Calculus of Variations (with notes on infinite-dimesional calculus)

J. B. Cooper
Johannes Kepler Universität Linz

Contents

1 Introduction

The calculus of variations deals with the problem of maximizing or minimizing **functionals** i.e. functions which are defined not on subsets of $\mathbf{R}^n$ but on spaces of functions. As a simple example consider all parameterized curves (defined on $[0,1]$) between two points x_0 and x_1 on the plane. Then the shortest such curve is characterized as that curve c with $c(0) = x_0$, $c(1) = x_1$ which minimizes the functional

$$L(c) = \int_0^1 \big|\dot{c}(t)\big| dt$$

Of course this example is not particularly interesting since we know that the solution is the straight line segment from x_0 to x_1. However, it does become interesting and non trivial if we consider points in higher dimensional space and consider only those curves which lie on a given curved surface (the problem of geodetics).

Further examples:

1. The Brachistone Problem. Given are two points O (the origin) and P with $y_P < 0$. Find the curve from O to P for which

   $$\int_c \frac{1}{\sqrt{-y}} ds = \int_0^1 \frac{1}{\sqrt{-c_2(t)}} \sqrt{\dot{c}_1^2(t) + \dot{c}_2^2(t)} dt$$

   is a minimum.

2. Minimal surfaces of revolution. Given are points $P = (x_0, y_0), Q = (x_1, y_1)$ with $x_0 < x_1, y_0 > 0, y_1 > 0$.

   We are looking for that function $f(x)$ so that $f(x_0) = y_0$, $f(x_1) = y_1$ which minimizes the functional

   $$\int_{x_0}^{x_1} f(x) \sqrt{1 + \big(f'(x)\big)^2} dx$$

3. Hamiltonian mechanics. We have a mechanical system, where positions can be defined by n generalized coordinates. The Lagrangian is a function of the form

   $$L(t, x, \dot{x}) = T(x, \dot{x}) - U(t, x)$$

   where T is the kinetic energy and U is potential energy (the particular form of L is dictated by the physics of the system). Then the dependency of the position of an object of the system is described by the fact that the corresponding n-dimensional curve c minimizes the functional

$$I(c) = \int_{t_0}^{t_1} L\big(t, c(t), \dot{c}(t)\big)\, dt.$$

4. The Dirchlet problem. U is a region in $\mathbf{R}^n$ with smooth boundary ∂U and a function g is given on the boundary. The Dirchlet problem is: find a harmonic function f on U with g as boundary values. Dirchlet's principle states that the solution is that function f which satisfies the boundary condition and minimizes the functional

$$D(f) = \int_U \sum_{k=1}^{n} \Big(\frac{\partial f}{\partial x_k}\Big)^2 dx_1 \ldots dx_n$$

(more precisely, is a stationary point).

Further problems which can be formulated in this way are: the isoperimetric problem, Platean's problem (soap bubbles!), Newton's problem, Fermat's principle.

Examples of classical optimization problems.

1. The base b and perimeter $a + b + c$ of a triangle ABC are given. Which triangle has the largest area (solution: the isosceles triangle with $|AB| = |AC|$).

2. Steiner's problem. Given an acute triangle ABC, determine that point P for which $|PA| + |PB| + |PC|$ is a minimum. (Solution: P is the point with $A\hat{P}B = B\hat{P}C = C\hat{P}A = 120°$).

3. A triangle ABC is given. Find P on BC, Q on CA, R on AB so that the perimeter of PQC is minimal. (Solution: PQR is the pedal triangle i.e. P is the foot of the altitude from A to BC etc.).

4. (The isoperimetric problem for polygons). Find that n-gon with fixed perimeter S so that the area is maximal. (Solution: the regular n-gon.)

1.1 Methods employed in the calculus of variations.

I. The direct method (cf. proofs of the Riemann mapping theorem, Radon-Nikodym theorem, existence of best approximations in closed convex sets.)
II. The method of Ritz. One approximates an infinite dimensional problem by a sequence of finite dimensional ones— with solution $x_1, x_2, x_3 \ldots$ Under suitable conditions, this sequence will have a cluster point (limit of a convergent subsequence) which is a solution to the original one.

Examples: We illustrate the latter method on some of the classical situations which lead naturally to variational problems:

I. The isoperimetric problem. In this case we approximate the original problem by the one for n-gons which, as we have seen, has the regular n-gon as solution. With care, a suitable limiting process leads to the circle as solution to the original problem.

II. The Dirichlet problem. Here we are looking for a smooth function ϕ on $\{x^2 + y^2 \leq 1\}$ which minimizes the functional

$$D\phi = \int\int (\phi_x^2 + \phi_y^2) dx dy$$

and satisfies the boundary condition

$$\phi(e^{i\theta}) = f(\theta)$$

for a suitable 2π-periodic function f.

We indicate briefly how to solve this problem. It is natural to work with polar coordinates where the functional D has the form: $D(\phi) = \int_0^{2\pi} \int_0^1 \Big(\phi_r^2 + \frac{1}{r^2}\phi_\theta^2\Big) r dr d\theta$.

We discretise this problem by using Fourier series, i.e., suppose that

$$f(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta)$$

Write $\phi = \frac{1}{2} f_0(r) + \sum_{n=1}^{\infty} \big(f_n(r) \cos n\theta + g_n(r) \sin n\theta\big)$ where the functions f_n and g_n satisfy the conditions $f_n(1) = a_n, g_n(1) = b_n$. Then

$$\begin{aligned} D(\phi) = \pi \int_0^1 f'(r)^2 r dr \quad &+ \quad \pi \sum_1^{\alpha} \int_0^1 \Big(f_n'(r)^2 + \frac{n^2}{r^2} f_n(r)\Big) r dr \\ &+ \quad \pi \sum_{n=1}^{\infty} \int_0^1 \Big(g_n'(r)^2 + \frac{n^2}{r^2} g_n(r)^2\Big) r dr \end{aligned}$$

Since each summand is independent we obtain a minimum by minimizing each term. Hence we consider the problem

$$\int_0^1 \Big(f_n'^2 + \frac{n^2}{r^2} f_n^2\Big) r dr = \min, \quad f(1) = a_n$$

This can be solved using the Ritz method—we consider the restriction of the functional to those functions of the form

$$f_n(r) = c_0 + c_1 r + \cdots + c_m r^m, c_0 + \cdots + c_m = a_m$$

for a fixed $m > n$ and so reduce to a classical optimization problem with solution $f_n(r) = r^n$.

Suummarising, this method leads to the well-known solution

$$\phi(r, \theta) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n r^n \cos n\theta + b_n r^n \sin n\theta)$$

for the Dirichlet problem.

2 The Riemann mapping theorem.

As an application of the principle of optimizing a functional on a space of functions we shall bring a sketch of a proof of the following famous result:

Proposition 1 *(The Riemann mapping theorem) Let Ω be a simply connected proper subset of $\mathbf{C}$. Then Ω is conformably equivalent to the unit disc $U = \{z \in \mathbf{C} : |z| < 1\}$.*

For the proof, we require the following facts.

1. Theorem of Ascoli-Arzela: A subset A of the Banach space $C(K)$ (K a compact metric space) is relatively compact if and only if it is uniformly bounded and equicontinuous. Hence every sequence in such a set has a uniformly convergent subsequence.

2. If Ω is an open subset of $\mathbf{C}$, then $H(\Omega)$, the space of holomorphic functions on Ω, is a linear space and the natural topology on this space, that of compact convergence, is a complete metric topology. A suitable metric can be defined as follows: Write $\Omega = \bigcup \Omega_k$ where each Ω_k is relatively compact in Ω_k and define

$$d(f, g) = \sum_k \frac{1}{2^k} \frac{||f - g||_k}{1 + ||f - g||_k}$$

   where $||f - g||_k = \sup\{|f(z) - g(z)| : z \in \Omega_k\}$.

3. A uniformly bounded subset of $H(\Omega)$ is relatively compact for the above topology and hence every uniformly-bounded sequence has a subsequence which is uniformly convergent on compacta. This follows from 1) (with the help of the diagonal process) since a uniformly-bounded sequence of holomorphic functions is equicontinuous on compacta (Cauchy integral formula).

We now proceed to the proof of the Riemann mapping theorem. We fix a point $z_0 \in \Omega$ and consider the family $\mathcal{F}$ of all functions f in $H(\Omega)$ which are

a) injective

b) such that $f(z_0) = 0$ and $f'(z_0) > 0$

c) map Ω into U.

We shall first show that $\mathcal{F}$ is non empty. Choose $a \in \mathbf{C}$ with $a \in \Omega$ and consider a branch of $\ln(z-a)$ on Ω (this exists since Ω is simply connected). This is, of course, injective. We even have that if it takes a value w_1 then it does not take any of the values $w_1 + 2\pi in$, (for if $\ln(z_1 - a) = \ln(z_2 - a) + 2\pi in$, then $z_1 - a = e^{\ln(z_1-a)} = e^{\ln(z_2-a)}e^{2\pi i} = z_2 - a$).

The function $\ln(z-a)$ is open and so there is a disc of the form $(w - w_0) < \epsilon$ around $(z_0 - a)$ which lies in the image of $\ln(z-a)$. By the above reasoning, the corresponding disc $|w - w_0 + 2\pi i| \leq \epsilon$ is *not* in the image of $\ln(z-a)$. Hence $w_0 = \ln(z-a)$ takes its values in the complement of the latter set. But this is conformably equivalent to U. This provides a univalent function $g : \Omega \to U$. The condition $f'(z_0) = 0$, $f'(z_0) > 0$ can be obtained by composing g with a suitable Blaschke factor $e^{i\theta}\dfrac{z - z_1}{1 - \bar{a}z_1}$ where $z_1 = g(z_0)$ (Exercise).

We now set $d = \sup\{f'(z_0) : f \in \mathcal{F}\}$. By the compactness, there is an $f \in \mathcal{F}$ with $f'(z_0) = d$. We claim that this f is surjective and so conformal from Ω to U. We do this by showing that if f were not surjective we can construct a $g \in \mathcal{F}$ with $g'(z_0) > f'(z_0)$ – contradiction.

Suppose that $w_1 \in U$ is not in the range of f. Once again, by composing with the Blaschke function $z \mapsto \dfrac{z - w_0}{1 - \bar{w}_0 z}$ we can assume that $w_0 = 0$. We let F denote a branch of $\ln f$ on Ω. This takes its values in the left half plane $\{\operatorname{Re} z < 0\}$. We compose with a Mobiustransformation with maps the latter onto U and takes $F(z_0)$ onto 0 to get function G from Ω into U i.e. the mapping

$$w \mapsto \frac{w - \overline{F(z_0)}}{w + \overline{F(z_0)}}.$$

One computes that

$$G'(z_0) = -d\frac{1 - |w_0|^2}{2w_0 \ln |w_0|}$$

Hence the function

$$g(z) = \frac{G(z)|G'(z_0)|}{G'(z_0)}$$

is in $\mathcal{F}$ and

$$g'(z_0) = d\frac{1-|w_0|^2}{2|w_0|\ln\left(\frac{1}{|w_0|}\right)}$$

which is $> d$ (contradiction). (Consider the function

$$h(t) = 2\ln\frac{1}{t} + t - \frac{1}{t}$$

$h(t) = 0$ and $h'(t) > 0$. Hence $2\ln\frac{1}{t} < \frac{1}{t} - t$ $(t < 1)$ and so $g'(z_0) > d$).

We remark that a similar method can be used to solve the Dirichlet problem in two dimensions (see Ahlfors, Complex analysis, p. 196).

3 Analysis in Banach spaces

Since the calculus of variations is concerned with maximising functionals on infinite dimensional space, we consider briefly the a bstract theory of differential calculus on such spaces.

3.1 Differentiation and integration of E-valued functions

Definition: Let Ω be an open or closed interval in $\mathbf{R}$. A function $x : \Omega \to E$ (E a Banach space) is **differentiable** at t_0 if

$$\lim_{t\to t_0}\frac{x(t)-x(t_0)}{t-t_0}$$

exists in E. If this is the case, its limit is the **derivative** of x at t_0, denoted by $x'(t_0)$. x is **differentiable on** Ω if it is differentiable at each $t \in \Omega$. The function $x' : t \mapsto x'(t)$ is then the **derivative** of x. If this derivative is a continuous function, then x is said to be C^1.

The following properties of the derivative are evident:

1. x is differentiable at t_0 with derivative a if and only if the function ρ defined on Ω by the equation

$$x(t) - x(t_0) = (t-t_0)a + (t-t_0)\rho(t)$$

has limit 0 as t tends to t_0;

2. if x_1 and x_2 are differentiable at t_0, then so is $x_1 + x_2$ and

$$(x_1+x_2)'(t_0) = x_1'(t_0) + x_2'(t_0);$$

3. if $T : E_1 \to E_2$ is continuous and linear and $x : \Omega \to E_1$ is differentiable at t_0, then so is $T \circ x$ and

$$(T \circ x)'(t_0) = T\big(x'(t_0)\big).$$

This concept of differentiability unifies some apparently unrelated notions of analysis as the following examples show:
I. Let x be a function from $[0, 1]$ into ℓ^∞. Then we can identify x with a bounded sequence (x_n) of functions on $[0, 1]$ (where $x(t) = (x_n(t))$. x is continuous if and only if for each t_0,

$$\lim_{h\to 0}\big\|x(t_0 + h) - x(t_0)\big\| = 0 \text{ i.e. } \lim_{h\to 0} \sup_n \big|x_n(t_0 + h) - x_n(t_0)\big| = 0$$

This means that the x_n are equicontinuous at t_0.

Similarly, one shows that x is differentiable at t_0 if and only if the following conditions hold:

a) each x_n is differentiable at t_0;

b) the sequence $(x'_n(t_0))$ is bounded;

c) the functions x_n are **uniformly differentiable** at t_0 i.e. for each $\epsilon > 0$ there is a $\delta > 0$ so that if $|h| < \delta$

$$\left|\frac{x_n(t_0 + h) - x_n(t_0)}{h} - x'_n(t_0)\right| < \epsilon$$

for each n.

II. Now let E be the space $C(J)$ (J a compact interval). A function $x : \Omega \to C(J)$ can be regarded as a function $\tilde{x}$ from $\Omega \times J$ into $\mathbf{R}$ (put $\tilde{x}(s, t) = (x(s))(t)$ for $x \in \Omega, t \in J$). Then one can check that

a) x is continuous if and only if $\tilde{x}$ continuous on $\Omega \times J$;

b) x is a C^1-function if and only if $D_1\tilde{x}(s, t)$ exists for each s, t, is continuous as a function on $\Omega \times J$ and the following condition holds: for each $s_0 \in \Omega$, the difference quotients $\dfrac{\tilde{x}(s_0 + h, t) - \tilde{x}(s_0, t)}{h}$ tend to $D_1\tilde{x}(s_0, t)$ uniformly in t (over J).

