C^2$ is defined as the matrix (2-tensor) valued function:*

$$H = \nabla\nabla y = \begin{pmatrix} \frac{\partial^2 y}{\partial x_1^2} & \frac{\partial^2 y}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_1 \partial x_n} \\ \frac{\partial^2 y}{\partial x_2 \partial x_1} & \frac{\partial^2 y}{\partial x_2^2} & \cdots & \frac{\partial^2 y}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 y}{\partial x_n \partial x_1} & \frac{\partial^2 y}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_n^2} \end{pmatrix}.$$

At the same time, we will occasionally need to work with vector valued functions, i.e.

$$\mathbf{u}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m,$$

where m may not be the same as n.

Definition 14.4.3. *Using the notation that $\mathbf{u} = (u_1, u_2, \ldots, u_m)^T$ then we can also consider the first derivative of $\mathbf{u}(\mathbf{x})$. In this case, we don't have a gradient, but instead have what is typically referred to as the Jacobian matrix:*

$$\nabla \mathbf{u} = D\mathbf{u} = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \cdots & \frac{\partial u_1}{\partial x_n} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \cdots & \frac{\partial u_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_m}{\partial x_1} & \frac{\partial u_m}{\partial x_2} & \cdots & \frac{\partial u_m}{\partial x_n} \end{pmatrix},$$

where the $\nabla \mathbf{u}$ notation is rarely (but occasionally) used in practice.

Definition 14.4.4. *Just as frequent as the use of the Jacobian (and something that you have definitely seen before), we also need to consider the divergence of a vector valued function which is given only when $m = n$ as:*

$$\nabla \cdot \mathbf{u} = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \cdots \frac{\partial u_n}{\partial x_n} = \sum_{k=1}^{n} \frac{\partial u_k}{\partial x_k}.$$

Notationally higher order derivatives that we make use of below are also defined in the following.

Definition 14.4.5. *The Laplacian of a scalar valued function is defined as the divergence of the gradient, i.e.*

$$\Delta y = \nabla^2 y = \nabla \cdot (\nabla y) = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2 y}{\partial x_2^2} + \cdots \frac{\partial^2 y}{\partial x_n^2} = \sum_{k=1}^{n} \frac{\partial^2 y}{\partial x_k^2}.$$

Note that there is more than one notational way to write out the Laplacian operator for a scalar valued function. We have included all of these versions here just in case you are going to look at other resources (despite the clear superiority of the presentation here, this is great idea).

Example 14.4.6. Consider the scalar valued function $y(x_1, x_2) = x_1^2 \sin(x_2)$. The gradient is given by

$$\nabla y = \begin{pmatrix} 2x_1 \sin(x_2) \\ x_1^2 \cos(x_2) \end{pmatrix},$$

and the Laplacian is given by

$$\Delta y = 2\sin(x_2) - x_1^2 \sin(x_2).$$

Definition 14.4.7. *We may also define the Laplacian of a vector valued function by applying the Laplace operator to each component of the corresponding vector field if the vector field $\mathbf{u} : \mathbb{R}^n \to \mathbb{R}^n$. This leads to the notational definition:*

$$\Delta \mathbf{u} = \nabla^2 \mathbf{u} = \nabla \cdot (\nabla \mathbf{u}) = \begin{pmatrix} \sum_{k=1}^n \frac{\partial^2 u_1}{\partial x_k^2} \\ \vdots \\ \sum_{k=1}^n \frac{\partial^2 u_n}{\partial x_n^2} \end{pmatrix}.$$

Note above that we have used the notation $\nabla \cdot (\nabla \mathbf{u})$ even though we already established that $\nabla \mathbf{u}$ is a tensor (matrix) for the vector valued function $\mathbf{u}$, and we have previously considered the operator ∇ as a vector valued differential operator, so the question remains what does it mean to take the dot product between a vector valued operator and a 2-tensor (matrix)? We address this question by instead defining the dot product between a vector and a matrix (2-tensor).

