Contents

Part II

Properties of minimizers and irregular problems

Chapter 1

Convexity and Geometric methods

1.1 Convexity

The best source for the theory of convexity is probably the book [**?**].

1.1.1 Definitions and inequalities

We start with definition of convexity.

Definition 1.1.1 The set Ω in R_n is convex, if the following property holds. If any two points x_1 and x_2 belong to the set Ω, all points x_h with coordinates $x_h = \lambda x_1 + (1-\lambda)x_2$ belong to Ω

Ellipsoid, cube, or paraboloid is a convex set, crescent is not convex. Convex sets are simply connected (do not have holes). The whole space R_n is a convex set, any hyperplane is also a convex set. The intersection of two convex sets is also a convex set, but the union of two convex sets may be not convex.

The boundary of a convex set has the following property: for each point of the boundary there is a plane that passes through this point but does not pass through any other interior point. Such plane is called *supporting plane*. One can define a convex body as a domain bounded by all supporting planes. This description is called the dual form of the definition of a convex body.

Next, we can define a convex function.

Definition 1.1.2 Consider a scalar function $f : \Omega \to R_1$ $\Omega \subset R_n$ of vector argument. Function F is called *convex* if it possesses the property

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \quad \forall x_1, x_2 \in R_n, \quad \forall \lambda \in [0,1] \quad (1.1)$$

Geometrically, the property (1.1) states that the graph of a convex function lies below any chord.

Figure 1.1: Basic property of convex function: The chord lies above the graph

Example 1.1.1 Function $f(x) = x^2$ is convex. Indeed, $f(\lambda x_1 + (1-\lambda)x_2)$ can be represented as follows

$$(\lambda x_1 + (1-\lambda)x_2)^2 = \lambda(x_1)^2 + (1-\lambda)(x_2)^2 - C$$

where $C = \lambda(1-\lambda)(x_1 - x_2)^2 \geq 0$ is nonnegative. Therefore, (1.1) is true and $f(x)$ is convex.

Properties of convex functions One can easily show (try!) that the function is convex if and only if

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{f(x_1) + f(x_2)}{2} \quad \forall x_1, x_2 \in R_n.$$

Example 1.1.2 (minimum of an even function) If a function of a vector argument x is convex and even, $f(-x) = f(x)$, then it reaches the minimum at $x = 0$,

$$f(0) \leq f(x) \quad \forall x \in R_n$$

The convex function $F(x)$ of a vector argument $x = (x_1, \ldots, x_n)$ is differentiable almost everywhere.

Definition 1.1.3 (Weierstrass function) If the first derivatives exist at a point $x \in R_n$, the following inequality holds

$$W_F(x, z) = F(z) - F(x) - (z - x)^T \frac{\partial F}{\partial x} \geq 0 \quad \forall z \in \mathcal{X} \tag{1.2}$$

where gradient $\frac{\partial F}{\partial x}$ is the vector with components $\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial x_n}$. The function $W(x, z)$ is called *Weierstrass function*. Function $F(x)$ is convex at a point x if (1.2) holds.

The inequality (**??**) compares the value $F(z)$ with the value of the hyperplane $P(z) = F(x) - (z - x)^T \frac{\partial F}{\partial x}$ that is tangent to the graph of F at the point x.

If the function is not differentiable at a point x, the inequality (1.2) must be modified. Instead of the tangent plane, we require that a plane exist that coincides with the graph of F at the point x and lies below this graph everywhere.

Definition 1.1.4 Function $F(x)$ is convex at a point x if

$$\exists A = (a_1, \ldots a_n): \quad F(z) - F(x) - (z - x)^T A \geq 0 \quad \forall z \in \mathcal{X}$$

Here, A does not need to be a tangent plane, but only a supporting plane.

Example 1.1.3 Proof of the convexity of the Euclidian norm

$$F(x_1, \ldots, x_n) = \sqrt{x_1^1 + \ldots + x_n^2}$$

First, assume that $x \neq 0$. Then the gradient exists and is equal to

$$\frac{\partial F}{\partial x} = \frac{1}{F(x)} x;$$

the left-hand side of (**??**) becomes

$$F(z) - F(x) - (z - x)^T \frac{\partial F}{\partial x} = \frac{1}{F(x)} \left(F(z)F(x) - z^T x \right)$$

Recall that $F(x)$ is the Euclidean norm, therefore

$$z^T x = F(x)F(z)\cos(\widehat{z, x}) \leq F(z)F(x).$$

Therefore, the left-hand side of (1.2) is nonnegative and F is convex everywhere.

At the point $x = 0$ the function is also convex, according to the Definition 1.1.4. It is enough to choose $A = 0$ and check that

$$F(z) - F(x) - (z - x)^T A = F(z) \geq 0 \quad \forall z \in R^n$$

Definition 1.1.5 (Convexity of a smooth function) If the second derivatives of a convex function exist at every point, the Hessian $He(f, x)$ is nonnegative everywhere

$$He(f, x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \ldots & \ldots & \ldots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix} \geq 0.$$

Particularly, the convex function of one variable has the nonnegative second derivative:

$$f''(x) \geq 0 \quad \forall x \in R_1. \tag{1.3}$$

Example 1.1.4 What are values of α for which

$$F_\alpha(x, y) = x^\alpha y^2, \quad x \geq 0$$

is convex with respect to x and y?

Compute the Hessian

$$H = \begin{pmatrix} \alpha(\alpha - 1)x^{\alpha-2}y^2 & 2\alpha x^{\alpha-1}y \\ 2\alpha x^{\alpha-1}y & 2x^\alpha \end{pmatrix}$$

and its determinant $\det H = -2x^{2\alpha-2}y^2\alpha(\alpha+1)$. The determinant is nonnegative and therefore function $F_\alpha(x, y)$ is convex if $\alpha \in [-1, 0]$. Notice, that if $F_\alpha(x, y)$ is convex for some x, y it is convex for all x, y.

Figure 1.2: Graph of nonconvex function $f(x) = \exp(-|x|)$

The nonnegativity of Hessian $He(F)$ everywhere at the domain of definition guarantees the convexity of F at that domain. However, nonnegativity of Hessian at a point is only a necessary condition for the convexity at this point, but not sufficient. Even if Hessian is positive at a point, the function can be nonconvex there as it is illustrated by the next example

Example 1.1.5 Function $f(x) = x^4 - 6x^2$ is convex if $x \notin [-3, 3]$. (see the graph in Figure ??). Indeed, the condition (1.2) of convexity reads

$$W_f(x, z) = f(z) - f(x) - (z - x)f'(x) = (3x^2 - 6 + 2zx + z^2)(x - z)^2 > 0 \quad \forall z$$

It is satisfied, when the first multiplier does not have roots, or if $x \notin [-\sqrt{3}, \sqrt{3}]$, because

$$W_f(x, z) = \left(2(x^2 - 3) + (x + z)^2\right)(x - z)^2 > 0 \quad \forall z$$

This condition should ne compared with inequality $f''(x) \geq 0$ that holds in a smaller interval $x \notin [-1, 1]$. At the intervals $x \in [-\sqrt{3}, -1]$ and $x \in [1, \sqrt{3}]$ the second derivative of F is positive, but F is not convex.

Convexity is a global property. If the inequality (1.1) is violated at one point, the function may be nonconvex everywhere.

Example 1.1.6 Consider, for example, $f(x) = \exp(-|x|)$. Its second derivative is positive everywhere, $f'' = \exp(-|x|)$ except $x = 0$ where it does not exist. This function is not convex, because

$$f(0) = 1 > \frac{1}{2}(f(x) + f(-x)) = \exp(-|x|) \quad \forall x \in R.$$

1.1.2 Jensen's inequality

The definition (1.1) is equivalent to the so-called Jensen's inequality

$$f(x) \leq \frac{1}{N}\sum_{i=1}^{N} f(x + \zeta_i) \quad \forall \zeta_i : \sum_{i=1}^{N} \zeta_i = 0 \tag{1.4}$$

for any $x \in \Omega$. (Show the equivalence!)