We denote by $C^1(\Omega; E)$ the space of continuously differentiable functions from Ω into E. Similarly, we can define recursively the concept of an r-**times continuously differentiable function** or C^r-**function**, by saying that x is C^r if x' exists and is C^{r-1}. The r-th derivative $x^{(r)}$ of x is then defined to be $(x')^{(r-1)}$.

Proposition 2 *(Mean value theorem) If $x \in C^1(\Omega; E)$, then*

$$\big\|x(t) - x(s)\big\| \leq |t-s| \sup\Big\{\big\|x'(u)\big\| : u \in]s,t[\Big\}$$

for $s, t \in \Omega$.

PROOF. By the Hahn-Banach theorem, there is, for each s, t, an $f \in E'$ so that $||f|| = 1$ and $f(x(t) - x(s)) = ||x(t) - x(s)||$. Then, by the classical mean value theorem applied to the scalar function $f \circ x$,

$$\begin{aligned} \big\|x(t) - x(s)\big\| &= f\big(x(t) - x(s)\big) \\ &= (t-s) f\big(x'(u_0)\big) \quad \big(u_0 \in]s,t[\big) \\ &\leq |t-s| \, ||f|| \sup\Big\{\big\|x'(u)\big\| : u \in]s,t[\Big\} \end{aligned}$$

■

Later in this chapter, we shall discuss the Bochner integral which is the analogue of the Lebesgue integral for vector-valued functions. In the meantime we introduce a very elementary integral which suffices for many purposes. We consider functions from a compact interval $I = [a, b]$ in $\mathbf{R}$ with values in E. A function $\alpha : I \to E$ is a **step function** if it has a representation $\sum a_i \chi_{I_i}$ $(a_1, \dots, a_n \in E)$ where the I_i are suitable *subintervals* (open, half-open or closed) of I. The step functions form a vector subspace St (I, E) of $\ell^\infty(I, E)$, the Banach space of bounded functions from I into E. We denote its closure therein by $R(I; E)$—the space of **regulated functions**.

We define a linear mapping $\lambda : \mathrm{St}(I; E) \to E$ by specifying that $\lambda(a_i \chi_i) = \mu(I_i) a_i$ ($\mu(I_i)$ is the length of I_i) and extending linearly (the usual difficulties concerning the well-definedness of this extension can be resolved as in the scalar case. Alternatively one can reduce the problem to the scalar one by using the Hahn-Banach theorem).

Lemma 1 *λ is a continuous linear mapping from $\mathrm{St}(I; E)$ into E and $||\lambda|| = \mu(I) = b - a$.*

PROOF. If $x \in \mathrm{St}(I; E)$ then it has a representation of the form $\sum_i a_i \chi_{I_i}$ where the I_i are disjoint. Of course, $||x|| = \max\{||a_k|| : k = 1, \dots, n\}$ and

$$\begin{aligned} \big\|\lambda(x)\big\| &= \Big\|\sum \mu(I_k) a_k\Big\| \leq \sum \mu(I_k) ||a_k|| \\ &\leq \mu(I) ||x|| \end{aligned}$$

and so $||\lambda|| \leq \mu(I)$. That the inequality is actually an equality is easy to see (consider the constant functions).

■

Hence λ has a unique extension to a continuous linear operator from the the **definite integral** of x (also written $\int_a^b x(t)dt$ or simply $\int_I x$). It is easy to check that if $x \in \mathrm{R}(I; E)$, $T \in L(E, F)$ then $T \circ x \in \mathrm{R}(I; F)$ and $\int_I T \circ x = T(\int_I x)$ (the formula holds trivially for step functions and the general result follows by approximation). We consider some examples:
I. Let x be a function from I into ℓ^∞, say

$$x(t) = \big(x_1(t), x_2(t), \dots\big).$$

Then x is regulated if and only if the (x_n) are uniformly regulated i.e. for each $\epsilon > 0$, we can find a partition $\{I_1, \dots, I_k\}$ of I so that each x_n can be approximated up to ϵ by step functions which are constant on the I_i. Then we have

$$\int x = \left(\int x_n\right)$$

Note that if x takes its values in c_0, then the above rather artificial condition is equivalent to the more natural one that each x_n be regulated.
II. Let $x : I \to C(J)$ be continuous. Then x is (of course) integrable and the integral is the classical parameterized integral

$$t \mapsto \int \tilde{x}(s, t) ds$$

($\tilde{x}(s, t) := x(s)(t)$).

Proposition 3 *(fundamental theorem of calculus) If $x \in C(I; E)$ then $X : s \mapsto \int_a^s x(t)dt$ is differentiable and $X' = x$ (X is a* **primitive** *of X).*

Proof.

$$\begin{aligned}
\big\|X(s) - X(s_0) &- (s - s_0)x(s_0)\big\| \\
&= \left\| \int_{s_0}^s x(t)dt - (s - s_0)x(s_0) \right\| = \left\| \int_{s_0}^s \big(x(t) - x(s_0)\big)dt \right\| \\
&\leq |s - s_0| \max\Big\{ \big\|x(u) - x(s_0)\big\| : u \in]s_0, s[\Big\}
\end{aligned}$$

and the expression in brackets tends to zero as s tends to s_0 (by the continuity of x).

■

As a simple corollary we have the formula

$$\int_a^b x(t)dt = X(b) - X(a)$$

for the definite integral where X is any primitive of x. For any two primitives differ by a constant.

Using the concepts of differentiation and integration of Banach space valued functions we can formulate and prove an abstract existence theorem for differential equations which unites many classical results. We consider equations of the form

$$\frac{dx}{dt} = f(t, x)$$

where the solution is a function on $\mathbf{R}$ with values in a Banach space. More precisely, U is open in $\mathbf{R} \times E$, f is a function from U into E and a C^1-function $x : I \to E$ (I an open interval in $\mathbf{R}$) is sought whereby the condition

$$(*) \quad \text{for each } t \in I, \quad \big(t, x(t)\big) \in U \text{ and } x'(t) = f\big(t, x(t)\big)$$

is to be satisfied.

If we specialize say to $E = \mathbf{R}^n$ we get the system of equations

$$\frac{dx_1}{dt} = f_1(t, x_1, \dots, x_n), \dots, \frac{dx_n}{dt} = f_n(t, x_1, \dots, x_n)$$

and so, by a standard trick, n-th degree equations of the form:

$$\frac{d^n x}{dt^n} = f(t, x, \dots, x^{(n-1)}).$$

For the special case $E = C(J) \times \mathbf{R}^n$ we get systems of equations with a parameter. Hence the following existence theorem contains several classical results as special cases:

Proposition 4 *Let $f : U \to E$ satisfy the LIPSCHITZ condition*

$$\big\| f(t_1, x_1) - f(t_2, x_2) \big\| \leq K\big(||x_1 - x_2|| + |t_1 - t_2|\big)$$

for (t_1, x_1), $(t_2, x_2) \in U$. Then for each $(t_0, x_0) \in U$ there is an $\epsilon > 0$ and a C^1-function $x :]t_0 - \epsilon, t_0 + \epsilon[\to E$ so that

$$\begin{aligned} x(t_0) &= x_0 \\ \big(t, x(t)\big) &\in U \text{ for } t \in]t_0 - \epsilon, t_0 + \epsilon[\\ x'(t) &= f\big(t, x(t)\big) \quad \big(t \in]t_0 - \epsilon, t_0 + \epsilon[\big). \end{aligned}$$

Proof. Since $||f(t,x)|| \leq ||f(t_0,x_0)|| + K(|t-t_0| + ||x-x_0||)$, f is bounded on bounded sets of U. Consider the operator

$$\pi : x \mapsto \left(t \mapsto \int_{t_0}^{t} f\big(\tau, x(\tau)\big) d\tau + x_0 \right)$$

on the space

$$\Big\{ x \in C\big([t_0-\epsilon, t_0+\epsilon], E\big) : x(t_0) = x_0 \Big\}.$$

Then if ϵ is small enough, π is a contraction (since

$$\begin{aligned} \big\|\pi(x) - \pi(y)\big\| &\leq \sup_t \left| \int_{t_0}^{t} \Big(f\big(\tau, x(\tau)\big) - f\big(\tau, y(\tau)\big) \Big) d\tau \right| \\ &\leq \epsilon K \max\Big\{ \big\|x(\tau) - y(\tau)\big\| : \tau \in [t_0-\epsilon, t_0+\epsilon]\big) \end{aligned}$$

and so has a fixed point x_0 by the Banach fixed point theorem. This x_0 is a local solution.

■

3.2 The Bochner integral

In this section we extend the Lebesgue integral to measurable Banach space valued functions on a measure space. This is the vector analogue of the Lebesgue integral and is called the **Bochner integral**. By a measure space we mean a triple (Ω, Σ, μ) where Ω is a set, Σ is a σ-algebra of subsets of Ω and μ is a finite, σ-additive non-negative measure on Σ.

Definition: Let (Ω, Σ, μ) be as above, E a normed space. A measurable E-valued step function is a function of the form

$$\sum_{i=1}^{n} \lambda_i \chi_{A_i}$$

where $\lambda_i \in \mathbf{C}$ and $A_i \in \Sigma$. A function $x : \Omega \to E$ is **measurable** if it is the pointwise limit (almost everywhere) of a sequence (x_n) of measurable step functions.

The following facts are then easy to prove:

1. The pointwise limit (almost everywhere) of a sequence of measurable functions is measurable;

2. (Egoroff's theorem) if (x_n) is a sequence of measurable functions and $x_n \to x$ pointwise almost everywhere then $x_n \to x$ almost uniformly (i.e. for every $\delta > 0$, there is an $A \in \Sigma$ with $\mu(A) < \delta$ and $x_n \to x$ uniformly in $X \setminus A$);

3. if $x : \Omega \to E$ is measurable and $T \in L(E, F)$ then $T \circ x$ is measurable;

4. if E is separable, x is measurable if and only if $x^{-1}(U) \in \Sigma$ for each open $U \subset E$.

It follows from 3. that if x is measurable, it is scalarly measurable i.e. for each $f \in E'$, $f \circ x$ is measurable. The converse is not true in general. However, surprisingly enough, the difference between scalar measurability and measurability is purely a question of the size of the range of x.

Proposition 5 *$x : \Omega \to E$ is measurable if and only if it is scalarly measurable and almost separably valued (i.e. $x(\Omega \setminus A)$ is separable for some negligible set A).*

PROOF. Suppose x is measurable. Then by the definition and Egoroff there is a sequence (x_n) of simple function and (A_n) of measurable sets with $\mu(A_n) \leq \dfrac{1}{n}$ so that $x_n \to x$ uniformly on $\Omega \setminus A_n$. Now $x(\Omega \setminus \bigcap A_n)$ is clearly separable and hence so is $x(\Omega \setminus A_n)$. Of course $\bigcap A_n$ is negligible. For the converse, we can assume that E is separable and so that it has a Schauder basis (x_n). This is because every separable space is isometrically isomorph to a subspace of a subspace with Schauder basis (e.g. $C([0,1])$). Then if (P_n) is the corresponding sequence of projection, $P_n \circ x$ is measurable and $P_n \circ x \to x$—hence x is measurable.

■

As we have seen, a function $x : \Omega \to E$ is **Bochner integrable** if and only if there is a sequence (x_n) of simple functions which converges a.e. to x and is such that for each $\epsilon > 0$, there exists an $N \in \mathbf{N}$ with $\int ||x_m - x_n|| d\mu < \epsilon$ for $m, n \geq N$. Then $(\int x_n d\mu)$ is a Cauchy sequence in E (the above integrals, being integrals of step functions, are defined in the obvious way) and we define

$$\int x d\mu = \lim \int x_n d\mu.$$

Proposition 6 *Let $x : \Omega \to E$ be measurable. Then x is Bochner integrable if and only if $\int ||x|| d\mu < \infty$.*

Proof. The necessity follows from the inequality

$$\int ||x|| d\mu \leq \int ||x - x_n|| d\mu + \int ||x_n|| d\mu$$

applied to the elements of an approximating sequence of simple functions.

For the sufficiency we can assume that the E is a space with a basis. We denote the corresponding projections once again by (P_n).

Now if $\int ||x|| d\mu < \infty$ we have that $||x - P_n \circ x||$ converges point wise to zero and can deduce that $\int ||x - P_n x|| d\mu \to 0$ by the Lebesgue theorem on dominated convergence. Now the $P_n \circ x$ take their values in finite dimensional subspaces and so can be approximated by simple functions in the L^1-norm.

■

Proposition 7 *Let $x : \Omega \to E$ be Bochner integrable, $T \in L(E, F)$. Then $T \circ x$ is Bochner integrable and*

$$\int T \circ x d\mu = T \int x d\mu.$$

Proof. The result is trivial if x is a simple function. The general result follows by continuity.

■

Proposition 8 *Suppose that $x \in L^1(\mu, E)$ and that C is a closed subset of E so that*

$$\int_A x d\mu \in \mu(A) C \quad (A \in \mathcal{A}).$$

Then $x(t) \in C$ for almost all $t \in \Omega$.

Proof. We can easily reduce to the case where E is separable. Now we show that if U is an open ball (with centre y and radius r) in $E \setminus C$, then $\mu(A) = 0$ where $A = \{t : x(t) \in U\}$. Indeed if $\mu(A) > 0$, then

$$\begin{aligned} \left\| \frac{1}{\mu(A)} \int_A x d\mu - y \right\| &= \frac{1}{\mu(A)} \left\| \int_A x d\mu - \int_A \bar{y} d\mu \right\| \\ &\leq \frac{1}{\mu(A)} \int_A ||x - y|| \leq r \end{aligned}$$

which is a contradiction (here we have used $\bar{y}$ to denote the constant function $t \mapsto y$).

Now $E \setminus C$ is a countable union of such U and so

$$\mu\{t : x(t) \in E \setminus C\} = 0.$$

■

The following Corollaries are easy consequences of this result.

Corollar 1 *Suppose that $x, y \in L^1(\mu, E)$. Then*

a) if $\int_A x d\mu = \int_A y d\mu$ for each $A \in \mathcal{A}$ then $x = y$ a.e.;

b) if $\int_A f \circ x d\mu = \int_A f \circ y d\mu$ for each f in a total subset of E' and each $A \in \mathcal{A}$ then $x = y$ almost everywhere;

c) if $|| \int_A x d\mu|| \leq k\mu(A)$ for each $A \in \mathcal{A}$ then $||x(t)|| \leq k$ for almost all t.

Proposition 9 *Let $x : \Omega \to E$ be Bochner integrable, A a measurable subset of Ω. Then*

$$\int_A x d\mu \in \mu(A)\bar{\Gamma}\big(x(A)\big)$$

(if $B \subset E, \bar{\Gamma}(B)$ is the closed convex hull of B).