Definition 14.4.8. *For the vector $\mathbf{a}$ and 2-tensor B (both of which may actually be functions of some independent variable x) defined pointwise as*

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix},$$

then the dot product $\mathbf{a} \cdot B$ is defined as the vector

$$\mathbf{a} \cdot B = \begin{pmatrix} \sum_{k=1}^n a_k b_{1k} \\ \vdots \\ \sum_{k=1}^n a_n b_{nk} \end{pmatrix}.$$

Example 14.4.9. To demonstrate the use of this notation, consider the 3-dimensional vector valued function

$$\mathbf{u}(\mathbf{x}) = \begin{pmatrix} u(x_1, x_2, x_3) \\ v(x_1, x_2, x_3) \\ w(x_1, x_2, x_3) \end{pmatrix}.$$

It follows that

$$\mathbf{u} \cdot \nabla \mathbf{u} = \begin{pmatrix} u u_{x_1} + v u_{x_2} + w u_{x_3} \\ u v_{x_1} + v v_{x_2} + w v_{x_3} \\ u w_{x_1} + v w_{x_2} + w w_{x_3} \end{pmatrix}.$$

With all of this notation in place we are almost ready to march on to the next extension of the Calculus of Variations, but we need to decide exactly what we will do with this notation. The first thing to recall is the Divergence Theorem (8.2.6) and its various Corollaries (8.2.7) which are restated here for completeness (omitting the time dependence as we aren't concerned with that quite yet here):

Theorem 14.4.10. *If the vector field $\mathbf{F}(\mathbf{x})$ is $C^1(\overline{\Omega}) \cap C^2(\Omega)$, then*

$$\int_\Omega \nabla \cdot \mathbf{F} \, d\mathbf{x} = \int_{\partial\Omega} \mathbf{F} \cdot \mathbf{n} \, dA.$$

Remark 14.4.11. Note that if $\Omega = [a, b] \subset \mathbb{R}$ then this reduces to the Fundamental Theorem of Calculus, i.e.

$$\int_a^b F'(x) dx = F(b) - F(a),$$

because in this case the boundary of the region Ω is just the set of points $\{a, b\}$, and the outward pointing normal at b is $+1$ and -1 at $x = a$.

Corollary 14.4.12. *If the scalar-valued function $u(\mathbf{x})$ is $C^1(\overline{\Omega}) \cap C^2(\Omega)$, then*

$$\int_\Omega u_{x_k} \, d\mathbf{x} = \int_{\partial\Omega} u \, \mathrm{n}_k \, dA, \;\; k = 1, 2, \ldots, n.$$

Corollary 14.4.13. *If scalar-valued functions $u(\mathbf{x})$ and $w(\mathbf{x})$ and a vector field $\mathbf{F}(\mathbf{x})$ are all $C^1(\overline{\Omega}) \cap C^2(\Omega)$, then there holds*

(i) an integration by parts formula,

$$\int_\Omega w u_{x_k} \, d\mathbf{x} = -\int_\Omega u w_{x_k} \, d\mathbf{x} + \int_{\partial\Omega} u w \, \mathrm{n}_k \, dA,$$

(ii) Green's First Identity,

$$\int_\Omega \big(u \Delta w + \nabla u \cdot \nabla w \big) \, d\mathbf{x} = \int_{\partial\Omega} u \frac{dw}{dn} \, dA,$$

(iii) and Green's Second Identity,

$$\int_\Omega u \Delta w \, d\mathbf{x} = \int_\Omega w \Delta u \, d\mathbf{x} + \int_{\partial\Omega} \left(u \frac{dw}{dn} - w \frac{du}{dn} \right) \, dA.$$

Remark 14.4.14. Just as the one-dimensional analogue of the Divergence Theorem is the Fundamental Theorem of Calculus, the one-dimensional analogue of these identities is integration by parts. For instance, the first identity above can be rewritten as

$$\int_a^b w(x)u'(x)dx = -\int_a^b u(x)w'(x)dx + u(b)w(b) - u(a)w(a).$$

14.4.2 Higher dimensional Calculus of Variations

Consider optimizing the functional

$$J[y] = \int_\Omega L[\mathbf{x}, y(\mathbf{x}), \nabla y(\mathbf{x})]d\mathbf{x},$$

where $\mathbf{x} \in \mathbb{R}^n$ is a vector, and the domain Ω is typically taken to be rectangular (or at least some smooth subset of $\mathbb{R}^n$), and $y(\mathbf{x}) = f(\mathbf{x})$ for $\mathbf{x} \in \partial\Omega$, i.e. $y(\mathbf{x})$ is fixed on the boundary. As always, an extremal of $J[y]$ is given by the condition $\delta J(y; h) = 0$ for all $h(\mathbf{x}) \in C^1(\Omega)$ with $h(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial\Omega$. Computing the first variation, we see that

