Chapter 1

A Bird's Eye View on Calculus of Variations

At face value, Calculus of Variations (CV) is about finding minima or maxima of certain expressions; in contradistinction to single- and multi-variable calculus (S&MVC), these expressions do not merely depend on a single or finitely many variables (numbers), but on one or several functions.

Example: *For which function(s)* $x \mapsto r(x)$*, subject to the constraints* $r(0) = r(1) = 2$*, will* $\int_0^1 r(x)\sqrt{1 + r'(x)^2}\,dx$ *be smallest possible?*

Like in S&MVC, where it is necessary that at a minimum certain equations are satisfied (namely the derivative must be 0), in CV, for a function r to realize a minimum of $\int_0^1 r(x)\sqrt{1 + r'(x)^2}\,dx$, this function must satisfy a certain equation, actually an ordinary differential equation, and the meaning behind this equation is again the vanishing of a derivative.

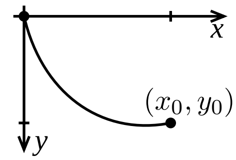
Unlike in S&MVC, where in practice we always proceed to solve these equations in order to find the minima (or at least candidates for the minima, as solutions might be minima, maxima, or saddle points), in CV, there are two common uses of the theory:

(a) solving the differential equation in order to find the minimum, analogous to the situation in S&MVC,
(b) studying the minimum problem by other means in order to learn something about the equation; this is the converse of what we are doing in S&MVC

The case (a) is the more classial one, and we will study it in more detail, but we will also get glimpses of case (b), which is more advanced and which is the reason why CV is still a theory of modern interest. It is good to study the analogies as well as the differences between S&MVC minimax problems and CV. Seeing the analogies between the two cases is definitely a 20th century paradigm and is therefore absent from many classical treatises (up to 19th century). But going beyond the classical results is what makes CV a powerful tool in the study of ordinary and partial differential equations. However, the mathematics to make this approach rigorous was only developed in the 20th century.

Examples

Example 1.1 *Among all curves —assumed for simplicity to be graphs* $y = f(x)$*— connecting* $(0,0$ *with* (x_0, y_0)*, find one (if any) along which a mass point slides most quickly from* $(0,0)$ *to* (x_0, y_0)*. We assume that the point slides without friction under the influence of gravity. This is the* **Brachystochrone Problem** *(Johann Bernoulli 1696). In formulas, we look for*



$$\min \left\{ I[y] := \int_0^{x_0} \sqrt{\frac{1+y'(x)^2}{y(x)}}\,dx \;\middle|\; y \in C^1[0,x_0]\,,\ y(0)=0\,,\ y(x_0)=y_0 \right\}$$

> **Warning:** In the form stated naively here, the Brachystochrone Problem does not have a solution. It turns out we should allow $y'(0)$ to be infinite, because the minimizing curve turns out to have vertical tangent at the start point.

Example 1.2 (Hanging chain (catenary)) *Among all curves that are graphs* $y = f(x)$ *and connect* (x_0, y_0) *to* (x_1, y_1)*, and that have a given length* L*, find one (if any) whose center of mass is lowest:*

$$\min \left\{ I[y] := \int_{x_0}^{x_1} y(x)\sqrt{1+y'(x)^2}\,dx \;\middle|\; y \in C^1[x_0,x_1]\,,\ \int_{x_0}^{x_1} \sqrt{1+y'(x)^2} = L\,,\ \begin{array}{c} y(x_0)=y_0\,, \\ y(x_1)=y_1 \end{array} \right\}$$

Example 1.3 (Rotational surface of least area)

$$\min \left\{ I[r] := 2\pi \int_{z_0}^{z_1} r(z)\sqrt{1+r'(z)^2}\,dz \;\middle|\; r \in C^1[z_0,z_1]\,,\ r(z_0)=r_0\,,\ r(z_1)=r_1 \right\}$$

Note: We have assumed that our functions are C^1 because their derivative is needed under the integral. We might ask for piecewise C^1 functions, so that the derivative would be allowed to have a few jump discontinuities. The technical question which functions precisely should be admissible turs out to be irrelevant in many of the classical problems, but is significant in more advanced studies. An optimal choice of the set of admissible functions also requires some advanced technical prerequisites, namely Lebesgue's theory of integration. We'll discuss this issue somewhat informally below, and we will see a few examples where the precise domain does make a difference.

Example 1.4 (Electrostatics) *On the boundary* $\partial\Omega$ *of an insulator represented by a bounded domain* Ω*, a given voltage distribution* $u_0 \in C^0(\partial\Omega)$ *is prescribed. Find the 'voltage' (potential) of the electric field inside* Ω*. This means we want to find* $u \in C^1(\Omega) \cap C^0(\bar\Omega)$ *that minimizes the electric field energy* $I[u] := \int_\Omega |\nabla u|^2\,dx$ *among all* u *that, on* $\partial\Omega$*, coincide with the given* u_0*. The same problem can, according to physics, also be described by a partial differetial equation* $\Delta u = 0$ *in* Ω *and* $u|_{\partial\Omega} = u_0$*. Both ways of viewing the problem are connected: The equation* $\Delta u = 0$ *describes the vanishing of the derivative of* $I[u]$ *with respect to* u *(whatever the derivative with respect to a function may mean). We'll explore this in more detail.*

> **Remark/Warning:** Again, in purpusefully ignorant innocence, I have hidden a severe problem in the preceding example. If u_0 is continuous on the boundary

and no better, not a single function u with boundary values u_0 will have finite $I[u]$. To assume u_0 continuous is a very good hypothesis in PDEs but is utterly inappropriate for the variational problem. On the other hand, to look only among $u \in C^1(\Omega)$ is asking too much. We don't need to have ∇u continuous; all we need is that its square is integrable. For the PDE on the other hand, asking that $u \in C^1$ would be asking too little, because we need 2 derivatives for Δu to be defined. We'll find it easy to resolve this kind of difficulty for ODEs, i.e., when the unknown function u is a single-variable functions. For PDEs, the same kind of problem can be resolved, but by methods beyond the scope of this course. Classical treatises mostly bypass this issue.

Review of Minimax problems in S&MVC

Theorem 1.5 *(a) If $u \in G \subset \mathbb{R}^n$ and $I : u \mapsto I(u) \in \mathbb{R}$ is differentiable and if I has a minimum or maximum at u_0 in the* interior *of G, then the derivative $DI(u_0) = 0$. [Equivalently, the gradient $\nabla I(u_0) = 0$.][1] Moreover, assuming I is actually C^2, the Hessian $D^2 I(u_0)$ is positive semidefinite if u_0 is an interior minimum, and negative semidefinite if u_0 is an interior maximum.*

Conversely, assuming $DI(u_0) = 0$ at an interior point, and $I \in C^2$: If the Hessian $D^2(u_0)$ is positive definite, then u_0 is a relative minimum; if the Hessian is negative definite, then u_0 is a relative maximum. notions reviewed below

(b) If $I : G \to \mathbb{R}$ is continuous and G is compact *(which in $\mathbb{R}^n$ is equivalent to 'closed and bounded'), then there exists a u_0 such that $I(u_0)$ is an absolute minimum. Likewise there exists a u_1 making $I(u_1)$ an absolute maximum.*

(c) (Lagrange multiplier theorem). If $I : G \to \mathbb{R}$ is C^1 and $G \subset \mathbb{R}^n$ is open, and u_0 is a relative minimum (or a relative maximum) of I subject to the constraints $K_1(u) = 0, \ldots, K_j(u) = 0$, with $j < n$, then there exist real numbers $\lambda_1, \ldots, \lambda_j$ such that $DI(u_0) - \lambda_1 DK_1(u_0) - \ldots - \lambda_j DK_j(u_0) = 0$, or else there exist $\lambda_1, \ldots, \lambda_j$ (not all 0), such that $\lambda_1 DK_1(u_0) + \ldots + \lambda_j DK_j(u_0) = 0$.

A 'good' selection of constraints is one that disallows the second possibility. Assuming that this 'or else' case does not occur is tantamount to assuming that the theorem of implicit functions can be applied to eliminate, locally, j of the n variables, by means of the j constraints.

It bears to remember the following Definition: u_0 is called a *relative* minimum of I, if $I(u_0) \leq I(u)$ for all u in a neighborhood of u_0, i.e., if there exists $\varepsilon > 0$ such that $I(u_0) \leq I(u)$ for all u satisfying $\|u - u_0\| < \varepsilon$. In contradistinction, u_0 is called an *absolute* minimum of I, if $I(u_0) \leq I(u)$ for all u for which $I(u)$ is defined. Similarly, relative and absolute maxima are defined.

[1] The gradient is the transpose of the derivative, i.e., the derivative is a row vector whose entries are the partial derivatives, whereas the gradient is a column vector whose entries are the partial derivatives. Many calculus textbook gloss over this distinction with impunity, but it is good for advanced purposes to maintain the distinction cleanly.

Convention: In normal usage, 'local minimum' and 'relative minimum' are synonymous, and likewise 'global minimum' and 'absolute minimum' are synonymous. For the purposes of this course however, in deviation from common usage, I will insist on only using the terms 'relative' vs. 'absolute' minimum in the sense of the given definition, and will reserve the words 'local' and 'global' for a later distinction that is different from the just defined 'relative' vs. 'absolute' distinction.

Some of you may need a review of the notions of definiteness of symmetric matrices and the Hessian:

Definition 1.6 *The Hessian $D^2I(u_0)$ of a C^2 function I in $\mathbb{R}^n$ at a point u_0 is the symmetric matrix whose entries are the 2nd partial derivatives: $(D^2I(u))_{ij} := \frac{\partial^2 I(u)}{\partial u_i \partial u_j}$.*

A symmetric $n \times n$ matrix A is called positive (resp. negative) semidefinite if for every vector $v \in \mathbb{R}^n$ it holds $v^T Av \geq 0$ (resp. $v^T Av \leq 0$). A symmetric $n \times n$ matrix is called positive (resp. negative) definite, if for every non-zero vector $V \in \mathbb{R}^n$, it holds $v^T Av > 0$ (resp. $v^T Av < 0$).

If $I : \mathbb{R}^n \to \mathbb{R}$ is a C^2 function and we 'slice' its graph along a line $u = u_0 + tv$ to get a single variable function $f(t) := I(u_0 + tv)$, with a given point u_0 and a given direction v, then the second derivative at $t = 0$ is $f''(0) = v^T D^2I(u_0)v$. So, to say

> "If u_0 (with $DI(u_0) = 0$) is a relative minimum, then $D^2I(u_0)$ is positive semidefinite"

is tantamount to saying:

> "If u_0 (with $DI(u_0) = 0$) is a relative minimum, then it satisfies the 2nd derivative test in every direction v"

Likewise, to say

> "If $DI(u_0) = 0$ and $D^2I(u_0)$ is positive definite, then u_0 is a relative minimum"

is tantamount to saying

> "If u_0 satisfies the sufficient conditions for a single-variable relative minimum in every direction v, then u_0 is a relative minimum"

The first statement (necessary condition) is trivial, but the second statement (sufficient condition) is not so obvious, and its analog in Calculus of Variations, may or may not be true, depending on a finer qualification of how 'relative minimum' is defined in this context.

Here is a theorem how to test positive definiteness of a matrix:

Theorem 1.7 (Hurwitz criterion) *A symmetric matrix $((a_{ij}))_{i,j=1,\dots n}$ is positive definite if and only if all of the following determinants are positive: $\det((a_{ij}))_{i,j=1,\dots k}$ for $k = 1, 2, \dots, n$. (Different ordering of indices are also possible: You can require $a_{ii} > 0$ for some i of your choice, then taking a 2×2 determinant made up of the ith and any further i'th row and column, and then adding in one more row and column at a time.)*

To test whether A is negative definite, you test whether $-A$ is positive definite; this amounts to alternating signs of the $\det((a_{ij}))_{i,j=1,\dots,k}$.

It is not *true that relaxing the inequalities* $\det > 0$ *to* $\det \geq 0$ *would test for positive semidefinite.*

The following is also good to know:

Theorem 1.8 *A symmetric matrix A is positive definite exactly if all of its eigenvalues are positive; it is positive semidefinite, if all of its eigenvalues are non-negative.*

Proof of the Lagrange multiplier theorem (Sketch only):

The 'or else' case $\sum \lambda_i DK_i(u_0) = 0$ is merely making room for the exceptional situation where the constraints do not leave a 'nice' set (a manifold) of admissible functions. Let us assume that $\sum \lambda_i DK_i(u_0) = 0$ happens only when all $\lambda_i = 0$. Then by the theorem of implicit functions, the set $\{u \mid K_i(u) = 0,\ i = 1, \dots, j\} \subset \mathbb{R}^n$ can be written as a graph, where j of the coordiantes are written as functions of the remaining $n - j$ coordinates For notational simplicity, let's assume that it's the first j coordinates which can be so written Carrying out this elimination (and using upper indices for the coordinates), $\bar{u}_0 = (u_0^{j+1}, \dots, u_0^n)$ minimizes the function

$$(u^{j+1}, \dots, u^n) \mapsto I(f_1(u^{j+1}, \dots, u^n), \dots, f_j(u^{j+1}, \dots, u^n), u^{j+1}, \dots, u^n) \ .$$

The standard derivative test for unconstrained minimizers yields the result after some calculation. Details are left as an exercise.

Here is a more intuitive argument for a *single* constraint in 2 or 3 variables.

We'll assume $DK(u_0) \neq 0$, because otherwise the 'or else' case in the theorem is satisfied; and this assumption guarantees that the level set $K^{-1}(\{0\})$ has a tangent at u_0. If $DI(u_0) = 0$, the conclusion of the theorem is satisfied with $\lambda = 0$; so we may now assume $DI(u_0) \neq 0$, giving the level set $I^{-1}(\{I(0)\})$ a smooth tangent at u_0, too.

If I is smallest at u_0 among all thse competitors u on the line/surface $\{u \mid K(u) = 0\}$, then the level set of I through u_0 cannot be crossed by the set $\{u \mid K(u) = 0\}$. So the level sets must be tangential, i.e., their normal vectors must be parallel: $\nabla K(u_0) \parallel \nabla I(u_0)$. But this means $DI(u_0) = \lambda DK(u_0)$.

Definition 1.9 (Derivatives in $\mathbb{R}^n$) *Let $G \in \mathbb{R}^n$ be open.*

(a) A function $I : G \to \mathbb{R}$ is differentiable at $u_0 \in G$ if a $1 \times n$ matrix $DI(u_0)$ exists such that

$$\frac{\|I(u) - I(u_0) - DI(u_0) \cdot (u - u_0)\|}{\|u - u_0\|} \to 0 \quad \textit{as} \quad u \to u_0$$

(b) The directional derivative of I at u_0 in direction v is defined as the single-variable derivative

$$\partial_v I(u_0) := \frac{d}{d\varepsilon} I(u_0 + \varepsilon v)\Big|_{\varepsilon=0}$$

If I is differentiable at u_0, then $\partial_v I(u_0) = DI(u_0) \cdot v$, and is sometimes written, using transposition and the dot product, as $v \cdot \nabla I(u_0)$.

(c) The partial derivatives are the directional derivatives in the coordiante directions $v_1 = [1, 0, \dots, 0]^T$, ..., $v_n = [0, \dots, 0, 1]^T$.

(d) Even if all directional derivatives exist, I may not be differentiable; however if all partial derivatives exist and *are continuous, then I is differentiable and $u \mapsto DI(u)$ is continuous.*

In Calculus of Variations, we typically have a vector space of functions, e.g., C^0 or C^1, instead of $\mathbb{R}^n$, together with a norm $\|\cdot\|$. We will also have the situation that this vector space is a Banach space, by which we mean that the completeness axiom "Every Cauchy sequence has a limit" holds. The same definition of differentiable applies as in $\mathbb{R}^n$, with the only modification that $DI(u_0)$ is to be understood as a continuous linear mapping from the Banach space into $\mathbb{R}$ instead of a $1 \times n$ matrix.[2]

The necessary conditions for a minimum 'derivative must vanish' in Calculus of Variations can be obtained from directional derivatives alone. The definition of directional derivatives carries over literally. Just as in $\mathbb{R}^n$, existence of directional derivatives is a weaker requirement than differentiability, much weaker actually.

We do not talk about partial derivatives in Calculus of Variations, simply because we do not have standard coordiante directions in the function spaces we are considering.

We do however often restrict the directions v in which we take directional derivatives $\partial_v I(u_0)$. Even if we require from u_0 do have only one derivative, we may still choose to restrict the directions to be infinitely differentiable functions. This usually suffices to extract the necessary conditions, just as in $\mathbb{R}^n$ we only choose directional derivatives in coordinate directions for practical calculation of minima.

In traditional texts, these 'directions in function space' are called variations, as they were invented earlier than were the function spaces that unified the finite and infinite dimensional case. So when we will study the directional derivative $\partial_v I(u_0)$, which is a limit $(I(u_0+\varepsilon v) - I(u_0))/\varepsilon$, traditional books write δu (a single symbol, not a product) for $\varepsilon \cdot v$ with δu called a variation of u. This is the source for the name 'Calculus of Variations'.

Frequently used variations are C^∞ with compact support (C^∞_{cpt}). One example of such a function (in a single variable) is given by $\phi(x) := \exp[-\frac{1}{1-x^2}]$ for $-1 < x < 1$ and $\phi(x) = 0$ for $|x| \geq 1$.

[2] If you see this definition of differentiability the first time, it takes a while to sink in. Math majors in this situation are encouraged to ponder the definition and check whether certain functionals I as given in the examples are indeed differentiable. Engineers needn't worry about this definition now. It should be mentioned that the choice of Banach space can have significant influence on whether a certain expression is differentiable or not.

Chapter 2

The Euler–Lagrange Equations of a Variational Problem, and the Legendre Condition.

Example 2.1 (cf Hwk. 2)
If u_* minimizes $I[u] := \int_0^1 (u'^2 - u^2 + 2u)\,dx$ among all $u \in C^1[0,1]$ satisfying $u(0) = 0$ and $u(1) = 0$, then for every fixed $v \in C^1[0,1]$ satisfying $v(0) = 0 = v(1)$, we are guaranteed that $\varepsilon = 0$ minimizes the single variable function $\varepsilon \mapsto I[u_* + \varepsilon v]$, and hence, from SV Calculus, we conclude that $\frac{d}{d\varepsilon} I[u_* + \varepsilon v]|_{\varepsilon=0} = 0$. Also we conclude from SV Calculus that $\frac{d^2}{d\varepsilon^2} I[u_* + \varepsilon v]|_{\varepsilon=0} \geq 0$.

Let's calculate this directional derivative.

$$
\begin{aligned}
I[u + \varepsilon v] &= \int_0^1 \{(u' + \varepsilon v')^2 - (u + \varepsilon v)^2 + 2(u + \varepsilon v)\}\,dx \\
&= I[u] + \varepsilon \int_0^1 (2u'v' - 2uv + 2v)\,dx + \varepsilon^2 \int_0^1 (v'^2 - v^2)\,dx \\
\frac{d}{d\varepsilon} I[u + \varepsilon v]\Big|_{\varepsilon=0} &= 2 \int_0^1 (u'v' - uv + v)\,dx
\end{aligned}
$$

So, for u_* to be a minimum, u_* must satisfy $\int_0^1 (u_*'v' - u_*v + v)\,dx = 0$ for every v as specified above.

Let's first cheat a bit an assume that u_* is actually C^2 (we wouldn't know this beforehand, so we'll have to reconsider the argument without this unwarranted hypothesis); then we can integrate by parts, and using $v(0) = 0 = v(1)$ will get rid of the boundary terms: we conclude $\int_0^1 (-u'' + u + 1)v\,dx = 0$ for every v as specified above. We claim, and will prove shortly, that this can only happen if $-u'' + u + 1 \equiv 0$. This equation is called the *Euler–Lagrange equation* of the variational problem $\min I[u]$. Together with the boundary conditions $u(0) = 0 = u(1)$, this equation has exactly one solution $u_*(x) = 1 - \cos x - (\tan \frac{1}{2}) \sin x$. By plugging this in, we find $I[u_*] = 1 - 2\tan\frac{1}{2} \approx -0.0926$.