Jensen's inequality enables us to define convexity in a point: The function f is convex at the point x if (1.4) holds.

Integral form of Jensen inequality Increasing the number N of vectors ζ_i in (1.4), we find the integral form of Jensen inequality:

Function $F(z)$ is convex if and only if the inequality holds

$$F(z) \le \frac{1}{b-a} \int_a^b F(z + \theta(x))dx \tag{1.5}$$

where

$$\int_a^b \theta(x)dx = 0 \tag{1.6}$$

and all integrals exist.

Remark 1.1.1 (Stability to perturbations) The integral form of the Jensen's inequality can be interpreted as follows: The minimum of an integral of a convex function corresponds to a constant minimizer. No perturbation with zero mean value can increase the functional.

Another interpretation is: The average of a convex function is larger then the function of an averaged argument.

Example 1.1.7 Assume that $F(u) = u^2$. We have

$$0 \le \frac{1}{b-a} \int_a^b (z + \theta(x))^2 dx = z^2 + \frac{2z}{b-a} \int_a^b \theta(x)dx + \frac{1}{b-a} \int_a^b \theta(x)^2 dx$$

The second integral in the right-hand side is zero because of (1.6), the third integral is nonnegative. The required inequality

$$z^2 \le \frac{1}{b-a} \int_a^b (z + \theta(x))^2 dx$$

(see (1.5) follows.

Next, we illustrate the use of convexity for solution of optimization problems. Being global property, convexity allow for establishing the most general between the optimal trajectory and any other trajectory.

1.1.3 Minimal distance at a plane, cone, and sphere

Let us start with the simplest problem with an intuitively expected solution: Find the minimal distance between the points (a, α) and (b, β) on a plane.

Consider any piece-wise differentiable path $x(t), y(t)$, $t \in [0.1]$ between these points. We set

$$x(0) = a, \quad x(1) = b, \quad y(0) = \alpha, \quad y(1) = \beta$$

The length of the path is

$$L(x, y) = \int_0^1 \sqrt{(x')^2 + (y')^2} dx$$

(We need the piece-wise differentiability of $x(t)$ and $y(t)$ to be able define the length of the pass) We have in mind to compare the path with the straight line (which we might expect to be a solution); therefore, we assume the representation

$$x(t) = a + t(b-a) + \int_0^t \psi_1(t)dt, \quad y(t) = \alpha + t(\beta - \alpha) + \int_o^t \psi_2(t)dt$$

the terms dependent on ϕ and ψ define the deviation from the straight path. The deviation in the beginning and in the end of the trajectory is zero, therefore we require

$$\int_0^1 \psi_1(t)dt = 0 \quad \int_0^1 \psi_2(t)dt = 0; \tag{1.7}$$

We prove that the deviation are identically zero at the optimal trajectory.

First, we rewrite the functional L in the introduced notations

$$L(\psi_1, \psi_2) = \int_0^1 \sqrt{((b-a) + \psi_1(t))^2 + ((\beta - \alpha) + \psi_2(t))^2} dx$$

where the Lagrangian $W((\psi_1, \psi_2)$ is

$$W((\psi_1, \psi_2) = \sqrt{((b-a) + \psi_1(t))^2 + ((\beta - \alpha) + \psi_2(t))^2}$$

and we use expressions for the derivatives x', y':

$$x' = (b-a) + \psi_1(t), \quad y' = (\beta - \alpha) + \psi_2(t).$$

The Lagrangian $W((\psi_1, \psi_2)$ is a convex function of its arguments ψ_1, ψ_2. Indeed, it is twice differentiable with respect to them and the Hessian He is

$$He(W) = \begin{pmatrix} y^2(x^2+y^2)^{-\frac{3}{2}} & xy(x^2+y^2)^{-\frac{3}{2}} \\ xy(x^2+y^2)^{-\frac{3}{2}} & x^2(x^2+y^2)^{-\frac{3}{2}} \end{pmatrix}$$

where $x = (b-a) + \psi_1(t)$ and $y = (\beta - \alpha) + \psi_2(t)$. The eigenvalues of the Hessian are equal to 0 and $(x^2+y^2)^{-\frac{1}{2}}$ respectively, and therefore it is nonnegative defined (as the reader can easily check, the graph of $W((\psi_1, \psi_2)$ is a cone).

Due to Jensen's inequality in integral form, the convexity of the Lagrangian and the boundary conditions (1.7) lead to the relation

$$L(\psi_1, \psi_2) \geq L(0,0) = \int_0^1 \sqrt{(b-a)^2 + (\beta - \alpha)^2} dx$$

and to the minimizer $\psi_1 = 0, \psi_2 = 0$.

Thus we prove that the straight line corresponds to the shortest distance between two points. Notice that (1) we compare all differentiable trajectories no matter how far away from the straight line are they, and (2) we used our correct guess of the minimizer (the straight line) to compose the Lagrangian. These features are typical for the global optimization.

Geodesic on a cone Consider the problem of shortest path between two points of a cone, assuming that the path should lie on the conical surface. This problem is a simplest example of geodesics, the problem of the shortest path on a surface discussed below in Section **??**.

Because of simplicity of the cone's shape, the problem can be solved by pure geometrical means. Firstly, we show that it exists a ray on a cone that does not intersect with the geodesics between any two point if none of then coincide with the vertex. If this is not the case, than a geodesics makes a whole spiral around the cone. This cannot be because one can shorten the line replacing spiral part of a geodesics by an interval if a ray.

Now, let us cut the cone along this ray and straighten the surface: It becomes a wedge of a plane with the geodesics lying entirely inside the wedge. Obviously, the straighten does not change the length of a path. The coordinates of any point of the wedge can be characterized by a pair r, θ where $r > 0$ is the distance from the vertex and θ, $0 \leq \theta \leq \Theta$ is the angle counted from the cut. Parameter Θ characterizes the cone itself.

The problem is reduced to a problem of a shortest path between two points that lies within a wedge. Its solution depends on the angle Θ of the wedge. If this angle is smaller that π, $\Theta < \pi$, the optimal path is a straight line

$$r = A \tan \theta + B \sec \theta \tag{1.8}$$

One can observe that the $r(\theta)$ is a monotonic function that passes through two positive values, therefore $r(\theta) > 0$ – the path never goes through the origin. This is a remarkable geometric result: *no geodesics passes through the vertex on a cone if* $\Theta < \pi$*!*: There always is a shorter path around the vertex.

At the other hand, if $\Theta > \pi$, then a family of the geodesics will path through the vertex and consist of two straight intervals. This happens if $\theta > \pi$. Notice that in this case the original cone, when cut, becomes a wedge with the angle larger than 2π and consist of at least two overtopping sheets.

Distance on a sphere: Columbus problem Consider the problem of geodesics on a sphere. Let us prove that a geodesics is a part of the great circle.

Suppose that geodesics is a different curve, or that it exists an arc that is a part of the geodesics but does not coincide with the arc of the great circle. This arc can be replaced with its mirror image – the reflection in the plane that passes through the ends of the arc and the center of the sphere. The reflected curve has the same length of the path and it lies on the sphere, therefore the new path remains a geodesics.