$$\begin{aligned}\delta J(y; h) &= \frac{\partial}{\partial\varepsilon}\int_\Omega L(\mathbf{x}, y(\mathbf{x}) + \varepsilon h(\mathbf{x}), \nabla y(\mathbf{x}) + \varepsilon\nabla h(\mathbf{x}))d\mathbf{x}\bigg|_{\varepsilon=0} \\ &= \int_\Omega \left[L_y h(\mathbf{x}) + \sum_{i=1}^n L_{y_{x_i}} h_{x_i}\right] dx\end{aligned}$$

where $y_{x_i} = \frac{\partial y}{\partial x_i}$. Now note that

$$L_{y_{x_i}} h_{x_i} = \partial_{x_i}(L_{y_{x_i}} h(\mathbf{x})) - \frac{\partial L_{y_{x_i}}}{\partial x_i} h(\mathbf{x}),$$

which can be shown by applying the chain rule to the first term on the right hand side, so that

$$\delta J = \int_\Omega \left(\left[L_y - \sum_{i=1}^n \frac{\partial L_{y_{x_i}}}{\partial x_i}\right] h(\mathbf{x}) + \sum_{i=1}^n \partial_{x_i}[L_{y_{x_i}} h(\mathbf{x})]\right) d\mathbf{x}.$$

Now at this point we could very carefully make use of some multivariable identities, but hopefully in the spirit of making things more transparent (not opaque) we will introduce the following notation which we refer to as a ‘pseudo-gradient operator(s)’:

$$\nabla' L = \begin{bmatrix} L_{y_{x_1}} \\ L_{y_{x_2}} \\ \vdots \\ L_{y_{x_n}} \end{bmatrix}, \quad \nabla y = \begin{bmatrix} y_{x_1} \\ y_{x_2} \\ \vdots \\ y_{x_n} \end{bmatrix}.$$

Using the Divergence Theorem (or one of its many Corollaries) combined with this notation, we arrive at

$$\int_\Omega \nabla \cdot \mathbf{u}(\mathbf{x})d\mathbf{x} = \int_{\partial\Omega} \mathbf{u}(\mathbf{x}) \cdot \mathbf{n}dA. \tag{14.20}$$

Combining this with the $'$ notation introduced above, we see that

$$\begin{aligned}\delta J &= \int_\Omega ([L_y - \nabla \cdot (\nabla' L)]h(\mathbf{x})d\mathbf{x} + \int_\Omega \nabla \cdot (\nabla' L h(\mathbf{x}))d\mathbf{x} \\ &= \int_\Omega ([L_y - \nabla \cdot (\nabla' L)]h(\mathbf{x})d\mathbf{x} + \int_{\partial\Omega} \nabla' L \cdot \mathbf{n} h(\mathbf{x})dA \\ &= 0,\end{aligned}$$

Now, because $h(\mathbf{x}) = 0$ for all $\mathbf{x} \in \partial\Omega$ (and using similar Lemmata to those used in the one-dimensional case), we see that $\delta J = 0$ for all $h(\mathbf{x}) \in C^1(\Omega)$ if the Euler-Lagrange equations are satisfied i.e.:

$$L_y - \nabla \cdot (\nabla' L) = 0, \qquad \text{or} \qquad L_y - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\frac{\partial L}{\partial y_{x_i}} = 0. \tag{14.21}$$

Example 14.4.15. This isn't a very explicit example, but it still kind of counts. Consider a 2D domain, i.e. $\Omega \subset \mathbb{R}^2$. It follows that the (EL) are given by

$$L_y - \frac{\partial L_{y_{x_1}}}{\partial x_1} - \frac{\partial L_{y_{x_2}}}{\partial x_2} = 0.$$

14.5 Examples of higher dimensional variational problems

14.5.1 Scalar valued functions of multiple independent variables

Now that you have seen the derivation and been reminded very intimately of the need for studying multivariable Calculus, we are prepared to consider some very interesting variational problems.

Example 14.5.1. Let $u(\mathbf{x})$ be the temperature of a certain material in the region Ω, with the appropriate boundary conditions (which we don't specify here). It turns out that for a given source $f(\mathbf{x})$ defined everywhere in Ω then, the steady state distribution of the temperature (independent of time) will minimize the cost functional

$$J[u] = \int_\Omega \left[\frac{\kappa}{2}|\nabla u|^2 + fu\right] d\mathbf{x},$$

where $|\nabla u|^2 = (\nabla u)\cdot(\nabla u)$. We can check this (at least one way) by seeing that the Euler-Lagrange equations for this cost functional will be exactly the steady state heat equation:

$$\kappa \Delta u = f,$$

with the appropriate boundary conditions prescribed.