Lessons from this Example:

- Setting the directional derivatives 0 gives conditions on a minimizer, namely a differential equation, which we can try to solve to get the only candidate(s) for a minimizer.

- The differential equation has *more* derivatives than the variational problem (actually twice as many); to get rid of the direction requires an integration by parts, which introduces the extra derivative. But this requires a justification since we do not want to assume that a minimizer has these extra derivatives. (We'll be able to give this justification easily in the ODE case, i.e., if the VP has a single-variable integral. In the PDE case, when the VP has multi-variable integrals, such a justification is usually possible, but is not as easy.)

Theorem 2.2 (Fundamental Lemma of Calculus of Variations) *Suppose $f : [a,b] \to \mathbb{R}^n$ is continuous and satisfies $\int_a^b f(x) \cdot v(x)\,dx$ for every $v \in C^\infty_{cpt}(]a,b[\to \mathbb{R}^n)$. Then $f \equiv 0$.*

PROOF: Choose $x_0 \in]a,b[$ and $i \in \{1, 2, \ldots, n\}$ arbitrarily. We want to show $f^i(x_0) = 0$. (Here f^i denotes the ith component of the vector f, not a power!) Assume the claim is false, say $f^i(x_0) > 0$. (The argument is similar if $f^i(x_0) < 0$.) By continuity, there is an interval $[x_0-\delta, x_0+\delta] \subset]a,b[$ on which $f^i(x) > 0$. Choose a nonnegative function $v^i \in C^\infty_{cpt}(]a,b[\to \mathbb{R})$ and choose the other components v^j to be 0. Then $\int_a^b f(x)\cdot v(x)\,dx = \int_{x_0-\delta}^{x_0+\delta} f^i(x)v^i(x)\,dx > 0$ in contradiction to the hypothesis. So we have now shown $f^i(x_0)$ for every $x_0 \in]a,b[$ and every component i. This implies $f \equiv 0$ in the open interval $]a,b[$ and then by continuity also on the closed interval $[a,b]$. ■

Note that we have used the $n = 1$ case of this lemma in our example.

The Fundamental Lemma is found in most books on Calculus of Variations. It is fundamental because with it, we can get rid of the direction v and obtain the EL equation. However, to use it, an unwarranted integration by parts is needed, assuming more differentiability of the solution than what is legitimate to assume. For this reason, Calculus of Variations should (and can) be done without the Fundamental Lemma of Calculus of Variations, as shown in the next lemma. Nevertheless, the Fundamental Lemma is commonly used in the theory of differential equations.

Theorem 2.3 (Lemma of DuBois–Reymond) *Let $f \in C^0([a,b] \to \mathbb{R}^n)$ and assume $\int_a^b f(x) \cdot v'(x)\,dx = 0$ for every $v \in C^1_0([a,b] \to \mathbb{R}^n) := \{v \in C^1([a,b] \to \mathbb{R}^n) \mid v(a) = 0 = v(b)\}$. Then f is a constant function.*

Before proving the theorem I should note that one can generalize the theorem to the weaker hypothesis that the condition is satisfied merely for all $v \in C^\infty_{cpt}([a,b] \to \mathbb{R}^n)$. This because every $v \in C^1_0$ can be viewed as the limit of $v_j \in C^\infty_{cpt}$ in such a way that $\lim \int f v_j'\,dx = \int f v'\,dx$, so the stronger hypothesis can be deduced from the weaker one. Proof details for this generalization are not important for us at present.

PROOF: We choose $v(x) := \int_a^x (f(t) - c)\,dt$ for a constant c yet to be determined. Clearly $v \in C^1$ and $v(a) = 0$. In order to get $v(b) = 0$ as well, we choose $c = \frac{1}{b-a}\int_a^b f(t)\,dt$. Then the hypothesis of the lemma can be used with this v, so $\int_a^b f(x) \cdot v'(x)\,dx = 0$. This means

$$0 = \int_a^b f(x) \cdot (f(x) - c)\,dx = \int_a^b |f(x) - c|^2\,dx + c \cdot \int_a^b (f(x) - c)\,dx = \int_a^b |f(x) - c|^2\,dx$$

Since the last integrand is nonnegative, and continuous, we conclude $f(x) - c \equiv 0$. ■

Let's now return to Example 2.1. With the DBR Lemma, we integrate the other term by parts: rather than moving the derivative off v', we move a derivative onto v. From $\int_0^1 \{u'v' - (u-1)v\}\,dx$, we conclude:

$$\int_0^1 \left\{ u'(x)v'(x) + \int_0^x (u(t)-1)\,dt\, v'(x) \right\} dx - \left[\int_0^x (u(t)-1)\,dt\, v(x) \right]_{x=0}^{x=1} = 0$$

The boundary term vanishes because v vanishes on the boundary. The DBR lemma implies that

$$u'(x) + \int_0^x (u(t)-1)\,dt \equiv c$$

Since u was assumed to be C^1, the integral is actually C^2 as an anti-derivative of a C^1 function, hence by the equation $u' \in C^2$, so actually $u \in C^3$ even. Repeating this argument gives that $u \in C^\infty$. Even if we had made weaker assumptions on u, the weakest reasonable assumption being that u has some kind of derivative (in a sufficiently generalized sense allowing eg. for jump discontinuities in the derivative) such that u'^2 is still integrable, these assumptions would at least have implied that u is continuous, hence the integral is C^1, hence $u' \in C^1$ and therefore $u \in C^2$, and the argument could still be salvaged, subject to proving a slightly generalized version of the DBR lemma.

The lesson from this example is that 'decent' variational problems tend to select minimizers that have more smoothness than is built into the original problem. We have to state precise hypotheses for this to happen. There do exist exceptions to this rule of thumb, when the hypotheses defining 'decent' VPs are violated.

Next let's look in our example what the 2nd derivative tells us:

$$\frac{d^2}{d\varepsilon^2} I[u_* + \varepsilon v]\Big|_{\varepsilon=0} = 2\int_1^1 (v'^2 - v^2)\,dx$$

By special coincidence in this example (more specifically, because I is quadratic in u), this second derivative doesn't depend on u_*; in general, it would of course depend on u_*.

It can be shown (but we'll not distract ourselves right now and skip the proof[1]) that this expression is strictly positive for every v other than $v \equiv 0$. So we are in a situation that is analogous to the 'Hessian positive definite' test in multi-variable calculus. But whether in the infinite dimensional case, this hypothesis still implies that u_* is a relative minimum, is a question that needs to be studied. There will be subtleties, so a clear-cut yes-or-no answer cannot be given right now. Sufficient conditions is when CV becomes interesting.

However, in this particular example, we are not relying on derivatives only, again because I is quadratic. We know, for *every* u that

$$I[u] = I[u_*] + \int_0^1 (2u_*'v' - 2u_*v + 2v)\,dx + \int_0^1 (v'^2 - v^2)\,dx$$

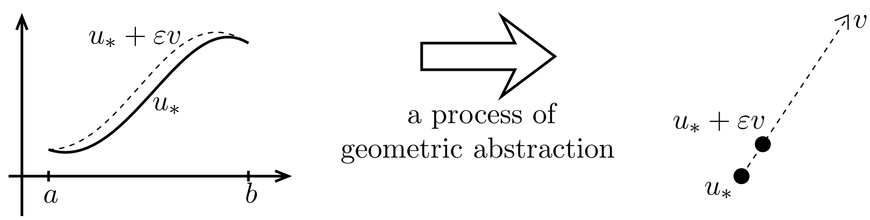
wth $v = u - u_*$. You can check this by mere algebra. Basically that is the fact that for a quadratic expression, the 2nd order Taylor approximation is already exact everywhere. The

[1]Ok, if you are curious: Writing v as a sine Fourier series $v = \sum b_k \sin k\pi x$ implies easily that $\int_0^1 v'^2(x)\,dx \geq \pi^2 \int_0^1 v^2(x)\,dx$. Alternatively, if you do not want to rely on Fourier series, you can get a slightly weaker result from the Cauchy-Schwarz inequality: For $0 \leq 1/2$, you have $|v(x)|^2 \leq x \int_0^x v'(t)^2\,dt$ (using $v(0) = 0$) and therefore $\int_0^{1/2} v(x)^2\,dx \leq \int_0^{1/2} x\,dx \int_0^{1/2} v'(t)^2\,dt$. Combined with a similar estimate starting at $v(1) = 0$, namely $\int_{1/2}^1 v(x)^2\,dx \leq \int_{1/2}^1 (1-x)\,dx \int_{1/2}^1 v'(t)^2\,dt$, we infer $\int_0^1 v^2(x)\,dx \leq (1/8) \int_0^1 v'(t)^2\,dt$.

first integral vanishes because u_* satisfies the EL equation, and the second integral is always > 0, except when $v \equiv 0$. So in this lucky example, we have actually verified that u_* is an absolute minimum.

Before we state a general theorem to the effect "If u_* is a relative minimum then the following equation must be satisfied by u_*", we'll study some examples to refine the meaning of 'relative' minimum. The definitions to be obtained will become even more crucial when we study sufficient conditions.

2.4 (History) One remark is appropriate: Before the 20th century, the idea to view functions as elements of an infinite dimensional vector space was unknown, even though a certain analogy was clearly perceived. Therefore, with the analogy between MVC and CV more vague than has been outlined in these notes, a notion like 'directional derivative' would not carry over formally. So when I call v a direction in function space, earlier one would have called εv a variation.



a function and a variation   a point and a direction in function space

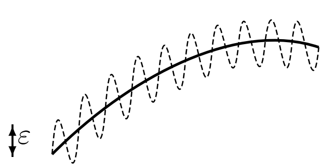
2.5 (Different Notions of Neighborhood and Relative Minimum)

1. A C^0 neigborhood (also called a wide neighborhood in CV) of a function $u_0 \in C^1[a,b]$ contains those functions $u \in C^1[a,b]$ for which $\|u - u_0\|_{C_0} < \varepsilon$ for some $\varepsilon > 0$; here

$$\|u - u_0\|_{C^0} := \max_{x\in[a,b]} |u(x) - u_0(x)|$$

2. A C^1 neigborhood (also called a narrow neighborhood in CV) of a function $u_0 \in C^1[a,b]$ contains those functions $u \in C^1[a,b]$ for which $\|u - u_0\|_{C_1} < \varepsilon$ for some $\varepsilon > 0$; here

$$\|u - u_0\|_{C^1} := \max\Big\{ \max_{x\in[a,b]} |u(x) - u_0(x)|, \max_{x\in[a,b]} |u'(x) - u_0'(x)| \Big\}$$

3. 

A variation that is of small size ε according to C^0 distance, but is large according to C^1-distance

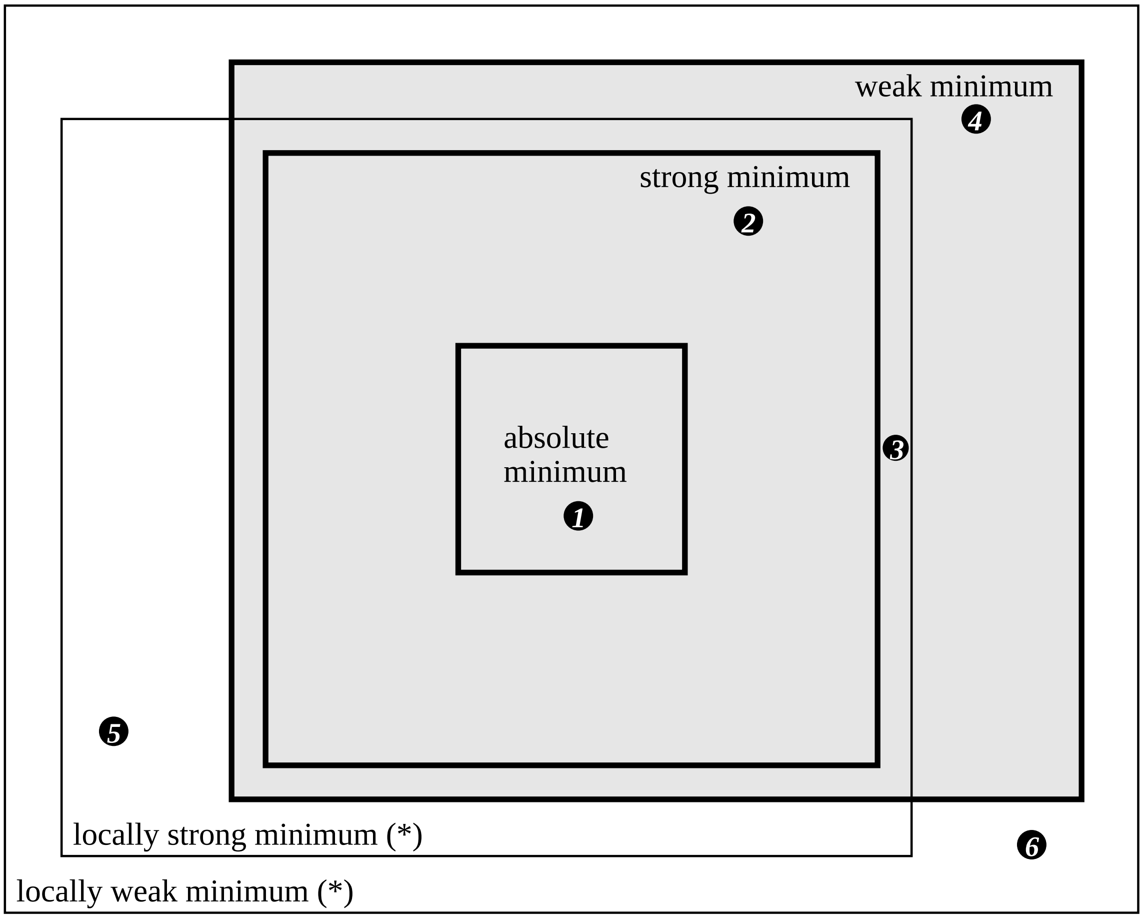
4. The notion of 'relative minimum' splits into two notions accordingly. Assume $I[u] := \int_a^b L(x, u(x), u'(x))\, dx$ for some function L.
We call u_* a *weak minimum* of a functional I if there exists $\varepsilon > 0$ such that $I[u_*] \le I[u]$ for all u such that $\|u - u_*\|_{C^1} < \varepsilon$.
We call u_* a *strong minimum* of a functional I if there exists $\varepsilon > 0$ such that $I[u_*] \le I[u]$ for all u such that $\|u - u_*\|_{C^0} < \varepsilon$.
We call u_* an *absolute minimum* of I if $I[u_*] \le I[u]$ for all u in the domain of I.

5. For the purpose of this class, I'll call the weak and strong minima also *'globally weak'* and *'globally strong'*. I will define the notions of 'locally weak' and 'locally strong' minima as follows: u_* defined on $[a, b]$ is called a *locally weak* (or *locally strong*) minimum if for each $x_0 \in]a, b[$, there is a small interval $[x_0 - \delta, x_0 + \delta] \subset [a, b]$, and an $\varepsilon > 0$, such that $I[u_*] \leq I[u]$ for all those u satisfying $\|u - u_*\|_{C^1} < \varepsilon$ (or $\|u - u_*\|_{C^0} < \varepsilon$) that coincide with u_* outside the interval $[x_0 - \delta, x_0 + \delta]$.

   The definitions 'locally weak' and 'locally strong' are not used in the literature, but will help us organize the sufficient conditions for minima. It should be noted that the analogy in $\mathbb{R}^n$ of locally weak and locally strong minima that are not globally weak or strong minima are saddlepoints, rather than minima, because the variations with small support $[x_0 - \delta, x_0 + \delta]$ represent directions in which u_* does look like a minimum, whereas the failure to be a globally weak or strong minimum implies the presence of variations (directions) in which the function I decreases.

6. Refer to Tables 2.1 and 2.2 for examples and comparisons of these notions

Table 2.1: **Comparing Various Notions of Minimality in Calculus of Variations**

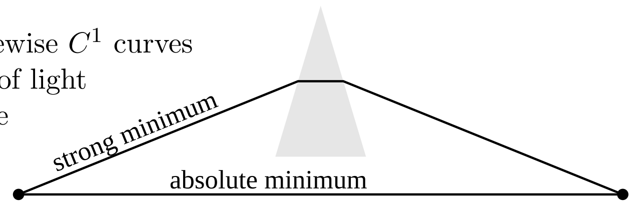


(*) Here we denote as *locally* strong or *locally* weak minima those extremals for which the strong or weak minimality property is only satisfied on sufficiently short segments. In other words every sufficiently short subsegment will be a strong or weak minimal. (I have made up these definitions of "locally strong" or "locally weak" ad hoc. They are not part of generally used mathematical language.)
Such extremals that are merely "locally weak minima" or "locally strong minima" are NOT relative minima in any functional analytic sense, but are saddle points. Genuine (relative) minima are found in the gray areas of the diagram.

Table 2.2: **Examples for Various Notions of Minimality**

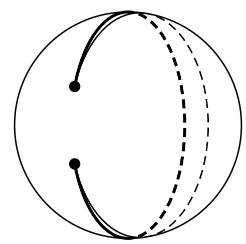
❶ The length functional, defined for all C^1 curves connecting two given points in the plane. The straight segment yields the absolute minimum of the length.

❷ Travel time for light, defined for all piecewise C^1 curves connecting two given points: The speed of light in glass is less than in air; this causes the refraction in the prism. The polygonal path is a strong minimum, but not the absolute minimum.



❹ The functional $I[y] := \int_{-1}^{1}(\dot{y}^2 - \dot{y}^4)\,dt$, defined for $y \in C^1[-1,1]$ with $y(-1) = y(1) = 0$: $y^* \equiv 0$ is a weak minimum, since for $|\dot{y}| < \varepsilon$ it holds: $\dot{y}^2 - \dot{y}^4 \geq (1-\varepsilon^2)\dot{y}^2$. But even short segments of $y^* \equiv 0$ aren't strong minimals; indeed, choose $\tilde{y}(t) := \varepsilon \sin^2 n(t - t_0)$ on $[t_0, t_0 + \pi/n]$ and $\tilde{y} = 0$ elsewhere: $I[\tilde{y}] = \frac{1}{8}\pi\varepsilon^2 n(4 - 3\varepsilon^2 n^2)$, which is negative for large n.

❺ The length functional, defined for all C^1 curves connecting two given points *on the sphere*: the great circle that connects the points 'on the back of the sphere' in the figure is not even a weak minimum. The smaller circle is shorter, but still ε-close in C^1. Nevertheless, sufficiently short segments of the great circle (short enough such as not to contain antipodes) are the absolutely shortest connections between their endpoints.



❻ Choose the functional $\int_0^\pi(\dot{y}^2 - \dot{y}^4 - 2y^2)\,dt$ on $C_0^1[0,\pi]$. Then $y \equiv 0$ is a critical point, but is not even weakly minimal. For $\tilde{y}(t) = \varepsilon \sin t$, it holds $I[\tilde{y}] = -\frac{1}{8}\pi\varepsilon^2(4 + 3\varepsilon^2)$. On short segments however, it is a weak minimum, but not a strong one; the reason is the same as in case ❹. To prove weak minimality, segments $[t_0, t_1] \subset [0,\pi]$ need to be so short that it always holds $\int_{t_0}^{t_1} \dot{y}^2\,dt > 2\int_{t_0}^{t_1} y^2\,dt$, and this happens if $(t_1 - t_0)^2 < \pi^2/2$.