At the other hand, the new path is broken in two points, and therefore cannot be the shortest path. Indeed, consider a part of the path in an infinitesimal circle around the point of breakage and fix the points A and B where the path crosses that circle. This path can be shorten by a arc of a great circle that passes through the points A and B. To illustrate this part, it is enough to imagine a human-size scale on Earth: The infinitesimal part of the round surface becomes flat and obviously the shortest path correspond to a straight line and not to a zigzag line with an angle.

The same consideration shows that the length of geodesics is no larger than π times the radius of the sphere or it is shorter than the great semicircle. Indeed, if the length of geodesics is larger than the great semicircle one can fix two opposite points – the poles of the sphere – on the path and turn around the axis the part of geodesics that passes through these points. The new path lies of the sphere, has the same length as the original one, and is broken at the poles, thereby its length is not minimal.

To summarize *geodesics on a sphere is a part of the great circle that joins the starting and end points and which length is less that a half of the equator.*

Remark 1.1.2 This geometric consideration, when algebraically developed and generalized to larger class of extremal problems, yields to the so-called Jacobi test, see below Section **??**. The Jacobi test is violated if the length of geodesics is larger than π times the radius of the sphere.

The argument that the solution to the problem of shortest distance on a sphere bifurcates when its length exceeds a half of the great circle was in fact famously used by Columbus who argued that the shortest way to India passes through the Western route. As we know, Columbus wasn't be able to prove or disprove the conjecture because he bumped into American continent discovering New World for better and for worst.

1.1.4 Minimal surface

A three-dimensional generalization of the geodesics is the problem of the minimal surface that is the surface of minimal area stretched on a given contour. If the contour is plane, the solution is obvious: the minimal surface is a plane. The proof is quite similar to the above proof of the minimal distance on the plane.

In general, the contour can be any closed curve in three-dimensional space; the corresponding surface can be very complicated, and nonunique. It may contain several smooth branches with nontrivial topology (see the pictures). The example of such surface is provided by a soap film stretched on a contour made from a wire: the surface forces naturally minimize the area of the film. Theory of minimal surfaces is actively developing area, see the books [**?**, **?**].

In contrast with the complexity of a minimal surface in the large scale, caused by the complexity of the supporting contour, the local feature of any minimal surface is simple; we show that any smooth segment of the minimal surface has zero mean curvature.

We prove the result using an infinitesimal (variational) approach. Let S be an optimal surface, and s_0 be a regular point of it. Assume that S is a smooth surface in the neighborhood of s_o and introduce a local Cartesian coordinate system ξ_1, ξ_2, Z so oriented that the normal to the surface at a point s_0 coincides with the axes Z. The equation of the optimal surface can locally be represented as

$$Z = D + A\xi_1^2 + 2C\xi_1\xi_2 + B\xi_2^2 + o(\xi_1^2, \xi_2^2) = 0$$

Here, the linear with respect to ξ_1 and ξ_2 terms vanish because of orientation of Z-axis. In cylindrical coordinates r, θ, Z, the equation of the surface $F(r, \theta)$ becomes

$$0 \le r \le \epsilon, \quad \pi \le \theta \le \pi,$$

and

$$F(r, \theta) = D + a\, r^2 + b\, r^2 \cos(2\theta + \theta_0) + o(r^2) \tag{1.9}$$

Consider now a cylindrical ϵ-neighborhood of s_0 – a part $r \le \epsilon$ of the surface inside an infinite cylinder with the cental axes Z. The equation of the contour Γ– the intersection of S with the cylinder $r = \epsilon$ – is

$$\Gamma(\theta) = F(r, \theta)|_{r=\epsilon} = D + \epsilon^2 a + \epsilon^2 b \cos(2\theta + \theta_0) + o(\epsilon^2) \tag{1.10}$$

If the area of the whole surface is minimal, its area inside contour Γ is minimal among all surfaces that passes through the same contour. Otherwise, the surface could be locally changed without violation of continuity so that its area would be smaller.

In other words, the coefficients D, a, b, θ_0 of the equation (1.9) for an admissible surface should be chosen to minimize its area, subject to restrictions following from (1.10): The parameters b, θ_0 and $D + \epsilon^2 a$ are fixed. This leaves only one degree of freedom – parameter a – in an admissible smooth surface. Let us show that the optimal surface corresponds to $a = 0$.

We observe, as in the previous problem, that the surface area

$$A = \int_0^{2\pi} \int_0^{\epsilon} \left(\sqrt{1 + \left(\frac{\partial F}{\partial r}\right)^2 + \left(\frac{1}{r}\frac{\partial F}{\partial \theta}\right)^2} \right) r\, dr d\theta$$

is a strictly convex and even function of a (which can be checked by substitution of (1.10) into the formula and direct calculation of the second derivative). This implies that the minimum is unique and correspond to $a = 0$.

Another way is to use the approximation based on smallness of ϵ. The calculation of the integral must be performed up to ϵ^3, and we have

$$A = \pi\epsilon^2 + \frac{1}{2} \int_0^{2\pi} \int_0^{\epsilon} \left(\left(\frac{\partial F}{\partial r}\right)^2 + \left(\frac{1}{r}\frac{\partial F}{\partial \theta}\right)^2 \right) r\, dr d\theta + o(\epsilon^3).$$

After substitution of the expression for F from (1.9) into this formula and calculation, we find that

$$A = \pi\epsilon^2 + \frac{8}{3}\pi\epsilon^3(a^2 + b^2) + 0(\epsilon^3)$$

The minimum of A corresponds to $a = 0$ as stated. Geometrically, the result means that the mean curvature of a minimal surface is zero in any regular point. The minimal surface area

$$A_{\min} = \pi\epsilon^2 + \frac{8}{3}\pi\epsilon^3 b^2 + 0(\epsilon^3)$$

depends only on the total variation $2b = (\max\Gamma - \min\Gamma)$ of Γ as expected.

In addition, notice that the minimal area between all surfaces enclosed in a cylinder that do not need to pass through a fixed contour is equal to the area $\pi\epsilon^2$ of a circle and corresponds to a flat contour $b = 0$, as expected.

Proof by symmetry Another proof does not involve direct calculation of the surface. We only states that the minimal surface S locally is entirely determined by the infinitesimal contour Γ. Therefore, a transform of the coordinate system that keeps the contour unchanged cannot change the minimal surface inside it. Observe, that the infinitesimal contour (1.10) is invariant to transform

$$Z' = -Z + 2(D + \epsilon^2 a), \quad r' = r, \quad \theta' = \theta + 90^\circ. \tag{1.11}$$

that consists of reverse of the direction of Z axes, shift along Z, and rotation on 90° across this axes. The minimal surface (1.9) must be invariant to this transform as well, which again gives $a = 0$.

Remark 1.1.3 This proof assumes uniqueness of the minimal surface.

Thin film model The equation of the minimal surface can be deduced from the model of a thin film as well. Assume that the surface of the film shrinks by the inner tangent forces inside each infinitesimal element of it, and there are no bending forces generated that is forces normal to the surface. The tangent forces at a point depend only on local curvatures at this point.

Separate again the cylindrical neighborhood and replace the influence of the rest of the surface by the tangential forces applied to the surface at each point of the contour. Consider conditions or equilibrium of these forces and the inner tangent forces in the film. First, we argue that the average force applied to the contour is zero. This force must be directed along the z-axes, because the contour is invariant to rotation on 180° degree around this axes. If the average force (that depends only on the geometry) had a perpendicular to z component, this component would change its sign. The z-component of the average force applied to the contour is zero too, by the virtue of invariance of the transform (1.11). By the equilibrium condition, the average z-component of the tangent force inside the surface element must be zero as well.

Look of the representation (1.9) of the surface: The average over the area force F depends on a and b: $F = F(a, b)$. This average force is independent of θ_0 because of symmetry. The dependence on b is even, because the change of sign of b corresponds to 90° rotation of the contour that leaves the force unchanged.