To be more precise, note that the (EL) for this cost functional become

$$L_u - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\frac{\partial L}{\partial u_{x_i}} = 0,$$
$$f - \nabla \cdot (\kappa \nabla u) = 0,$$

which upon re-arranging yields the steady state heat equation given above.

Not to belabor this example too much, but we will also here verify this result in the specific 2D case. In two independent spatial variables we rewrite the cost-functional as

$$J[u] = \int_\Omega \left[\frac{\kappa}{2}\left(u_{x_1}^2 + u_{x_2}^2\right) + fu\right] d\mathbf{x}.$$

Applying the Euler-Lagrange equations in this case leads to

$$\begin{aligned} L_u - \sum_{i=1}^{2} \frac{\partial}{\partial x_i}\frac{\partial L}{\partial u_{x_i}} &= 0, \\ f - \frac{\partial}{\partial x_1}\kappa u_{x_1} - \frac{\partial}{\partial x_2}\kappa u_{x_2} &= 0, \\ f &= \kappa\left(u_{x_1x_1} + u_{x_2x_2}\right). \end{aligned}$$

Remark 14.5.2. This example is noteworthy because it also illustrates where natural boundary conditions may come into play for a particular PDE. For example, if you already have a PDE, but are unclear on the exact type of boundary conditions to use, you can see if the solution of your PDE satisfies just such an optimization principle (this happens more often than you would believe), in which case you can try to impose a natural or even perhaps a transversality boundary condition that coincides with whatever the optimization principle is that you are enforcing.

Example 14.5.3. Just to demonstrate that the PDEs you have already seen are indeed often derived as providing the solution to an optimization principle, consider minimization of

$$\frac{1}{2}\int_{\Omega\times[0,T]} \left(u_t^2 - c^2|\nabla u|^2\right) d\mathbf{x}dt,$$

where $\Omega \times [0,T]$ indicates that we are integrating over the entire spatial domain Ω and the time interval $[0,T]$. The (EL) for this problem become

$$u_{tt} = c^2\Delta u,$$

which is precisely the wave equation in d dimensions.

Just as we did in the previous example, we will verify this (EL) in 2D. In two spatial dimensions the functional becomes

$$\frac{1}{2}\int_{\Omega\times[0,T]} \left(u_t^2 - c^2\left[u_{x_1}^2 + u_{x_2}^2\right]\right) d\mathbf{x}dt,$$

which has (EL)

$$\begin{aligned} L_u - \frac{\partial}{\partial t}\frac{\partial L}{\partial u_t} - \frac{\partial}{\partial x_1}\frac{\partial L}{\partial u_{x_1}} - \frac{\partial}{\partial x_2}\frac{\partial L}{\partial u_{x_2}} &= 0, \\ u_{tt} - c^2\left(u_{x_1x_1} + u_{x_2x_2}\right) &= 0. \end{aligned}$$

This example of course brings up a very interesting point that we have not addressed yet. Up to this point we have primarily focused on optimization over spatial variables only, but this isn't necessary. In fact this example demonstrates that the wave equation satisfies an optimization principle in space and time simultatneously. Note that we have also ignored the boundary conditions of $u(\mathbf{x},t)$ in time here. To really understand how this works we need to be much more careful with our Gateaux differential.

Example 14.5.4. Return to the previous example, i.e. optimization of

$$\frac{1}{2}\int_{\Omega\times[0,T]}\left(u_t^2 - c^2|\nabla u|^2\right) d\mathbf{x}dt,$$

but now suppose that $\mathbf{u}(\mathbf{x}) = A$ is constant for $\mathbf{x} \in \partial\Omega$ (this means we don't need to worry about the integration by parts/Divergence Theorem in the spatial derivatives), and also enforce the condition that $u(\mathbf{x}, t = 0) = u_0(\mathbf{x})$. What does this mean for the optimal solution?

Following the derivation of the (EL) in this case leads to the Gateaux differential

$$\delta J = \int_\Omega \left[c^2\nabla\cdot\nabla u - \partial_t u_t\right] h(\mathbf{x}, t) d\mathbf{x}dt + \int_\Omega u_t(\mathbf{x}, t = T)h(\mathbf{x}, t = T)d\mathbf{x} = 0,$$

where the variations $h(\mathbf{x}, t)$ are constrained to vanish for $\mathbf{x} \in \partial\Omega$ and $t = 0$, but not necessarily for $t = T$. As this Gateaux differential must vanish for all such potential $h(\mathbf{x}, t)$ then it must vanish for even those where $h(\mathbf{x}, t = T) = 0$ and hence the same (EL) are satisfied as in the past example. It follows that because the (EL) are satisfied even for those variations $h(\mathbf{x}, t)$ that do not vanish for $t = T$ then we see that $u_t(\mathbf{x}, t = T) = 0$ must hold which gives the final natural boundary condition in this case.