❸ This case is a curiosity for which I could only find a somewhat artificial example: A weak minminal such that the functional can be made smaller by means of strong oscillations, but only if these strong oscillations occur on long segments. I am giving an example with vector-valued $y = (y_1, y_2)$, in which the functional itself is also more complicated than a simple integral expression:

$$I[y_1, y_2] := \int_0^\pi \left(\dot{y}_1^2 - \frac{1}{2}y_1^2\right)\,dt + \left(\int_0^\pi (\dot{y}_1^2 - 2y_1^2)\,dt\right)\left(\int_0^\pi \dot{y}_2^2\,dt\right)$$

on $C_0^1([0,\pi] \to \mathbb{R}^2)$. We conclude $I[y_1, y_2] \geq \left(\frac{1}{2} - \int \dot{y}_2^2\right)\left(\int \dot{y}_1^2\right)$, because $C_0^1[0,\pi]$ functions satisfy $\int \dot{y}^2 \geq \int y^2$; hence $(y_1, y_2) \equiv (0,0)$ is weakly minimal.

It is not strongly minimal, because for $\tilde{y}_1(t) = \varepsilon \sin t$, $\tilde{y}_2 = \varepsilon \sin nt$ it holds $4I[\tilde{y}_1, \tilde{y}_2] = \varepsilon^2\pi - n^2\varepsilon^4 n^2$, which is negative for large n.

However, if we consider short segments $[t_0, t_1]$ only, then the term $\int_{t_0}^{t_1}(\dot{y}_1^2 - 2y_1^2)\,dt$ will become ≥ 0, and oscillations in $\int \dot{y}_2^2$ cannot do harm to the minimum property.

Theorem 2.6 *Consider the functional*

$$I : u \mapsto \int_{t_0}^{t_1} L(t, u(t), u'(t))\, dt$$

where $L \in C^1([t_0, t_1] \times G \times \mathbb{R}^n \to \mathbb{R})$ and $G \subset \mathbb{R}^n$ is open. Suppose $u_* \in C^1([t_0, t_1] \to \mathbb{R}^n)$ is locally a weak minimum (possibly subject to the boundary conditions $u(t_0) = u_0$, $u(t_1) = u_1$). Then u_* satisfies the Euler–Lagrange equation

$$\begin{aligned} \int^t L_{u_i}(t, u_*(t), u_*'(t))\, dt - L_{u_i'}(t, u_*(t), u_*'(t)) &= const \\ L_{u_i}(t, u_*(t), u_*'(t))\, dt &= \tfrac{d}{dt} L_{u_i'}(t, u_*(t), u_*'(t)) \end{aligned} \tag{2.1}$$

Proof: For the moment, let's assume that u_* is globally a weak minimum (we'll remove this extra hypothesis later). Choose any $v \in C_0^1[t_0, t_1]$. Then $u_* + \varepsilon v$ still satisfies the same boundary conditions as u_*, if boundary conditions were required, and $u_* + \varepsilon v$ is close to u_* in C^1-norm. Moreover, if the values $u_*(t)$ are in the open set G, then the values of $(u_* + \varepsilon v)(t)$ are also still in G, provided ε is suffiicently small. The weak minimum property implies that $\varepsilon \mapsto I[u_* + \varepsilon v]$ has a relative minimum at $\varepsilon = 0$. So, $\partial_v I[u_*] = \frac{d}{d\varepsilon} I[u_* + \varepsilon v]\Big|_{\varepsilon=0} = 0$. Let's drop the $*$ subscript for simplicity of notation:

$$\begin{aligned} 0 = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} I[u_* + \varepsilon v] &= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int_{t_0}^{t_1} L(t, u(t) + \varepsilon v(t), u'(t) + \varepsilon v'(t))\, dt \\ &= \int_{t_0}^{t_1} \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} L(t, u(t) + \varepsilon v(t), u'(t) + \varepsilon v'(t))\, dt \\ &= \int_{t_0}^{t_1} \sum_i \Big\{ L_{u_i}(\cdots) v_i(t) + L_{u_i'}(\cdots) v_i'(t) \Big\}\, dt \Big|_{\varepsilon=0} \\ &= \int_{t_0}^{t_1} \sum_i \Big\{ L_{u_i}(t, u(t), u'(t)) v_i(t) + L_{u_i'}(t, u(t), u'(t)) v_i'(t) \Big\}\, dt \end{aligned}$$

We'll justify the exchange of $\frac{d}{d\varepsilon}$ with $\int \ldots dt$ a bit later. Next, doing the integration by parts that converts the v into v', we conclude

$$\begin{aligned} 0 = \sum_i & \left[\int_{t_0}^{t} L_{u_i}(s, u(s), u'(s))\, ds\, v_i(t) \right]_{t_0}^{t_1} \\ &+ \int_{t_0}^{t_1} \sum_i \Big\{ - \int_{t_0}^{t} L_{u_i}(s, u(s), u'(s))\, ds\, v_i'(t) + L_{u_i'}(t, u(t), u'(t)) v_i'(t) \Big\}\, dt \end{aligned}$$

The integrated term vanishes because $v(t_0) = 0 = v(t_1)$. By the Lemma of DuBois–Reymond, we conclude the integrated version of the Euler–Lagrange equation (2.1). From the hypothesis, the $L_{u_i}(\cdots)$ term under the integral is continuous, and therefore the fundamental theorem of calculus allows differentiation, and we get the second form of (2.1).

We still have to justify the exchange of the limits $\frac{d}{d\varepsilon}$ and $\int \ldots dt$. This is easy here: The difference quotients $(L(\cdots \varepsilon_1 \ldots) - L(\cdots \varepsilon \ldots))/(\varepsilon_1 - \varepsilon)$ converge uniformly as $\varepsilon_1 \to \varepsilon$, because they extend to a continuous function on the compact set $[t_0, t_1] \times [-\hat{\varepsilon}, \hat{\varepsilon}]$, hence a uniformly continuous function.

We had originally assumed that u_* is *globally* a weak minimizer. Now assume merely that u_* is *locally* a weak minimizer. So for every $\tau \in\,]t_0, t_1[$, there is an interval $J := [\tau - \delta, \tau + \delta]$ such that $I[u_*] \le I[u_* + \varepsilon v]$ for all those $v \in C^1$ whose support is in J. These can be viewed as

$v \in C_0^1(J)$ that have vanishing derivative at the endpoints of v. As mentioned in the remark after the Lemma of DuBois-Reymond, this extra rectriction on v does not affect the validity of the lemma; so the same argument as before guarantees that the (second version of the) Euler-Lagrange equation is satisfied on the interval J, and therefore at every point $\tau \in]t_0, t_1[$ (and then by continuity in $t_{0,1}$ as well). ■

2.7 (Note) Differential equations are *local* in the sense that in order to check whether a function satisfies the DE, it suffices to check that it satisfies the DE in a neighborhood of each point. The examples in Table 2.2 have shown that minimality is *not* merely a local phenomenon: In order for a function u_* to minimize a functional I, it is not sufficient that short restrictions of u_* to short subintervals minimize the functional. The example with the great circles on a sphere will be a hint: When different solutions of the EL equation (close to each other) intersect twice, the segments have become long to be minima. See also the Hwk problem #14

Theorem 2.8 (Erdmann's corner condition) *Under the same hypotheses as in Thm. 2.6, except that now we allow $u_* : [t_0, t_1] \to \mathbb{R}^n$ to be piecewise C^1 (i.e., $C^0[t_0, t_1]$, and C^1 except at finitely many points —called corners— where the left and the right derivatives may differ). Then at each corner $\hat{t}$, the condition*

$$L_{u_i'}(\hat{t}, u_*(\hat{t}), u_*'(\hat{t}-)) = L_{u_i'}(\hat{t}, u_*(\hat{t}), u_*'(\hat{t}+))$$

is satisfied.

Proof: Without loss of generality, we may assume that u_* has just one corner $\hat{t}$. Otherwise, we take shorter segments of u_* that do have a single corner and restrict variations v to be supported on these shorter segments only. Redo the calculation from the proof of Thm. 2.6 with $\int_{t_0}^{t_1} = \int_{t_0}^{\hat{t}} + \int_{\hat{t}}^{t_1}$. You may do the integration by parts along the lines of the fundamental lemma rather than along the lines of DBR (why?). Fill in the details in Hwk. #10. ■

Theorem 2.9 (Natural Boundary Conditions) *Under the same hypotheses as in Thm. 2.6, except that there are no boundary conditions prescribed, a minimal satisfies the EL equations with the* natural boundary conditions

$$p(t_0) = 0 = p(t_1)\,, \quad \textit{where} \;\; p_i(t) := L_{u_i'}(t, u_*(t), u_*'(t))$$

Proof: Variations $v \in C_0^1[t_0, t_1]$ are still allowed and give the EL equations. But now, we can also take $v \in C^1[t_0, t_1]$ with nonvanishing $v(t_{0,1})$. Consider the boundary terms in the integration by parts. Fill in the details in Hwk. #12. ■

Example 2.10 (Pedestrian Calculation for the Brachystochrone)
We have $L = \sqrt{1 + y'^2}\, y^{-1/2}$. So

$$\begin{aligned} L_y &= \sqrt{1 + y'^2}\,(-\tfrac{1}{2})y^{-3/2} \\ L_{y'} &= y'(1 + y'^2)^{-1/2}\, y^{-1/2} \end{aligned}$$

It should be understood that these are partial derivatives of the 3-variable function L, so y and y' are treated as separate variables in this step. The EL equation is therefore $\frac{d}{dx}L_{y'} = L_y$, where now we have inserted $y(x)$ and $y'(x)$ in the y and y' slots of L and its partial derivatives, and $\frac{d}{dx}$ is a total derivative with respect to x. So

$$\frac{d}{dx}L_{y'} = \frac{d}{dx}\frac{y'}{\sqrt{1+y'^2}}\,y^{-1/2} = \frac{y''\sqrt{1+y'^2} - y'\,(y'/\sqrt{1+y'^2})\,y''}{1+y'^2}y^{-1/2} + \frac{y'}{\sqrt{1+y'^2}}(-\tfrac{1}{2})y^{-3/2}y'$$
$$= \frac{y''y^{-1/2}}{(1+y'^2)^{3/2}} - \frac{y'^2y^{-3/2}}{2(1+y'^2)^{1/2}}$$

Multiplying $\frac{d}{dx}L_{y'} = L_y$ with $y^{3/2}(1+y'^2)^{1/2}$, we conclude

$$-\frac{1}{2}(1+y'^2) = \frac{y''y}{1+y'^2} - \frac{y'^2}{2}\,, \qquad \text{or} \qquad -\frac{1}{2} = \frac{y''y}{1+y'^2}$$

We will have to return to this calculation yet, because it has assumed, rather than proved, that indeed y'' exists. Let us first try to solve the ODE obtained here; this is not trivial, but fortunately it can be done. (We will learn a smart shortcut later.) Ingenuity finds that y'/y is an integrating factor: multiplying the DE with it, we conclude

$$\frac{y''y'}{1+y'^2} = -\frac{y'}{2y} \quad \text{and therefore} \quad \frac{1}{2}\ln(1+y'^2) = -\frac{1}{2}\ln y + \mathit{const}$$

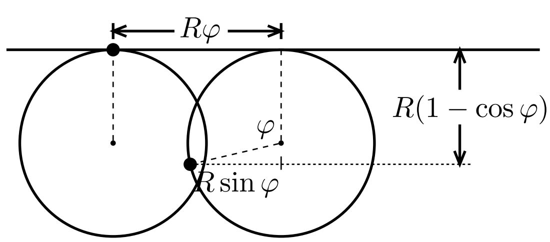
This implies $1 + y'^2 = 2R/y$ (where the choice to call the constant $2R$ will be convienient later). Then $y' = \pm\sqrt{2R/y - 1}$, a separable equation, which can be integrated to

$$\pm\int \frac{y\,dy}{\sqrt{2Ry - y^2}} = \int dx$$

In view of $\sqrt{2Ry - y^2} = \sqrt{R^2 - (R-y)^2}$, a trigonometric substitution $R - y = R\cos\varphi$, $\pm\sqrt{R^2-(R-y)^2} = R\sin\varphi$ allows to evaluate the integral onthe left: $\int R(1-\cos\varphi)\,d\varphi = \int dx$ and therefore $R(\varphi - \sin\varphi) = x + \mathit{const}$. Since we want $y = 0$ (hence $\varphi = 0$) when $x = 0$, the constant of integration must be 0. We have obtained $y(x)$ implicitly, in parametrized form:

$$x = R(\varphi - \sin\varphi)\,, \qquad y = R(1-\cos\varphi)$$

This curve describes a cycloid: the curve traced by a point on the circle of radius R as this circle rolls along a straight line.



We conclude that the only candidates for the brachystochrone are cycloids. We'll later show that there is a unique cycloid connecting given points, and that this cycloid is indeed an absolute minimum.

Let's first justify that $y \in C^2$. We used the usual rules from calculus to evaluate $\frac{d}{dx}\frac{y'}{\sqrt{1+y'^2}}y^{-1/2}$ to an expression involving y''. The EL equation merely implies that $L_{y'}(t, y(t), y'(t)) =$

$y'y^{-1/2}/\sqrt{1+y'^2}$ is C^1. So as long as $y > 0$, we know $z := \frac{y'}{\sqrt{1+y'^2}}$ is C^1. This can be solved for y', namely $y' = z/\sqrt{1-z^2}$. So y' is a C^1 function of z (as long as $|z| < 1$, which is the case), and z is a C^1-function of t. We relied on being able to invert the relation $y' \mapsto L_{y'}(\cdots, y')$.

With the experience from this example, we can show this regularity result for solutions of the EL equation:

Theorem 2.11 (Regularity) *Let $L \in C^2([t_0, t_1] \times G \times \mathbb{R}^n \to \mathbb{R})$. If $u_* \in C^1$ solves the EL equation and the Hessian $L_{u'u'}$ is invertible along u_* (i.e., for each t, the matrix valued function $t \mapsto L_{u'u'}(t, u_*(t), u_*'(t))$ is invertible), then $y_* \in C^2$. Moreover, if $L \in C^n$ for $n \geq 2$, then $u_* \in C^n$.*

The theorem is not quite optimal, and using the Hamilton formalism below could give a stronger result.

PROOF: Since the Jacobi matrix (total derivative) of the function

$$u' \mapsto p := L_{u'}(t, u, u') \;, \quad \mathbb{R}^n \to \mathbb{R}^n$$

is invertible at the point $(t, u_*(t), u_*'(t))$ (and by continuity in a neighborhood thereof), namely this Jacobi matrix is the Hessian mentioned in the theorem, the theorem of implicit functions guarantees that we can write u' as a C^1 function of p, namely $u' = \Phi(t, u, p)$ with $\Phi \in C^1$. Plugging the C^1 functions $t \mapsto u(t)$ and $t \mapsto p(t)$ in, results in u' being a C^1 function of t. For higher derivatives, the argument can be carried on inductively. ■

In the example of the Brachystochrone, we were able to reduce the 2nd order EL equation to a 1st order equation by means of an integrating factor. This was no coincidence, but is a general feature of integrands L that do not explicitly contain the integration variable:

Theorem 2.12 (The Energy Integral) *If $L \in C^2$ does not depend on the integration variable explicitly, so we study $I[u] := \int_{t_0}^{t_1} L(\not{t}, u, u')\, dt$, then every C^2 solution to the EL equation satisfies $L - u'L_{u'} \equiv const$. (If u is vector valued, the term $u'L_{u'}$ is to be understood as a dot product).*

PROOF:

$$\frac{d}{dt}(L - u'L_{u'}) = L_t + L_u u' + L_{u'}u'' - u''L_{u'} - u'\frac{d}{dt}L_{u'} = 0$$

since L_t vanishes by hypothesis, the 2nd term cancels the last by use of the EL equation. In the vector valued case, expressions like $L_u u'$ are to be understood as $\sum_i L_{u_i} u_i'$ and the proof is the same. ■

Note that for the scalar valued case, the energy integral retains almost all information from the EL equation (except that multiplication of the EL equation with the integrating factor u' has introduces spurious constant solutions. In the vector valued case however, the energy equation is significantly weaker than the EL equations: we dot-multiplied the EL equations with u', thus obtaining a scalar valued equation from a vector valued equations.

The naming comes from Newtonian mechanics. In this example, the forces are gradients of a scalar function $-V$ (where V is called potential and the $-$ sign is a convention customarily made in physics, so that the force pulls 'downhill' whereas the gradient points 'uphill'). So the Newton equations

$$m_i \vec{u}_i''(t) = -\vec{\nabla}_i V(\vec{u}_1, \ldots, \vec{u}_n)$$

arise as EL equations of the functional

$$I[\vec{u}_1, \ldots, \vec{u}_n] := \int_{t_0}^{t_1} \left(\sum_i \frac{1}{2} m_i |\vec{u}_i'(t)|^2 - V(\vec{u}_1(t), \ldots, \vec{u}_n(t)) \right) dt$$

The Lagrange function L is kinetic energy minus potential energy, and the quantity $u' \cdot L_{u'} - L$ can easily be seen to be kinetic energy plus potential energy, i.e., the total menchanical energy.

This is the appropriate opportunity to move towards a second derivative test. The following Legendre condition is a local condition and as such weaker than any analog of a 2nd derivative test in MV Calculus. It is more akin to checking the diagonal of the Hessian. For a symmetric matrix A to be positive definite (semidefinite), it is necessary (but not sufficient) that the diagonal elements are positive (nonnegative).

Theorem 2.13 (Legendre Condition — "Diagonal of the Hessian") *Suppose that $u_* \in C^1$ is locally a weak minimal of $I[u] := \int_{t_0}^{t_1} L(t, u(t), u'(t))\, dt$, where $L \in C^2$. Then $L_{u'u'}(t, u_*(t), u_*'(t)) \geq 0$ for all $t \in [t_0, t_1]$. In the vector valued case, the notation $L_{u'u'} \geq 0$ refers to the positive semidefiniteness of the matrix.*

Proof: Globally weak minimality would imply

$$\frac{d^2}{d\varepsilon^2}\Big|_{\varepsilon=0} I[u_* + \varepsilon v] \geq 0$$

for every $v \in C_0^1$. Locally weak minimality still implies this same estimate for many $v \in C_{cpt}^\infty$; namely for every $\tau \in \,]t_0, t_1[$, there exists $\delta > 0$ such that the estimate applies for all $v \in C_{cpt}^\infty(]\tau - \delta, \tau + \delta[)$. Differentiating under the integral sign yields

$$\int_{t_0}^{t_1} \sum_{i,j} \Big(L_{u_i u_j} v_i v_j + 2 L_{u_i u_j'} v_i v_j' + L_{u_i' u_j'} v_i' v_j' \Big)\, dt \geq 0 \ ,$$

or in vector notation

$$\int_{t_0}^{t_1} \big(\langle v, L_{uu} v \rangle + 2 \langle v', L_{uu'} v \rangle + \langle v', L_{u'u'} v' \rangle \big)\, dt \geq 0 \ .$$

It is understood that $(t, u_*(t), u_*'(t))$ is the argument of the 2nd partials of L occurring here.