Proof. If the above does not hold, then by the Hahn-Banach theorem there is an $f \in E'$ with

$$f\left(\int_A x d\mu\right) > K$$

where $K = \sup\{f(y) : y \in x(A)\}$. (To simplify the notation, we are assuming that the measure of A is one). Then

$$\int_A f \circ x d\mu > K \sup\big\{f(y) : y \in x(A)\big\}$$

which is impossible because of standard estimates for integrals of real-valued functions.

■

3.3 The Orlicz-Pettis Theorem

Using the machinery of the Bochner integral, we can give a short proof of a famous result on convergence.

Proposition 10 *(Orlicz-Pettis Theorem) Let (x_n) be a sequence in a Banach space E so that for each subsequence (x_{n_k}) there is an element x with $\sum_{k=1}^{\infty} f(x_{n_k})$ converging to $f(x)$ for each $f \in E'$. Then $\sum x_n$ converges (unconditionally) to x. (i.e. weak unconditional convergence of a series implies norm unconditional convergence).*

PROOF. We use the fact that if $f : \Omega \to E$ is Bochner integrable then the family

$$\left\{ \int_A f d\mu : A \in \mathcal{A} \right\}$$

is relatively compact in E. (Exercise! Hint if f is a measurable step function then this set is bounded and finite dimensional – hence relatively compact. For the general case use an approximation argument).

■

We let $\Omega = \{-1, 1\}^{\mathbf{N}}$ be the Cantor set with Haar-measure. We define a function $\Phi : \Omega \to E$ by defining $\Phi(\epsilon_k x_k)$ to be that element of E to which $\sum(\epsilon_k x_k)$ converges weakly. Φ is measurable, since it is clearly weakly measurable (even weakly continous). Since it is bounded (uniform boundedness theorem) it is Bochner integrable. Hence its range is norm compact. But on norm compact sets, weak and norm convergence concide – hence $\sum x_{n_k}$ converges in the norm for each subsequence.

Remark: We use here the result from "Elementare Topologie", that if (K, τ) is a compact space and τ_1 is a weaker T_2-topology on K, then $\tau = \tau_1$. In the above case τ is the norm topology and τ_1 is the weak topology i.e. the initial topology on E induced by the funcitonals of E^1.

3.4 Holomorphic E-valued functions

We now introduce the concept of (complex) differentiablitity for functions x defined on an open subset U of $\mathbf{C}$ with values in a Banach space E. This topic is not directly relevant to the calculus of variations but since it is so similar in nature to the above and is important in other branches (e.g. the spectral theory of operators on infinite dimensional spaces) we include a brief treatment here. For obvious reasons, we shall consider only complex Banach spaces in this context. It will also be convenient to use the notation U_r for the set $\{\lambda \in \mathbf{C} : |\lambda| < r\}$.

A function $x : U \to E$ is **complex-differentiable** at $\lambda \in U$ if

$$\lim_{\lambda \to \lambda_0} \frac{x(\lambda) - x(\lambda_0)}{\lambda - \lambda_0}$$

exists. Then the value of this limit is denoted by $x'(\lambda_0)$ – the **derivative** of x at λ_0. If x is differentiable at each $\lambda_0 \in U$ it is **analytic** or **holomorphic**. $H(U; E)$ denotes the vector space of E-valued analytic functions on U. As is easily seen, if $T : E \to F$ is continuous and linear then $T \circ x$ is analytic whenever x is and we have the relation: $(T \circ x)' = T \circ x'$. Just as in the scalar case, holomorphicity can be characterized using a Cauchy integral, as we now show.

Definition: Suppose that $x : U \to E$ is continuous and Γ is a smooth curve with parameterization

$$c : [a, b] \to U.$$

We then define the curvilinear integral

$$\int_\Gamma x(\lambda)d\lambda = \int_a^b x\big(c(t)\big)\dot{c}(t)dt$$

(where the integral used is the one defined at the beginning of this section).

(It is no problem to extend this definition to integrals over piecewise smooth or even rectifiable curves but the above definition will suffice for our purposes.) Then it follows that if $T : E \to F$ is continuous and linear

$$\int_\Gamma T \circ x(\lambda)d\lambda = T \int_\Gamma x(\lambda)d\lambda.$$

Proposition 11 *Let $x : U \to E$ be holomorphic and let Γ be a smooth, closed nullhomotopic curve in U. Then*

$$\int_\Gamma x(\lambda)d\lambda = 0.$$

Proof. If $f \in E'$, then the scalar function $f \circ x$ is analytic and so

$$f\left(\int_\Gamma x(\lambda)d\lambda\right) = \int_\Gamma f \circ x(\lambda)d\lambda = 0.$$

Since this holds for each $f \in E'$, the integral must be zero. Exactly as in the scalar case, this implies:

■

Corollar 2 *(Cauchy formula) Let x be as above, λ_0 a point of U and Γ a smooth closed curve with winding number 1 with respect to Γ. Then*

$$x(\lambda_0) = \frac{1}{2\pi i}\int_\Gamma \frac{x(\lambda)}{\lambda - \lambda_0}d\lambda.$$

Using the techniques we have developed, together with the classical results, it is now easy to obtain alternative characterizations of complex differentiability—in particular, the fact that it is equivalent to representability locally by Taylor series. As we shall now see it follows from the uniform boundedness theorem that $x : U \to E$ is analytic if it satisfies the apparently much weaker condition that the scalar function $f \circ x$ be holomorphic for each $f \in E'$ (weak analyticity).

Proposition 12 *For a function $x : U_r \to E$, the following are equivalent:*

1. *x is holomorphic on U_r;*
2. *for each $f \in E'$, $f \circ x$ is holomorphic on U_r;*
3. *there is a sequence (c_n) in E so that*

$$\limsup ||c_n||^{1/n} \leq 1/r$$

and $x : \lambda \mapsto \sum_{n=0}^{\infty} c_n \lambda^n$.

If these condition are satisfied, then x' is also differentiable and hence x is infinitely differentiable. Also we have $c_n = x^{(n)}(0)/n!$

PROOF. (3) implies (1): Just as in the scalar case, we can show that if (c_n) satisfies the above condition, then the given series is absolutely convergent on U_r and uniformly convergent on each compact subset. The series obtained by formal differentiation also have the same convergence properties and from this it follows that x is infinitely differentiable and the formulae for the c_n hold.

(1) implies (2) is trivial

(2) implies (3): First we note that if $\rho < r$ then $f \circ x$ is bounded on the compact set $\bar{U}_\rho$ for each $f \in E'$ and so by the uniform boundedness theorem, x is also norm-bounded on $\bar{U}_\rho$. Let

$$M_\rho := \sup\Big\{\big\|x(\lambda)\big\| : |\lambda| \leq \rho\Big\}.$$

If $f \in E'$ we define

$$c_n(f) = (f \circ x)^{(n)}(0)/n! = \frac{1}{2\pi i} \int_{\Gamma_\rho} \frac{(f \circ x)(\lambda)}{\lambda^{n+1}} d\lambda$$

where Γ_ρ is the boundary of $\bar{U}_\rho$. As is easy to see $f \mapsto c_n(f)$ is a linear form on E'.

We can estimate as follows:

$$\big|c_n(f)\big| \leq ||f|| M_\rho / \rho^n.$$

Hence c_n is a bounded form (i.e. $c_n \in E''$) and $||c_n|| \leq M_\rho/\rho^n$. Since this holds for each $\rho < r$ we have $\limsup ||c_n||^{1/n} \leq 1/r$. Now using the implication (3) implies (1) we know that

$$\tilde{x}(\lambda) := \sum_{n=0}^{\infty} c_n \lambda^n$$

defines a holomorphic function from U_r into E''. But it follows immediately form the definition of the c_n that

$$f \circ \tilde{x} = f \circ x$$

for each $f \in E'$. Hence $x = \tilde{x}$ and so x is holomorphic and each c_n actually lies in E.

■

This result has the following global form:

Proposition 13 *For a function $x : U \to E$, the following are equivalent:*

1. *x is holomorphic;*
2. *for each $f \in E'$, $f \circ x$ is holomorphic;*
3. *for each $\lambda_0 \in U$, there is a $r > 0$ and a sequence (c_n) in E so that*

$$x(\lambda) = \sum_{n=0}^{\infty} c_n (\lambda - \lambda_0)^n$$

in $\lambda_0 + U_r \subset U$ where convergence is absolute and uniform on compact subsets. x is then infinitely often differentiable.

Proposition 14 *(Morera's theorem) Let $x : U \to E$ be continuous. Then x is analytic if*

$$\int_\Gamma x(\lambda) d\lambda = 0$$

for each smooth, closed, nullhomotopic curve Γ in U.

Proof. If x satisfies the given condition, then so does $f \circ x (f \in E')$ and so $f \circ x$ is holomorphic by the classical form of Morea's theorem. Hence x is holomorphic by the above Proposition.

■

Proposition 15 *(Liouville's theorem) If $x : \mathbf{C} \to E$ is holomorphic and bounded, then it is constant.*

Proof. If x is not constant, choose $\lambda_1, \lambda_2 \in \mathbf{C}$ with $x(\lambda_1) \neq x(\lambda_2)$. There is an $f \in E'$ with

$$f\big(x(\lambda_1)\big) \neq f\big(x(\lambda_1)\big).$$

Then $f \circ x$ is a non-constant, bounded, entire function which contradicts the classical form of Liouville's theorem.

■

3.5 Differentiability of functions on Banach spaces

We now discuss the concept of differentiability for function defined on Banach spaces. The definitions for function of several variables—i.e. via local approximation by linear operators—can be carried over without any problems. The resulting concept is called **Freéchet differentiability**. We shall also discuss a weaker one—that of **Gateaux differentiability**).

In the following E, F, G etc. will be real Banach spaces. Letters such as U, V, W will denote open subsets of Banach spaces.

Fréchet differentiability A function $f : U \to F$ ($U \subset E$) is Fréchet differentiable at $x_0 \in U$ if there is a $T \in L(E, F)$ so that

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - Th}{||h||} = 0$$

(equivalently $f((x_0 + h) - f(x) = Th + \rho(h)$ where $\rho(h)/||h||$ goes to zero with h).

T is uniquely determined by this condition and is called the **(Fréchet) derivative** of f at x_0, denoted by $(Df)_{x_0}$. f is **differentiable** on U if $(Df)_x$ exists for each $x \in U$. f is a **C^1-function** on U if the function

$$Df : x \mapsto (Df)_x$$

from U into $L(E, F)$ is continuous (in symbols $f \in C^1(U; F)$).

Gateaux differentiability There is a weaker concept of differentiability which is sometimes useful—that of Gateaux differentiability. This means that the restriction of f to lines through x_0 are differentiable in the sense of 1.1 and the corresponding derivatives are continuous and linear as functions of the direction of differentiation. More precisely, $f : U \to E$ is **Gateaux differentiable** (or G-differentiable) at $x_0 \in U$ if

1. for each $h \in E$, the derivative

   $$Df(x_0, h) = \lim_{t \to 0} \frac{f(x_0 + th) - f(x_0)}{t}$$

   exists and

2. the mapping $h \mapsto Df(x_0, h)$ from E into F is continuous and linear.

Note that the limit in 1) can exists for each $h \in E$ without the mapping in 2) being linear (the standard example is

$$(s,t) \to \begin{cases} 0 & (s,t) = (0,0) \\ \frac{st^2}{s^2+t^2} & \text{otherwise} \end{cases}$$

from $\mathbf{R}^2$ into $\mathbf{R}$). Also a function can be G-differentiable at x_0 without being continuous at x_0 (and so certainly not differentiable). The standard example is

$$(s,t) \to \begin{cases} 0 & t^2 \geq s^4 \\ 1 & \text{otherwise} \end{cases}$$

Of course, if f is differentiable it is G-differentiable. A useful relation between the two concepts is established in the next Proposition.

Proposition 16 *$f : U \to F$ is differentiable at $x_0 \in U$ if it is G-differentiable there and*

$$\lim_{t \to 0} \left\| \frac{f(x_0 + th) - f(x_0)}{t} - Df(x_0, h) \right\| = 0$$

uniformly for h is the unit sphere of E.

This is just a reformulation of the definition of differentiability.

Corollar 3 *If $f : U \to E$ is G-differentiable and the function*

$$x \mapsto \big(h \mapsto Df(x,h)\big)$$

from U into $L(E,F)$ is continuous, then f is differentiable and

$$(Df)_x : h \mapsto Df(x,h) \quad (x \in U).$$

Proof. The continuity of the above mapping means that for every $\epsilon > 0$ there is a $\delta > 0$ so that

$$\big|Df(x,h) - Df(x_0,h)\big| \leq \epsilon ||h||$$

for $||x - x_0|| < \delta, h \in E$.

Now there is a $t \in [0,1]$ so that

$$\big\|f(x+h)) - f(x_0) - Df(x_0,h)\big\| = \big\|Df(x_0 + th, h) - Df(x_0,h)\big\| \leq \epsilon ||h||$$

(applying the mean value theorem to the function

$$t \mapsto f(x_0 + th)\,).$$

Thus the difference quotient converges to $Df(x_0, h)$ uniformly in h and so f is Fréchet differentiable by the above.

■

We bring some simple examples of differentiable functions:

If $T : E \to F$ is continuous and linear, then T is differentiable and $(DT)_x = T$ for each x.

If E is the one-dimensional space $\mathbf{R}$ and $U \subset \mathbf{R}$ is open, then the above concepts of differentiability both coincide with that of IIa).

If $B : E \times F \to G$ is continuous and bilinear, then B is differentiable and $(DB)_{(x,y)}$ is the linear mapping $(h, k) \to B(h, y) + B(x, k)$ on $E \times F$.

Polynomials If E, F are Banach spaces, a **homogeneous polynomial** of degree n is a mapping $Q : E \to F$ of the form

$$Q(x) = B(x, \dots, x)$$

where $B \in L^n(E, F)$. For convenience we write Bx^n for the right hand side.

Note that we can and shall assume that B is **symmetric** i.e.

$$B(x_1, \dots, x_n) = B(x_{\pi(1)}, \dots, x_{\pi(n)})$$

for each permutation in S_n. For B induces the same polynomial as its symmetrisation.

A **polynomial** is a function from E into F which is a finite sum of such homogeneous polynomials. For example if $k_i : [0, 1]^i \to \mathbf{R}$ are continuous symmetric functions, the mapping

$$\begin{aligned} x \quad \mapsto & \int_0^1 k_1(s)x(s)ds + \int_0^1\int_0^1 k_2(s_1, s_2)x(s_1)x(s_2)ds_1 ds_2 + \cdots + \\ & \int_0^1 \dots \int_0^1 k_n(s_1, \dots, s_n)x(s_1) \dots x(s_n)ds_1, \dots, ds_n \end{aligned}$$

is a polynomial on $C(I)$.