These two examples demonstrate the power of using such optimization principles when working with PDEs and ODEs. Not only does this provide us with another avenue to explore the derivation of different types of auxiliary conditions, but the variational framework provided here is also very useful when considering questions of existence and uniqueness of solutions (a very important question for applications of PDEs, not just a mathematical red herring). Sometimes it is easier to show that the variational optimization problem has a unique optimizer for specific prescribed auxiliary conditions than it is to determine that the solution of the corresponding (EL) are indeed unique. Such a fascinating convergence of these two topics is not coincidental. Solutions of PDEs are very often the solutions of a specific optimization problem, i.e. every time you are studying optimization there is a differential equation hidden deep inside.

14.5.2 Vector valued functions of multivariable independent variables

We will consider only one example here, not because this setting isn't important but contrarily because despite being extremely important, it is significantly more complicated and several examples would require several days to work through.

Example 14.5.5. Similarly to the previous couple of examples, we would like to consider minimization of the following cost functional

$$J[\mathbf{u}] = \frac{1}{2}\int_\Omega \left[(\nabla\mathbf{u}) : (\nabla\mathbf{u})\right] d\mathbf{x},$$

subject to the constraint that $\nabla \cdot \mathbf{u} = 0$ for a vector field $\mathbf{u}$, where the tensor inner product $A : B$ refers to pointwise multiplication of the two tensors A and B and then evaluating the sum of all entries of the resultant matrix. Often we will see this as $A : A = |A|^2$ which is the sum of all the squared entries of the matrix A. We don't yet know what to do with the constraint $\nabla \cdot \mathbf{u} = 0$ (that comes in the next Chapter) so we will temporarily ignore it and see that the cost functional can be rewritten as

$$J[\mathbf{u}] = \frac{1}{2} \int_\Omega \left[\sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial u_i}{\partial x_j} \right)^2 \right] d\mathbf{x}.$$

Gratefully only derivatives of $\mathbf{u}$ appear in this cost functional so the (EL) are significantly simplified. Even so, they become:

$$-\sum_{j=1}^n \frac{\partial^2 u_i}{\partial x_j^2} = 0,$$

for every $i = 1, \ldots n$ which is equivalent to stating that

$$\Delta \mathbf{u} = 0.$$

Keeping with tradition (be that as it may), we note that specifically in 2D this cost functional is written as:

$$J[\mathbf{u}] = \frac{1}{2} \int_\Omega \left[\left(\frac{\partial u_1}{\partial x_1} \right)^2 + \left(\frac{\partial u_1}{\partial x_2} \right)^2 + \left(\frac{\partial u_2}{\partial x_1} \right)^2 + \left(\frac{\partial u_2}{\partial x_2} \right)^2 \right] dx_1 dx_2,$$

which in turn leads to the (EL)

$$\frac{\partial^2 u_1}{\partial x_1^2} + \frac{\partial^2 u_1}{\partial x_2^2} = 0,$$
$$\frac{\partial^2 u_2}{\partial x_1^2} + \frac{\partial^2 u_2}{\partial x_2^2} = 0.$$

14.5.3 Further Generalizations

There are of course several other generalizations of the simplest problem that we do not explore here, some of which are considered in the homework. The general format for the derivation of the corresponding Euler-Lagrange equations is quite standard. First, the variational optimization problem must be stated, and then the class of increments must be specified to ensure that the optimization procedure remains in the appropriate space. Then the first variation is calculated as the difference between two evaluations of the functional, incremented by some small (infinitesimal) perturbation of the base state. Setting the first variation to zero for all admissible values of the increment $h(\mathbf{x})$ then yield the relevant Euler-Lagrange equations.

Making these derivations completely rigorous is an onerous task, as we saw earlier, but generically integration by parts or invoking different versions of Green's (Divergence) Theorem (necessary for the multi-dimensional setting) will lead to the proper form of the (EL). This highlights one of the great tools of applied mathematics, as integration by parts is really just adding zero in an appropriately re-labeled way, i.e. the formal derivation of most results in the Calculus of Variations, and indeed for any type of optimization process boils down to adding zero appropriately, so now you know why your elementary school teacher was so concerned that you learn how to add zero to something. If only we could figure out why they were so focused on some other topics...