We evaluate this for v with *small* support in $]\tau - \delta, \tau + \delta[$. To this end, we scale a standard function $\psi \in C_{cpt}^\infty(]-1, 1[\to \mathbb{R})$ and choose a fixed vector $\xi \in \mathbb{R}^n$, letting $v(t) := \xi\, \psi(N(t - \tau))$ with N large (so that $1/N < \delta$). So we get

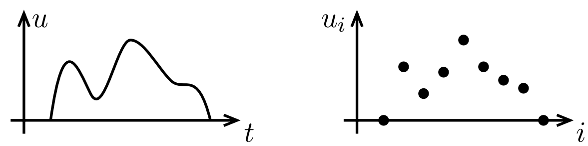
$$\begin{aligned} 0 &\leq \int_{\tau-1/N}^{\tau+1/N} \Big(\psi^2(N(t-\tau)) \langle \xi, L_{uu} \xi \rangle + 2N \psi \psi' \langle \xi, L_{uu'} \xi \rangle + N^2 \psi^2 \langle v', L_{u'u'} v' \rangle \Big)\, dt \\ &= \int_{-1}^{1} \Big(\psi^2(s)) \langle \xi, L_{uu} \xi \rangle + 2N \psi \psi' \langle \xi, L_{uu'} \xi \rangle + N^2 \psi^2 \langle v', L_{u'u'} v' \rangle \Big) \frac{ds}{N} \end{aligned}$$

Dividing by N and letting $N \to \infty$ makes only the $L_{u'u'}$ contribution survive and implies

$$\Big\langle \xi, L_{u'u'}(\tau, u_*(\tau), u_*'(\tau))\xi \Big\rangle \int_{-1}^{1} \psi'^2(s)\, ds \geq 0$$

for every $\tau \in]t_0, t_1[$ and every $\xi \in \mathbb{R}^n$. This is the claimed Legendre condition for $\tau \in]t_0, t_1[$. For $\tau \in \{t_0, t_1\}$ follows by continuity. ■

Remark: The condition "$L_{u'u'}$ positive semidefinite (pointwise along the solution)" is much weaker than a condition "$D^2 I[u_*]$ positive semidefinite" (regardless of the precise definition of the latter in function space). Remember the coordinate analog between vector spaces of functions and $\mathbb{R}^n$:



$D^2 I[u_*]$ pos semidef (any direction v)	Hessian pos semidef (any direction v)
$\Downarrow$ $\not\Uparrow$	$\Downarrow$ $\not\Uparrow$
$L_{u'u'}$ pos semidef (only v with tiny supp)	diagonal of Hessian pos semidef (only coord vec's v)

The style of calculation we have done here, together with a rigorous, but 'routine' application of Banach space formalism, can be made into a theorem:

Theorem 2.14 *Let $L \in C^2([a,b] \times G \times \mathbb{R}^n$ and assume $u_* \in C^1([a,b] \to \mathbb{R}^n)$ satisfies the Euler–Lagrange equation $L_u = \frac{d}{dt} L_{u'}$ and the strict Legendre condition holds, namely $L_{u'u'}(t, u_*(t), u_*'(t))$ is positive definite. Then u_* is locally a weak minimum.*

The proof of this theorem that follows comes with different advice fro different audience:

- Those that are not familiar with normed vector spaces or Banach spaces and do not aspire to learn it, may ignore the proof and should view the above analogy as an intuitive substitute for the proof. But they should learn the content of the theorem.
- Those who have an idea about derivatives of functions in Banach spaces, but are not sufficiently prolific with using them in action, should use the proof as an example of such use and can deepen their understanding of the formalism with this example.
- Those that have mastered the formalism can simply verify that the proof is correct.

2.15 (Proof of Thm. 2.14) We define the Banach space $X := C^1([a,b] \to \mathbb{R}^n)$ with the norm

$$\|u\| := \max\{\max_{t\in[a,b]} |u(t)|, \max_{t\in[a,b]} |u'(t)|\} \ .$$

For $G \in \mathbb{R}^n$ open, we have the open subset $X_G := \{u \in X \mid u(t) \in G \text{ for all } t \in [a,b]\} \subset X$. Then, given $L \in C^2([a,b] \times G \times \mathbb{R}^n \to \mathbb{R})$, the functional I is defined on X_G by

$$I[u] := \int_a^b L(t, u(t), u'(t))\, dt \ , \quad I : X_G \to \mathbb{R} \tag{2.2}$$

We first claim that $I : X_G \to \mathbb{R}$ is actually a C^2-functional, and that its first and second derivative are given by the formulas

$$DI[u]v = \int_a^b \Big\{ L_u(t,u(t),u'(t))v(t) + L_{u'}(t,u(t),u'(t))v'(t) \Big\}\, dt \tag{2.3}$$

and

$$\begin{aligned} D^2I[u](v,w) = \int_a^b \Big\{ & v(t)^T L_{uu}(t,u(t),u'(t))w(t) + \\ & + v(t)^T L_{uu'}(t,u(t),u'(t))w'(t) + w(t)^T L_{uu'}(t,u(t),u'(t))v'(t) + \\ & + v'(t)^T L_{u'u'}(t,u(t),u'(t))w'(t) \Big\}\, dt \end{aligned} \tag{2.4}$$

respectively.

• To this end, we first show that the linear mapping $v \mapsto DI[u]v$ defined by (2.3) is *continuous* from the space $X := C^1([a,b] \to G)$ to $\mathbb{R}$. This we need to show because the very definition of 'derivative', requires $DI[u]$ to be a *continuous* linear map as the analog of the Jacobi matrix. This step is therefore preliminary to differentiability. It has nothing to do with *continuous* differentiability, which is about continuity of the map $u \mapsto DI[u]$, $X_G \to \mathcal{L}(X \to \mathbb{R})$.

Note that the range $R := \{(t,u(t),u'(t)) : a \le t \le b\} \subset [a,b] \times G \times \mathbb{R}^n$ is compact for each fixed function u. For sufficiently small δ, a closed δ-neighbourhood $R_\delta := \{(t,y,y') : a \le t \le b\,,\ |y - u(t)| \le \delta\,,\ |y' - u'(t)| \le \delta\}$ of this range will therefore still lie in $[a,b] \times G \times \mathbb{R}^n$, and R_δ is still a compact set. In particular, given any such function u, the set

$$\{(t,y(t),y'(t)) \mid a \le t \le b\,,\ \|y - u\| \le \delta\}$$

lies in the compact set R_δ. Here $\|\cdot\|$ of course refers to the norm in X. The *continuous* functions L_u, $L_{u'}$, L_{uu}, $L_{uu'}$, $L_{u'u'}$ are *uniformly* continuous on the compact set R_δ.

The continuity of the linear map $DI[u]$ follows from

$$|DI[u]v| \le |b-a| \Big(\max_{R_\delta} |L_u| + \max_{R_\delta} |L_{u'}| \Big) \|v\|$$

(and the modulus of continuity is locally uniform).

The continuity (with locally uniform modulus of continuity) of the bilinear map $(v,w) \mapsto D^2I[u](v,w)$ that is defined by (2.4) follows analogously.

• At this stage, we have not shown yet that $DI[u]$ and $D^2I[u]$ indeed are the derivatives which by name they claim to be, even though it is clear from the directional derivative arguments that they are the only candidates for the job.

• We next show the continuity of the maps $u \mapsto DI[u]$ and $u \mapsto D^2I[u]$.

By uniform continuity of L_u and $L_{u'}$ on R_δ, we conclude that, given any ε, we can find $\eta < \delta$ such that $|L_u(t,u(t),u'(t)) - L_u(t,y(t),y'(t))| < \varepsilon$ (and a similar formula with $L_{u'}$) provided $\|u - y\| < \eta$. Then $|DI[u]v - DI[y]v| < 2\varepsilon|b-a|\|v\|$, i.e. the norm of the linear maps satisfies $\|DI[u] - DI[y]\| < 2\varepsilon|b-a|$, provided $\|u - y\| < \eta < \delta$. An analogous argument can be made for D^2I.

• Now we show that $DI[u]$ is indeed the derivative of I at u, i.e., that

$$\Big| I[u+v] - I[u] - DI[u]v \Big| / \|v\| \to 0 \ \text{ as } \ \|v\| \to 0$$

Indeed

$$
\begin{aligned}
I[u+v] - I[u] - DI[u]v = & \\
= \int_a^b \Big(& L(t,u(t)+v(t),u'(t)+v'(t)) - L(t,u(t),u'(t)) \\
& -L_u(t,u(t),u'(t))v(t) - L_{u'}(t,u(t),u'(t))v'(t) \Big)\, dt \\
= \int_a^b \int_0^1 \frac{d}{ds} \Big(& L(t,u(t)+sv(t),u'(t)+sv'(t)) \\
& -sL_u(t,u(t),u'(t))v(t) - sL_{u'}(t,u(t),u'(t))v'(t) \Big)\, ds\, dt \\
= \int_a^b \int_0^1 \Big\{ & \Big(L_u(\ldots u+sv\ldots) - L_u(\ldots u\ldots)\Big) v(t) \\
& + \Big(L_{u'}(\ldots u+sv\ldots) - L_{u'}(\ldots u\ldots)\Big) v'(t) \Big\}\, ds\, dt
\end{aligned}
$$

Now if $\|v\|$ is sufficiently small, then all occurring arguments to L_u and $L_{u'}$ lie in the set R_δ, where the functions L_u and $L_{u'}$ are uniformly continuous. Then for every ε, there exists some η such that the differences in the big parentheses will be uniformly smaller than ε, provided only $\|v\| < \eta$. We conclude that $|I[u+v] - I[u] - DI[u]v| \le 2(b-a)\varepsilon\|v\|$, which was to be shown.

By the same method it can be shown that $D^2I[u]$ is indeed the second derivative it claims to be. Namely, the estimate needs to prove that

$$\Big| DI[u+w]v - DI[u]v - D^2I[u](w,v) \Big| / \|v\|\,\|w\| \to 0$$

as $\|w\| \to 0$, and two invocations of the fundamental theorem of calculus and the uniform continuity of the 2nd derivatives of L on R_δ do the trick.

• So we have seen that $I \in C^2(X_G \to \mathbb{R})$, where X_G is the open subset of the Banach space X defined above. This implies the same approximation of I by a 2nd degree Taylor 'polynomial' as in multivariable calculus, with basically the same proof, namely:

$$\Big| I[u+v] - I[u] - DI[u]v - \tfrac{1}{2}D^2I[u](v,v) \Big| / \|v\|^2 \to 0 \quad \text{as} \quad \|v\| \to 0\,. \tag{2.5}$$

Proof of (2.5):

$$I[u+v] - I[u] = \int_0^1 \frac{d}{ds} I[u+sv]\, ds = \int_0^1 DI[u+sv]v\, ds$$

Hence

$$
\begin{aligned}
I[u+v] - I[u] - DI[u]v &= \int_0^1 \Big(DI[u+sv] - DI[u] \Big) v\, ds \\
&= \int_0^1 \int_0^s \frac{d}{d\sigma} DI[u+\sigma v]v\, d\sigma\, ds \\
&= \int_0^1 \int_\sigma^1 D^2I[u+\sigma v](v,v)\, ds\, d\sigma \\
&= \int_0^1 (1-\sigma) D^2I[u+\sigma v](v,v)\, d\sigma
\end{aligned}
$$

and

$$
\begin{aligned}
I[u+v] - I[u] - DI[u]v - \tfrac{1}{2}D^2I[u](v,v) = & \\
= \int_0^1 (1-\sigma) \Big(& D^2I[u+\sigma v](v,v) - D^2I[u](v,v) \Big)\, d\sigma
\end{aligned}
\tag{2.6}
$$

Using formula (2.4) for D^2I and the uniform continuity of L_{uu}, $L_{uu'}$, $L_{u'u'}$ on R_δ again, claim (2.5) is immediate.

We therefore have a rigorous analog of the MV-Calculus argument for relative minima. Having shown that, if $L \in C^2([a,b] \times G \times \mathbb{R}^n \to \mathbb{R})$, then $I \in C^2(X \to \mathbb{R})$, we obtain from (2.5) the 1st part of the following result immediately:

Theorem:
(a) If $L \in C^2([a,b] \times G \times \mathbb{R}^n \to \mathbb{R})$ and u_* satisfies the EL-eqn and if $D^2I[u_*](v,v) \geq c\|v\|^2$ with $c > 0$ for all $v \in C_0^1([a,b] \to \mathbb{R}^n)$ in case of fixed boundary conditions, or else for all $v \in C^1([a,b] \to \mathbb{R}^n)$ in case of free boundary, then u_* is a weak minimum of I given in (2.2).

(b) If however we only have the positive definiteness of $L_{u'u'}(t, u_*(t), u_*'(t))$ for all $t \in [a,b]$ (which implies the uniform positive definiteness in a neighbourhood R_δ), then u_* is locally *a weak minimum, i.e., sufficiently short segments of u_* are weak minima.*

As far as proving the 2nd part is concerned, we suppose that the support of v has (small) length h, and recall that by continuity, there is an upper bound M for L_{uu} and $L_{uu'}$, and by uniform positive definiteness, there is a lower bound $m > 0$ such that $v'(t)^T L_{u'u'}(\ldots)v'(t) \geq m|v'(t)^2|$.

For such v, we estimate taking a worst-case (negative) scenario for L_{uu} and $L_{uu'}$ relying only on $L_{u'u'}$ to give something positive in the end (more explanations after the formula):

$$
\begin{aligned}
D^2I[u_*](v,v) &\geq -M\int v^2 - 2M\int |v|\,|v'| + m\int v'^2 \\
&\geq -M\int v^2 - M\Big(\frac{2M}{m}\int v^2 + \frac{m}{2M}\int v'^2\Big) + m\int v'^2 \\
&\geq -\hat{M}\int v^2 + \frac{m}{2}\int v'^2 \geq \Big(\frac{m}{2} - \frac{\hat{M}h^2}{8}\Big)\int v'^2 \\
&\geq \Big(\frac{m}{2} - \frac{\hat{M}h^2}{8}\Big)(1 + h^2/8)^{-1}\|v\|^2_{W^{1,2}}
\end{aligned}
\tag{2.7}
$$

In line 2, we have estimated $2|v|\,v'|$ under the integral using the famous and immensely useful inequality $2pq \leq Ap^2 + q^2/A$. This inequality is true because $(p\sqrt{A} - q/\sqrt{A})^2 \geq 0$

In line 3 we have introduced the abbreviation $\hat{M} := M + 2M^2/m$ and then used the fact that for v supported on an interval of length h, the estimate $\int v(t)^2\,dt \leq h^2/8\int v'(t)^2\,dt$ holds (as mentioned in the footnote on pg 9.

In line 4, we have introduced a NEW norm $\|v\|^2_{W^{1,2}} := \int v^2 + \int v'^2$.

We need to (and can) take h so small that $\frac{m}{2} - \frac{\hat{M}h^2}{8} > \frac{m}{4} > 0$, and then (3), together with the EL eqn $DI[u_*]v = 0$, *ought to imply* $I[u_* + v] > I[u_*]$ (but doesn't, yet...).

The problem is that the Legendre condition has allowed us to prove $D^2I[u_*](v,v) \geq \frac{m}{4}\|v\|^2_{W^{1,2}}$, but the error term in the 2nd order Taylor formula was estimated as $\varepsilon\|v\|^2$ with the C^1 norm rather than the $W^{1,2}$ norm. We need to have a better remainder estimate than (2.5), namely

$$
\Big|I[u+v] - I[u] - DI[u]v - \tfrac{1}{2}D^2I[u](v,v)\Big|/\|v\|^2_{W^{1,2}} \to 0 \text{ as } \|v\| \to 0\,. \tag{2.8}
$$

Note new norm ↗ ↖ still old C^1 norm!

This is much stronger than (2.5) because we divide by a potentially much smaller expression. Luckily the proof follows readily from (2.6): For all $\varepsilon > 0$ there exists $\eta < \delta$ such that

$$
|I[u+v] - I[u] - DI[u]v - \frac{1}{2}D^2I[u](v,v)| \leq \int_0^1 (1-\sigma)(\varepsilon v^2 + 2\varepsilon|vv'| + \varepsilon v'^2)d\sigma \leq \varepsilon\|v\|^2_{W^{1,2}}
$$

if $\|v\| < \eta < \delta$. Now we can indeed use (2.7) with (2.8) instead of (2.5) and argue $I[u_*+v] > I[u_*]$ for v with small C^1 norm and small support. Hence u_* is locally a weak minimum, as claimed. ■

We will subsequently study further conditions (necessary as well as sufficient ones) for locally strong, globally weak, and globally strong minima, and will encounter similar patterns, where a particular type of condition corresponds to a particular type of feature (locally vs globally, weak vs strong), and nonstrict inequality giving a necessary condition, strict inequality giving a sufficient condition.

Before doing this, let's observe how the Legendre condition gives us a glimpse of the role of convexity in minimization. For future reference, let's have a look at a result from multivariable calculus:

Theorem 2.16 *Suppose K is an open convex subset of $\mathbb{R}^n$ and $I \in C^2(K \to \mathbb{R})$ satisfies $D^2I(u) > 0$ (which by definition means that the symmetric matrix $D^2I(u)$ is positive definite), for all $u \in K$. Then I is strictly convex, i.e., for all $x, y \in K$ with $x \neq y$ and all $\lambda \in]0,1[$, it holds*

$$I(\lambda x + (1-\lambda)y) < \lambda I(x) + (1-\lambda)I(y) .$$

Moreover, the gradient $\nabla I = (DI)^T : K \to \mathbb{R}^n$ is one-to-one.

Also $I(y) > I(x) + DI(x) \cdot (y - x)$ for every $y \neq x$.

Proof:
(a) Let's consider the case $n = 1$ first: For $I : \mathbb{R} \to \mathbb{R}$, we can argue that

$$h(\lambda) := \lambda I(x) + (1-\lambda)I(y) - I(\lambda x + (1-\lambda)y)$$

satisfies $h(0) = 0 = h(1)$ and

$$h''(\lambda) = -I''(\lambda x + (1-\lambda)y)\,(x-y)^2 < 0 .$$

Therefore the function h, which does have a minimum on the compact interval $[0,1]$, cannot have its minimum in the interior (as that would require $h'' \geq 0$ at the minimum). So the minimum is 0, taken on only at $\lambda = 0, 1$. We conclude $h(\lambda) > 0$ for $\lambda \in]0,1[$.

(b) The case $n > 1$ is analogous: With the same definition of h we get

$$h''(\lambda) = (x-y)^T D^2I(\lambda x + (1-\lambda)y)\,(x-y) > 0$$

because the Hessian D^2I is positive definite. The conclusion $h(\lambda) > 0$ follows as before. So we have proved the convexity.

(c) We calculate

$$DI(y) - DI(x) = \int_0^1 \frac{d}{d\lambda} DI(x + \lambda(y-x))\,d\lambda = \int_0^1 (y-x)^T D^2I(x + \lambda(y-x))\,d\lambda .$$

Therefore

$$(DI(y) - DI(x))(y-x) = \int_0^1 (y-x)^T D^2I(x + \lambda(y-x))\,(y-x)\,d\lambda > 0$$

provided $y - x \neq 0$. In particular, this implies $DI(y) \neq DI(x)$, hence $\nabla I(y) \neq \nabla I(x)$. This takes care of the injectivity of the gradient.