If $f : E \to F$ is a homogeneous polynomial of the form $x \mapsto Bx^n$ (B symmetric in $L^n(E, F)$) then f is differentiable and

$$(Df)_x : h \to nB(x^{n-1}, h)$$

(the last symbol denotes $B(x, x, \dots, x, h)$). The proof is based on the **binomial formula**

$$B(x + h) - B(x) = \sum_{r=0}^{n} \binom{n}{r} B(x^r, h^{n-r})$$

(where $B(x^r, h^{n-r})$ has the obvious meaning) which can be proved just as in the classical case.

Thus we have

$$f(x+h) - f(x) = nB(x^{n-1}, h) + \frac{n(n-1)}{2} B(x^{n-1}, h^2) + \dots$$

and the result follows.

For example, the derivative of the above polynomial on $C(I)$ is

$$(Df)_x : h \mapsto \int_0^1 k_1(s)ds + 2 \int_0^1\int_0^1 k_2(s_1, s_2)x(s_1)h(s_2)ds_1 ds_2 + \\ \dots + \int_0^1 \dots \int_0^1 k_n(s_1, \dots, s_n)x(s_1) \dots x(s_{n-1})h(s_n)ds_1 \dots ds_n$$

We now show that the Fréchet derivative possesses the basic properties required for efficient calculation. We begin with the chain rule and then show that suitable forms of the inverse function theorem hold.

Proposition 17 *(the chain rule) If $f : U \to V, g : V \to G$ where f is differentiable at $x \in U$ and g is differentiable at $f(x)$, then $g \circ f$ is differentiable at x and*

$$D(g \circ f)_x = (Dg)_{f(x)} \circ (Dg)_x.$$

PROOF. Let k be the function $h \mapsto f(x+h) - f(x)$ so that

$$k(h) = (Df)_x(h) + ||h||\rho_1(h)$$

where $\rho_1(h) \to 0$. Also

$$\begin{aligned} g \circ f(x+h) - g \circ f(x) &= g\big(f(x+h)\big) - g\big(f(x)\big) \\ &= g\big(f(x) + k(h)\big) - g\big(f(x)\big) \\ &= (Dg)_{f(x)}\Big(k(h) + ||k(h)||_{\rho_2}\big(k(h)\big)\Big) \end{aligned}$$

where $\rho_2(k) \to 0$ as $k \to 0$.

Then

$$g \circ f(x+h) - g \circ f(x) = (Dg)_{f(x)} \circ (Df)_x(h)$$

plus the remainder term $||h||(Df)_{f(x)}\rho_1(h) + ||k(h)||\rho_2(k(h))$ which, as it is easy to check, has the correct growth properties.

■

Using the chain rule, it is easy to prove the following results:

1. if $f, g \in C^1(U; E)$ then so does $f + g$ and $D(f+g)_x = (Df)_x + (Dg)_x$ (consider the chain
$$U \to E \times E \to E$$
$$x \mapsto \big(f(x), g(x)\big) \mapsto f(x) + g(x)\ );$$

2. if $f \in C^1(U; E)$, $g \in C^1(U; F)$ and $B : E \times F \to G$ is bilinear and continuous, then the mapping $x \mapsto B(f(x), g(x))$ is in $C^1(U; G)$ and its derivative at x is the mapping
$$h \mapsto B\big(Df_x(h), g(x)\big) + B\big(f(x), Dg_x(h)\big);$$

3. (partial derivatives) let U (resp. V) be open in E (resp. F) $f : U \times V \to G$. Then if f is differentiable at (x_0, y_0), f_1 (resp. f_2) is differentiable at x_0 (resp. y_0) where $f_1 : x \mapsto f(x, x_0)$, $f_2 : y \mapsto f(x_0, y)$ and then
$$(Df)_{(x_0,y_0)}(h, k) = (Df_1)_{x_0}(h) + (Df_2)_{y_0}(k).$$

We now prove an analogue of the classical inverse function theorem for function between Banach spaces. The proof is based on an inverse function theorem for Lipschitz functions. Recall the definition:

If (X, d) and (Y, d_1) are metric spaces, a mapping $f : X \to Y$ is **Lipschitz** if there exists a $K > 0$ so that
$$d_1\big(f(x), f(y)\big) \le Kd(x, y) \quad (x, y \in X).$$
We write $\mathrm{Lip}(f)$ for the infimum of the K which satisfy the condition. For example if E and F are Banach spaces and $T \in L(E, F)$ then T is Lipschitz and $\mathrm{Lip}(T) = ||T||$.

Proposition 18 *If E is a Banach space and $T : E \to E$ is a linear isomorphism and $f : E \to E$ is Lipschitz with $Lip(f) \le ||T^{-1}||^{-1}$ then $T + f$ is a bijection and $(T + f)^{-1}$ is Lipschitz with constant*
$$Lip(T + f) \le \big(||T^{-1}||^{-1} - Lip(f)\big)^{-1}.$$

PROOF. If $x \in E$, then

$$||z|| = ||(T^{-1} \circ T)(z)|| \leq ||T^{-1}|| \, ||Tz||$$

and so $||Tz|| \geq ||T^{-1}||^{-1}||z||$.

Hence we can estimate

$$\begin{aligned} \big(||T^{-1}||^{-1} - \mathrm{Lip}(f)\big) &= ||T^{-1}||^{-1}||x-y|| - \mathrm{Lip}(f)||x-y|| \\ &\leq \big\|T(x-y)\big\| - \big\|f(x) - f(y)\big\| \\ &\leq \big\|(T+f)x - (T+f)y\big\| \end{aligned}$$

Thus $T + f$ is injective and its inverse from $(T + f)E$ into E is Lipschitz with constant at most $(||T^{-1}||^{-1} - \mathrm{Lip}(f))^{-1}$. We now show that $T + f$ is surjective: take $y \in E$ and let $x_0 := T^{-1}(y)$. We are looking for an $h \in E$ so that

$$(T + f)(x_0 + h) = y \text{ i.e. } Th + f(x_0 + h) = 0$$

i.e. h is a fixed point of the mapping

$$\rho : h \mapsto -T^{-1}(x_0 + h).$$

Now ρ is easily seen to be a contraction and so it has a fixed point which gives the result. The inverse function theorem says that if the derivative of a function at a given point x_0 is an isomorphism, then f is invertible in a neighborhood of this point. Now by the very definition of the derivative, f (up to a constant) is a small perturbation of its derivative in a neighborhood of x_0 and so satisfies the conditions of Proposition 19. However, in order to apply Proposition 19 we must construct a function which is defined on all of E and Lipschitz there, but agrees with f in a neighborhood of x. To do this we use a "bell function" i.e. a function $\rho : \mathbf{R} \to [0, 1]$ which is infinitely differentiable, is equal to 1 on $[-1, 1]$ and equal to 0 outside of $[-2, 2]$.

Lemma 2 *Let U be an open neighborhood of zero in the Banach space E, $f \in C^1(U; E)$, with $f(0) = 0 = (Df)_0$. Then for every $\epsilon > 0$, there is an open neighborhood $V \subset U$ of zero and $\tilde{f} : E \to E$ so that*

1. $\tilde{f}\big|_V = f\big|_V$;
2. *$\tilde{f}$ is bounded;*
3. *$\tilde{f}$ is Lipschitz with Lip $(f) < \epsilon$.*

PROOF. Since Df is continuous, there is an $\alpha > 0$ so that $2\alpha B_E \subset U$ and

$$\|Df(x)\| < \frac{\epsilon}{1 + 2||\rho'||}$$

$(||x|| \leq 2\alpha)$ where $||\rho'||$ is the supremum norm of ρ'. Let $\tilde{f}$ be the mapping $x \to (\rho(\frac{||x||}{\alpha})f(x)$ for $||x|| \leq 2\alpha$ and $\tilde{f}(x) = 0$ outside of $2\alpha B_E$. Clearly $\tilde{f} = f$ on αB_E. Also $\tilde{f}$ is bounded since f is bounded on $2B_E$ (by the mean value theorem). To check the Lipschitz condition, it is sufficient to consider x, y in $2\alpha B_E$. Then

$$\begin{aligned}
|\tilde{f}(x) - \tilde{f}(y)| &= \left|\rho\Big(\frac{||x||}{\alpha}\Big)f(x) - \rho\Big(\frac{||y||}{\alpha}\Big)f(y)\right| \\
&\leq \left|\rho\Big(\frac{||x||}{\alpha}\Big) - \rho\Big(\frac{||y||}{\alpha}\Big)\right| \|f(x)\| + \rho\Big(\frac{||y||}{\alpha}\Big)\|f(x) - f(y)\| \\
&\leq \frac{1}{\alpha}||\rho'||\,||x-y||\|f(x)\| + \rho\Big(\frac{||y||}{\alpha}\Big) \sup_{z \in 2\alpha B(E)} \|Df(z)\|||x-y|| \\
&\leq \frac{1}{\alpha}||\rho'||||x-y|| \sup\|Df(z)\|||x|| + \sup\|Df(z)\|||x-y|| \\
&\leq \epsilon||x-y||
\end{aligned}$$

■

Using this we can prove the following version of the inverse function theorem:

Proposition 19 *Let U be an open neighborhood of zero in E, $f \in C^1(U; E)$ with $f(0) = 0$. If $(Df)_0$ is an isomorphism, then f is local diffeomorphism i.e. there exists a neighborhood $V \subset U$ of zero so that $f(V)$ is open and $f\big|_V : V \to f(V)$ is a C^1-isomorphism.*

PROOF. Put $g = f - (Df)_0$. We can apply the above Lemma to get a neighborhood V of zero and a $\tilde{g} : E \to E$ so that

1. $\tilde{g} = g$ on V;
2. $\tilde{g}$ is bounded and Lipschitz continuous with

$$\mathrm{Lip}(g) < \left\|(Df)_0^{-1}\right\|^{-1}$$

Then by the Lipschitz inverse function theorem $(Df)_0 + \tilde{g}$ is a homomorphism of E with Lipschitz continuous inverse. Hence $f\big|_V = (\tilde{g} + (Df)_0)\big|_V$ is a bijection from V onto an open set. We show that its inverse is differentiable. (From this it follows easily that the inverse mapping is C^1.)

Choose $y, y+k$ in $f(V)$. Then if $x = f^{-1}(y)$,

$$f^{-1}(y+k) - f^{-1}(y) - (Df)_x^{-1}k = (Df)_x^{-1}\Big((Df)_x\big(f^{-1}(y+k) - f^{-1}(y)\big) - k\Big).$$

Now we put $h = f^{-1}(y+k) - f^{-1}(y)$ i.e. $k = f(x+h) - f(x)$. Then the right hand side is

$$(Df)_x h - \big(f(x+h) - f(x)\big)$$

and this is $||h||\rho(h)$ where $\rho(h) \to 0$ as $||h|| \to 0$. Now since f and f^{-1} are Lipschitz, $||h|| \to 0$ if and only if $||k|| \to 0$ and so f^{-1} is differentiable at x and

$$(Df^{-1})_x = \big((Df)_x\big)^{-1}.$$

■

As in the finite dimensional case we can deduce the following corollary:

Proposition 20 *(implicit function theorem) Let U be an open neighborhood of (x_0, y_0) in $E \times F$, $f : U \to G$ a C^1-function with $f(x_0 - y_0) = 0$. Suppose that $(D_2 f)_{(x_0,y_0)}$; $F \to G$ is an isomorphism, where $(D_2 f)_{(x_0,y_0)}$ is the derivative of the function*

$$y \mapsto f(x_0, y)$$

at y_0.

Then there are open sets $W \subset E$, $W' \subset U$ with $x_0 \in W$, $(x_0, y_0) \in W'$ and a C^1-mapping $g : W \to F$ so that for $x, y \in W'$,

$$f(x,y) = 0 \text{ if and only if } x \in W \text{ and } y = g(x).$$

PROOF. Write π for the C^1-mapping

$$(x,y) \mapsto \big(x, f(x,y)\big).$$

Then the derivative $(D\pi)_{(x_0,y_0)}$ is the operator

$$\begin{bmatrix} \mathrm{Id}_E & 0 \\ (D_1 f)_{(x_0,y_0)} & (D_2 f)_{(x_0,y_0)} \end{bmatrix}$$

Here we are using an obvious matrix notation to describe mappings from $E \times F$ into $E \times F$ (i.e. the matrix

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$$

denote the maps

$$(x, y) \mapsto (T_{11}x + T_{12}y, T_{21}x + T_{22}y)$$

where $T_{11} \in L(E, E)$, $T_{12} \in L(F, E)$, $T_{21} \in L(E, F)$, $T_{22} \in L(F, F)$).

$(D_1 f)_{(x_0, y_0)}$ is the derivative (at x_0) of the mapping

$$x \mapsto f(x, y_0).$$

Now $(D\phi)_{(x_0,y_0)}$ is an isomorphism and so there is an open neighborhood $W' \subset U$ of (x_0, y_0) so that $\phi|_{W'}$ is a diffeomorphism. Denote its inverse by ψ. Then if $(x, y) \in W'$ $f(x, y) = 0$ if and only if $(x, f(x, y)) = (x, 0)$ or, equivalently, $\phi(x, y) = (x, 0)$ i.e. $(x, y) = \psi(x, 0)$.

So the required g is the mapping

$$x \mapsto \pi\big(\psi(x, 0)\big)$$

where π is the natural projection from $E \times F$ and g is defined on

$$W := \big\{x : (x, 0) \in \phi(W')\big\}.$$

We now consider higher derivatives. In contrast to the situation of function on $\mathbf{R}$ a new complication arises due to the fact that the derivative of a function f from U into F takes its values not in F but in the Banach space $L(E, F)$. Hence, the derivative of Df, if it exists, will have its values in $L(E, L(E, f))$. By the time we get to say the fourth derivative the range space is a rather complicated nested operator space. Fortunately, the above space is naturally isometric to $L^2(E, F)$. More generally, we have

$$L^k(E, F) \cong L\big(E, L^{k-1}(E, F)\big).$$

This puts us in the position to define recursively the notion of a C^r-function and its higher derivatives: $f : U \to F$ is **of class** C^r (or $f \in C^r(U; E)$) if f is differentiable and $Df \in C^{r-1}(U; L(E, F))$. We then define $D^r f$ by the equation $(D^r f)_x(h_1, \dots, h_r) = (D(D^{r-1} f)_x(h_1)(h_2, \dots, h_r)$.

■

Proposition 21 *(Taylor's theorem) Let $f : U \to E$ be a C^r-function and let $x \in U, h \in E$ be such that $\{x + th : t \in [0, 1]\} \in U$. Then*

$$f(x + h) = f(x) + (Df)_x(h) + \cdots + \frac{1}{(r-1)!}(D^{r-1}f)_x h^{r-1} + R_r(h)$$

where the remainder term R_r is

$$\int_0^1 \frac{(1-t)^{r-1}}{(r-1)!}(D^r f)_{x+th}(h^r)dt$$

and so satisfies the growth property $\lim_{||h|| \to 0} \dfrac{R_r(h)}{||h||^r} = 0$.