The dependence on a is odd, because the change of the direction of the force correspond to change of the sign of a.

$$F = \text{ constant}(\theta_0), \quad F(a,b) = F(a,-b) = -F(-a,b) \quad \forall \theta_0, a, b.$$

Therefore, zero average force corresponds to $a = 0$, as stated.

The direction of average along the contour and over the surface forces cannot depend on b because the 180° degree rotation of the contour leaves is invariant, therefore the force remains invariant, too.

1.2 Convex envelope

1.2.1 Shortest path around an obstacle: Convex envelope

A helpful tool in the theory of extremal problem is the convex envelope. Here, we introduce the convex envelope of a finite set in a plane as the solution of a variational problem about the minimal path around an obstacle. The problem is to find the shortest closed contour that contains finite not necessarily connected domain Ω inside. This path is called the convex envelope of the set Ω.

Definition 1.2.1 (Convex envelope of a set) The convex envelope $\mathcal{C}\Omega$ of a finite closed set Ω is the minimal of the sets that (i) contain Ω inside, $\mathcal{C}\Omega \supset \Omega$ and (ii) is convex.

We argue that the minimal path Γ is convex, that is every straight line intersects its boundary not more than twice. Indeed, if a component is not convex, we may replace a part of it with a straight interval that lies outside of Γ thus finding another path Γ' that encircles a larger set but has a smaller perimeter. Perimeter of a convex set is decreased only when the encircled set Γ is lessen.

Also, the strictly convex (not straight) part of the path coincides with the boundary of Ω. Otherwise, the length of this boundary can be decreased by replacing an arc of it with the chord that lies completely outside of Ω.

We demonstrated that a convex envelope consists of at most two types of lines: the boundary of Ω and straight lines (shortcuts). The convex envelope of a convex set coincide with it, and the convex envelope of the of the set of finite number of points is a convex polygon that is supported by some of the points and contains the rest of them inside.

Properties of the convex envelope The following properties are geometrically obvious and the formal proofs of then are left to the interested reader.

1. Envelope cannot be further expanded.

$$\mathcal{C}(\mathcal{C}(\Omega)) = \mathcal{C}(\Omega)$$

2. Conjunction property:

$$\mathcal{C}(\Omega_1 \cup \Omega_2) \supseteq \mathcal{C}(\Omega_1) \cup \mathcal{C}(\Omega_2)$$

3. Absorbtion property: If $\Omega_1 \subset \Omega_2$ then

$$\mathcal{C}(\Omega_1 \cup \Omega_2) = \mathcal{C}(\Omega_2)$$

4. Monotonicity: If $\Omega_1 \subset \Omega_2$ then

$$\mathcal{C}(\Omega_2) \subseteq \mathcal{C}(\Omega_1)$$

Shortest trajectory in a plane with an obstacle

Find the shortest path $p(A, B, \Omega)$ between two points A and B on a plane if a bounded connected region (an obstacle) Ω in a plane between them cannot be crossed.

- First, split a plane into two semiplanes by a straight line that passes through the connecting points A and B.

- If the interval between A and B does not connect inner points of Ω, this interval is the shortest pass. In this case, the constraint (the presence of the obstacle) is *inactive*, $p(A, B, \Omega) = \|A - B\|$ independently of Ω.

- If the interval between A and B connects inner points of Ω, the constraint becomes *active*. In this case, obstacle Ω is divided into two parts Ω_+ and Ω_- that lie in the upper and the lower semiplanes, respectively, and have the common boundary – an interval ∂_0 – along the divide; ∂_0 lies inside the original obstacle Ω.

  Because of the connectedness of the obstacle, the shortest path lies entirely either in the upper or lower semiplane, but not in both; otherwise, the path would intersect ∂_0. We separately determine the shortest path in the upper and lower semiplanes and choose the shortest of them.

- Consider the upper semiplane. Notice that points A and B lie on the boundary of the convex envelope $\mathcal{C}(\Omega_+, A, B)$ of the set Ω and the connecting points A and B.

  The shortest path in the upper semiplane $p_+(A, B, \Omega)$ coincides with the upper component of the boundary of $\mathcal{C}(\Omega_+, A, B)$, the component that does not contains ∂_0. It consists of two straight lines that pass through the initial and final points of the trajectory and are tangents to the obstacle, and a part that passes along the boundary of the convex envelope $\mathcal{C}\Omega$ of the obstacle itself.

- The path in the lower semiplane is considered similarly. Points A and B lie on the boundary of the convex envelope $\mathcal{C}(\Omega_-, A, B)$. Similarly to the shortest path in the upper semiplane, the shortest path in the lower semiplane $p_-(A, B, \Omega)$ coincides with the lower boundary of $\mathcal{C}(\Omega_-, A, B)$.

- The optimal trajectory is the one of the two pathes $p_+(A, B, \Omega)$ and $p_-(A, B, \Omega)$; the one with smaller length.

  Analytical methods cannot tell which of these two trajectories is shorter, because this would require comparing of non-close-by trajectories; a straight calculation is needed.

If there is more than one obstacle, the number of the competing trajectories quickly raises.

Convex envelope supported at a curve Consider a slightly different problem: Find the shortest way between two points around the obstacle assuming that the these points lie on a curve that passes through the obstacle on the opposite sides of it. The points are free to move along the curve it this would decrease the length of the path. Comparing with the previous problem, we asking in addition where the points A and B are located. The position of the points depends on the shape of the obstacle and the curve, but it is easy to establish the conditions that must be satisfied at optimal location.

Problem: Show that an optimal location of the point A is either on the point of intersection of the line and an obstacle, or the optimal trajectory $p_-(A, B, \Omega)$ has a straight component near the point A and this component is perpendicular to the line at the point A.

Lost tourists

Finally, we consider a variation of the theme of convex envelope, the problem of the lost tourists. Crossing a plain, tourists have lost their way in a mist. Suddenly, they find a pole with a message that reads: "A straight road is a mile away from that pole." The tourists need to find the road; they are shortsighted in the mist: They can see the road only when they step on it. What is the shortest way to the road even if the road is most inauspiciously located?

The initial guess would suggest to go straight for a mile in a direction, then turn $90°$, and go around along the one-mile-radius circumference. This route meets any straight line that is located at the one mile distance from the central point. The length of this route is $1 + 2\pi \approx 7.283185$ miles.

However, a detailed consideration shows that this strategy is not optimal. Indeed, there is no need to intersect each straight line (the road) at the point of the circle but at any point and the route does not need to be closed. Any route that starts and ends at two points A and B at a tangent to a circle and goes around the circle intersects all other tangents to that circle. In other words, the convex envelope of the route includes a unit circle. The problem becomes: Find the curve that begins and ends at a tangent AB to the unit circle, such that (i) its convex envelope contains a circle and (ii) its length plus the distance $0A$ from the middle of this circle to one end of the curve is minimal.

The optimal trajectory consists of an straight interval OA that joints the central point O with a point A outside of the circle C and the convex envelope (ACB) stretched on the two points A and B and circle C.

The boundary of the convex envelope is either straight or coincide with the circle. More exactly, it consists of two straight intervals $A\,A_1$ supported by the point A and a point A_1 at the circumference and $A\,A_1$ supported by the end point B and a point B_1 of circumference. These intervals are tangent to the circumference at the points A_1 and B_1, respectively. Finally, line AB touches the circumference a point V.