(d) Finally,

$$\begin{aligned} I(y) - I(x) - DI(x)(y-x) &= \int_0^1 \frac{d}{d\lambda}\Big[I(\lambda y + (1-\lambda)x) - \lambda DI(x)(y-x)\Big]\,d\lambda \\ &= \int_0^1 \Big[DI(\lambda y + (1-\lambda)x) - DI(x)\Big](y-x)\,d\lambda \\ &= \int_0^1 \int_0^1 \frac{d}{dt} DI\Big(t(\lambda y + (1-\lambda)x) + (1-t)x\Big)(y-x)\,dt\,d\lambda \\ &= \int_0^1 \int_0^1 \lambda (y-x)^T D^2 I\Big(\ldots\Big)(y-x)\,dt\,d\lambda > 0\ . \end{aligned}$$

This proves the theorem. ■

It is interesting to give a variant proof for (d) that relies merely on the convexity hypothesis proved in (b), rather than the stronger 2nd derivative hypothesis: From

$$I(x + \lambda(y-x)) = I(\lambda y + (1-\lambda)x) \le \lambda I(y) + (1-\lambda)I(x)$$

we immediately infer, by subtracting $I(x)$, that

$$\frac{I(x+\lambda(y-x)) - I(x)}{\lambda} \le I(y) - I(x)$$

Taking the limit $\lambda \to 0+$ proves a non-strict version of (d), namely $I(y)-I(x) \ge DI(x)(y-x)$. We lost the strict iniequality when taking the limit, but we can recover it: Assume for some y we had equality in our estimate, namely $I(y) - I(x) = DI(x)(y-x)$. Then we look at $I(\frac{y+x}{2})$ and get, from our nonstrict estimate, that

$$I(\tfrac{y+x}{2}) \ge I(x) + DI(x)(\tfrac{y+x}{2} - x) = I(x) + DI(x)\tfrac{y-x}{2}\ .$$

On the other hand, we have $I(\frac{y+x}{2}) < \frac{1}{2}(I(y)+I(x))$ from strict convexity (when $y \neq x$). Then, with the assumed equality, we infer

$$I(\tfrac{y+x}{2}) < \tfrac{1}{2}(I(y)+I(x)) = \tfrac{1}{2}\Big(I(x) + DI(x)(y-x) + I(x)\Big) = I(x) + DI(x)\tfrac{y-x}{2},$$

a contradiction.

In the situation where $DI(x) = 0$, convexity implies that x is an absolute minimum.

2.17 (Remarks on Convexity) The strict Legendre condition states that the Lagrangian L, considered as a function of the derivative u' only, is strictly convex, in a neighborhood of any point $(t, u_*(t), u_*'(t))$. It is sufficient for local weak minimality of a solution to the EL equations; its nonstrict version is necessary for local weak minimality. It will transpire later, that strict convexity of L in the derivative u' alone, but for all u' not just those in a neighborhood of u_*', is sufficient for locally strong minimality. However the theorem to that effect will be worded differently, giving a weaker condition that is still sufficient.

Many interesting variational problems satisfy the Legendre condition, and even feature convexity of L in the derivative variable. Notable exceptions are problems arising from nonlinear elasticity, where such convexity is usually not satisfied.

In contrast, convexity of the functional I is a much less frequent situation to be encountered. It would follow if the Lagrangian L were convex as a function of both variables (u, u'): Indeed, if it were true that

$$L(t, \lambda u_1 + (1-\lambda)u_2, \lambda u_1' + (1-\lambda)u_2') < \lambda L(t, u_1, u_1') + (1-\lambda)L(t, u_2, u_2')$$

it would immediately follow by integration that $I[\lambda u_1 + (1-\lambda)u_2] < \lambda L[u_1] + (1-\lambda)I[u_2]$. If this convexity hypothesis is verified, then a solution u_* where the directional derivatives $DI[u_*]v$ all vanish will automatically be an absolute minimum, by the same reasoning as in the finite dimensional case.

Unfortunately, many problems of interest do not have a convex Lagrangian (i.e., not convex in both variables). For instance, the two-variable function $L(y, y') = \sqrt{(1+y'^2)/y}$ from the brachystochrone problem has a Hessian $\begin{bmatrix} L_{yy} & L_{yy'} \\ L_{yy'} & L_{y'y'} \end{bmatrix}$ whose determinant can be calculated, with labor, to be $(3-y'^2)/4(1+y'^2)y^3$. It is negative when $y'^2 > 3$, so the Hessian cannot be positive definite in this part of the domain.

On the flipside, it is sometimes possible, by a change of variables, to get a different, but equivalent, Lagrange function, for which convexity does hold. This phenomenon can be observed in the single variable case: $I(x) = \sqrt[4]{1+x^2}$ is not convex, but $I^2 = \sqrt{1+x^2}$ is. Also, substituting $x = y^3$, we get $I(x) = \tilde{I}(y) = \sqrt[4]{1+y^6}$, and $\tilde{I}$ is convex. Of course the minimization problems for I, I^2, and $\tilde{I}$ are trivially equivalent.

We will see that such a convexity-inducing coordinate transformation is possible in the brachystochrone problem.

Chapter 3

Classical Examples

Example 3.1 (Brachystochrone) We have already found the EL equation $yy''/(1+y'^2) = -\frac{1}{2}$ and its solution in parametric form

$$\begin{aligned} x &= R(\varphi - \sin\varphi) \\ y &= R(1 - \cos\varphi) \end{aligned}$$

which already incorporates the first boundary condition $y(0) = 0$. Using the energy method 2.12, we could have gotten this result with less work. We still have to solve the boundary value problem: given any pair $x_1 > 0, y_1 \geq 0$, we claim there is exactly one choice of R such that $y(x_1) = y_1$. Indeed, the expression $\frac{y}{x} = \frac{1-\cos\varphi}{\varphi-\sin\varphi}$ is increasing as a function of φ, since

$$\frac{d}{d\varphi}\frac{y}{x} = \frac{\sin\varphi\,(\varphi-\sin\varphi) - (1-\cos\varphi)^2}{(\varphi-\sin\varphi)^2} = \frac{4\sin(\frac{\varphi}{2})}{(\varphi-\sin\varphi)^2}\,(\tfrac{\varphi}{2}\cos\tfrac{\varphi}{2} - \sin\tfrac{\varphi}{2}) < 0$$

since $\frac{\varphi}{2} < \tan\frac{\varphi}{2}$. As $\frac{y}{x} \to \infty$ as $\varphi \to 0+$ and $\frac{y}{x} \to 0$ as $\varphi \to 2\pi$, there is for every slope $\frac{y_1}{x_1} \in [0, \infty[$ exactly one $\varphi_1 \in]0, 2\pi]$ corresponding to the end point, and then the parameter R can be determined from the cycloid equation. The BVP for the brachystochrone problem has exactly one solution.

Now we calculate $L_{y'y'} = y^{-1/2}(1+y'^2)^{-3/2} > 0$. So the strict version of Legendre's condition is satisfied (even without restricting the arguments to such (x, y, y') as arise from solutions to the EL equation). We also conclude from this that $L_{y'}$ is strictly increasing as a function of y', and by Erdmann's corner condition, this rules out corners.

From the Legendre condition, we can conclude that short segments of the cycloid are weak minimizers, but this is a far shot from the absolute minimality of the unique cycloid connecting the given points that we want to conclude.

In line with Remark 2.17, there is an ad-hoc convexity argument available here, which helps us answer the problem without developing a full theory of sufficient conditions. We use the substitution $v := \sqrt{2y}$, which is a sort of coordinate transformation in function space. This transforms

$$I[y] = \int_0^{x_1} \sqrt{\frac{1+y'^2}{y}}\,dx = \int_0^{x_1} \sqrt{2}\,\sqrt{\frac{1+v^2v'^2}{v^2}} =: \sqrt{2}\,\tilde{I}[v]$$

Now $\tilde{L}(v, v') := \sqrt{v'^2 + \frac{1}{v^2}}$ is strictly convex on $\mathbb{R} \times \mathbb{R}^+$ because the Hessian

$$\begin{bmatrix} \tilde{L}_{vv} & \tilde{L}_{vv'} \\ \tilde{L}_{vv'} & \tilde{L}_{v'v'} \end{bmatrix} = (v'^2 + v^{-2})^{-3/2} \begin{bmatrix} 2v^{-6} + 3v'^2v^{-4} & v'v^{-3} \\ v'v^{-3} & v^{-2} \end{bmatrix}$$

is positive definite. Therefore $v \mapsto \tilde{I}[v]$ is stricly convex, and this implies that a v_* where the directional derivatives vanish is an absolute minimizer.

The book by Troutman "Variational Calculus with Elementary Convexity" has more material on this method.

Example 3.2 (The Catenoid) *See Homework Problems*

Example 3.3 (The Catenary) Given $(x_1, y_1) \in \mathbb{R}^2$ and $(x_2, y_2) \in \mathbb{R}^2$ as well as $\ell > \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$, we ask for

$$\min\Big\{ I[y] \;\Big|\; y \in C^1[x_0, x_1]\,,\; y(x_1) = y_1\,,\; y(x_2) = y_2\,,\; K[y] = \ell \Big\}$$

where

$$I[y] = \int_{x_1}^{x_2} y\sqrt{1 + y'^2}\, dx \quad \text{and} \quad K[y] = \int_{x_1}^{x_2} \sqrt{1 + y'^2}\, dx$$

If we trust that the Lagrange multiplier method from MV calculus carries over (and we will see shortly why it does), we need to find the EL equation corresponding to $I - \lambda K$ and determine the Lagrange multiplier λ using the constraint $K[y] = \ell$. With $L(x, y, y') := (y - \lambda)\sqrt{1 + y'^2}$, the energy form of the EL equation is $L - y' L_{y'} \equiv \mathit{const}$, i.e.,

$$\frac{y - \lambda}{\sqrt{1 + y'^2}} \equiv E$$

The case $E = 0$ can be discarded, because $y \equiv \lambda$ does not solve the EL equation, but is a spurious constant solution. The relevant solutions are (as can be seen by separating variables)

$$y = \lambda + E \cosh \frac{x - x_0}{E}$$

These represent hanging chains in the case $E > 0$ (we expect these to be minimal), and (standing) arches in the case $E < 0$. The latter we expect to be maximal. The two boundary conditions $y(x_1) = y_1$, $y(x_2) = y_2$, and the cnstraint $K[y] = \ell$ can be used to determine E, λ, x_0, giving rise to a unique chain and a unique arch for each set of boundary conditions and length $\ell > \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$.

It remains for later to actually prove minimality or maximality in this case. Sec. 3.6 of Troutman's book gives a proof of minimality via a convexity generating coordinate transformation.

Let's quickly do the calculus exercise to show that the data in the catenary problem determine a unique chain and a unique arch: It is straightforward to calculuate the length

$$\ell = \int_{x_1}^{x_2} \sqrt{1 + \left(\frac{d}{dx}\Big(\lambda + E\cosh\frac{x - x_0}{E}\Big)\right)^2}\, dx = \ldots = E\left(\sinh\frac{x_2 - x_0}{E} - \sinh\frac{x_1 - x_0}{E}\right)\,.$$

We have to solve, for given $x_{1,2}$, $y_{1,2}$ and ℓ, the equations

$$\begin{aligned} y_1 &= \lambda + E\cosh\tfrac{x_1 - x_0}{E} \\ y_2 &= \lambda + E\cosh\tfrac{x_2 - x_0}{E} \\ \ell &= E\left(\sinh\tfrac{x_2 - x_0}{E} - \sinh\tfrac{x_1 - x_0}{E}\right) \end{aligned}$$

for λ, E, x_0. Equivalently, introducing $\xi_1 = \frac{x_1 - x_0}{E}$, $\xi_2 = \frac{x_2 - x_0}{E}$, $\Delta y = y_2 - y_1$ and $\Delta x = x_2 - x_1$, we have to solve, for given Δx, Δy and ℓ, the system of equations

$$
\begin{aligned}
E(\xi_2 - \xi_1) &= \Delta x \\
E(\cosh \xi_2 - \cosh \xi_1) &= \Delta y \\
E(\sinh \xi_2 - \sinh \xi_1) &= \ell
\end{aligned}
$$

for ξ_1, ξ_2, E. These equations imply that

$$
\frac{\ell^2 - \Delta y^2}{\Delta x^2} = \frac{2(\cosh \xi_2 \cosh \xi_1 - \sinh \xi_2 \sinh \xi_1) - 2}{(\xi_2 - \xi)^2} = \frac{\sinh^2 \frac{\xi_2 - \xi_1}{2}}{(\frac{\xi_2 - \xi_1}{2})^2}
$$

The function $t \mapsto \frac{\sinh t}{t}$ is even, and is (by monotonicity) bijective from $]0, \infty[$ to $]1, \infty[$. Our hypothesis $\ell^2 > \Delta x^2 + \Delta y^2$ on the data guarantees that there is exacly one $t_* > 0$ such that $\frac{\sinh t}{t} = \sqrt{\ell^2 - \Delta y^2}/\Delta x$, and the equations imply that $\Delta \xi := \xi_2 - \xi_1 = \pm 2 t_*$. Then $E = \pm \Delta x/(2t_*)$ from the first equation. We now determine $\xi_{1,2}$ separately from the 2nd equation, which becomes $\cosh(\xi_1 + \Delta \xi) - \cosh \xi_1 = \Delta y / E$, or equivalently,

$$
\cosh \xi_1 \frac{\cosh \Delta \xi - 1}{\sinh \Delta \xi} + \sinh \xi_1 = \frac{\Delta y}{E \sinh \Delta \xi} \; .
$$

The fraction next to $\cosh \xi_1$ equals $\tanh \frac{\Delta \xi}{2}$ and has absolute value < 1. This makes the left hand side a strictly increasing function of ξ_1, mapping $\mathbb{R} \to \mathbb{R}$ bijectively. The equation therefore has exactly one solution ξ_1 for each of the two pairs $(\Delta \xi, E)$ previously retrieved. It is now routine to show that the values thus found indeed satisfy the three equations.

3.4 (Lagrange multipliers in Calculus of Variations) As with the EL equation, where we did not really use the notation of a (total) derivative $DI[u_*]$ in function spaces, but instead reduced the problem to single variable derivatives $\frac{d}{d\varepsilon} I[u_* + \varepsilon v]$, we are using a finite dimensional version of the Lagrange multiplier theorem, rather than attempting to generalize the theorem to function spaces.

The $\mathbb{R}^n$ version allowed for the exceptional case when the gradient of the constraint function K vanished (or more generally the rank of $DK[u_*]$ wasn't maximal). Such points would occur if we were to attempt an unconstrained minimax problem on the the constraint K; the Lagrange multiplier method doesn't apply in these points.

In the catenary example, K is the length functional, and there always exists a variation v such that $\partial_v K[u_*] \neq 0$, *except* when u_* is the shortest connection (straight line). The exceptional case will be ruled out when we assume $K[u] = \ell$ with ℓ larger than the straight distance between the endpoints.

Returning tho the general case of a functional I to minimize subject to a constraint $K[u] = \ell$, we assume u_* is a constrained minimal and fix a v such that $\partial_v K[u_*] \neq 0$ (assuming such a v exists; if not, u_* satisfies the EL equation for K). We now consider 2-parameter variations of u_* for I, i.e., the function

$$
(a, b) \mapsto I[u_* + av + bw] =: f(a, b)
$$

with w an arbitrary ariation in C_0^1 and v the one fixed previously. Of this 2-parameter family, there is a one-parameter family of functions that satisfies the constraint $K[u] = \ell$. Letting

$g(a,b) := K[u_* + av + bw] - \ell$, we conclude that $(0,0)$ minimizes f subject to the constraint $g(a,b) = 0$, where

$$\nabla g(0,0) = \begin{bmatrix} \partial_a g(0,0) \\ \partial_b g(0,0) \end{bmatrix} = \begin{bmatrix} \partial_v K[u_*] \\ \partial_w K[u_*] \end{bmatrix} = \begin{bmatrix} \neq 0 \\ \text{whatever} \end{bmatrix} \neq \vec{0}\,,$$

Therefore there exists a Lagrange multiplier λ (as of yet it might depend on w) such that $\nabla(f - \lambda g)(0,0) = \vec{0}$, i.e.,

$$\begin{aligned} \partial_v I[u_*] - \lambda \partial_v K[u_*] = 0 \\ \partial_w I[u_*] - \lambda \partial_w K[u_*] = 0 \end{aligned}$$

The first equation tells us that $\lambda = \partial_v I[u_*]/\partial_v K[u_*]$. While this is not helpful to find λ practically until we have actually found u_* (and then we don't care to find λ any more), it does guarantee that λ does not depend on w.

As the second equation is required for every w, with the same λ, it is tantamount to u_* satisfying the EL equation for $I - \lambda K$.

Technique in general: If we have k constraints $K_1[u] = 0, \ldots, K_k[u] = 0$, we take k linearly independent variations $v_1, \ldots v_k$ such that the matrix $((\partial_{v_i} K_j[u_*]))_{i,j}$ is invertible at a presumed constrained minimizer u_*; if such v_i cannot be found, we are in the exceptional case (on which we elaborate below). We take an arbitrary variation w and use the Lagrange multiplier theorem on the function $f(a_1, \ldots, a_k, b) := I[u_* + a_1 v_1 + \ldots + a_k v_k + bw]$, which is minimal at $(\vec{0}, 0)$ under the constraints $g_i(a_1, \ldots, a_k, b) := K_i[u_* + a_1 v_1 + \ldots + a_k v_k + bw] = 0$. By hypothesis, the matrix $((\partial g_i/\partial a_j))_{i,j=1\ldots n}$ is invertible. So there is a 1-parameter family of parameters $a_1(b), \ldots, a_k(b), b)$ on which the g_i vanish. There exist Lagrange multipliers λ_i such that $\nabla(f(\vec{a}, b) - \sum \lambda_i g_i(\vec{a}, b))$ vanishes at $(\vec{a}, b) = (\vec{0}, 0)$. As of yet the Lagrange multipliers might depend on the w, but we'll see immediately that they don't. Namely the a-derivatives of $f - \sum \lambda_i g_i$, namely $\partial_{a_j} f(\vec{0}, 0) - \sum_i \lambda_i \partial_{a_j} g_i(\vec{0}, 0) = 0$, allow us to calculate λ_i in principle independently of w, since the matrix $D_a g$ is invertible. The b-derivative, namely $\partial_b f(\vec{0}, 0) - \sum_i \lambda_i \partial_b g_i(\vec{0}, 0) = 0$ is tantamount to $\partial_w (I - \sum_i \lambda_i K_i)[u_*] = 0$, for every variation w and with the λ_i independent of w. So The EL equations for $I - \sum_i \lambda_i K_i$ must hold.

Now we show that the exceptional case means that there exist λ_i (not all 0) such that the EL equations for $\sum_i \lambda_i K_i$ are satisfied. This is an induction over the number of constraints. If K_1 does not allow for any v such that $\partial_v K_1[u_*] \neq 0$, then the EL equation for $K_1 = 1 \cdot K_1 + 0 \cdot K_2 + \ldots 0 \cdot K_k$ are satisfied. Otherwise choose v_1 such that $\partial_{v_1} K_1[u_*] \neq 0$. If there does not exist a w such that the matrix $\begin{bmatrix} \partial_{v_1} K_1[u_*] & \partial_w K_1[u_*] \\ \partial_{v_1} K_2[u_*] & \partial_w K_2[u_*] \end{bmatrix}$ is invertible, then the 2nd row must be a multiple of the first row; i.e., for every w there must exist λ such that $\partial_{v_1} K_2[u_*] = \lambda \partial_{v_1} K_1[u_*]$ and $\partial_w K_2[u_*] = \lambda \partial_w K_1[u_*]$. The first equation shows that λ is independent of w. The second equation, being valid vor every w, shows that the EL equation for $K_2 - \lambda K_1$ must be satisfied. On the other hand if there does exist such a w that makes the matrix invertible, choose one and call it v_2. Continue inductively. ■

We briefly referred to the next example in the context of the energy theorem. Let us now study it in more detail

Example 3.5 (Newtonian Mechanics) Let $\vec{x}_i$ be the position of particle number i, with mass m_i, under the influence of a force $\vec{F}_i = -\nabla V(\vec{x_1}, \ldots, \vec{x}_n)$. Newton's equations $m_i \vec{x}_i'' = \vec{F}_i$

are the Euler-Lagrange equations of the variational problem

$$I[\vec{x}_1,\dots,\vec{x}_n] := \int_{t_0}^{t_1} \left(\sum_i \frac{1}{2} m_i |\vec{x}_i|^2 - V(\vec{x}_1,\dots,\vec{x}_n) \right) dt \; .$$

The minimizing property is not actually the issue here, simply the vanishing of the derivative. However, checking the Legendre condition, we see that $L_{x'x'}$ is the diagonal matrix $\mathrm{diag}[m_1, m_1, m_1, m_2, m_2, m_3, \dots, m_n, m_n, m_n]$, which is clearly positive definite, since the masses (and hence the kinetic energies of each particle) are positive. So the extremals (i.e., solutions of the EL equation) of the variational principle are indeed locally weak minimizers.