PROOF. Let g be the function $t \mapsto f(x+th)$. Then

$$g^{(k)}(t) = (D^k f)_{x+th} h^k.$$

(Exercise.) Now

$$\frac{d}{dt}\left(g(t) + (1-t)g'(t) + \cdots + \frac{1}{(r-1)!}(1-t)^{n-1}g^{(r-1)}(t)\right) = \frac{1}{(r-1)!}(1-t)^{r-1}g^{(r)}(t).$$

Integrating both sides from 0 to 1 and substituting for the terms of the form $g^{(k)}(t)$ gives the result.

■

As in the classical case, the first and second derivatives give information on the extrema of functionals on Banach spaces:

Proposition 22 *Let $f : U \to \mathbf{R}$ ($U \subset E$) be a C^2-function, x_0 a point in U. Then*

1. *if $f(x_0)$ is a local minimum or maximum for f, $(Df)_{x_0} = 0$;*
2. *if $(Df)_{x_0} = 0$ and the second derivative $(D^2 f)_{x_0}$ satisfies the condition: there is a $k > 0$ so that $(D^2 f)_{x_0}(h^2) \geq k||h||^2$ (resp. $(D^2 f)_{x_0}(h^2) \leq -k||h||^2$), then $f(x_0)$ is a local minimum (maximum).*

The proof is exactly as in the finite dimensional case.

We now return to the main subject:

4 The calculus of variations

4.1 The Euler equations – application

In this section we shall use the above theory to derive the Euler equation for a variety of problems of the calculus of variations. We begin with a simple case.

Example A: Let L be a smooth function of three variables (t, x, z) we calculate the derivative of the functional:

$$I(c) = \int_{t_0}^{t_1} L\big(t, c(t), \dot{c}(t)\big)dt.$$

We assume that L is analytic in the three variables t, x, z. Then I is continuous (us a functional say on $E = C^1([t_0, t_1])$) and so, in order to show that it is analytic, it suffices to demonstrate that its restrictions to one dimensional affine subspaces of E are analytic. Also, in order to calculate the derivative $(DI)_c(h)$ at h it suffices to calculate $\lim_{s\to 0} \frac{1}{s}[I(c+sh)-I(c)]$. But

$$\begin{aligned}\frac{1}{s}\big[I(c+sh)-I(c)\big] &= \frac{1}{s}\left[\int_{t_0}^{t_1}\Big\{L\big(t,c(t)+sh(t),\dot c(t)+s\dot h(t)\big)-L\big(t,c(t),\dot c(t)\big)\Big\}dt\right]\\ &= \frac{1}{s}\left[\int_{t_0}^{t_1}\big[sh(t)L_x(t,c,\dot c)+sh'(t)L_z(t,c,\dot c)\big]dt+o(s)\right]\\ &\to \int_{t_0}^{t_1}\big[h(t)L_x(t,c,\dot c)+h'L_z(t,c,\dot c)\big]dt\end{aligned}$$

This means that $(DI)_c$ is the linear form

$$h\mapsto \int_{t_0}^{t_1} h(t)L_x\big(t,c(t),\dot c(t)\big)dt+\int_{t_0}^{t_1} h'(t)L_z\big(t,c(t),\dot c(t)\big)dt.$$

Hence the vanishing of the derivative at x means that this integral must vanish for each $h\in C^1([t_0,t_1])$.

Problems with fixed endpoints: In many concrete problems, we wish to specify a maximum or minimum of the functional not on the whole Banach space but on an affine subspace (i.e. a translate of a subspace). Thus in the above case we often have a problem **with fixed endpoints** that is values x_0 and x_1 are given and we consider the extremal values of the functional on the subspace

$$\begin{aligned}E_1 &= \Big\{c\in C^1\big([t_0,t_1]\big): c(t_0)=x_0 \quad\text{and}\quad c(t_1)=x_1\Big\}\\ &= c+E_0\end{aligned}$$

where $c\in E_1$ and E_0 is the corresponding space with the homogeneous conditions $c(t_0)=0=c(t_1)$. In the order to calculate the derivatives we compute exactly as above, but using test functions $h\in E_0$. In this case the final result can be simplified, using integration by parts, as follows:

$$\begin{aligned}(DI)_c(h) &= \int_{t_0}^{t} h(t)L_x\big(t,c(t),\dot c(t)\big)dt-\int_{t_0}^{t_1} h(t)\frac{d}{dt}L_z\big(t,c(t),\dot c(t)\big)dt\\ &= \int_{t_0}^{t} h(t)\Big[L_x\big(t,c(t),\dot c(t)\big)-\frac{d}{dt}L_z\big(t,c(t),\dot c(t)\big)\Big]dt\end{aligned}$$

Since this holds for each $h\in E_0$ we have the **necessary** condition:

$$L_x\big(t,c(t),\dot c(t)\big)-\frac{d}{dt}L_z\big(t,c(t),\dot c(t)\big)=0$$

for c to be an extremum of I. This is known as Euler's equation (it is an ordinary differential equation for c – of the most general form i.e. implicit and non-linear). In a similar manner we have:
B. The Euler equations for the functional:

$$I(c) = \int_{t_0}^{t_1} L\big(t, c_1, \ldots, c_n(t), \dot{c}_1(t), \ldots, \dot{c}_n(t)\big)dt$$

with kernel L a smooth function of the variables $t, x_1, \ldots, x_n, z_1, \ldots, z_n$ are:

$$\frac{d}{dt} L_{z_k}\big(t, c(t), \dot{c}(t)\big) - L_{x_k}\big(t, c(t), \dot{c}(t)\big) = 0$$

$(k = 1, \ldots, n)$.
C. The Euler equations for the functional:

$$I(c) = \int_{t_0}^{t_1} L\big(t, c(t), \dot{c}(t), \ldots, c^{(n)}(t)\big)dt$$

with smooth kernel $L = L(t, x, z_1, \ldots, z_n)$ (and boundary conditions

$$\begin{aligned} c^{(k)}(t_0) &= \alpha_k \quad (k = 0, \ldots, n-1) \\ c^k(t_1) &= \beta_k \quad (k = 0, \ldots, n-1) \end{aligned}$$

is

$$L_x - \frac{d}{dt} L_z + \frac{d^2}{dt^2} L_{z_2} - \cdots + (-1)^n \frac{d^n}{dt^n} L_{z_n} = 0$$

D. The Euler equation for functions of several variables. Consider the functional

$$I(\phi) = \iint_U L\left(u, v, \phi(u,v), \frac{\partial \phi}{\partial u}, \frac{\partial \phi}{\partial v}\right) du\, dv$$

with boundary conditions

$$\phi(u,v) = f(u,v) \quad \text{on} \quad \partial U.$$

$$\begin{aligned} \frac{1}{s}\big\{I(\phi + sh) - I(\phi)\big\} &= \iint_U \big[L(u, v, \phi + sh, \phi_1 + sh_1, \phi_2 + sh_2) - L(u, v, \phi, \phi_1, \phi_2)\big] du\, dv \\ &= \iint hL_x du\, dv + \iint \phi_1 h_1 L_z du\, dv + \int \phi_2 h_2 L_w du\, dv \end{aligned}$$

But, by Green's theorem:

$$\iint_U h_1 L_z du\, dv + \iint_U h_2 L_w du\, dv = \int_{\partial U} h(L_z dv - L_w d_u) - \iint_U h\left(\frac{\partial}{\partial x} L_z + \frac{\partial}{\partial y} L_w\right) du\, dv$$

(Consider the differential form

$\emptyset = h(u,v)L_z\big(u,v,\phi(u,v),\phi_1(u,v),\phi_2(u,v)\big)dv - h(u,v)L_w\big(u,v,\phi(u,v),\phi_1(u,v)\phi_2(u,v)\big)du$

$d\emptyset = (h_1 L_z + \frac{\partial}{\partial u}L_z) + (h_2 L_w + \frac{\partial}{\partial v}L_w))$. Hence $(DI)_\phi(h) = \iint h\{L_x - \frac{\partial}{\partial u}L_z - \frac{\partial}{\partial v}L_w\}du\,dv$ and so the Euler equation for

$$I(\phi) = \iint_U L(u,v,\phi,\phi_1,\phi_2)du\,dv$$

with smooth kernel $L = L(u,v,x,z,w)$ is

$$\frac{\partial}{\partial u}L_z + \frac{\partial}{\partial u}L_w - L_x = 0$$

In the case where higher derivatives occur, for example second derivatives, we have the following form for the Euler equations: In this case we have

$$L = L(u,v,x,z_1,z_2,z_2,z_1w_1,w_2).$$

(z_s is a place-holder for the second partial derivative with respect to u, z_1w_1 for the mixed partial derivative (*not* the product of z_1 and w_1). Then the Euler equation is:

$$L_x - \frac{\partial}{\partial u}L_{z_1} - \frac{\partial}{\partial v}L_{w_1} + \frac{\partial^2}{\partial u^2}L_{z_2} + 2\frac{\partial^2}{\partial u \partial v}L_{z_1w_1} + \frac{\partial^2}{\partial v^2}L_O w_2 = 0.$$

For example, if

$$I(\phi) = \int\int \frac{1}{2}\left[\left(\frac{\partial^2\phi}{\partial u^2}\right)^2 + \frac{\partial^2\phi}{\partial u^2}\frac{\partial^2\phi}{\partial v^2} + \left(\frac{\partial^2\phi}{\partial v^2}\right)^2\right] ldudv,$$

then in this case Euler's equation is

$$\Delta\Delta\phi = 0.$$

Examples: I. The shortest line between (t_0,α_0) and (t_0,α_1), $(\alpha_1 > \alpha_0)$ in $\mathbf{R}^2$: Here

$$I(c) = \int_{t_0}^{t_1} \sqrt{1+\dot{c}(t)^2}dt$$

$L(t,x,z) = \sqrt{1+z^2}$
$L_x = 0, \quad L_z = \dfrac{z}{\sqrt{1+z^2}}$. Euler's equation:

$$\frac{d}{dt}\frac{\dot{c}(t)}{\sqrt{1+\dot{c}(t)^2}} = 0$$

Hence$\dfrac{\dot{c}(t)}{\sqrt{1+\dot{c}(t)^2}}$ and so also $\dfrac{\dot{c}(t)^2}{1+\dot{c}(t)^2}$ is constant. From this it follows easily that $\dot{c}(t)$ is constant i.e. $c(t)$ has the form $at+b$ for suitable a, b which can be determined from the boundary conditions.

II. As above in 3 dimensions:

$$I(c) = \int_{t_0}^{t_1} \sqrt{1+\dot{c}_1(t)^2+\dot{c}_2(t)^2}dt$$

$$L(t,x,z) = \sqrt{1+z_1^2+z_2^2}$$

Euler equations:

$$\frac{d}{dt}\frac{\dot{c}_1(t)}{\sqrt{1+\dot{c}_1(t)^2+\dot{c}_2(t)^2}} = 0$$

$$\frac{d}{dt}\frac{\dot{c}_2(t)}{\sqrt{1+\dot{c}_1(t)+\dot{c}_2(t)^2}} = 0$$

Once again, one can deduce that c is affine.

III. Surfaces of revolution with minimal area:

$$I(c) \int_{t_0}^{t_1} c(t)\sqrt{1+\dot{c}(t)^2}dt \quad c(t_0) = x_0, c(t_1) = x_1.$$

$L(t,x,z) = x\sqrt{1+z^2}$

$L_x = \sqrt{1+z^2}, L_z = \dfrac{xt}{\sqrt{1+z^2}}$

Euler's equation: $\dfrac{d}{dt}\dfrac{c(t)\dot{c}(t)}{\sqrt{1+\dot{c}(t)^2}} = \sqrt{1+\dot{c}(t)^2}$

This simplifies to the equation $c(t)\ddot{c}(t) = 1+\dot{c}(t)^2$ which tacitly assumes that c is C^2. However this follows from the form of the equation. For if $y = L_z(t,x,z) = \dfrac{xz}{\sqrt{1+z^2}}$, simple algebra shows that $z = \dfrac{y}{\sqrt{x^2-y^2}}$.

Hence if $d(t) = L_z(t, c(t), \dot{c}(t))$ we have

$$d(t) = \frac{c(t)\dot{c}(t)}{\sqrt{1+\dot{c}(t)^2}}$$

and so $\dot{c}(t) = \dfrac{d(t)}{\sqrt{c(t)^2-d(t)^2}}$.

Hence c is C^2 (since d is C^1).

The original Euler's equation is equivalent to the system:

$$\begin{aligned}\dot{c}(t) &= \frac{d(t)}{\sqrt{c(t)^2 - d(t)^2}} \\ \dot{c}(t) &= \frac{c(t)}{\sqrt{c(t)^2 - d(t)^2}}\end{aligned}$$

By standard methods one obtains the explicit solution

$$c(t) = b\cosh\frac{t - t_0}{b} \quad (b > 0).$$

The constants b and t_0 are determined by the boundary conditions. (Note that such a solution cannot be found for arbitrary boundary conditions.)
IV. Geodetics in $\mathbf{R}^2$ for the metric tensor

$$\begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}$$

(where the g's are smooth functions of two variables with $g_{11} > 0$, $g_{11}g_{12} - g_{12}^2 > 0$). In this case:

$$I(c) = \int_{t_0}^{t} \sqrt{g_{11}(t, c(t)) + 2g_{12}(t, c(t))\dot{c}(t) + g_{22}(t, c(t))\dot{c}^2(t)dt}$$

$$L(t, x, z) = \sqrt{g_{11}(t, x) + 2g_{12}(t, x)z + g_{22}(t, x)z^2}$$

The Euler equation is then

$$\begin{aligned}\frac{d}{dt} \quad & \frac{g_{12}(t, c(t)) + g_{22}(t.c(t))\dot{c}(t)}{\sqrt{g_{11}(t, c(t)) + 2g_{12}(t, c(t))\dot{c}(t) + g_{22}(t, c(t))\dot{c}(t)^2}} \\ - \quad & \frac{\frac{\partial g_{11}}{\partial x}(t, c(t)) + 2\frac{\partial g_{12}}{\partial x}(t, c(t))\dot{c}(t) + \frac{\partial^2 g_{22}}{\partial x}(t, c(t))\dot{c}(t)^2}{2\sqrt{g_{11}(t, c(t)) + 2g_{12}(t, c(t))\dot{c}(t) + g_{22}(t, c(t))\dot{c}(t)^2}}\end{aligned}$$

Sometimes written as

$$\frac{d}{dt}\frac{F + G\dot{c}}{\sqrt{E + 2F\dot{c} + G\dot{c}^2}} - \frac{E_2 + 2F_2\dot{c} + G_2\dot{c}^2}{2\sqrt{E + 2F\dot{c} + H\dot{c}^2}} = 0$$

(with $E = g_{11}, F = g_{12} = g_{21}, G = g_{22}$). We can consider the same problem in 3 dimensions.