14.6 Numerical Solutions

After having such a fun time the last couple of sections developing a general theory for various changes in the simplest problem, we are going to now move back to the simplest problem once again to consider numerically approximated solutions to the corresponding (EL). Given a functional

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

we have determined that if $\bar{y}$ is a minimum of J, then $\bar{y}$ satisfies the Euler Lagrange (EL) equation. We also found that there are boundary conditions that y_0 must satisfy. Most of the time, these (EL) are not analytically solvable but the solution must be approximated numerically.

We begin by numerically solving the BVP

$$y''(x) = -y(x)$$

subject to the boundary conditions $y(0) = 1$ and $y(\pi/2) = 2$. To begin we need to write the problem as a first order system

$$\begin{aligned} y_0' &= y_1 \\ y_1' &= -y_0. \end{aligned}$$

We define a function for this first order system first, then define a function for the specified boundary conditions. Finally, as shown in

We now call the solver to obtain our numerical approximation of the solution. In the block of code below, x is the grid on which we provide an initial guess to the solution, and y0 is the guess we provide on that grid. Since our ODE is two-dimensional, our initial guess y0 has 2 rows and the same number of columns as the length of x. Finally we call solve_bvp from scipy.integrate. The input is the ode function, the boundary condition function, the initial grid, and the solution guess on the initial grid.

The solution to this BVP is $y(x) = \cos x + 2\sin x$. In the code below, we compare the true solution against the numerical approximation.

There are frequently circumstances where we have parameters in the (EL) that come from the variational problem. In many cases the parameter is actually a part of the problem itself, such as identifying the eigenvalue for an eigenvalue problem. The great thing about Python's solve_bvp function is that it treats these unknown parameters as just another set of parameters in the solution, i.e. we don't have to do anything special to include them in the solution.

Unexample 14.6.1. Consider the boundary value problem

$$\alpha y''(x) = \beta y(x),$$

```python
import numpy as np
from scipy.integrate import solve_bvp
import matplotlib.pyplot as plt

#Defining the right hand side of the ODE
def ode_fun(x, y):
    # y'' = -y
    return np.vstack((y[1], -y[0]))

#Defining the boundary conditions
def bc(ya, yb):
    # y(0) = 1, y(pi/2) = 2
    return np.array([ya[0]-1, yb[0]-2])

#Determining the x-values to use in the solution
x = np.linspace(0, np.pi/2, 5)

#the initial guess for the solution
y0 = np.zeros((2, x.size))

#the actual solution...check the format in the documentation
res = solve_bvp(ode_fun, bc, x, y0)

#the rest of this is plotting only
x_plot = np.linspace(0, np.pi/2, 100)
y_plot = res.sol(x_plot)[0]

plt.plot(x_plot, y_plot, '-b', label='y')
plt.plot(x_plot, np.cos(x_plot)+2*np.sin(x_plot),'--r',label='cos(x)+2sin(x)')
plt.legend()
plt.xlabel("x")
plt.ylabel("y")
plt.show()
```

Algorithm 14.1: Python implementation of a very simple boundary value problem.

with free parameters α and β, and boundary conditions $y(0) = 0$, $y'(0) = 1$, and $y(\pi) = 0$. Note that the solution is $y(x) = \sin(x)$ and that any choice of α and β satisfying $\beta/\alpha = -1$ would admit this solution. Without going into the details, we note that in approximating a solution to the BVP, we are using Newton's method to solve for the zeros of a large nonlinear system. However, the zeros of this system are not isolated because any α and β correspond to a zero of the system as long as $\alpha = -\beta$. The code in Algorithm 14.2 implements the solve_bvp routine directly for this problem with both unknown parameters α and β. You should try running this code and see what happens. Why shouldn't you expect this to work?

```python
import numpy as np
from scipy.integrate import solve_bvp
import matplotlib.pyplot as plt

def bc3(ya, yb, p):
    alpha = p[0] # free parameter

    # y_0(0) = 1, y_1(0) = 0, y_0(1) = 0
    return np.array([ya[0]-1, ya[1], yb[0]])

def ode_fun3(x, y, p):
    alpha = p[0] # the free parameter
    beta = p[1]

    # y_0' = y_1, y_1' = -alpha*y_0
    return np.vstack((y[1], -(alpha/beta) * y[0]))

x = np.linspace(0, 1, 5)
y = np.ones((2, x.size))
sol3 = solve_bvp(ode_fun3, bc3, x, y, p=[2,1])
```

Algorithm 14.2: Python implementation of a very simple boundary value problem where we have more unknown parameters than boundary conditions. Pay particular attention to how the unknown parameters are included in the function and identified by solve_bvp.