Calculation The length L of the trajectory is

$$L = \mathcal{L}(O\,A) + \mathcal{L}(A\,A_1) + \mathcal{L}(A_1\,B_1) + \mathcal{L}(B_1\,B)$$

where $\mathcal{L}$ is the length of the corresponding component. These components are but straight lines and an circle's arch; the problem is thus parameterized. To compute the trajectory, we introduce two angles α and $-\beta$ from the point V there the line AB touches the circle. Because of symmetry, the points A_1 and B_1 correspond to the angles 2α and -2β, respectively, and we compute

$$\begin{aligned} \mathcal{L}(O\,A) &= \frac{1}{\cos\alpha}, \quad \mathcal{L}(A\,A_1) = \tan\alpha \\ \mathcal{L}(A_1\,B_1) &= 2\pi - 2\alpha - 2\beta, \quad \mathcal{L}(B_1\,B) = -\tan\beta, \end{aligned}$$

plug these expressions into the expression for L, solve the conditions $\frac{dL}{d\alpha} = 0$ and $\frac{dL}{d\beta} = 0$, and find optimal angles:

$$\alpha = \frac{\pi}{6}, \quad \beta = -\frac{\pi}{4},$$

The minimal length L equal to $L = \frac{7}{6}\pi + \sqrt{3} + 1 = 6.397242$.

Solution without calculation One could find solution to the problem without any trigonometry but with a bit of geometric imagination. Consider the mirror image C_m of the circle C assuming that the mirror is located at the tangent AB. Assume that the optimal route goes around that image instead of original circle; this assumption evidently does not change the length of the route. This new route consists of three pieces instead of four: The straight line $O\,A'_m$ that passes through the point O and is tangent to the circumference C_m, the part $A'_m\,B'_m$ of this circumference, and the straight line $B'_m B$ that passes through a point B on the line and is tangent to the circumference C_m.

The right triangle $O_m A'_m O$ has the hypothenuse $O'O$ equal to two and the side $O_m A'_m$ equal to one; the length of remaining side OA' equals to $\sqrt{3}$ and the angle $O_m O\,A'_m$ is $\frac{\pi}{3}$. The line $B'_m B$ is perpendicular to AB, therefore its length equals one. Finally, the angle of the arch $A'_m B'_m$ equals to $\frac{7}{6}\pi$. Summing up, we again obtain $L = \frac{7}{6}\pi + \sqrt{3} + 1$.

Generalization The generalization of the concept of convex envelope to the three-dimensional (or multidimensional) sets is apparent. The problem asks for set of minimal surface area that contains a given closed finite set. The solution is again given by the convex envelope, definition (2.1.1) is applicable for the similar reasons.

Consider the three-dimensional analog of the problem 1.2.1 assuming in addition that the obstacle Ω is convex. Repeating the arguments for the plane problem, we conclude that the optimal trajectory belongs to the convex envelope $\mathcal{C}(\Omega, A, B)$. The envelope is itself a convex surface and therefore the problem is reduced to geodesics on the convex set – the envelope $\mathcal{C}(\Omega, A, B)$. The variational analysis of this problem allows to disqualify as optimal all (or almost all) trajectories on the convex envelope one by comparing near-by trajectories that touch the obstacle in close-by points.

If the additional assumption of convexity of obstacle is lifted, the problem becomes much more complex because the passes through "tunnels" and in folds in the surface of Ω should be accounted for. If at least one of the points A or B lies inside the convex envelope of a nonconvex obstacle, the minimal path partly goes inside the convex envelope $\mathcal{C}\Omega$ as well. We leave this for the interested reader.

1.3 Symmetrization

An interesting geometric method, symmetrization, is based on convexity inequality; it allows for solution of several isoperimetric problem. The detailed discussion can be found in the books by Blaschke [], Pólya and Szegö [**?**]

The idea of symmetrization Consider a plane finite domain Ω and a straight line A. The transformation of Ω is called a symmetrization with respect to A if it moves each interval that crosses Ω and is orthogonal to A parallel to itself so that the middle of the interval belongs to A.

One can easily see that the symmetrization of a polygon is a polygon with equal or larger number of angles than the original one.

1.3.1 Symmetrization of a triangle and quadrangle

Let us prove that unilateral triangle has minimal perimeter among all triangles with given area.

Consider an arbitrary nonunilateral triangle ABC and apply symmetrization to it. Generally, the symmetrization transforms a triangle into a quadrangle; the triangle remains a triangle only if the axis of symmetrization is orthogonal to one of the side. In this case, an arbitrary triangle becomes an isosceles triangle, the base a and the hight h remain unchanged. This implies that symmetrization leaves the area A of the triangle unchanged.

Let us show that symmetrization decreases the perimeter. Let the coordinates of the vertexes be

$$A = (a, 0), \quad B = (-a, 0), \quad C = (c, h)$$

and let the axes of symmetrization be the Y axes. After symmetrization, the coordinates A and B remain the same, and the vertex C moves to $C' = (0, h)$.

The sum of the two sides' lengths equal to

$$L = \sqrt{h^2 + (a-c)^2} + \sqrt{h^2 + (a+c)^2}$$

becomes

$$L_S = 2\sqrt{h^2 + a^2}.$$

We prove that

$$L_S \le L \tag{1.12}$$

and the equality sign corresponds to the only case $c = 0$.

Consider the length of a side as a function of c:

$$L(c) = \sqrt{h^2 + (a+c)^2}$$

the function f is strictly convex since

$$L''(c) = \frac{h^2}{(h^2 + (a+c)^2)^{\frac{3}{2}}} > 0$$

and is even. The inequality of convexity (**??**) implies that

$$L(c) + L(-c) \ge 2f(0)$$

that is, the inequality (1.12)

If the obtained rectangle is not unilateral, the symmetrization procedure can be repeated, using one of the equal legs as the base. The new triangle has the same area and smaller perimeter.

Consider now a sequence of symmetrizations applied to an arbitrary triangle. On each step, symmetrization preserves one side, makes two other sides equal to each other, and decreases their total length. The area of the triangle is preserved, its perimeter decreases and is obviously bounded from below, say by zero. Therefore the sequence of symmetrizations is a monotone bounded sequence and it must have a unique stable point: A triangle that is stable against symmetrization. This is of course a unilateral triangle. We have proved the theorem:

Theorem 1.3.1 Among all triangles with equal area, the unilateral triangle has the smallest perimeter.

Symmetrization of a quadrangle Let is apply symmetrization to an arbitrary quadrangle, requiring that the quadrangle remains quadrangle after the symmetrization.

At the first step, we have to perform symmetrization orthogonal to one of two diagonals. The resulting quadrangle has two pairs of neighboring sides of equal lengths. At the second step, we symmetrize orthogonally to the other diagonal, the resulting figure is a rhombus of equal area but of smaller perimeter than the original quadrangle.

Now we may start a two-steps sequence of symmetrizations. Firstly, we transform the rhombus into a rectangle using the side as an axis of symmetrization. Secondly, we transform the rectangle back to rhombus, using the diagonal as the axis of symmetrization. The obtained rhombus has smaller ratio of the larger diagonal to the smaller one (compute the change of this ratio!) and the smaller perimeter, but its area stays unchanged. This monotonic sequence has a stable point. The stable point is the square, which enables us to formulate the next theorem.

Theorem 1.3.2 Among all quadrangles with equal area, the square has the smallest perimeter.

1.3.2 Circle: A stable manifold

Symmetrization can be applied to an arbitrary finite bounded domain $F(x, y) \geq 0$ with the boundary $F(x, y) = 0$. For definiteness let assume that the y-axis is the axis of symmetrization.

Dissect the plane by a family $\{y_k\}$ of equidistant parallel lines

$$y_0, \quad y_1 = y_0 + \Delta, \quad y_2 = y_0 + 2\Delta, \ldots, \quad y_N = y_0 + N\Delta.$$

Assume that this division covers the figure $F(x, y) = 0$ and that the number N is arbitrary large so that the distance between two neighboring parallel lines is infinitesimal.