Here $\begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix}$ is a positive definite matrix whose entries are smooth functions of those variables (t, x_1, x_2).

In this case:

$$I(c) = \sqrt{g_{11} + 2g_{12}\dot{c}_2 + 2g_{13}c_3 + g_{22}\dot{c}_2^2 + 2g_{23}\dot{c}_2\dot{c}_3 + g_{33}\dot{c}_3^2}$$

$$L(t,x,z) = \sqrt{g_{11}(t,x_2,x_3) + 2g_{12}z_2 + 2g_{13}z_3 + g_{22}\dot{c}_2^2 + 2g_{23}z_2z_3 + g_{33}z_3^2}$$

Euler's equation:

$$\frac{d}{dt}\left(\frac{g_{12} + g_{22}\dot{c}_2 + g_{23}\dot{c}_3}{L}\right) - \frac{1}{2L}\left(\frac{\partial}{\partial x_2}g_{11} + 2\frac{\partial}{\partial x_2}g_{12}\dot{c}_2 + \qquad\right) = 0$$

V. Fermat's principle:

$$I(c) = \int_{t_0}^{t_1} \mu\big(t, c_1(t), c_2(t)\big)\sqrt{1 + \dot{c}_1(t)^2 + \dot{c}_2(t)^2}dt \quad \mu = \mu(t, x_1, x_2)$$

the coefficient of refraction)

$$L(t, x_1, x_2, z_1, z_2) = \mu(t,x)\sqrt{z_1^2 + z_2^2}$$

$$L_{x_1} = \frac{\partial\mu}{\partial x_1}\sqrt{1 + z_1^2 + z_2^2} \qquad L_{z_1} = \frac{\mu z_1}{\sqrt{1 + z_1^2 + z_2^2}}$$

$$L_{x_2} = \frac{\partial\mu}{\partial x_2}\sqrt{1 + z_1^2 + z_2^2} \qquad L_{z_2} = \frac{\mu z_2}{\sqrt{1 + z_1^2 + z_2^2}}$$

Equations

$$\frac{d}{dt}\mu\big(t, c(t)\big)\frac{\dot{c}_k(t)}{\sqrt{1 + \dot{c}_1(t)^2 + \dot{c}_2(t)^2}} = \mu_{x_k}\big(t, c(t)\big)\sqrt{1 + \dot{c}_1(t)^2 + \dot{c}_2(t)^2} \quad (k = 1, 2).$$

VI. Dirichlet's problem. Recall that we minimized the functional

$$I(\phi) = \iint \left[\left(\frac{\partial\phi}{\partial u}\right)^2 + \left(\frac{\partial\phi}{\partial v}\right)^2\right] du\, dv$$

Thus $L = \frac{1}{2}(z^2 + w^2)$ and Euler's equation is $\Delta\phi = \dfrac{\partial^2\phi}{\partial x^2} + \dfrac{\partial^2\phi}{\partial y^2} = 0$

Similarly if

$$I(\phi) = \iint \frac{1}{2}\left[\left(\frac{\partial^2\phi}{\partial u^2}\right)^2 + 2\left(\frac{\partial^2\phi}{\partial u^2}\right)\left(\frac{\partial^2\phi}{\partial y^2}\right) + \left(\frac{\partial^2\phi}{\partial v^2}\right)^2\right] du\, dv$$

then the Euler equation is

$$\Delta\Delta\phi = \frac{\partial^4\phi}{\partial u^4} + 2\frac{\partial^4\phi}{\partial u^2\partial v^2} + \frac{\partial^4\phi}{\partial v^4} = 0$$

VII. Minimal surfaces. Here

$$I(\phi) = \int \sqrt{1+\phi_u^2+\phi_v^2}$$

$L(u,v,x,z,w) = \sqrt{1+z^2+w^2}$

Euler's equation is

$$\phi_{uu}(1+\phi_v^2) - 2\phi_{uv}\phi_u\phi_v + \phi_{vv}(1+\phi_u^2) = 0$$

This can be interpreted geometrically as stating that the mean curvative vanishes (cf. Vorlesung Differentialgeometrie).

Remark: Consider the implicit curves

$$\{f = c\}$$

Then their curvatures are given by the formula:

$$\kappa = \frac{-f_{11}f_2^2 + 2f_1f_2f_{12} - f_{22}f_1^2}{|\mathrm{grad}f|^3}$$

(cf. Differentialgeometrie) and so the equation hat the form:

$$f_{11} + f_{22} = \kappa|\mathrm{grad}f|^3$$

Hence if the level curves are straight lines, then

$$f_{11} + f_{22} = 0$$

i.e. f is harmonic.

Example: f is linear. Then it is a plane.

f is $a\theta(x,y)+b$. Then the graph is the helicoid

$$f(u,v) = (u\cos v, u\sin v, av)$$

Scherk's example: Scherk solved this equation for f of the form $f(x,y) = g(x) + h(y)$

This leads to the equation

$$\frac{-g''(x)}{1+g^1(x)^2} = \frac{h''(y)}{1+h^1(y)^2} = \text{const}$$

Solutions: $g(x) = \dfrac{1}{a}\ln\cos ax, h(x) = -\dfrac{1}{a}\ln\cos ay$

$$f(u,v) = \frac{1}{a}\ln\left(\frac{\cos au}{\cos av}\right)$$

(Catalan showed that the only ruled surfaces which are minimal are the helicoid or the plane).

Surfaces of revolution: We rotate the curve $(t, c(t), 0)$ about the x-axis.

$$\phi(u,v) = \big(u, c(u)\cos v, c(u)\sin v\big).$$

The equation $H = 0$ then becomes

$$-c\ddot{c} + 1 + \dot{c}^2 = 0$$

with solution

$$c(t) = a\cosh\left(\frac{t}{a} + b\right).$$

4.2 The Weierstraß representation:

We now show how to construct a large class of minimal surfaces by using results from function theory. We consider a parameterized surface M in $\mathbf{R}^3$ i.e. M is the image of a smooth function $\phi = (\phi^1, \phi^2, \phi^3)$ from $U(\subset \mathbf{R}^2)$ into $\mathbf{R}^3$. We denote by $\mathbf{N}$ the normal vector to the curve i.e. $\mathbf{N}(u,v) = \dfrac{\phi_1(u,v)\times\phi_2(u,v)}{|\phi_1(u,v)\times\phi_2(u,v)|}$. We suppose that the parameterization is conformal (or isothermal). This means that there is a scalar function $\lambda(u,v)$ so that $(\phi_1|\phi_1) = \lambda(u,v) = (\phi_2|\phi_2)$, $(\phi_1|\phi_2) = 0$. Then we have the equation $\Delta\phi = \phi_{11} + \phi_{22} = 2\lambda^2 H\mathbf{N}$ where H is the mean curvature of the surface. From this it follows that the surface is minimal iff $\Delta\phi = 0$ i.e. the components ϕ^1, ϕ^2, ϕ^3 are all harmonic. Then if we introduce the functions f_1, f_2, f_3 where $f_i = \dfrac{\partial\phi^i}{\partial z}$, these are analytic functions.

Remark: $\dfrac{\partial}{\partial z}$ is the operator $\dfrac{1}{2}(\dfrac{\partial}{\partial u} - i\dfrac{\partial}{\partial v})$.

A simple calculation shows that if ϕ is harmonic, then $\dfrac{\partial}{\partial z}\phi$ satisfies the Cauchy-Riemann equations and so is holomorphic). Furthermore we have the relations by

$$f_1^2 + f_2^2 + f_2^2 = 0$$

between the f's. For

$$\begin{aligned} f_1^2 + f_2^2 + f_3^2 &= \frac{1}{4}\sum_i \left(\frac{\partial\phi^i}{\partial u} - i\frac{\partial\phi^i}{\partial v}\right)^2 \\ &= \frac{1}{4}\sum_i \left(\frac{\partial\phi^i}{\partial u}\right)^2 - \sum_i \left(\frac{\partial\phi^i}{\partial v}\right)^2 - 2i\sum_i \frac{\partial\phi^i}{\partial u}\frac{\partial\phi^i}{\partial v} \\ &= \frac{1}{4}\left[|\phi_u|^2 - |\phi_v|^2 - 2i(\phi_u|\phi_v)\right] = 0 \end{aligned}$$

A similar calculation shows that

$$|f_1|^2 + |f_2|^2 + |f_3|^2 = 2\lambda^2 > 0$$

Conversely, we have the following. Suppose that we have three holomorphic functions (f_1, f_2, f_3) on U so that

a) $f_1^2 + f_2^2 + f_3^2 = 0$

b) $|f_1|^2 + |f_2|^2 + |f_2|^2 \neq 0$

c) each f_i has a primitive F_i.

Then the function $\phi = (\phi^1, \phi^2, \phi^3)$ where $\phi^i = \Re F_i$ is a parameterization of a minimal surface.

Examples. I. The helicoid. Here

$$f_1 = \cosh z, f_2 = -i\sinh z, f_3 = -i.$$

$$\begin{aligned} \phi^1 &= \Re\sinh z = \cos v \sinh u \\ \phi^2 &= \Re(-i\cosh z + i) = \sin v \sinh u \\ \phi^3 &= \Re(-iz) = v \end{aligned}$$

This is the helicoid as one sees by making the change of variable $t = \sinh u$.

II. The Catenoid. Here we take

$$f_1 = \sinh z, \quad f_2 = i\cosh z, \quad f_3 = 1$$

$$\begin{aligned} \phi^1 &= \cos v \cos u - 1 \\ \phi^2 &= \sin v \cosh u \\ \phi^3 &= u \end{aligned}$$

This is a parameterization of the surface of rotation of the catenary.
III. Scherk's surface. Take

$$\begin{aligned} f_1(z) &= \frac{i}{z+i} - \frac{i}{z-i} = \frac{2}{1+z^2} \\ f_2(z) &= \frac{i}{z+1} - \frac{i}{z-1} = \frac{2i}{1-z^2} \\ f_3(z) &= \frac{4z}{1-z^4} = \frac{2z}{z^2+1} - \frac{2z}{z^2-1} \end{aligned}$$

(defined on $U = \{|z| < 1\}$). Then

$$\begin{aligned} \phi^1 &= \arg \frac{z+i}{z-i} \\ \phi^2 &= \arg \frac{z+1}{z-1} \\ \phi^3 &= \ln \left| \frac{z^2+1}{z^2-1} \right| \end{aligned}$$

IV. Enneper's surface:

$$\begin{aligned} f_1 &= \frac{1}{2}(1-z^2) \\ f_2 &= \frac{i}{2}(1+z^2) \\ f_3 &= z \end{aligned}$$

$$\phi(u,v) = \frac{1}{2}\left(u - \frac{u^3}{3} + uv^2, -v + \frac{v^3}{3} - u^2 v, u^2 - v^2\right).$$

V. Henneberg's surface:

$$\begin{aligned} f_1(z) &= \left(-\frac{1}{z^4} + \frac{1}{z^2} + 1 - z^2\right) \\ f_2(z) &= i\left(-\frac{1}{z^4} - \frac{1}{z^2} + 1 - z^2\right) \\ f_3(z) &= 2\left(z - \frac{1}{z^3}\right) \end{aligned}$$

$$\phi(u,v) = \Big(\frac{u^3(1-u^2-v^2)^3 - 3uv^2(1-u^2-v^2)(1+u^2+v^2)^2}{3(u^2+v^2)^3}$$
$$\frac{3u^2v(1+u^2+v^2)^2(1-u^2-v^2) - v^3(1-u^2-v^2)}{3(u^2+v^2)^3}$$
$$\frac{(1-u^2-v^2)^2u^2 - (1+u^2+v^2)v^2}{(u^2+v^2)^2}\Big)$$

In the case of problems

$$I(c) = \int_{t_0}^{t_1} L(c_1, \ldots, c_n, \dot{c}_1, \ldots, \dot{c}_n)dt$$

which are defined on parameterized curves it is natural to ask: under which conditions is the integral independent of the parameterization? This will be the case if L is homogeneous in z i.e.

$$L(x, \rho z) = \rho L(x,z) \quad (\rho > 0)$$

As an example we have

$$I(c) = \int \sqrt{\sum g_{ij}\dot{c}_i(t)\dot{c}_j(t)}dt$$

Remark: In the case of a parameterization problem we have L is homogeneous of degree 1 – hence L_{x_i} is homogeneous of degree 1 and L_{z_i} is homogeneous of degree 0. A result of Euler states that if ϕ is homogeneous of degree n, then

$$\sum \xi_i \frac{\partial \phi_i}{\partial \xi_i} = n\phi$$

From this it follows that we have the following relationships:

$$\begin{aligned} \sum L_{z_k} z_k &= L \\ \sum L_{x_i z_k} z_k &= L_{x_i} \\ \sum L_{z_i z_j} z_j &= 0 \end{aligned}$$

Hence the equations: $\frac{d}{dt}L_{z_k}\big(c(t), \dot{c}(t)\big) - L_{x_k}\big(c(t), \dot{c}(t)\big) = 0 \quad (k = 1, \ldots, n)$ are not independent. This means that if we add the further equation

$$\sum \dot{c}_i(t)^2 = 1,$$

the system will not be overdetermined. This is useful in simplifying calculations.

Example: Fermat's principle.

$$L(t,z) = \mu(x_1,x_2,x_3)\sqrt{z_1^2+z_2^2+z_3^2}.$$

If we choose the curve with arc-length parameterization, we have

$$\frac{d}{ds}(\mu\gamma_k') = \mu_{x_k} \quad \text{(explicitly:} \quad \frac{d}{ds}\Big(\mu\big(\gamma(s)\big)\gamma_k'(s) = \mu_{x_k}\big(\gamma(s)\big)\Big)$$

or $\dfrac{d}{ds}(\mu(\gamma(s))\gamma_k'(s)) = \operatorname{grad}\mu(\gamma(s))$.

The Brachistone:

$$I(c) = \int_{t_0}^{t_1} \frac{\sqrt{\dot c_1(t)^2+\dot c_2(t)^2}}{\sqrt{c_2(t)}}$$

with boundary conditions $c(t_0) = (a,0)$, $c(t_1) = (b_1,b_2)$ with $b_1 > a$, $b_2 > 0$.

$$L(x,z) = \frac{\sqrt{z_1^2+z_2^2}}{\sqrt{x_2}}$$

$$L_{x_1} = 0, \quad L_{x_2} = -\frac{\sqrt{z_1^2+z_2^2}}{2x_2\sqrt{x_2}}$$

$$L_{z_i} = \frac{1}{\sqrt{x_2}}\frac{z_i}{\sqrt{z_1^2+z_2^2}}$$

We assume that the curve is parametrized by arc length i.e. that $\dot c_1^2+\dot c_2^2 = 1$. In accordance with the notation of "Differentialgeometrie" we write $\gamma(s)$ for $c(t)$. The first equation is:

$$\frac{d}{ds}\frac{\gamma_1'(s)}{\sqrt{\gamma_2(s)}} = 0.$$

Hence $\dfrac{\gamma_1'(s)}{\sqrt{\gamma_2(s)}}$ is constant.