We may be tempted to throw out Unexample 14.6.1 entirely and claim this is just a physically untenable problem. We would be remiss if this were the case. One must never throw an entire problem out because of poorly phrased parameters (kind of like the problem of small children remaining in bath water). Instead, recall that we haven't considered the dimensional structure of the problem yet. Even without quantifying precisely what the quantities x and y stand for, we can easily see that the boundary value problem from Unexample 14.6.1 can be rewritten as

$$y''(x) = \gamma y(x), \tag{14.22}$$

where $\gamma = \beta/\alpha$, subject to the boundary conditions $y(0) = 0$, $y'(0) = 1$, and $y(\pi) = 0$, which can be solved numerically via solve_bvp.

Remark 14.6.2. We have only dealt with one-dimensional boundary value problems here, but the generalization to higher dimensional problems should be rather apparent (although certainly not simple). We haven't talked about the actual methods behind solve_bvp as there are several different options built in to the basic method that are selected dependent on the type of boundary value problem to be solved. At this point we suffice it to say that solve_bvp uses a modified Newton solver to identify the discretized function that best approximates the solution of the specified problem. There are a variety of ways to reach this setup, but the default method in solve_bvp is to utilize a finite difference style discretization and then employ a lot of linear algebra with Newton's method doing the heavy lifting.

Exercises

Note to the student: Each section of this chapter has several corresponding exercises, all collected here at the end of the chapter. The exercises between the first and second line are for Section 1, the exercises between the second and third lines are for Section 2, and so forth.

You should **work every exercise** (your instructor may choose to let you skip some of the advanced exercises marked with *). We have carefully selected them, and each is important for your ability to understand subsequent material. Many of the examples and results proved in the exercises are used again later in the text. Exercises marked with ⚠ are especially important and are likely to be used later in this book and beyond. Those marked with † are harder than average, but should still be done.

Although they are gathered together at the end of the chapter, we strongly recommend you do the exercises for each section as soon as you have completed the section, rather than saving them until you have finished the entire chapter.

14.1. Which curve minimizes the integral
$$\int_0^1 \left(\frac{1}{2}y'^2 + yy' + y' + y\right) dx,$$
when $y(0) = \frac{1}{2}$ and $y(1)$ is free?

14.2. Set up a cost functional whose optimizer will describe Tom Sawyer's optimal route for crossing the Mississippi river if he starts at point $(x, y) = (0, 0)$ and wants to reach the opposite bank where $x = a$ (he doesn't care what value y is). Let the velocity of the river be given by $V(x)$. How do you incorporate the river's velocity?

14.3. Modify the cost functional from the previous problem to account for Tom trying to avoid a swamp on the other side of the river for $y \in [-1, 1]$. Give an explanation for your choice of cost functional.

14.4. Find the Euler-Lagrange equations for both of the previous two problems with appropriate boundary conditions. You do NOT need to solve either of these, just derive the relevant differential equations and boundary conditions.

14.5. Now set up a problem wherein Tom has to worry about the current, and is aiming for a non-vertical river bank on the other side. Choose a different function $\phi(x)$ for the opposite river bank than was used above, but make sure you can justify it is reasonable (at least locally). Make sure that you include the effects of the river's current.

14.6. Derive the (EL) and relevant boundary conditions for the previous problem.

14.7. Following Example 14.2.3, what is Tom's optimal path to cross the Mississippi if the opposite river bank is actually shaped like $\phi(x) = x^3 - 5$, but all other aspects of the problem are the same?

14.8. Find the extremals of the functional

$$J[y, z] = \int_0^{\pi/2} (y'^2 + z'^2 + 2yz)dx,$$

subject to the boundary conditions

$$y(0) = 0, \quad y(\pi/2) = 1, \quad z(0) = 0, \quad z(\pi/2) = 1.$$

14.9. Consider an elliptic paraboloid given by $z(x, y) = Ax^2 + By^2$. Compute the Euler-Lagrange equations that would yield the geodesics on this surface (the path along the surface which minimizes the distance traveled) from the origin $(0, 0)$ to the point given by (x_0, y_0). If you're really ambitious, compute the solution...it isn't pretty. *HINT: in 3 dimensions* $ds^2 = dx^2 + dy^2 + dz^2$ *and in this case* $dz = 2Axdx + 2Bydy$.