An infinitesimal part of the domain $F(x, y) = 0$ located between two close-by parallel lines can be approximated by a trapezoid. Symmetrization replaces this trapezoid by a equilateral trapezoid of equal area, parallel sides of equal length, but with smaller total length of the non-parallel sides (show this!). We can formulate

Theorem 1.3.3 The total area of the symmetrized domain remains constant, but its perimeter (equal to the sum of the lengths of the sides of the trapezoids) decreases.

Now consider the sequence of symmetrization with variable axis. The sequence of the transformed figures tends to a circle: The only figure that is stable against any symmetrization. Indeed, this sequence tends to its unique stable point, and the circle is that point.

We came to the theorem

Theorem 1.3.4 Among all plane domains with equal finite area, the circle has the smallest perimeter.

Geometric proof of the theorem An independent geometric proof of the theorem is elegant and does not require any infinitesimal operation. However, we need to assume the existence of an optimal shape which we do not need to do in the previous consideration.

The proof requires the following steps:

1. We show that the optimal domain is convex. If it is not convex, we pass to the convex envelope increasing the area and decreasing the perimeter at the same time.

2. We cut the optimal domain by a straight line so that both parts have the same area. This is always possible by moving a line across the domain and keeping it parallel to itself. The two cut parts must have the same perimeter too, otherwise the perimeter could be decreased replacing the part with larger perimeter with the mirror image of the part with smaller perimeter. The replacement of one of the domain with the mirror image of the second one changes neither the area nor the perimeter.

3. Consider the half of the optimal shape with the straight base. Choose an arbitrary point C on it surface and connect it with the ends of the base by two straight intervals. The domain is thus divided into two outer shapes and a triangle.

4. We may change the length of the base without changing the perimeter. This change keeps the areas of two outer domains constant but varies the angle and by the area of the triangle. The maximal area of the triangle corresponds to the angle C opposite the base being equal $90°$. Indeed, by the geometric theorem the area A equals to

$$A = \frac{1}{2} a\, b\, \sin C$$

where the lengths a and b of the intervals are constant to the motion and the angle arbitrary varies. The maximal area corresponds to $C = 90°$.

5. Because the point C was arbitrarily chosen, the angle between any point of the surface and the base is equal $90°$. The figure must be a circle: the set of points from which an interval (diameter) is visible on a right angle.

1.3.3 Dido problem

Probably the first extremal problem known from the antic time is the Dido problem. The problem is based on a passage from Virgil's Aeneid (cited from []):

> "The Kingdom you see is Carthage, the Tyrians, the town of Agenor;
> "But the country around is Libya, no folk to meet in war.
> Dido, who left the city of Tyre to escape her brother,
> Rules here – a long a labyrinthine tale of wrong
> Is hers, but I will touch on its salient points in order
> ...
> Dido, in great disquiet, organized her friends for escape.

> They met together, all those who harshly hated the tyrant
> Or keenly feared him: they seized some ships which chanced to be ready
> ...
> They came to this spot, where to-day you can behold the mighty
> Battlements and the rising citadel of New Carthage,
> And purchased a site, which was named 'Bull's Hide' after the bargain
> By which they should get as much land as they could enclose with a bull's hide."

According to the legend, the Trojans arrived in the North African shore of Mediterranean Sea after the defeat by Greeks. Here, their leader, wise queen Dido, purchased from the local tribe a piece land on the shore "that can be covered by the bull's hide." Sophisticated Trojans had a much more advanced technology than the locals; in particular, they knew how to use sharp knifes to cut hides into thin strips (and they knew some math, too!). So, they made a long leather rope out of the hide and encircled by it enough land to build the city of Carthage who later become a mighty rival of Rome.

Dido brilliantly solved the following extremal problem: Given a curve of a given length (the rope) and a straight line (the sea shore), encircle the domain of maximal area (place for future Carthage). This problem, known as *Dido problem*, inspirited many generations by its cleverness; it influenced the development of theory of extremal problems, demonstrated usefulness of mathematics, and accustomed people to respect political leaders able to use brains instead of brutal force.

Dido problem can be solved by symmetrization together with the following trick: Assume that the seashore is a mirror and consider the domain Ω of the enclosed land and its mirror image; obviously, the perimeter and area of Ω is twice larger than the perimeter and area of the enclosed domain, respectively. The symmetrization tells that Ω is a circle; thereby, the answer to Dido problem is a semicircle with the shore as a diameter and the rope as a semi-circumference.

The reference of how to use Maple to work on Dido problem:
http://www.mapleapps.com/powertools/engineeringmath/html/Section

1.3.4 Formalism of symmetrization

The considered symmetrization of a plane domain can be formalized as following: Assume for simplicity that the boundary of the set F is y-simple: The set $F(x, y) \geq 0$ described as

$$f_-(x) \leq y \leq f_+(x), \quad a \leq x \leq b$$

The area A of the domain is equal to

$$A = \int_a^b (f_+ - f_-)dx \tag{1.13}$$

and the perimeter P is

$$P = \int_a^b \left(\sqrt{1 + (f'_+)^2} + \sqrt{1 + (f'_-)^2} \right) dx$$

The symmetrized domain is described as

$$-\frac{1}{2} \left(f_+(x) - f_-(x) \right) \le y \le \frac{1}{2} \left(f_+(x) - f_-(x) \right), \quad a \le x \le b$$

Its area A_S of the symmetrized domain is obviously given by the formula (1.13) and its perimeter P_S is

$$P_S = 2 \int_a^b \sqrt{1 + \frac{1}{4}(f'_+ - f'_-)^2}\, dx$$

If remains to prove that $P \ge P_S$ or

$$\int_a^b \left(\sqrt{1 + (f'_+)^2} + \sqrt{1 + (f'_-)^2} - 2\sqrt{1 + \frac{1}{4}(f'_+ - f'_-)^2} \right) dx \ge 0$$

We show that the integrant is nonnegative in each point. Starting with the inequality

$$\sqrt{1 + (f'_+)^2} + \sqrt{1 + (f'_-)^2} \ge 2\sqrt{1 + \frac{1}{4}(f'_+ - f'_-)^2}$$

we square its left- and right-hand sides, cancel equal terms, and obtain an equivalent inequality

$$\sqrt{\left(1 + (f'_+)^2\right) \left(1 + (f'_-)^2\right)} \ge 1 - f'_+ f'_-$$

If the right-hand side is negative, the inequality is true, otherwise square it one more time and obtain the true equivalent inequality

$$(f'_+ + f'_-)^2 \ge 0$$

The result is proved.

Remark 1.3.1 If the contour is not y-simple, the result remain the same.

3D symmetrization

Consider a bounded body

$$F(x, y, z) \ge 0$$

in three-dimensional space with the boundary

$$F(x, y, z) = 0.$$

Dissect it by a family of equidistant parallel planes

$$z = z_0, z = z_0 + \Delta, z = z_0 + 2\Delta, \ldots, z = z_0 + N\Delta.$$

Replace a part of the body located between two planes by a conical surface, replacing each closed contour $F(x, y, z_0 + k\Delta) = 0$ by the circle of equal area, all centered at the z-axis

$$x^2 + y^2 = r_k^2, \quad \text{where } \pi r_k^2 = \text{Area of } F(x, y, z_0 + k\Delta)$$

Doing this, we obtain a body of revolution defined by the curve $r(z)$ that revolves around the z-axis.

We can show (do it yourself or look into [**?**]) that this transformation called Schwartz symmetrization (i) conserves the volume of the body and (ii) decreases its surface area.