Case 1. The constant $= 0$. Then $\gamma_1'(s) = 0$ and so γ_1 is constant. But this is impossible, since $b_1 > a$.

Case 2. The constant $k \neq 0$.

$$\gamma_2(s) = k\,\gamma_1'(s)^2 \quad (k > 0)$$

It follows from results of "Differentialgeometrie" that we can write $\gamma'(s) = (\cos\theta(s), \sin\theta(s))$ for a smooth function θ. Then

$$\gamma_2(s) = k\cos^2\theta(s).$$

Differentiating: $\sin\theta(s) = 2k\cos\theta(s)\cdot -\sin\theta(s)\theta'(s)$ i.e. $-2k\cos\theta(s)\theta'(s) = 1$. Hence $\dfrac{ds}{d\theta} = -2k\cos\theta$ and so $\dfrac{d\gamma_1}{d\theta} = \gamma'(s)\dfrac{ds}{d\theta} = -2k\cos^2\theta$.

Hence $\gamma' = k_1 - k(\theta + \sin\theta\cos\theta)$. If we use $t = \pi - 2\theta$ as a new parameter; we have

$$\cos^2\theta = \frac{1}{2}(1+\cos 2\theta) = \frac{1}{2}(1-\cos t)$$

$$\theta + \sin\theta\cos\theta = \theta + \frac{1}{2}\sin 2\theta = \frac{\pi - t}{2} + \frac{1}{2}\sin t.$$

This leads to the parameterization

$$\gamma_1 = d + \frac{k}{2}(t - \sin t), \quad \gamma_2 = \frac{k}{2}(1-\cos t) \quad \Big(d = k_1 - k\frac{\pi}{2}\Big)$$

i.e. the solution is a cycloid.

Geodetics: We consider a Riemann manifold with metric

$$ds^2 = \sum g_{ij}(x)dx_i dx_j$$

(cf. Differentialgeometrie). Then

$$I(c) = \int_{t_0}^{t_1} \Big[\sum g_{ij}\big(c_1(t), \ldots, c_n(t)\big)\dot{c}_i(t)\dot{c}_j(t)\Big]^{\frac{1}{2}}\, dt$$

with side conditions $c(t_0) = x_0, c(t_1) = x_1$.

$$L(x,z) = \left(\sum_{i,j} g_{ij}(x)z_i z_j\right)^{\frac{1}{2}}$$

$$L_{x_k} = \frac{1}{2}\frac{\sum \frac{\partial g_{ij}}{\partial x_k} z_i z_j}{\sqrt{\sum g_{ij} z_i z_j}}$$

$$L_{z_k} = \frac{\sum_l g_{kl} z_l}{\sqrt{\sum g_{ij} z_i z_j}}.$$

If we assume that we have parametrized by arc length, i.e. $\sum g_{ij}(c(t))\dot{c}_i(t)\dot{c}_j(t) = 1$ we have:

$$\frac{d}{ds}\Big(L_{z_k}\big(\gamma(s), \gamma'(s)\big)\Big) = \sum_l g_{kl}\big(\gamma(s)\big)\gamma_l''(s) + \sum_{i,j}\frac{\partial g_{kj}}{\partial x_i}\gamma_i'(s)\gamma_j'(s)$$

$$= \sum_l g_{kl}\gamma_l'' + \frac{1}{2}\sum_{i,j}\left(\frac{\partial g_{kj}}{\partial x_i} + \frac{\partial g_{ki}}{\partial x_j}\right)\gamma_j'(s)\gamma_i'(s)$$

This leads to the equation:

$$\sum_l g_{kl}\gamma_l'' + \sum_{i,j}\gamma_{ij,k}\big(\gamma(s)\big)\gamma_j'(s)\gamma_i'(s) = 0 \quad k = 1, \dots, n$$

where

$$\Gamma_{ij,k} = \frac{1}{2}\left(\frac{\partial g_{jk}}{\partial x_i} + \frac{\partial g_{ik}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_k}\right)$$

(cf. Differentialgeometrie).

4.3 Applications to Physics:

The Euler equations of suitable variational problems arise in Physics in two situations:
I. Equilibrium states – Principle of minimization of potential energy.
II. Dynamics – Variational principle of Hamilton. The corresponding Euler equations are the differential equations of the corresponding physical systems.
We consider a system with n degrees of freedom. Then the "position" is a function of n parameters

$$q_1, \dots, q_n$$

$T(\dot{q}_1, \dots, \dot{q}_n, q_1, \dots, q_n)$ is the kinetic energy. It has the form

$$T = \sum P_{ik}(q_1, \dots, q_n, t)\dot{q}_i\dot{q}_k$$

The potential energy U is a function of position and time i.e. $U(q_1, \dots, q_n, t)$.

Hamilton's principle: This states that the system minimizes the functional:

$$J = \int_{t_0}^{t_1}(T - U)dt$$

Then the Euler equations are:

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_i} - \frac{\partial}{\partial q_i}(T - U) = 0$$

Now at an equilibrium point, we have $T = \sum a_{ik}\dot{q}_i\dot{q}_k$ (where (a_{ik}) is a positive definite matrix and $U = \sum b_{ik}q_iq_k$ (where (b_{ik}) is a positive definite matrix). This leads to the system of equations:

$$\sum a_{ik}\ddot{q}_k + \sum b_{ik}q_k = 0$$

Examples: I. The vibrating string: Kinetic energy

$$T = \frac{1}{2}\int_0^1 \rho u_t^2 dx \quad (\rho \,\text{the density function})$$

$$U = \frac{1}{2}\int_0^1 \mu u_x^2 dx.$$

(μ is the coefficient of elasticity). Then

$$\int_{t_0}^{t_1} (T-U)dt = \frac{1}{2}\int_{t_0}^{t_1}\int_0^1 (\rho u_t^2 - \mu u_x^2) dx\, dt.$$

This gives the equation

$$\rho u_{tt} - \mu u_{xx} = 0$$

If a force $f(x,t)$ is acting, then we give an additional term $\int_0^1 f(x,t)u\, dx$ and so the equation

$$\rho u_{tt} - \mu u_{xx} + f(x,t) = 0$$

For stable equilibrium we minimize the functional

$$\int_0^1 \left(\frac{\mu}{2}u_x^2 + fu\right) dx.$$

This leads to equation $\mu u_{xx} - f(x) = 0$. The corresponding terms for a beam are:

$$\rho u_{tt} + \mu u_{xxxx} + f(x,t) = 0$$

resp.

$$\mu u_{xxxx} + f(x) = 0$$

Further examples: I. A free particle. In this case

$$L = \frac{1}{2}mv^2 2 = \frac{m}{2}(v_1^2 + v_2^2 + v_3^2).$$

i.e. $L = z_1^2 + z_2^2 + z_3^2$.
II. A system of non-reacting particles:

$$L = T = \sum \frac{m_\alpha v_\alpha^2}{2}$$

where $v_\alpha^2 = (z_1^\alpha)^2 + (z_2^\alpha)^2 + (z_3^\alpha)^2$. If the particles react with each other (e.g. by gravity or electricity), then this is modified by a term U which is a function of the position vectors of the particles e.g. $L = T - U$ where T is as above and

$$U = -\frac{1}{2}\sum \frac{\gamma m_a m_b}{|\mathrm{tr}_\alpha - \mathrm{tr}_\beta|}.$$

The special cases of two resp. three particles are the famous Kepler problem of the motion of the Sun-Earth system, respectively the three body problem.
III. Double pendulum swinging in a plane (cf. Landau and Lifchitz):

$$\begin{aligned} L &= \frac{m_1 + m_2}{2}[l_1^2\dot\phi_1^2 + l_2^2\dot\phi_2^2 + m_2 l_1 l_2 \dot\phi_1\dot\phi_2\cos(\phi_1 - \phi_2)] + \\ &\quad +(m_1 + m_2)g l_1 \cos\phi_1 + m_2 l_2 \cos\phi_2. \end{aligned}$$

In this context it is illuminating to consider the invariance of the Euler equation under changes of coordinates. Suppose that we introduce new coordinates $(\tilde q_\alpha, \tilde v_\alpha, \tilde t)$ where

$$\tilde q_\alpha = \tilde q(q, t) \qquad \tilde t = t.$$

Then

$$\tilde v_\alpha = \sum_j \frac{\partial \tilde q_i}{\partial q_j} v_j + \frac{\partial q_i}{\partial t}.$$

Then if $F = F(q, v, t)$,

$$\frac{d}{dt}\left(\frac{\partial F}{\partial v_i} - \frac{\partial F}{\partial q_i}\right) = \sum_j \frac{\partial \tilde q_j}{\partial q_i}\left[\frac{d}{dt}\left(\frac{\partial F}{\partial \tilde q_j}\right) - \frac{\partial F}{\partial \tilde q_j}\right].$$

Example: Consider a single particle moving under gravity. We calculate its equations of motion in spherical coordinates i.e.

$$q_1 = x, q_2 = y, q_3 = z, U = mgy, T = \frac{m}{2}(\dot x^2 + \dot y^2 + \dot z^2)$$

resp.

$$\tilde q_1 = r, \tilde q_2 = \theta, \tilde q_3 = \phi.$$

Then

$$L = \frac{1}{2}m(\dot r^2 + r^2\dot\theta^2 + r^2\sin^2\theta\dot\phi^2) - mgr\cos\theta.$$

Hence the equations of motion are

$$0 = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot r} - \frac{\partial L}{\partial r}\right) = m\ddot r - mr(\dot\theta^2 + \sin^2\theta\dot\phi^2) + mg\cos\theta$$

(corresponding to the dependency on r). From the dependency on θ, we have

$$0 = mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta} - r^2 \sin\theta\cos\theta\dot{\phi}^2 - mgr\sin\theta,$$

while from the dependency of ϕ, we get

$$\frac{d}{dt}(mr^2\sin^2\theta\dot{\phi}) = 0.$$

(The last equation means that angular momentum about the z-axis is preserved. This is a g general fact. If we have a coordinate system in which L is independent of the coordinate $\tilde{q}_i$, then the corresponding momentum $p_l = \dfrac{\partial L}{\partial v_l}$ will be conserved.

Example: We consider the transformation from stationary Cartesian coordinates to rotating ones i.e.

$$q_1 = \tilde{q}_1\cos\omega t - \tilde{q}_2\sin\omega t, q_2 = \tilde{q}_1\sin\omega t + \tilde{q}_2\cos\omega t, q_3 = \tilde{q}.$$

Solving the Euler equations: The general Euler equation can be rather intractable. However, in certain cases where L has a particularly simple form, one can apply standard methods of ordinary differential equations. Examples

a) L is independent of x. Then the Euler equation is

$$\frac{d}{dt}L_z\big(t, \dot{c}(t)\big) = 0$$

and so $L_z(t, \dot{c}(t))$ is constant. In principle we can solve this to get $\dot{c}(t)$ explicitly as a function of t and obtain the solution by interpolation.

b) L is independent of t Then we have

$$\begin{aligned}\frac{d}{dt}\big(\dot{c}(t)L_z c(t), \dot{c}(t)\big) - L(c(t), \dot{c}(t)) &= \ddot{c}(t)L_z + \dot{c}(t)\frac{d}{dt}L_z - L_z\ddot{c}(t) - L_x\dot{c}(t)\\ &= \dot{c}(t)\left(\frac{d}{dt}L_z - L_x\right)\\ &= 0\end{aligned}$$

Hence we have the equation:

$$L\big(c(t), \dot{c}(t)\big) - \dot{c}(t)L_z\big(c(t), \dot{c}(t)\big) = \text{const}$$

which we again solve for $\dot{c}(t)$ as function of $c(t)$ and the constant. We can solve this by standard methods (rewriting as a differential equation in t!).

Examples. I. The Isoperimetric problem. First we simplify the problem by assuming that the curve is convex. We choose the x-axis so that it halves the area. We thus reduce to the following variational problem: Maximize:

$$\int_0^{t_0} c(t)dt$$

under the conditions $c(0) = c(t_0) = 0$, $\int_0^{t_0} \sqrt{1+\dot{c}(t)^2}dt = 1$ (where t_0 is also unknown). If we choose $s = \int_0^t \sqrt{1+\dot{c}(u)^2}du$ as new independent variable, then we reduce to the following:

$$\int_0^1 y(s)\sqrt{1-y'(s)^2}ds$$

where $y(s) = c(t)$. One solves this equation. The solution to the original problem is then the parameterized curve

$$\big(x(s), y(s)\big)$$

where

$$x(s) = \int_0^s \sqrt{1-\Big(\frac{dy}{ds}\Big)^2}ds$$

In our concrete case we have

$$L(t,x,z) = x\sqrt{1-z^2}$$

Hence $\dfrac{-y(s)}{\sqrt{1-\dot{y}(s)^2}} = -\dfrac{1}{d}$. This has solution $y = \dfrac{1}{d}\sin(ds+c_1)$

$$x(s) = \int \sqrt{1-\dot{y}(s)^2}ds = \int \sin(ds+c_1)ds = -\frac{1}{d}\cos(ds+c_1)+d_1.$$

From this it follows that the solution of the isoperimetric problem is a circle.
II. The Brachistochrone:

$$\int_{t_0}^{t_1} \frac{\sqrt{1+\dot{c}(t)^2}}{\phi(t,c(t))}dt$$

$$L(t,x,z) = \frac{\sqrt{1+z^2}}{\phi(t,x)} = \psi(t,x)\sqrt{1|z^2}$$

$$\psi\ddot{c}(t) = \psi_x\big(t,c(t)\big) - \psi_t\big(t,c(t)\big)\dot{c}(t)\big(1+\dot{c}(t)^2\big).$$

Special case:

$$\phi(t,x) = \frac{1}{\sqrt{x}} \quad \text{i.e.} \quad L(t,x,z) = \sqrt{\frac{1+z^2}{x}}$$

Then the equation is $\dfrac{-1}{\sqrt{c(t)(1+\dot{c}(t)^2)}} = \dfrac{1}{d}$. (The solution is a cycloid).
II. Minimal surface of variation.

$$\int_{t_0}^{t_1} c(t)\sqrt{1+\dot{c}(t)^2}dt$$

This is the special case of the above example with

$$L(t,x,z) = x\sqrt{1+z^2} \quad \text{i.e.} \quad \psi(t,x) = x$$

Then we have

$$\frac{-c(t)}{\sqrt{1+\dot{c}(t)^2}} = -\frac{1}{d}$$

This has solution $c(t) = \dfrac{1}{d}\cosh(dt + c_1)$ – i.e. the solution is the surface of rotation of a catenary i.e. a catenoid.