There are advantages to writing the Newtonian equations in variational form:

(1) Suppose you transform the equations into curved coordiantes (e.g, spherical coordinates). It is easier to transform the Lagrange function, since only first derivatives occur, whereas for the EL equations, second derivatives would need to be transformed.

(2) Even if the variables $\vec{x}_i$ are constrained to some surface, the same variational principle prevails, and it is possible to work with intrinsic coordinates on this surface. For instance, you might think of two particles attached to a rod that restricts their movement to a sphere (constant distance from the point wher the rod is hinged). Instead of having six cartesian coordinates (x_1, x_2, x_3) and (x_4, x_5, x_6) to describe the first and second particle respectively, but subject to constraints that $x_1^2 + x_2^2 + x_3^2 = \ell^2$ (with ℓ the length of the first rod), and a similar constraint for the second particle, we only need four coordinates, say (y_1, y_2) and (y_3, y_4) for the georaphical latitude and longitude of the first and second particle respectively. Allowing even for the sphere to be moved around by some external force, we have given functions $x_i = g_i(y_1, \dots, y_j, t)$, where the t dependence allows for the moving around of the sphere.

In such a system we may encounter two kinds of forces: explicitly known and given forces F_i, and unknown forces Z_i that are supplied by the rod in response to the movement of the particle. These forces are nota-priori known, but rather they are determined by achieving the desired effect (namely to keep the particle at the fixed distance from the hinge point, i.e., on the sphere).

We might assume that the internal forces Z_i are orthogonal to the surface, i.e., their dot product with each tangent vector to the surface vanishes. This means (eg., for the first particle) that $\sum_{i=1}^{3} Z_i \frac{\partial g_i}{\partial y_j} = 0$ for each j. Summing over all n particles, this implies

$$\sum_{i=1}^{3n} Z_i \frac{\partial g_i}{\partial y_j} = 0$$

This equation can physically be interpreted as the forces Z_i not doing any work on the system, and it is weaker than the original hypothesis that the forces are orthogonal to the surface. For instance it allows the motion of the particles being confined by connecting rods.

In such a situation, we'd want to eliminate the functions g_i and the internal forces Z_i from Newton's equation of motion $m_i x_i'' = F_i + Z_i$. (Here I have made a slight change in indexing, so that $m_1 = m_2 = m_3$, rather than only m_1, would be the mass of the first particle, etc.) Dotting Newton's equation with $\frac{\partial g}{\partial y_j}$, the Z_i drop out and we conclude

$$\sum_i m_i x_i'' \frac{\partial g_i}{\partial y_j} = \sum_i F_i \frac{\partial g_i}{\partial y_j}$$

Since $x_i = g_i(y,t)$, we obtain

$$x_i' = \sum_j \frac{\partial g_i}{\partial y_j} y_j' + \frac{\partial g_i}{\partial t} =: h_i(y, y', t)$$

and therefore $\frac{\partial g_i}{\partial y_j} = \frac{\partial h_i}{\partial y_j'}$. We can also write the terms on the left side of Newton's equation as

$$m_i x_i'' \frac{\partial g_i}{\partial y_j} = \frac{d}{dt}\left(m_i x_i' \frac{\partial g_i}{\partial y_j}\right) - m_i x_i' \frac{d}{dt}\frac{\partial g_i}{\partial y_j} = \frac{d}{dt}\left(m_i x_i' \frac{\partial h_i}{\partial y_j'}\right) - m_i h_i \frac{\partial h_i}{\partial y_j} = \left(\frac{d}{dt}\frac{\partial}{\partial y_j'} - \frac{\partial}{\partial y_j}\right)\frac{1}{2} m_i x_i'^2$$

where, in the last step we have used $\frac{d}{dt}\frac{\partial g_i}{\partial y_j} = \frac{\partial h_i}{\partial y_j}$, because both sides are equal to $\sum_l \frac{\partial^2 g_i}{\partial y_j \partial y_l} y_l' + \frac{\partial^2 g_i}{\partial y_j \partial t}$. Now Newton's equations are in the form

$$\sum_i \left(\frac{d}{dt}\frac{\partial}{\partial y_j'} - \frac{\partial}{\partial y_j}\right)\frac{1}{2} m_i h_i^2 = -\sum_i \frac{\partial}{\partial x_i} V(g(y,t)) \frac{\partial g_i}{\partial y_j} = -\frac{\partial}{\partial y_j}\tilde{V}(y,t)$$

where we used the MV chain rule and $\tilde{V}$ is the potential V expressed in terms of y and t, rather than x. Since the potential doesn't depend on the velocities y', it is gratuitous to write the right hand side as $\left(\frac{d}{dt}\frac{\partial}{\partial y_j'} - \frac{\partial}{\partial y_j}\right)\tilde{V}(y,t)$. In other words, they are the EL equations for the Lagrangian (kinetic energy − potential energy), expressed merely in intrinsic coordinates y.)

In this calculation, we have assumed (as is verified in mechanics) that the forces are the gradient of a potential, and also are independent of the velocities. However, it turns out that the Lorentz forces (which are velocity dependent forces on charged particles in a magnetic field) van also be written in the form $\left(\frac{d}{dt}\frac{\partial}{\partial y_j'} - \frac{\partial}{\partial y_j}\right)\tilde{V}(y,t)$ for a velocity dependent expression $\tilde{V}$. This may be understood as an indication that the variational principle should be viewed as more fundamental than is indicated merely by the calculation given here for Newtonian mechanics.

The next example doesn't fit quite in the theory developed so far, because it involves a multi-variable integral, but it is of quite fundamental importance, and many ideas carry over:

Example 3.6 (Dirichlet's principle) Suppose $\Omega \subset \mathbb{R}^n$ is a bounded domain with 'decent' boundary $\partial\Omega$. For instance, Lipschitz boundary[1] is sufficient, so polyhedra and smoothly bounded domains are included.

We now ask the question whether we can minimize the expression

$$I[u] := \int_\Omega |\nabla u|^2 \, dx$$

among all functions $u \in W^{1,2}(\Omega)$ satisfying $u = f$ on the boundary $\partial\Omega$ where f is a given fucntion on the boundary. The space $W^{1,2}(\Omega)$ consists of all those functions that have a first derivative $Du = (\nabla u)^T$ in a certain sense that generalizes the classical definition of a

[1]Lipschitz boundary means that every boundary point has a neighborhood and a cartesian coordinate system $(x_1, \ldots, x_n)$ such that within that neighborhood, the boundary can be described as the graph of a function h, namely $x_n = h(x_1, \ldots, x_{n-1})$, with $x_n > h(x_1, \ldots, x_{n-1})$ characterising the interior of Ω and $x_n < h(x_1, \ldots, x_{n-1})$ the exterior. The function h needs to be a Lipschitz function, i.e., it must satisfy $|h(x') - h(x)| \le L|x' - x|$ for some constant L. The hypothesis $h \in C^1$ is sufficient for Lipschitz.

derivative, and such that the sqaure $|\nabla u|^2$ has still a finite integral over Ω; the notion of integral is also appropriately generalized beyond Riemann's definition, namely we are using Lebesgue's definition of the integral. [You may ignore these technicalities if you are not familiar with them already.]

In this situation, we can argue similarly as for single variable integrals: **If** a minimum exists, then we can argue

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} I[u+\varepsilon v] = 0 \qquad \text{for every } v \in W_0^{1,2}(\Omega)$$

where the index 0 in $W_0^{1,2}$ indicates boundary data 0. Now

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int_\Omega |\nabla(u+\varepsilon v)|^2\,dx = 2\int_\Omega \nabla u\cdot\nabla v\,dx = 0$$

and **if** the minimizer u is actually a C^2 function up to the boundary (slightly weaker hypotheses may suffice, but we must assume the existence of 2nd derivatives in some sense), then we can integrate by parts (i.e., use Gauss' divergence theorem) to continue

$$0 = \int_\Omega \big(\operatorname{div}(v\nabla u) - v\Delta u\big)\,dx = \int_{\partial\Omega} v\partial_\nu u\,dS(x) - \int_\Omega v\Delta u\,dx$$

Since v vanishes on the boundary, the boundary integral drops out; the fundamental lemma of Calculus of variations generalizes to the multi-variable integral setting, and we conclude that u must satisfy $\Delta u = 0$, the EL equation for this variational problem.

Historically, the problem to solve $\Delta u = 0$ in Ω under the boundary condition $u = f$ on $\partial\Omega$ arose in electrostatics. An existence 'proof' that this equation does have a solution would be by Dirichlet's principle: Find the minimum of the electric field energy $\int_\Omega |\nabla u|^2\,dx$ among 'all' functions that satisfy the prescribed boundary data f. The minimizer is the desired solution to the equation $\Delta u = 0$.

Weierstrass criticized that the existence of a minimizer in variational problems is not trivial, but needs proof. So using Dirichlet's principle to 'prove' the existence of a solution to the equation $\Delta u = 0$ was not a valid argument.

Tools for proving the existence of a minimizer became available only in the 20th century, and this salvaged Dirichlet's principle from disreputable oblivion. Such an existence proof does not work in a space like $C^2(\bar\Omega)$, but must of necessity look for u in a much large space $W^{1,2}$ of functions that need to have only one derivative, and even this in a generalized sense, so that even the continuity of $u \in W^{1,2}$ is not ascertained everywhere. For instance a function like $\ln\ln\frac{2}{|x|}$ is in $W^{1,2}$ on the unit disk in $\mathbb{R}^2$, even though it goes to infinity as $|x| \to 0$.

The proof that the minimizer is indeed in $C^2(\Omega)$ and maybe $C^0(\bar\Omega)$ is now another trouble spot. The simple DuBois Raymond technique does not generalize; another piece of subtle analysis is needed.

But with this modern machinery (existence and smoothness) in place, Calculus of Variations has become ('again') a powerful tool that certain partial differential equations indeed have a solution. A simpler version of this theory can be used that boundary value problems for certain ODEs have solutions.

In the next chapter, we will see glimpses and key ingredients of such existence proofs. Even as full technical details are beyond this course, the basic ideas can be appreciated.

3.7 (Illustrated Facts on Dirichlet's Principle (without proofs)) (1) If $\Omega \subset \mathbb{R}^n$ is a bounded Lipschitz domain and $f \in C^0(\partial\Omega)$, then there exists exactly one solution to the boundary value problem $\Delta u = 0$ in Ω and $u = f$ on $\partial\Omega$. This fact is proved by methods of Partial Differential Equations and Multi-variable Calculus, without reference to Calculus of Variations. If Ω happens to be a ball, the solution u can be given by an explicit integral formula in terms of f.

(2) In the case of the unit disk in $\mathbb{R}^2$, we can specify an example $f(\varphi) = \sum_{n=1}^{\infty} \frac{1}{n^2} \cos(2^n \varphi)$ of a continuous function $f \in C^0(\partial\Omega)$ such there is no function u interpolating f into the disc that would have finite $\int_\Omega |\nabla u|^2\, dx$. So the solution u whose existence and uniqueness is asserted could not be found by means of the Dirichlet variation principle. This solution can even be calculated explicitly: $u(r,\varphi) = \sum_{n=1}^{\infty} \frac{1}{n^2} r^{(2^n)} \cos(2^n\varphi)$.

(3) The following function u satisfies $\Delta u = 0$ in the unit disk with boundary data f that are not even continuous; this solution can be obtained as a minimizer of the Dirichlet integral $I[u]$. In complex notation $z = x + iy$, the function is $u(z) = \operatorname{Re} \ln \ln \frac{3}{z+1}$, which in real coordinates is

$$u(x,y) = \frac{1}{2} \ln \left[\left(\ln \frac{3}{\sqrt{(x+1)^2 + y^2}} \right)^2 + \left(\arctan \frac{3y}{(x+1)^2 + y^2} \right)^2 \right] .$$

The boundary data are

$$f(\varphi) = u(\cos\varphi, \sin\varphi) = \frac{1}{2} \ln \left[\left(\ln \frac{3}{2|\cos \frac{\varphi}{2}|} \right)^2 + \left(\arctan \frac{3}{2} \tan \frac{\varphi}{2} \right)^2 \right]$$

These examples show that $f \in C^0(\partial\Omega)$ is neither necessary nor sufficient for CV to be applicable to the problem. The methods referred to as PDE methods above desire $f \in C^0$ as a natural hypothesis. CV methods want f to be boundary values of $W^{1,2}$ functions. Lipschitz functions f are special cases in the overlap of the two methods.

(4) It is true that for every f defined on $\partial\Omega$ that can be interpolated by a function u with finite $I[u] := \int_\Omega |\nabla u|^2\, dx$ the variational problem to minimize $I[u]$ does have a solution (this is routine provided some high tech theory is available), and that such a minimum is unique (that's just convexity of I). And it is also true that such a minimizer u is automatically C^∞ inside Ω. Of course u cannot be C^2 up to the boundary unless f itself is C^2 and the boundary itself needs to be suffiicently smooth.

To *prove* that u is smooth inside Ω is subtle and belongs to an area called 'elliptic regularity theory' that covers more generality than just this example. The principal idea is that one goes a long way without any integration by parts. The EL equations stays in the integrated form $\int \nabla u \cdot \nabla v\, dx = 0$ for 'every' direction v, and we do not try to eliminate v. The wisdom is now to concoct special choices for v, in terms of the yet unknown function u, and to use these v in the EL equation to draw conclusions about u. Appropriately generalized notions of derivative ('weak derivative') and integral (Lebesgue's integral rather than Riemann's integral) are key ingredients necessary to carry this idea out in practice.

Chapter 4

Key Ideas of Direct Methods

In this chapter, we will study some ideas of proving the existence of absolute minima in calculus of variations, prior to obtaining a differential equation that a possible minimizer would have to satisfy. These methods are called 'direct methods' and can be used to prove that certain differential equations do have a solution.

As has been our custom, we'll first look how 'direct methods' would look in the case of MV Calculus. Then we will have a look at the crucial and significant changes that are needed to carry these methods over to CV.

Theorem 4.1 *Let $K \subset \mathbb{R}^n$ be compact (i.e., closed and bounded), $K \neq \emptyset$, and let $I : K \to \mathbb{R}$ be continuous. Then there exists $u_* \in K$ such that $I[u_*] \leq I[u]$ for all $u \in K$. Likewise there exists $u^* \in K$ such that $I[u^*] \geq I[u]$ for all $u \in K$.*

PROOF: Let $a := \inf_K I$ where $a \in [-\infty, \infty[$. This means a is the largest possible number such that $a \leq I[u]$ for all $u \in K$. If no such real number a exists, then we define a to be $-\infty$. Advanced calculus builds on the fact that such an a always exists, and this is a property of the real number system and has nothing to do with I.

There exists a sequence (u_n) such that $I[u_n] \to a$ as $n \to \infty$. In a compact set, every sequence (u_n) has a convergent subsequence. Let's call this subsequence (u_{n_j}), and its limit u_*. Since I is continuous and $u_{n_j} \to u_*$ as $j \to \infty$, we conclude $I[u_{n_j}] \to I[u_*]$. But by construction, $I[u_{n_j}] \to a$. So $I[u_*] = a$, and in particular $a \neq -\infty$. So we have found a minimizer u_*.

The proof that there is a maximizer u^* is analogous, with $b \in]-\infty, +\infty]$ being the supremum of I. ■

This method is called 'direct method' because we prove the existence of a minimizer directly, without first calculating promising candidates for a minimizer and then subsequently verifying that they are indeed minimizers.

We usually use Thm. 4.1 in a context in which I is defined on a set that is not even compact, by cutting off 'hopeless' candidates for minimizers first:

Example 4.2 *Show that $I[u] := u^4 + u + \sin u$ where $u \in \mathbb{R}$ has an absolute minimum.*
Solution: *We cannot solve the equation $I'[u] = 0$ by practical algebra. So we use the direct method. Since $I[0] = 0$, any u for which $I[u] > 0$ may be disregarded beforehand. Clearly*

$I[u] > 0$ if $u > 0$. Moreover, for $u \le 1$, we have $I[u] \ge u^4 + u - 1 \ge u^4 + 2u \ge u^4 - 2u^2 = (u^2-1)^2 - 1$. So if $u < -\sqrt{2}$, we conclude $I[u] > 0$.

By Thm. 4.1, I restricted to $[-\sqrt{2}, 0]$ has an absolute minimum at some u_*, and its value is ≤ 0 because $I[0] = 0$. And since $I[u] > 0$ for u outside $[-\sqrt{2}, 0]$, the same u_* is the absolute minimum of $I[u]$ for all $u \in \mathbb{R}$.

Now we could look for solutions to $I'[u] = 0$ in the interval $[-\sqrt{2}, 0]$ numerically, and we know *that the absolute minimum can be found among these solutions. A 2nd derivative test is not needed.*

4.3 *(Problems in carrying over Thm. 4.1 to Calculus of Variations)* The theorem "A continuous function on a compact set takes on a minimum and a maximum" is true in generality, if the notion compact is properly defined, eg. as the possibility to extract a convergent subsequence (not equivalent to 'closed and bounded'). But this principle ceases to be useful, and there are various ways to see this:

(a) The same principle, when it is applicable, establishes the existence of both a minimum and a maximum. However, in most CV problems, we know beforehand that a maximum does not exist. (No longest curve from A to B; no slowest track in the brachystochrone problem.) So clearly the Thm. 4.1 cannot apply in these situations. The way out of this dilemma is to split the notion of continuity in halves, one of which is good to prove the existence of minima, the other to prove the existence of maxima.

(b) In vector spaces of functions, the notion 'bounded and closed' is not equivalent to the notion of 'compact'. For instance, the sequence u_n given by the formula $u_n(x) = \sin nx$ is a bounded sequence in the space $C^0[0, 2\pi]$; it lies in the closed ball $\|u\| \le 1$. But this sequence does not have a convergent subsequence. Cutting off some obviously disqualified candidates u, like letting $K := \{u \mid I[u] \le c\}$ for some c does *not* result in a compact set K. The way out of this dilemma is to relax the compactness condition. But this means, we cannot argue any more 'Every sequence has a convergent subsequence'.

(c) There are actually different notions of convergence in vector spaces of functions. For instance $\frac{1}{n} \sin nx \to 0$, if we define distance in terms of the C^0 norm, but $\frac{1}{n} \sin nx \not\to 0$ if we define distance in terms of the C^1 norm. This subtlety works to our benefit. By selecting a sufficiently weak ('easy to achieve') notion of convergence we may salvage the argument of finding a convergent subsequence: we may find a 'weakly convergent subsequence' (yet to be defined precisely).

(d) Selecting an 'easy-to-achieve' notion of convergence however comes at a price. If we make it easier for u_n to converge to u, we get a 'harder-to-achieve' notion of continuity, because we need to prove $I[u_n] \to I[u]$ in more cases. As it turns out, none of the expressions $I[u]$ that we considered will remain continuous as defined by the property $I[u_n] \to I[u]$ whenever $u_n \rightharpoonup u$, where $\rightharpoonup$ refers to our weak notion of convergence. What we can salvage, and this brings us back to part (a), is the useful 'half' of continuity that will still give us a minimum, but not a maximum along the lines of Thm. 4.1.

This originally quite ingenious modification of the ideas of Thm 4.1 has now become a routine method in CV. We will first implement it in a simple example, where its main ideas can be studied without the advanced formalism of integrals and derivatives. The functional I in which we will study the method will therefore not be defined in terms of an integral at all.