Particularly, consider the domain bounded by the plane $z = 0$ and a non-negative surface $z = u(x, y) \geq 0$ such that $u(x, y) = 0$ if $(x, y) \in \Gamma = \partial\Omega$. The symmetrization

1. Replaces the base Ω with a circle of equal area:

$$\Omega_S = \text{ A circle: } |\Omega| = |\Omega_S|$$

2. Conserves the volume:

$$\int_\Omega u \, dx \, dy \quad \text{is stable to symmetrization}$$

3. Decreases the surface area:

$$\int_\Omega \sqrt{1 + (\nabla u)^2} \, dx \, dy \quad \text{decreases by symmetrization} \tag{1.14}$$

Using symmetrization, we may deduct some inequalities for the functionals different from the volume or the area. For example, assuming that $u(x, y) \ll 1$, we notice that (1.14) implies the decrease of the Dirichlet integral:

$$\int_\Omega (\nabla u)^2 \, dx \, dy \quad \text{decreases by symmetrization}$$

Here $u(x, y)$ is a differentiable function such that $u = 0$ on the contour $\partial\Omega$.

Extremal property of the sphere As in two-dimensional case, one applies the series of symmetrization around all axes, look into the resulting stable point and arrive at the theorem:

Theorem 1.3.5 Among all three-dimensional bodies with equal finite volume, the sphere has the smallest surface area.

Limits of the method

The method of symmetrization operates with special type of functionals (area, perimeter, volume).

It cannot handle any additional constraints besides the fixed area, such the requirement that a part of the boundary stays unchanged. In particular, it does not preserve the number of edges in polygons of more than fourth order.

1.3.5 Summary

The sufficient conditions are the most elegant statements in the theory extremal problems. In these methods, the guessed optimal solution is directly compared with all admissible solutions; thus the global optimum of the functional is proven. By its nature, a sufficient conditions technique is irregular and the area of its applicability is limited.

Symmetrization shows that is many problem a symmetric solution is better than a nonsymmetric one. This principle is reflected in an intuitive preference to symmetric designs which are often considered to be more elegant or beautiful that nonsymmetric ones.

1.4 Problems

1. Use Jensen inequality to prove the relation between arithmetic and harmonic means:
$$\frac{a_1 + \ldots + a_N}{N} \geq (a_1 \cdot \ldots \cdot a_N)^{\frac{1}{N}} \quad \forall a_1 \geq 0, \ldots a_N \geq 0$$

2. Describe the area of a symmetrized ellipse.

Chapter 2

Nonconvex Lagrangians

2.1 Irregular solutions

The classical approach to variational problems assumes that the optimal trajectory is a differentiable curve – a solution to the Euler equation that, in addition, satisfies the Weierstrass and Jacobi tests. In this chapter, we consider the variational problems which solutions do not satisfy necessary conditions of optimality. Either the Euler equation does not have solution, or Jacobi or Weierstrass tests are not satisfied at any stationary solution; in any case, the extremal cannot be found from stationarity conditions. We have seen such solution in the problem of minimal surface (Goldschmidt solution, Section **??**).

A minimization problem always can be solved by *a direct method* that is by constructing a corresponding minimizing sequence, the functions $u^s(t)$ with the property $I(u^s) \geq I(u^{s+1})$. The functionals $I(u^s)$ form a monotonic sequence of real number that converges to a real or improper limit. In this sense, every variational problem can be solved, but the limiting solution $\lim_{s\to\infty} u^s$ may be irregular; in other terms, it may not exist in an assumed set of functions. Especially, derivation of Euler equation uses an assumption that the minimum is a differentiable function. This assumption leads to complications because the set of differentiable functions is open and the limits of sequences of differentiable functions are not necessary differentiable functions themselves.

We recall several types of sequences that one meets in variational problems

Example 2.1.1 (Various limits of functional sequences)

- The limit $\delta(x)$ of the sequence of infinitely differentiable function

$$\phi_n(x) = \frac{n}{2\pi} \exp\left(-\frac{x^2}{2n}\right)$$

is not a function but a distribution - the δ function.

- The limit $H(x)$ of the sequence of antiderivatives of these infinitely differentiable functions is a discontinuous Heaviside function,

$$H(x) = \int_{-\infty}^{x} \phi_n(t)dt = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 1 \end{cases}$$

- The limit of the sequence of oscillating functions

$$\lim_{n\to\infty} \sin(nx)$$

  does not exist for any $x \neq 0$.

- The sequence $\{\phi_n(x)\}$, where $\phi_n(x) = \frac{1}{\sqrt{n}}\sin(nx)$ converges to zero pointwise, but the sequence of the derivatives $\phi_n(x)' = \sqrt{n}\cos(nx)$ does not converges and is unbounded everywhere.

These or similar sequences can represent minimizing sequences in variational problems. Here we give a brief introduction to the methods to deal with such "exotic" solutions.

As always, we try to find an analogy of irregular solutions in finite-dimensional minimization problems. Consider such a problem $\min_{x\in R_n} F(x)$. The minimum may either correspond to the regular stationary point where $\nabla F(x) = 0$, or to an irregular point where $\nabla F(x)$ is not defined or its norm is unbounded, or to improper x. It is natural to expect, that in variational problems where the minimizing functions $u(x)$ belong to more complex than R_n sets and and are bounded by additional requirements of differentiability, the number of irregular cases growths and causes for these cases are more diverse.

How to deal with irregular problems The possible nonexistence of minimizer poses several challenging questions. Some criteria are needed to establish which problems have a classical solution and which do not. These criteria analyze the type of Lagrangians and result in existence theorems.

There are two alternative ideas in handling problems with nondifferentiable minimizers. The admissible class of minimizers can be enlarged and closed in such a way that the "exotic" limits of minimizers would be included in the admissible set. This *relaxation* procedure, underlined in the Hilbert's quotation, motivated the introduction of distributions and the corresponding functional spaces, as well as development of relaxation methods. Below, we consider several ill-posed problems that require rethinking of the concept of a solution.

Alternatively, the minimization problem can be constrained so that the "exotic" behavior of the solutions is penalized and the minimizer will avoid it; this approach called *regularization*, forces the problem to select a classical solution at the expense of increasing the value of the objective functional. When the penalization decreases, the solution tends to the solution of the original problem, remaining conventional. An example of this approach is the *viscosity solution* developed for dealing with the shock waves.

Existence of a differentiable minimizer We formulate here a list of conditions guarantying the smooth classical solution to a variational problem.

1. The Lagrangian grows superlinearly with respect to u':

$$\lim_{|u'|\to\infty} \frac{F(x,u,u')}{|u'|} = \infty \quad \forall x, u(x) \tag{2.1}$$

   This condition forbids any finite jumps of the optimal trajectory $u(x)$; any such jump leads to an infinite penalty in the problem's cost.

2. The cost of the problem increases when $|u| \to \infty$. This condition forbids a blow-up of the solution.

3. The Lagrangian is convex with respect to u':

$$F(x,u,u') \quad \text{is a convex function of } u' \quad \forall x, u(x)$$

   at the optimal trajectory u. This condition forbids infinite oscillations because they would increase the cost of the problem.

Let us outline the idea of the proof:

1. First two conditions guarantee that the limit of any minimizing sequence is bounded and has a bounded derivative. The cost of the problem unlimitedly grows when either the function or its derivative tend to infinity at a set of nonzero measure.

2. It is possible to extract a weakly convergent subsequence $u^S \rightharpoonup u^0$ from a weakly bounded minimizing sequence. Roughly, this means that the subsequence $u^\epsilon(x)$ in a sense approximates a limiting function u^0, but may wiggle around it infinitely often.

3. Next, we need the property of *lower weakly semicontinuity* of the objective functional $I(u)$. The lower weakly semicontinuity states that

$$\lim_{u^S \rightharpoonup u^0} I(u^s) \geq I(u^0)$$

   We illustrate this property on the following examples.