Problems without fixed endpoints: Here we consider the problem of finding an optimal value of the functional

$$I = \int_{t_0}^{t_1} L(t,c,\dot{c})dt$$

without conditions on the endpoints. Since a solution is automatically a solution of the problem with fixed endpoints and so satisfies the Euler equation, the expression for its derivative simplifies to

$$(DI)_c(h) = L_z\big(t, c(t), \dot{c}(t)\big)\Big|_{t_0}^{t_1}.$$

This leads to the additional condition: $L_z(t, c(t), \dot{c}(t)) = 0$ at t_0 and t_1 for the solution. In a similar way for the case of optimizing the functional

$$I = \int_{t_0}^{t_1} L(t, x_1, \dots, x_n, z_1, \dots, z_n)dt$$

without conditions on the endpoints, we get the additional equations

$$L_{z_i}\big(t, c_1(t), \dots, c_n(t), \dot{c}_1(t), \dot{c}_2(t), \dots, \dot{c}_n(t)\big) = 0 \quad (i = 1, \dots, n)$$

at t_0 and t_1. For

$$I = \int\!\!\int L\big(u, v, \phi(u,v), \phi_1(u,v), \phi_2(u,v)\big)du\,dv,$$

we get the additional equation

$$L_{z_1}\frac{d\phi}{ds} - L_w\frac{d\phi}{ds} = 0$$

on ∂U.

Problems with side conditions: Recall the proof of the spectral theorem: In the finite dimensional case, we consider the problem of maximizing the function of $(Ax|x) = \sum a_{ij}\xi_i\xi_j$ under the side condition $(x|x)_2 = \sum \xi_i^2 = 1$. The method of Lagrange multipliers leads to the fact that the maximum occurs at an eigenvector and hence such an eigenvector exists. In the case of a compact, self-adjoint operator, we consider the problem of optimizing $I(x) = (Ax|x)$ under the condition $(x|x) = 1$. Then

$$\begin{aligned}(DI)_x(h) &= 2(Ax|h)\\ (DI^*)_x(h) &= 2(x|h)\end{aligned}$$

Hence if a x_0 is a maximum, there is a λ so that

$$(Ax_0|h) = \lambda(x_0|h)$$

for each $h \in H$. This again means that x_0 is an eigenvector. The general situation is as follows. I and I^* are functionals on an affine subspace of a Banach space. Then if c is a solution of the optimisation problem: $I(c)$ is maximum (or minimum) with side condition $I^*(c) = 0$ then there is a $\lambda \in \mathbf{R}$ so that

$$(DI)_c = \lambda DI_c^*.$$

We can reduce to the finite dimensional case by the following trick. Firstly, we assume that E is separable (the general case is proved similarly, with a slightly more complicated terminology). Then we can express E as the closed hull of the union $\bigcup E_n$ of finite dimensional subspaces. Suppose that c is a solution of our optimisation problem. Without loss of generality we can assume that c lies in E_1 and hence in each E_n. c is then *a fortiori* a solution of the optimisation problem for I restricted to each E_n. Since $(DI^*)_c \neq 0$, we can also, without loss of generality, assume that the restriction of the latter to each E_n is non-zero. Hence by the finite dimensional version of the result, there is, for each n a scalar λ_n so that

$$(DI_n)_c + \lambda_n(DI_n^*)_c = 0,$$

where the subscript denotes the restriction to E_n. But it then follows easily that the λ_n are all equal and so we have the equation

$$(DI)_c + \lambda(DI^*)_c = 0$$

where we have denoted by λ the common value.

Example: The isoperimetric problem. In this case

$$\begin{aligned}
L(x,z) &= \frac{1}{2}(x_1 z_2 - x_2 z_1), \quad L^*(x,t) = \sqrt{z_1^2 + z_2^2} \\
L_{x_1} &= \frac{1}{2} z_2, \quad L^*_{x_1} = 0 \\
L_{x_2} &= = -\frac{1}{2} z_1, \quad L^*_{x_2} = 0 \\
L_{z_1} &= -\frac{1}{2} x_2, \quad L^*_{z_1} = \frac{z_1}{\sqrt{z_1^2 + z_2^2}} \\
L_{z_1} &= \frac{1}{2} x_1, \quad L^*_{z_2} = \frac{z_2}{\sqrt{z_1^2 + z_2^2}}
\end{aligned}$$

Assume: $z_1^2 + z_2^2 = 1$. Then we have the equation

$$\frac{d}{ds}\left(-\frac{1}{2}\gamma_2(s) + \mu\gamma'{}_1(s)\right) - \frac{1}{2}\gamma_2'(s) = 0$$

$$\frac{d}{ds}\left(\frac{1}{2}\gamma_1(s) + \mu\gamma_2'(s)\right) + \frac{1}{2}\gamma_1'(s) = 0$$

Hence:

$$\mu\gamma_1'' - \gamma_2' = 0, \quad \mu\gamma_2'' + \gamma_1' = 0$$

Hence:

$$\mu(\gamma_1'\gamma_2'' - \gamma_2'\gamma_1'') = \gamma_1'^2 + \gamma_2'^2 = 1$$

Hence κ is constant – thus the solution is a circle.

Example: The hanging chain: This corresponds to minimising the functional

$$I(c) = \int_{t_o}^{t_1} c(t)\sqrt{1 + \dot{c}(t)^2}\, dt$$

under the side condition

$$I^*(c) = \int_{t_0}^{t_1} \sqrt{1 + \dot{c}(t)^2}\, l dt - 1 = 0.$$

Then

$$L + \lambda L^* = (x + \lambda)\sqrt{1 + z^2}$$

and so

$$(L + \lambda L^*)_x = \sqrt{1 + z^2}, (L + \lambda L^*)_z = (x + \lambda)\frac{z}{\sqrt{1 + z^2}}.$$

The Euler equation is thus

$$\frac{d}{dt}(c(t)+\lambda)\frac{\dot{c}(t)}{\sqrt{1+\dot{c}(t)^2}}) = \sqrt{1+\dot{c}(t)^2}.$$

This is the same as the equation for the minimal surface of rotation (with $c(t)$ replaced by $c(t)+\lambda$). Hence the solution is

$$c(t) = -\lambda + d\cosh\frac{t-\bar{t}}{d}$$

with boundary conditions

$$-\lambda + d\cosh\frac{t-\bar{t}}{d} = \alpha_1, -\lambda + d\cosh\frac{t_1-\bar{t}}{d} = \alpha_2$$

and further condition

$$d(\sinh\frac{t_1-\bar{t}}{d} - \sinh\frac{t_0-\bar{t}}{d}) = 1.$$

(The constants d and $\bar{t}$ are then determined by the equations

$$d = \frac{t_1-t_0}{2\sigma}, \bar{t} = \frac{1}{2}[(1-\frac{\tau}{\sigma})t_0 + (1+\frac{\tau}{\sigma})t_1]$$

where $\tanh\tau = b-a$ and

$$\frac{\sinh\sigma}{\sigma} = \frac{\sqrt{1-(b-a)^2}}{t_1-t_0}$$

with $\sigma > 0$).

In our treatment, we have tacitly assumed that the Euler equation, which in the case of the simplest functional

$$I(c) = \int_{t_0}^{t_1} L(t, c(t), \dot{c}(t))\, dt$$

is an implicit ordinary differential equation of order 2 can be solved to give an explicit equation. If we regard this equation in the following form

$$L_x(t, c(t)\dot{c}(t)) - L_{zt}(t, c(t)\dot{c}(t)) - c(t)L_{zx}(t, c(t)\dot{c}(t)) - \ddot{c}(t)L_{zz}(t, c(t), \dot{c}(t)) = 0$$

we see that the condition for this to be the case is that $L_{zz}(t, c(t)\dot{c}(t)) \neq 0$ along the solution c. We consider here the case which is the extreme opposite of this namely where the Euler equation is such that *every* suitable curve is a solution. If we consider the above form of the equation, we must clearly have that L_{zz} vanishes identically. This means that L has the form

$$A(t,x) + zB(t,x).$$

The Euler equation is then

$$\frac{d}{dt}L_z - L_x = 0$$

which reduces to the equation $B_t - A_x = 0$. i.e. the well-known integrability condition for the differential form $Adt + Bdx$. In this case we have

$$\int_{t_0}^{t_1} L(t, c(t)\dot{c}(t))\, dt = \int_{t_0}^{t-1} [A(t, c(t)) + \dot{c}(t)B(t, c(t))]\, dt = \int_{\bar{c}} Adx + Bdy$$

where $\bar{c}$ is the curve with parametrisation $t \mapsto (t, c(t))$ and this is independent of the curve when the above integrability condition holds.

The second variation: As in the case of finite dimensional optimisation, the second variation of the functional can provide more detailed information on the nature of an extremum. We shall use this method here to show that if c is a solution of the Euler equations for the functional

$$I(c) = \int_{t_0}^{t_1} L(t, c(t), \dot{c}(t))\, dt$$

with fixed endpoints, then

$$L_{zz}(t, c(t), \dot{c}(t)) \geq 0$$

is necessary for c to be a minimum of I. (We remark here that the non-vanishing of this expression is the Lagrangian condition for the fact that the Euler equation can be written as an explicit differential equation of second order. In the case of a system of equations i.e. where

$$L = L(x_1, \dots, x_n, z_1, \dots, z_n),$$

then the corresponding condition is that the matrix $L_{z_i z_j}$ be positive semi-definite).

In order to prove the above fact, consider the Taylor expansion

$$I(c + tH) = I(c) + (DI)_c(th) + (D^2I)_c(th, th) + o(||th||^2).$$

Since $(DI)_v$ vanishes, we see that

$$\lim_{s \to 0} \frac{1}{s^2}(I(c+sh) - I(c)) = (D^2 I)_c(h,h).$$

Hence, by the one-dimensional result, if c is a minimum, we must have $(D^2 I)_c(h,h) \geq 0$ for each h. But a simple calculation shows that in this case,

$$\lim_{s \to 0} \frac{1}{s^2}(I(c+sh) - I(c)) = \int_{t_0}^{t_1} \left[L_{xx} H^2(t) + 2 L_{xz} h(t) h'(t) + L_{zz} h'(t)^2 \right] dt.$$

By choosing for h concrete "peak functions", it follows that $L_{zz} = 0$ along c.

Invariance of the Euler equation under coordinate transformations: It follows from the coordinate-free description of the optimisation problem, that the Euler equations are invariant under suitable coordinate transformations. The explicit calculation of this fact is particularly important in two concrete cases: changes in the independent variable t (reparametrisations) and changes in x (new generalised coordinates for physical systems). For the detail, see the lectures.

Remarks on existence: In this course we have generally avoided the difficulties involved in the demonstration of the *existence* of a solution to the optimisation problem. Three exceptions were the proofs of the Riemann mapping theorem and the spectral theorem for compact operators (where we used a compactness argument) and the existence of a best approximation from a compact subset of a Hilbert space (where we used a geometrical argument). In general, the existence question is one of considerable depth. For example it embraces such significant problems as the solution of the Dirichlet problem and Plateau's problem on the existence of minimal surfaces bounded by given closed, space curves, resp. the existence of geodetics on curved manifolds (theorem of Hopf-Rinow). In fact, it is not difficult to supply examples where simple optimisation problems fail to have a solution or, even when they do, the solution is not obtainable as a limit of a minimising sequence.

Examples: 1. Firstly, we can obtain a large collection of examples where a smooth functional need not have a minimal by considering a well-behaved manifold such as the plane or the sphere where any pair of points can always be joined by a minimal geodetic and removing a point. Then it is easy to construct pairs in the adjusted manifold which *cannot* be so joined.

2. Consider the points $(-1, 0)$ and $(1, 0)$ in the plane and the problem of finding a smooth curve c of minimal length joining them and satisfying the condition that $\dot{c}(0) = (0, 1)$. It is clear that the minimal length is 2 but that this is not attained by any appropriate curve.
3. The problem of minimising the functional

$$\int_{-1}^{1} x^4 y'^2 \, dx$$

for the set of all curves $y = y(x)$ with $y(-1) = -1, y(1) = 1$. The minimal value is 0 and one again, it is not attained by any smooth curve.
4. The following is an example where the optimisation problem trivially has a solution. However, we display a minimal sequence which is such that no subsequence converges to the solution. We consider the Dirichlet problem i.e. that of minimising the functional

$$D(\phi) = \int \int_G (\phi_x^2 + \phi_y^2) dx dy$$

in the trivial case where $G = \{x^2 + y^2 \leq 1\}$ and the boundary value is the zero function. Of course, the solution is the constant function zero. However we can construct an explicit minimal sequence which does not converge to the zero function (cf. Lectures). We conclude the course with a brief mention of abstract methods which are used to attack the problem of existence.

Weak topologies: It is often advantageous to replace the norm topology on the Banach space upon which the functional is defined by the so-called **weak topology** (i.e. the initial topology induced by the elements of the dual of E). This has the advantage that larger sets can be compact for the weak topology. For example in the L^p-spaces $(1 < p < \infty)$ and the corresponding Sobolew spaces (see below), the unit ball is weakly compact (but not, of course, norm compact). The disadvantage is that, this topology being weaker than the norm one, it is harder for the functional ϕ to b e continuous.

Semicontinuity. The previous objection can often be met by noting that for a function on a compact set to be bounded below and attain its minimum, it suffices for it to be lower semi-continuous. Thus the norm on an infinite dimensional Banach spaces is not weakly continuous but it *is* weakly lower semi-continuous.

Sobolew spaces: In applying abstract methods to functionals given by integrals of suitable kernels, it is crucial to choose the correct Banach spaces on which the functional acts. These are determined by the growth and differentiability conditions required to give meaning to the integral expression involved. In order to get spaces with suitable properties, it is usually necessary use a generalised form of derivative, the distributional derivative (see Vorlesung, Funktionalanalylsis II, or a course on Distribution theory). The resulting spaces are called Sobolew spaces and are defined as spaces of functions for which suitable distributional derivatives are L^p-functions.

As an example of an abstract result of the type which can be used to obtain existence statements in a general situation, we mention without proof:

Proposition 23 *Let E be a reflexive Banach space, $\phi : E \to \mathbf{R}$ a norm continuous functional. Then if ϕ is convex, it is weakly lower semi=continuous. If ϕ is weakly semi-continuous and coercive i.e. such that $\lim_{x\to\infty} \dfrac{\phi(x)}{||x||}.0$, then ϕ attains its minimum on E.*