The geodesic problem (a low-tech implementation)

4.4 (Def. of geodesic problem) Consider a surface that is given by a C^1-function $f : \mathbb{R}^2 \to \mathbb{R}$, $z = f(x, y)$. We want to find the shortest connection within the surface between $(x_0, y_0, f(x_0, y_0))$ and $(x_1, y_1, f(x_1, y_1))$. More precisely we want to prove that such a shortest connection *exists*. This question can be written as a problem where a certain arclength integral is to be minimized, but for our purposes at the moment, a different definition of the length will be more expedient:

A parametrized curve in $\mathbb{R}^3$ is a continuous mapping from an interval $[a, b]$ to $\mathbb{R}^3$. The length of a parametrized curve $\gamma \in C^0([a, b] \to \mathbb{R}^3)$ is defined as

$$\ell[\gamma] := \sup \left\{ \sum_{i=1}^{n} |\gamma(t_i) - \gamma(t_{i-1})| \;\middle|\; a = t_0 < t_1 < \ldots < t_n = b \right\}$$

The curve is called rectifiable if its length is finite.

If $h : [a', b'] \to [a, b]$ is continuous, one-to-one and onto, with continuous inverse function, such that $h(a') = a$ and $h(b') = b$, then the parametrized curves γ and $\tilde{\gamma} := \gamma \circ h$ represent the same geometric curve, only with different parametrizations, $\tilde{\gamma}(s) = \gamma(h(s)) = \gamma(t)$. It is easy to see that $\ell[\gamma] = \ell[\tilde{\gamma}]$, so the length of the geometric curve does not depend on the chosen parametrization.

It is always possible to reparametrize a rectifiable curve over the unit interval $[0, 1]$ and proportional to arclength, i.e., $\ell(\gamma|_{[t_1,t_2]}) = (t_2 - t_1)\ell[\gamma]$. This is not quite trivial, but the proof is of no concern to us here.

We will always choose a parametrization over $[0, 1]$ proportional to arclength.

Now the geodesic problem consists of minimizing $\ell[\gamma]$ among all curves on the graph of a given function f:

$$\min \left\{ \ell[\gamma] \;\middle|\; \gamma \in C^0([0, 1] \to \mathbb{R}^3)\,,\; \begin{array}{l} \gamma(0) = (x_0, y_0, f(x_0, y_0)) \\ \gamma(1) = (x_1, y_1, f(x_1, y_1)) \end{array}\,,\; \gamma(s)_3 = f(\gamma(s)_{1,2}) \right\}$$

4.5 The existence of a shortest geodesic is not trivial. Consider as surface the sphere with its south pole punched out. If on the full sphere, the shortest connection from point A to point B passes through the south pole, then on the sphere with the south pole punched out, there is no shortest connection between A and B.

This obvious example can be camouflaged a bit by means of stereographic projection (which sends the south pole to infinity). With x and y stereographic projection coordinates on the sphere, the arclength of a curve $(x(t), y(t))$ on the sphere could be expressed as an integral according to the formula in calculus, $I[x, y] := \int \frac{\sqrt{x'^2+y'^2}}{1+x^2+y^2}\, dt$, which looks unconspicuous, but of course the existence of a minimum depends very subtly on the boundary data.

In the problem as defined in 4.4, this issue is avoided, because the function f whose graph determines the surface was assumed to be defined on all of $\mathbb{R}^2$, so no points are punched out.

Moreover we use the supremum definition of the length, rather than the integral definition from calculus, even though the two definitions could be shown to be equivalent, provided a sufficiently advanced definition of the integral is adopted. (With the Riemann integral, the calculus definition is a restriction of the supremum definition to sufficiently nice curves.)

Theorem 4.6 (Arzelà–Ascoli) *Consider a sequence of continuous functions* $f_n : [a, b] \to \mathbb{R}^k$*. Assume the sequence is bounded and equicontinuous. ('Bounded' means that there exists an* M *such that* $\|f_n\|_{C^0} \le M$ *for all* n*. 'Equicontinuous' means that in the* $\varepsilon - \delta$ *definition of continuity, the* δ *can be chosen independent of* n*: For every* $t_0 \in [a, b]$ *and every* $\varepsilon > 0$*, there exists a* $\delta > 0$*, such that* $|t - t_0| < \delta$ *implies* $|f_n(t) - f_n(t_0)| < \varepsilon$ *for all* n*.)*

Then there exists a subsequence f_{n_j} *and a limit function* $f_* \in C^0([a, b] \to \mathbb{R}^k)$ *such that* $\|f_{n_j} - f_*\|_{C_0} \to 0$.

The proof of this theorem can be found in advanced calculus texts. It plays the same role in C^0 as does Bolzano-Weierstrass in $\mathbb{R}^n$, namely it provides hypotheses sufficient for the existence of a convergent subsequence.

4.7 (Existence Proof) On the set X consisting of all rectifiable curves

$$\gamma : t \mapsto (x(t), y(t), f(x(t), y(t))) \,,\ [0, 1] \to \mathbb{R}^3$$

that satisfy the boundary conditions $x(0) = x_0, y(0) = y_0$ and $x(1) = x_1, y(1) = y_1$ and are parametrized proportional to arclength, the functional $\ell[\gamma]$ is bounded below, because $\ell[\gamma] \ge 0$ trivially. Also the set X is non-empty, if $f \in C^1$ (a simple argument that we are skipping for the moment).

So let $\ell_* := \inf\{\ell[\gamma] \mid \gamma \in X\}$. Take a sequence of curves γ_n such that $\ell[\gamma_n] \to \ell_*$. For each γ_n, we have $|\gamma_n(t) - \gamma_n(s)| \le \ell[\gamma_n|_{[s,t]}] = |t - s|\, \ell[\gamma_n] \le L|t - s|$ for some constant L. This means that the sequence γ_n is equicontinuous (take $\delta = \varepsilon/L$). The sequence is clearly bounded since eache $\gamma_n(t)$ is within distance $\le \ell[\gamma_n] \le L$ from the point $(x_0, y_0, f(x_0, y_0))$, and therefore $\|\gamma_n\|_{C^0} \le L + \sqrt{x_0^2 + y_0^2 + f(x_0, y_0)^2}$.

By the theorem of Arzelà–Ascoli, we can find a limit curve γ_* and a subsequence γ_{n_j} converging uniformly (i.e., in terms of the C^0-norm) to γ_*.

Now does $\|\gamma_{n_j} - \gamma_*\|_{C^0} \to 0$ imply $\ell[\gamma_{n_j}] \to \ell[\gamma_*]$? If so, we would conclude $\ell[\gamma_*] = \ell_*$, because we already know that $\ell[\gamma_{n_j}] \to \ell_*$. And then we would find γ_* to be a minimizer.

But the answer is a resounding NO! $\gamma_{n_j} \to \gamma_*$ in the sense of $\|\gamma_{n_j} - \gamma_*\|_{C^0} \to 0$ does NOT imply $\ell[\gamma_{n_j}] \to \ell[\gamma_*]$. – Nevertheless, we can still salvage the reasoning. (Proof to be finished below)

We interrupt the proof to study the issue.

The failure of the continuity property $\ell[\gamma_{n_j}] \to \ell[\gamma_*]$ is obvious in the following example:

The figure shows a zigzag curve γ_n consisting of n straight segments with slope 1 and -1 alternatingly, and a limit curve γ_*, which is the straight line. The common endpoints are distance 1 apart. As $n \to \infty$, $\gamma_n \to \gamma_*$ in the C^0 norm. However, $\ell[\gamma_n] = \sqrt{2}$ for all n, but $\ell[\gamma_*] = 1$.

The crucial observation is the following: whereas $\ell[\gamma_*]$ may not be the limit of $\ell[\gamma_n]$, it can only be smaller, not larger than the limit of $\ell[\gamma_n]$. This inequality applies generally for all sequences γ_n converging to γ_*.

Definition 4.8 (Lower Semicontinuity) *A real valued function I (defined on $\mathbb{R}^n$ or some function space X, or any metric space for that matter) is continuous if $u_n \to u$ implies* $\lim I[u_n] = I[u]$.

It is called lower semicontinuous (lsc), if $u_n \to u$ implies $\liminf I[u_n] \geq I[u]$.

It is called upper semicontinuous (usc), if $u_n \to u$ implies $\limsup I[u_n] \leq I[u]$.

Theorem 4.9 *The length functional is lower semicontinuous with respect to uniform convergence, i.e., if $\|\gamma_j - \gamma_*\|_{C^0} \to 0$, then $\ell[\gamma_*] \leq \liminf \ell[\gamma_j]$.*

PROOF: For each partition $a = t_0 < t_1 < t_2 < \ldots < t_r = b$ of $[a, b]$, we can write

$$\sum_{i=1}^{r} |\gamma_*(t_i) - \gamma_*(t_{i-1})| = \lim_{j\to\infty} \sum_{i=1}^{r} |\gamma_j(t_i) - \gamma_j(t_{i-1})| \leq \liminf_{j\to\infty} \ell[\gamma_j]$$

On the left hand side, take now the supremum over all partitions to get the conclusion. ■

4.10 (Existence proof finished) We had constructed a sequence γ_{n_j} converging to some curve γ_* where $\ell[\gamma_{n_j}] \to \ell_* = \inf\{\ell[\gamma] \mid \gamma \text{ admissible }\}$. We now can conclude from the lower semicontinuity property of the length functional that $\ell[\gamma_*] \leq \liminf \ell[\gamma_{n_j}] = \ell_*$. On the other hand, it is trivial that also $\ell[\gamma_*] \geq \ell_*$, because the limit curve γ_* is still admissible.

Together this implies $\ell[\gamma_*] = \ell_*$, so γ_* is a minimizer. ■

4.11 *Remark for those who know point set topology: The set $\{]\alpha, \infty[\mid \alpha \in \mathbb{R} \cup \{\pm\infty\}\} =: \tau$ is a (non-Hausdorff) topology on $\mathbb{R}$, coarser than the usual metric topology on $\mathbb{R}$. Now $I : X \to \mathbb{R}$ is lower semicontinuous exactly if $I : X \to (\mathbb{R}, \tau)$ is continuous, i.e., if and only if $I^{-1}(]\alpha, \infty[)$ is open in X for all α.*

Parts missing – see handwritten notes

Chapter 5

Discerning weak and strong minima

5.1 (Outline / Introduction) We assume now $I[y] := \int_{t_-}^{t_+} L(t, y(t), y'(t))\, dt$ where $L \in C^3$ (the 3rd derivative will be conveninet below) and y real-valued (scalar-valued). So far, we have only a theory for *locally* weak minima:

(a) If y_* is locally a weak min,
then the EL eqn holds and $L_{y'y'}(t, y_*(t), y_*'(t)) \geq 0$ for all $t \in [t_-, t_+]$.

(b) If y_* satisfies the EL equation and $L_{y'y'}(t, y_*(t), y_*'(t)) > 0$ for all $t \in [t_-, t_+]$,
then y_* is locally a weak minimum.

Recall that "locally weak minimum" means that short segments $y_*|_{[t_0-\delta, t_0+\delta]}$ are weak minimizers. The entire segment may not be a genuine relative minimizer.

We will now assume that $y_* \in C^1$ solves the EL equation and satisfies the strict Legendre condition $L_{y'y'}(t, y_*(t), y_*'(t)) =: L^*_{y'y'} > 0$. To discern genuine weak minima (on the whole segment), we study as before the 2nd variation

$$\frac{d^2}{d\varepsilon^2}\Big|_{\varepsilon=0} I[y_* + \varepsilon v] = \int_{t_-}^{t_+} (L^*_{yy} v^2 + 2L^*_{yy'} vv' + L^*_{y'y'} v'^2)\, dt$$

where the L^*_{yy} and $L^*_{yy'}$ arise by plugging $y_*(t)$ and $y_*'(t)$ into L_{yy} and $L_{yy'}$ (just as we had defined it already for $l^*_{y'y'}$).

We had previously drawn the conclusion that **if** this quantity is $\geq c \int_{t_-}^{t_+} v'^2\, dt$ for some positive c and all v with boundary data 0, **then** y_* is a weak minimizer. (See discussion after Thm. 2.14). At the time we continued arguing that on short segments $[t_0 - \delta, t_0 + \delta]$, the lst term $L^*_{y'y'} v'^2$ dominates the other two terms, so that only the positivity of $L^*_{y'y'}$ was needed to conclude the positivity of the 2nd derivative. But as we now relinquish the 'short segment' hypothesis, it is not easy to check whether indeed

$$\int_{t_-}^{t_+} (L^*_{yy} v^2 + 2L^*_{yy'} vv' + L^*_{y'y'} v'^2)\, dt \geq c \int_{t_-}^{t_+} v'^2\, dt\ .$$

Using an integration by parts on the middle term, we can write the left hand side as

$$A[v] := \int_{t_-}^{t_+} (Qv^2 + Pv'^2)\, dt$$

where $P = L^*_{y'y'} > 0$ and $Q = L^*_{yy} - \frac{d}{dt}L^*_{yy'}$. If $Q > 0$, the task is trivial, but in most cases Q will have some negative values on $[t_-, t_+]$. We will call the task of minimizing $A[\cdot]$ the 'accessory variational problem'. There are only two possibilities: Either $\min A[\cdot] = 0$ (taken on for $v = 0$ at least), or else $\inf A[\cdot] = -\infty$. This is because, if $A[\tilde{v}] < 0$ for some $\tilde{v}$, then $A[n\tilde{v}] \to -\infty$ when $n \to \infty$.

We could try to find c by first normalizing, say $\int_{t_-}^{t_+} v^2\,dt$ to 1 and use direct methods to minimize $\int (P-\varepsilon)v'^2 + Qv^2$, hopefully a positive λ_1, and then conclude $\int (Pv'^2 + Qv^2)\,dt \geq \varepsilon \int v'^2\,dt + \lambda_1 \int v^2\,dt \geq \varepsilon \int v'^2\,dt$. Historically, such an approach was not available since direct methods are younger than the study of (globally) weak minimizers; even now, this approach would not be very helpful since the actual solution of the EL equation and determination of λ_1 may not be practical. But the study of the EL for the accessory variational problem is nevertheless useful, and its purpose will become clear shortly.

Definition 5.2 *The Euler-Lagrange equation of the accessory variational problem, namely the equation*

$$\frac{d}{dt}(L^*_{y'y'}v' + L^*_{yy'}v) = L^*_{yy'}v' + L^*_{yy}v \qquad \textit{equivalently} \qquad \frac{d}{dt}(Pv') = Qv$$

is called Jacobi equation. If a nontrivial solution v to the Jacobi equation with initial condition $v(t_-) = 0$, $v'(t_-) \neq 0$ has onother zero t_c, then $y_*(t_c)$ is called a conjugate point to $y_*(t_-)$ on the extremal[1] *y_*.*

It turns out that conjugate points on an extremal are obstructions to minimality. In the problem about geodesics on the sphere, on any extremal (= great circle), the antipode of a point P_0 would be conjugate to P_0. Informally speaking, a conjugate point is one where nearby extremals (that start at P_0 as well, but with slightly different slope) meet the original extremal again.

Lemma 5.3 *Let $y(\cdot, s)$ for $|s| < 1$ be a family of solutions to the EL equation $\frac{d}{dt}L_{y'}(t, y(t), y'(t)) = L_y(t, y(t), y'(t)$ (regardless of BCs). Then $v(\cdot) = \frac{d}{ds}\Big|_{s=0} y(\cdot, s)$ solves the Jacobi equation at $y_* = y(\cdot, 0)$.*

PROOF: For each s, we have (with $'$ still referring to $\frac{\partial}{\partial t}$)

$$L_y(t, y(t,s), y'(t,s)) = \frac{\partial}{\partial t} L_{y'}(t, y(t,s), y'(t,s))$$

Taking the derivative with respect to s, we get

$$L_{yy}(t, y(t,s), y'(t,s))\frac{\partial}{\partial s}y(t,s) + L_{yy'}(\ldots)\frac{\partial}{\partial s}y'(t,s) = \frac{\partial}{\partial t}\Big(L_{y'y}(\ldots)\frac{\partial}{\partial s}y(t,s) + L_{y'y'}(\ldots)\frac{\partial}{\partial s}y'(t,s)\Big)$$

At $s = 0$, we get from this

$$L^*_{yy}v + L^*_{yy'}v' = \frac{d}{dt}\Big(L^*_{y'y}v + L^*_{y'y'}v'\Big)$$

which is the Jacobi equation. ■

[1] The word extremal refers to a solution of the EL equation of a variational problem, regardless of minimality.

Theorem 5.4 *Assume* $L \in C^3$, *and* $y_* \in C^1$ *satisfies the EL equation, and* $L^*_{y'y'} > 0$ *on* $[t_-, t_+]$.
(a) If y_* *is a weak minimum, then there cannot be a conjugate point to* $y_*(t_-)$ *in the interior of the segment* y_*, *i.e., for* $t_- < t < t_+$.
(b) If there is no conjugate point to $y_*(t_-)$ *in the closed segment* $t_- \leq t \leq t_+$, *then* y_* *is a weak minimum.*

We'll see two proofs for part (b), the second of which will also shed light on the issue of strong minima.

5.5 (Proof of Thm. 5.4(a))
Suppose we have a conjugate point in the interior of the segment y_*, say at $t_c < t_+$. So we have v a solution to the Jacobi equation, satisfying $v(t_-) = 0 = v(t_c)$. Then we consider the variation w given by $w(t) = v(t)$ for $t_- \leq t \leq t_c$ and $w(t) = 0$ for $t_c \leq t \leq t_+$. We conclude using that v satisfies the Jacobi equation, that

$$A[w] := \partial_w^2 I[y_*] = \int_{t_-}^{t_+} (Pw'^2 + Qw^2)\, dt = \int_{t_-}^{t_c} (Pv'^2 + v\frac{d}{dt}(Pv'))\, dt = \int_{t_-}^{t_c} (Pv'^2 - Pv'^2)\, dt = 0$$

We note that $v'(t_c) \neq 0$ because otherwise the uniqueness theorem for the Initial value problem $\frac{d}{dt}(Pv') = Qv$, $v(t_c) = 0$, $v'(t_c) = 0$ would imply that $v \equiv 0$. We have used $L \in C^3$, hence $P \in C^1$, to apply the simplest version of the uniqueness theorem, but somewhat weaker hypotheses would still work. So if the accessory VP $A[\cdot]$ were nonnegative, then w would be a minimizer with a corner, hence would need to satisfy Erdmann's corner condition $P(t_c)w'(t_c-) = P(t_c)w'(t_c+)$, i.e., $P(t_c)v'(t_c) = 0$; but since $P(t_c) > 0$, the corner condition is violated. So we conclude that $A[\cdot]$ does not have a minimum at w, i.e., it is negative for some variation $\tilde{w}$, and therefore y_* is not a weak minimum.

Before tackling part (b), let us recall key results for the Theory of ODEs:

Theorem 5.6 *Given an ODE* $y' = f(t, y)$ *where* $f \in C^1$, *with an initial condition* $y(t_0) = a$, *there exists exactly one solution to this initial value problem, and it exists on some interval* $]t_{min}, t_{max}[$ *around* t_0. *This interval may or may not have infinite length.*

If the ODE is linear, $y' = a(t) + b(t)y'$, *then the existence and uniqueness result is global, i.e.,* $t_{min} = -\infty$ *and* $t_{max} = +\infty$ *(provided only* $a(t)$ *and* $b(t)$ *are continuous and defined on all of* $\mathbb{R}$*).*

The same results hold for ODEs of vector-valued functions. Writing a 2nd order equation $y'' = f(t, y, y')$ *as a 1st order system for* $Y = \begin{bmatrix} y \\ y' \end{bmatrix}$, *namely* $Y' = \begin{bmatrix} f_1(t,Y) \\ f_2(t,Y) \end{bmatrix}$ *with* $f_1(t, Y) = Y_2$ *and* $f_2(t, Y) = f(t, Y_1, Y_2)$ *proves therefore a uniqueness result for an IVP* $y'' = f(t, y, y')$, $y(t_0) = a$, $y'(t_0) = b$.