14.10. Derive the Euler-Lagrange equation that will determine the extrema of a functional of the form:

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

by supposing that all the integration by parts are possible and that $y(x)$ is restricted so that $y(a)$, $y(b)$, $y'(a)$, and $y'(b)$ are all fixed.

14.11. Derive the natural boundary conditions that arise for the extrema of the functional given in the previous example when the function $y(x)$ (and it's first derivative) is not fixed at the boundaries.

14.12. Use definition 14.4.8 and the vector valued operator definition of ∇ to arrive at definition 14.4.7.

14.13. Using the divergence theorem or one of its Corollaries, show that

$$\int_\Omega p(\mathbf{x})(\nabla \cdot \mathbf{u}(\mathbf{x}))d\mathbf{x} = -\int_\Omega (\nabla p(\mathbf{x})) \cdot \mathbf{u}(\mathbf{x})d\mathbf{x} + \int_{\partial\Omega} p(\mathbf{x})\mathbf{u}(\mathbf{x}) \cdot \mathbf{n}dA,$$

for $p(\mathbf{x}) : \mathbb{R}^3 \to \mathbb{R}$ and $\mathbf{u}(\mathbf{x}) : \mathbb{R}^3 \to \mathbb{R}^3$ both being C^2.

14.14. The following integral for a smooth region $\Omega \subset \mathbb{R}^3$ is a well known geometric property of that region. What is it?

$$\int_{\partial\Omega} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \mathbf{n}dA.$$

14.15. Show that the area of a surface given by the graph of $z = f(x, y)$ above a region Ω in the plane is given by the integral

$$I[f] = \int_\Omega \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} dxdy.$$

HINT: The area of a parallelogram is the cross product of the vectors that define the two sides of the parallelogram. Also recognize that the surface of the graph is given by the coordinates $(x, y, f(x, y))$ *and then the infinitesimal sides of a small piece of the surface are defined by the* x *and* y *derivatives of those values.*

14.16. For the functional $I[f]$ in the previous problem, determine the Euler-Lagrange equation that is necessarily satisfied if $z = f(x, y)$ represents a minimal surface, i.e. $I[f]$ is minimized.

14.17. Show that a plane is a solution to the Euler-Lagrange equation derived in the previous problem.

14.18. *Show that the helicoid, parameterized below is also a solution of the (EL) for the previous two problems, where for $-\pi/2 < z < \pi/2$:

$$\begin{aligned} x &= \rho\cos(\theta) \\ y &= \rho\sin(\theta) \\ z &= \theta. \end{aligned}$$

14.19. Find the Euler-Lagrange equations for the following functional, assuming that $y(\mathbf{x})$ is fixed at the boundaries and $q(\mathbf{x})$ and $p(\mathbf{x})$ are sufficiently smooth that all the necessary operations are kosher:

$$J[y] = \int_\Omega \left[p(\mathbf{x})\left(\nabla y \cdot \nabla y\right) + q(\mathbf{x})y(\mathbf{x})^2\right] d\mathbf{x}.$$

14.20. Let $u(x, y) \in [0, 1]$ represent a two-dimensional black and white image so that each value of (x, y) is mapped to a grayscale value $u(x, y)$ between 0 (black) and 1 (white). Suppose that you obtain a very pixelated image $v(x, y)$ and you want to recreate the full non-pixelated image. Come up with a cost function for $u(x, y)$ that will minimize the difference between $u(x, y)$ and $v(x, y)$, but will also not allow for the grayscale to change to rapidly, i.e. you don't want $u(x, y)$ to have large gradients anywhere either. What are the Euler-Lagrange equations you end up with?

14.21. Modify Algorithm 14.2 to solve the modified boundary value problem so that you numerically find the solution of (14.22) with the specified boundary conditions.

14.22. Use solve_bvp to approximate the solution to the following BVP.

$$\begin{aligned} y_0' &= \alpha y_1 \\ y_1' &= y_0^2 + y_1^2 + y_0/2 - 1 \end{aligned}$$

subject to the boundary conditions $y_0(0) = 1$, $y_1(0) = 0$, and $y_0(\pi) = 0$. Specifically, find a solution such that y_0 is decreasing and concave down. Note that α is a free variable.

14.23. Come up with a numerical solution of problem 14.2 from section 14.1, giving an explicit value of the width of the river, and for a 'reasonable' river velocity $V(x)$. How do you include the boundary conditions?

Notes