If f *also depends continuously on a parameter* ρ, *then the unique solution to the IVP depends continuously on* ρ *and on the initial condition* a, *and it does so uniformly on every compact time interval.*

5.7 (Proof of Thm. 5.4(b))
In a first step we show simply $A[v] \geq 0$. The only reason how $A[v]$ could be negative is that $Q < 0$ for some t, since $P > 0$ was assumed. Note that for any function W, it holds

$$\int_{t_-}^{t_+} (Pv'^2 + Qv^2)\, dt = \int_{t_-}^{t_+} \left(Pv'^2 + \frac{d}{dt}(Wv^2) + Qv^2 \right)\, dt$$

We may try to find W in such a way that the integrand becomes pointwise a perfect square:

$$Pv'^2 + 2Wvv' + (W' + Q)v^2 = P\Big(v' + \frac{W}{P}v\Big)^2 + \Big(W' + Q - \frac{W^2}{P}\Big)v^2$$

If we can solve the ODE $W' = \frac{1}{P}W^2 - Q$ on the interval $[t_-, t_+]$ then we'd prove $A[v] \geq 0$ at least. But in general, if the interval $[t_-, t_+]$ is loo long, a solution W may not exist: while for each initial condition $W(t_-) = a$, there is a unique solution W, this solution may go to ∞ for some $t < t_+$ already. This equation is a *Riccati* equation, for which routine cookbook method exists: namely introducing a new unknown function U by the substitution $W = -PU'/U$ reduces the Riccati equation to a linear second order ODE:

$$W' = \frac{1}{P}W^2 - Q \quad \textit{becomes} \quad -P'\frac{U'}{U} - P\frac{U''U - U'^2}{U^2} = P\frac{U'^2}{U^2} - Q$$

i.e.,

$$-P'U' - PU'' = -QU \ , \qquad \text{or} \qquad \frac{d}{dt}(PU') = QU \ ,$$

our good old Jacobi equation. Its solutions exist for all times t, because it is a *linear* equation, but we need nonvanishing solutions U in order for $W = -PU'/U$ to be meaningful.

By hypothesis, we have a solution U with $U(t_-) = 0$, $U'(t_-) \neq 0$ (w/o l.o.g., $U'(t_-) = 1$) that does not vanish anywhere on $]t_-, t_+]$. By continuity, for small $\varepsilon > 0$, the solution with initial conditions $U(t_-) = \varepsilon$, $U'(t_-) = 1$, doesn't vanish anywhere on $[t_-, t_+]$ and it gives therefore rise to a well-defined solution $W = -PU'/U$ to the Riccati equation, thus proving that $A[v] = \int P(v' + \frac{W}{P}v)^2\,dt \geq 0$. Equality holds only if $v' + \frac{W}{P}v = 0$ on the entire interval, and this means, by the uniqueness theorem, and since $v(t_-) = 0$, that $v \equiv 0$. We need a bit more than $A[v] \geq 0$, namely $A[v] \geq c\int v'^2$, and here is how we get it:

Method 1: By direct methods, the minimum of $A[v]$ under the constraint $\int v^2\,dt = 0$ exists. This minimum is positive because $A[v] > 0$ unless $v = 0$. Denoting the value of this minimum to be $\alpha > 0$, we conclude

$$A[v] = \int P\Big(v' + \frac{W}{P}v\Big)^2\,dt \geq \alpha\int v^2\,dt$$

Note that we still don't have v'^2 on the right, only v^2. We couldn't have used $\int v'^2\,dt = 1$ as a constraint, because under weak convergence of the v', this constraint may not pass to the limit. Now we choose a small positive constant ρ, such that at least $\rho \leq \frac{1}{2}\min P$, and we argue

$$(P-\rho)v'^2 + 2Wvv' + \frac{W^2}{P}v^2 = (P-\rho)\Big(v' + \frac{W}{P-\rho}v\Big)^2 - \rho\frac{W^2}{P(P-\rho)}v^2$$

Hence, with $M := \max 2W^2/P^2$,

$$\begin{aligned} A[v] &= \int \rho v'^2\,dt + \int (P-\rho)\Big(v' + \frac{W}{P-\rho}v\Big)^2\,dt - \int \rho\frac{W^2}{P(P-\rho)}v^2\,dt \\ &\geq \int \rho v'^2\,dt + 0 - \rho M\int v^2\,dt \geq \rho\int v'^2\,dt - \frac{\rho M}{\alpha}A[v] \end{aligned}$$

and this implies $A[v] \geq \frac{\rho}{1+\rho M/\alpha}\int v'^2\,dt$.

Method 2: The same idea, from scratch rather than relying on direct methods. We want to show that $A_\rho[v] := \int\Big((P-\rho)v'^2 + \frac{d}{dt}(Wv^2) + Qv^2\Big)\,dt \geq 0$. Our new W needs to satisfy the Riccati equation $W' = \frac{1}{P-\rho}W^2 - Q$, and our new U (where $W = -(P-\rho)U'/U$) must then solve $QU = \frac{d}{dt}((P-\rho)U')$. With a nonzero solution U_0 on $[t_-, t_+]$ for $\rho = 0$, we still get a nonvanishing solution U_ρ for ρ sufficiently small and positive, because solutions of ODEs depend continuously on parameters. So we prove $A_\rho[v] \geq 0$, which is $A[v] \geq \rho \int v'^2\, dt$.

So we have shown a property that amounts to the positive definiteness of the second derivative of I, uniformly in all directions, as measured in the $W^{1,2}$-norm. We still need to show that the deviation of I from its 2nd order Taylor approximation is small compared to our estimate, provided $\|v\|_{C^1}$ is sufficiently small. This is done as in the case of the locally weak minimum, namely in eqn. (2.6) we saw that

$$I[y_* + v] - I[y_*] - DI[y_*]v - \tfrac{1}{2}D^2I[y_*](v,v) =$$
$$= \int_0^1 (1-\sigma)\Big(D^2I[y_* + \sigma v](v,v) - D^2I[y_*](v,v)\Big)\, d\sigma$$

One typical term in the big paranthesis looks like

$$\Big(L_{yy}(t, y_*(t) + \sigma v(t), y_*'(t) + \sigma v'(t)) - L_{yy}(t, y_*(t), y_*'(t))\Big)v(t)^2$$

Depending on the modulus of continuity for L_{yy}, for any $\varepsilon > 0$, we can find $\delta > 0$ so that if $\|v'\|_{C^1} < \delta$, then this quantity is bounded by $\varepsilon v(t)^2$. There are similar terms $(L_{yy'}(t, y_* + \sigma v, y_*' + \sigma v') - L_{yy'}(t, y_*, y_*'))vv'$ and $(L_{y'y'}(t, y_* + \sigma v, y_*' + \sigma v') - L_{y'y'}(t, y_*, y_*'))v'^2$ for which the same reasoning can be applied. So we conclude that

$$|I[y_* + v] - I[y_*] - \tfrac{1}{2}D^2I[y_*](v,v)A| \leq 2\varepsilon \int (v^2 + v'^2)\, dt \leq K\varepsilon \int v'^2\, dt$$

where K depends on the length of the integration interval. If we choose ε small enough for $K\varepsilon$ to be less than the ρ from the lower estimate of the accessory variational problem, we infer that $I[y_* + v] > I[y_*]$. ■

While we have done this reasoning only for the case of scalar-valued functions, it can be generalized to vector-valued functions with minor modifications. (Sec. 29 of Gelfand-Fomin). We will now study an approach that can also handle strong extrema. This is the approach of extremal fields, which means that we cover a neighborhood of our extremal y_* with other solutions of the EL equation, none of which intersect. Let's first look at a simple paradigm:

Example 5.8 (Paradigm for Extremal Field) *Suppose we want to minimize* $\int_0^1 y'^2\, dt$ *subject to the boundary conditions* $y(0) = 0, y(1) = 1$*. The EL equation is* $y'' = 0$ *with the general solution* $y(t) = a + bt$*. The only solution satisfying the boundary conditions is* $y(t) = t$*. We want to show that this is an absolute minimim. In this simple example we could argue by convexity, but we will forego such a shortcut to pursue a more generally applicable method. We write*

$$\int_0^1 y'^2\, dt = \int_0^1 \Big((y'-1)^2 + 2y' + 1\Big)\, dt = \int_0^1 (y'-1)^2\, dt + [2y(t) + t]_0^1 = \int_0^1 (y'-1)^2\, dt + 1$$

There are two crucial features in this decomposition: The second term under the integral is a total derivative, so its integral depends only on the boundary data, not on the curve connecting between them. The first term yields an obviously non-negative integral, because it *is nonnegative pointwise. So this integral is clearly minimal (with value 0) exactly if* $y' \equiv 1$. *This condition* $y' \equiv 1$ *determines a 1-parameter family of curves* $y = a + t$ *(all of which are extremals). The extramal* y_* *is contained in this 1-parameter family, and is the only one that satisfies the boundary conditions. So clearly* y_* *is absolutely minimal.*

We can redo the same reasoning with another 1-parameter subfamily of the 2-parameter family of extremals, provided the selected subfamily of extremals has no intersections. For instance we could take the family $y = a + (1 - \frac{a}{2})t$. *(These curves intersect at* $t = 2$*, which is insignificant, because only* $0 \le t \le 1$ *is in view.)*

By differentiatinf $y = a + (1 - \frac{a}{2})t$ *with respect to* t *we get* $y' = 1 - \frac{a}{2}$. *Eliminating* a *from these equations, we find* $y' = \frac{2-y}{2-t}$. *This 1st order ODE has the given 1-parameter family of functions as general solution. We can now write*

$$\int_0^1 y'^2\,dt = \int_0^1 \Big(y' - \frac{2-y}{2-t}\Big)^2\,dt - \int_0^1 \frac{d}{dt}\frac{(2-y)^2}{2-t}\,dt$$

Again, the second integral depends only on the boundary conditions, and equals 1 if we use the ones that were prescribed. The first integral is nonnegative and vanishes exactly if y *is an extremal in the selected 1-parameter family (among which only* y_* *satisfies the boundary conditions.*

We will now perform this construction in full generality:

5.9 (General Construction of Extremal Fields) Now y may be vector valued, and $I[y] = \int_{t_-}^{t_+} L(t,y,y')\,dt$ is as before. At every point (t,y), we want to define a 'cost-free direction' $\psi(t,y)$. And together with it we want to find a scalar function $S(t,y)$, such that L can be decomposed as

$$L(t,y(t),y'(t)) = \tilde{L}(t,y(t),y'(t)) + \frac{d}{dt}S(t,y(t))$$

where the function $\tilde{L}(t,y,\cdot)$ (viewed as a function of the third variable only) has an absolute minimum at $y' = \psi(t,y)$, with value 0. If we can achieve this, then the $\frac{d}{dt}S$ term contributes to the integral only via the boundary conditions, whereas the $\int \tilde{L}\,dt$ term is always nonnegative, and vanishes only if y solves the 1st order ODE $y' = \psi(t,y)$. We also intend for the extremal y_* satisfying the boundary conditions to be among the solutions of $y' = \psi(t,y)$.

For the mapping

$$y' \mapsto \tilde{L}(t,y,y') = L(t,y,y') - S_t(t,y) - S_y(t,y)y'$$

to have a minimum when $y' = \psi(t,y)$, we need the partial wrt y' to vanish at ψ:

$$L'_y(t,y,\psi(t,y)) = S_y(t,y)$$

For the value of this minimum to be 0, we need

$$L(t,y,\psi(t,y)) - \psi(t,y)L_{y'}(t,y,\psi(t,y)) = S_t(t,y)$$

For a function S with prescribed partial derivatives S_t and S_y to exist, it is necessary that integrability conditions are satisfied: the crosswise partial derivatives must coincide: $\partial_y S_t = \partial_t S_y$. If y is vector-valued, the crosswise partial derivatives for different y-components must coincide as well: $\partial_{y_i} S_{y_j} = \partial_{y_j} S_{y_i}$. In other words, ψ must satisfy

$$\partial_t L_{y'_i} = \sum_k L_{y'_i y'_k}\psi^k_{y_j} = L_{y'_j y_i} + \sum_k L_{y'_j y'_k}\psi^k_{y_i}$$

or, equivalently,

$$L_{y_i' y_j} - L_{y_j' y_i} = \sum_k \Big(L_{y_j' y_k'} \psi^k_{y_i} - L_{y_i' y_k'} \psi^k_{y_j} \Big)$$

We now show that the integrability conditions require that solutions of $y' = \psi(t, y)$ satisfy the EL equation: Indeed (with explanations below),

$$\begin{aligned}
\frac{d}{dt} L_{y_i'} &= \frac{d}{dt} L_{y_i'}(t, y(t), \psi(t, y(t))) = \partial_t L_{y_i'} + \sum_j L_{y_i' y_j} y_j'(t) + \sum_{j,k} L_{y_i' y_k'} \psi^k_{y_j} y_j'(t) \\
&=_* \partial_{y_i} \Big(L - \sum_k \psi^k L_{y_k'} \Big) + \sum_j L_{y_i' y_j} y_j'(t) + \sum_{j,k} L_{y_i' y_k'} \psi^k_{y_j} \psi^j \\
&= L_{y_i} + \sum_k L_{y_k'} \psi^k_{y_i} - \sum_k \psi^k_{y_i} L_{y_k'} - \sum_k \psi^k L_{y_k' y_i} - \sum_{k,j} \psi^k L_{y_k' y_j'} \psi^j_{y_i} + \sum_j L_{y_i' y_j} \psi^j + \sum_{j,k} L_{y_i' y_k'} \psi^k_{y_j} \psi^j \\
&= L_{y_i} - \sum_k \psi^k L_{y_k' y_i} - \sum_{k,j} \psi^k L_{y_k' y_j'} \psi^j_{y_i} + \sum_j L_{y_i' y_j} \psi^j + \sum_{j,k} L_{y_i' y_k'} \psi^k_{y_j} \psi^j \\
&=_{**} L_{y_i} + \sum_j (L_{y_i' y_j} - L_{y_j' y_i}) \psi^j + \sum_{j,k} \psi^j (L_{y_i' y_k'} \psi^k_{y_j} - L_{y_j' y_k'} \psi^k_{y_i}) = L_{y_i}
\end{aligned}$$

In "$=_*$", the y-t integrability was used, and $y' = \psi$; in "$=_{**}$", terms 2 and 4 were combined, and also terms 3 and 5, each time with a renaming of a summation variable. Finally, all terms but the L_y term canceled in view of the y-y integrability condition.

We conclude: *Our program was to decompose the 'cost' I into a system specifying a 'cost-free' direction ψ at every point (t, y) and a path-independent fixed 'cost'. In order for this program to be successful, it is necessary that the ψ are directions along a familiy of nonintersecting extremals (i.e., that the solutions to $y' = \psi(t, y)$ satisfy the EL equation), i.e,* if we have *a direction field ψ satisfying all the integrability conditions,* then *solution curves of $y' = \psi(t, y)$ satisfy the EL equation. Conversely we claime for the* scalar-valued *case: If the integral curves of the direction field ψ (i.e., the solutions to $y' = \psi(t, y)$) satisfy the EL equations, then the (one and only) integrability condition is satisfied.*

The sufficiency is easily seen if we redo the calculation in the scalar case and read it backwards:

$$\begin{aligned}
\frac{d}{dt} L_{y'} &= \frac{d}{dt} L_{y'}(t, y(t), \psi(t, y(t))) = \partial_t L_{y'} + L_{y'y} y'(t) + L_{y'y'} \psi_y y'(t) \\
&=_* \Big[\partial_t L_{y'} - \partial_y \Big(L - \psi L_{y'} \Big) \Big] + \partial_y \Big(L - \psi L_{y'} \Big) + L_{y'y} y' + L_{y'y'} \psi_y y' \\
&= [\dots] + L_y + L_{y'} \psi_y - \psi_y L_{y'} - \psi L_{y'y} - \psi L_{y'y'} \psi_y + L_{y'y} y' + L_{y'y'} \psi_y y' \\
\frac{d}{dt} L_{y'} - L_y &= [\dots] + (L_{y'y} + \psi_y L_{y'y'})(y' - \psi)
\end{aligned}$$

So indeed if the integral curves satisfy the EL equations, the left side as well as $y' - \psi$ are 0 and the bracketed term has to vanish, which is the t-y integrability condition. — Similarly, for $n > 1$, if we assume only the EL equation and the y-y integrability conditions, the y-t integrability condition follows.

Definition 5.10 *An extremal field in $[t_-, t_+] \times G$, where $G \subset \mathbb{R}^n$ is a simply connected domain, is a family of extremals (solutions to the EL equation) such that through each point of $[t_-, t_+] \times G$, there passes exactly one extremal. If $n > 1$, a Mayer field is an extremal field* *satisfying the also the y-y integrability conditions (with the y-t integrability conditions then following because it is an extremal field).*

If we have an extremal field in a relatively open subset of $]t_-, t_+] \times \mathbb{R}^n$ *such that all extremals pass through one point* (t_-, y_0), *we call it a central field (similarly central Mayer field).*

We will see soon that the embeddability of one given extremal y_* into an extremal field covering some neighborhood of the graph of y_* in $[t_-, t_+] \times \mathbb{R}^n$ is equivalent to the absence of conjugate points along y_*.

First we notice that the integrability conditions are not only necessary for the existence of the function S, but in a simply connected domain they are also sufficient. On the other hand we notice that our program has only used the derivative condition on $\tilde{L}$, so we still have to determine if $\tilde{L}$ has an absolute or relative minimum at $y' = \psi$.

Lemma 5.11 *Suppose it is possible to define a vector field* $\psi : (t, y) \mapsto \psi(t, y)$ *in some simply connected neighborhood of an extremal* $t \mapsto y_*(t)$, $[t_-, t_+] \to \mathbb{R}^n$, *in such a way that* $a_i(t, y) := L_{y'_i}(t, y, \psi(t, y))$ *and* $b(t, y) := L(t, y, \psi(t, y)) - \sum_j \psi^j(t, y) L_{y'_j}(t, y, \psi(t, y))$ *satisfy the integrability conditions* $\partial_{y_k} a_i = \partial_{y_i} a_k$ *and* $\partial_t a_i = \partial_{y_i} b$, *then there exists a function* $(t, y) \mapsto S(t, y)$ *in that same neighborhood, satisfying* $\partial_{y_i} S = a_i$ *and* $\partial_t S = b$, *and one can write*

$$\int_{t_-}^{t_+} L(t, y(t), y'(t))\, dt = \int_{t_-}^{t_+} \mathfrak{E}(t, y(t), \psi(t, y(t)), y'(t))\, dt + [S(t, y(t))]_{t_-}^{t_+}$$

where

$$\mathfrak{E}(t, y, \psi, y') := L(t, y, y') - L(t, y, \psi) - (y' - \psi) L_{y'}(t, y, \psi)$$

Proof: The existence of S (i.e., the sufficiency of the integrability conditions) is known from Calculus. We write $L = (L - ay' - b) + (ay' + b)$, and the first term is immediately seen to be $\mathfrak{E}$, whereas the second term is $y' \partial_y S + \partial_t S = \frac{d}{dt} S(t, y(t))$. ■

5.12 (Notes) (a) $\mathfrak{E}$ is called the Weierstrass excess function, or Weierstrass $\mathfrak{E}$ function. It describes the excess of $L(t, y, y')$ over the first order Taylor approximation at $y' = \psi$. Convexity of L is the y' variable alone is sufficient for $\mathfrak{E} \geq 0$. — Some authors swap the 3rd and 4th variable in the definition of $\mathfrak{E}$.
(b) The quantity S is called Hilbert's invariant integral.
(c) we will be able to show the existence of the vector field ψ under hypotheses that can be checked easily.