   Example 2.1.2 The weak limit of the sequence $u^s = \sin(s\,x)$ is zero.

$$\sin(s\,x) \rightharpoonup 0 \quad s \to \infty$$

   Compute the limit of the functional

$$I_1(u^s) = \int_0^1 (u^s)^2 dx$$

We have

$$\lim_{s\to\infty}\int_0^1 \sin^2(s\,x)dx = \frac{1}{2}\lim_{s\to\infty}\int_0^1 (1-\cos(2s\,x)dx = \frac{1}{2}$$

and we observe that

$$\lim_{u^S\to u^0} I_1(u^s) > I(u^0) = 0$$

The limit of the functional

$$I_2(u^s) = \int_0^1 \left((u^s)^4 - (u^s)^2\right) dx$$

is smaller than $I_2(0)$. Indeed,

$$\lim_{s\to\infty}\int_0^1 \left(\sin^4(s\,x) - \sin^2(s\,x)\right) dx = -\frac{1}{4}$$

or

$$\lim_{u^S\to u^0} I_2(u^s) < I(u^0) = 0$$

The wiggling minimizing sequence u^s increases the value of the first functional and decrease the value of the second. The fist functional corresponds to convex integrand and is weakly lower semicontinuous.

The convexity of Lagrangian eliminates the possibility of wiggling, because the cost of the problem with convex Lagrangian is smaller for a smooth function than on any close-by wiggling function by virtue of Jensen inequality. The functional of a convex Lagrangian is lower weakly semicontinuous.

2.1.1 Formalism of convex envelopes

In dealing with nonconvex variational problems, the central idea is to *relax* them replacing the nonconvex Lagrangian with its convex envelope. We already introduced the convex envelope of sets in R^n. Here we transform the notion of convex envelope from sets to functions.

A graph of any function $y = f(x)$ divides the space into two sets, and the convex envelope of a function is the convex envelope of the set $y > f(x)$. It the function is not defined for all $x \in R^n$ (like $\log x$ is defined only for $x \geq 0$), we extend the definition of a function assigning the improper value $+\infty$ to function of in all undefined values arguments.

There are two dual description of the convex envelope. One can either define it as a unity of all planes that lie below the graph of the function, or as a unity of all intervals that join two points on that graph

They are formalized as follows.

Figure 2.1: Left: Convex envelope as a unity of lines, Right: Convex envelope as a unity of intervals

Definition 2.1.1 (Convex envelope of a function) The convex envelope $\mathcal{C}f(xv)$ of a function $f : R^n \to R^1$ is the maximal of the set of affine function $g(v) = a^T v + b$ that do not surpass $f(v)$ everywhere [?]:.

$$\mathcal{C}F(\boldsymbol{v}) = \max_{a,b} \phi(\boldsymbol{v}) : \ \phi(\boldsymbol{v}) \leq F(\boldsymbol{v}) \ \forall \boldsymbol{v} \quad \text{and } \phi(\boldsymbol{v}) \text{ is convex.} \tag{2.2}$$

Remark 2.1.1 In the above definition, one can replace the set of affine functions with convex functions.

The Jensen's inequality produces the following definition of the convex envelope:

Definition 2.1.2 The *convex envelope* $\mathcal{C}F(v)$ is a solution to the following minimal problem:

$$\mathcal{C}F(\boldsymbol{v}) = \inf_{\boldsymbol{\xi}} \frac{1}{l} \int_0^l F(\boldsymbol{v} + \boldsymbol{\xi}) dx \quad \forall \ \boldsymbol{\xi} : \int_0^l \boldsymbol{\xi} dx = 0. \tag{2.3}$$

This definition determines the convex envelope as the minimum of all parallel secant hyperplanes that intersect the graph of F; it is based on Jensen's inequality (**??**).

To compute the convex envelope $\mathcal{C}F$ one can use the Carathéodory theorem (see [**?**, **?**]). It states that the argument $\boldsymbol{\xi}(x) = [\xi_1(x), \ldots, \xi_n(x)]$ that minimizes the right-hand side of (2.3) takes no more than $n+1$ different values. This theorem refers to the obvious geometrical fact that the convex envelope consists of the supporting hyperplanes to the graph $F(\xi_1, \ldots, \xi_n)$. Each of these hyperplanes is supported by no more than $(n+1)$ points. For example, a line $(x \in R^1)$ is supported by two points, a plane $(x \in R^2)$ – by three points. These points are called supporting points.

The Carathéodory theorem allows us to replace the integral in the right-hand side of (2.3) in the definition of $\mathcal{C}F$ by the sum of $n+1$ terms; the definition (2.3) becomes:

$$\mathcal{C}F(\boldsymbol{v}) = \min_{m_i \in M} \min_{\boldsymbol{\xi}_i \in \boldsymbol{\Xi}} \left\{ \sum_{i=1}^{n+1} m_i F(\boldsymbol{v} + \boldsymbol{\xi}_i) \right\}, \tag{2.4}$$

where

$$M = \left\{ m_i : \quad m_i \geq 0, \quad \sum_{i=1}^{n+1} m_i = 1 \right\} \tag{2.5}$$

and

$$\boldsymbol{\Xi} = \left\{ \ \boldsymbol{\xi}_i : \quad \sum_{i=1}^{n+1} m_i \boldsymbol{\xi}_i = 0 \ \right\}. \tag{2.6}$$

The convex envelope $\mathcal{C}F(\boldsymbol{v})$ of a function $F(\boldsymbol{v})$ at a point $\boldsymbol{v}$ coincides with either the function $F(\boldsymbol{v})$ or the hyperplane that touches the graph of the function F. The hyperplane remains below the graph of F except at the tangent points where they coincide.

The position of the supporting hyperplane generally varies with the point $\boldsymbol{v}$. A convex envelope of F can be supported by fewer than $n+1$ points; in this case several of the parameters m_i are zero. Generally, the are only n parameters that vary, some them are coordinates of of the supporting points, other are coordinates of the points

Example 2.1.3 Obviously, the convex envelope of a convex function coincides with the function itself, so all m_i but m_1 are zero in (2.26) and $m_1 = 1$; the parameter $\boldsymbol{\xi}_1$ is zero because of the restriction (2.6).

The convex envelope of a "two-well" function,

$$\Phi(\boldsymbol{v}) = \min\left\{F_1(\boldsymbol{v}),\ F_2(\boldsymbol{v})\right\}, \tag{2.7}$$

where $F_1,\ F_2$ are convex functions of $\boldsymbol{v}$, either coincides with one of the functions $F_1,\ F_2$ or is supported by no more than two points for every $\boldsymbol{v}$; supporting points belong to different wells. In this case, formulas (2.26)–(2.6) for the convex envelope are reduced to

$$\mathcal{C}\Phi(\boldsymbol{v}) = \min_{m,\boldsymbol{\xi}}\left\{mF_1(\boldsymbol{v} - (1-m)\boldsymbol{\xi}) + (1-m)F_2(\boldsymbol{v} + m\boldsymbol{\xi})\right\}. \tag{2.8}$$

Indeed, the convex envelope touches the graphs of the convex functions F_1 and F_2 in no more than one point. Call the coordinates of the touching points $\boldsymbol{v} + \boldsymbol{\xi}_1$ and $\boldsymbol{v} + \boldsymbol{\xi}_2$, respectively. The restrictions (2.6) become $m_1\boldsymbol{\xi}_1 + m_2\boldsymbol{\xi}_2 = 0,\ m_1 + m_2 = 1$. It implies the representations $\boldsymbol{\xi}_1 = -(1-m)\boldsymbol{\xi}$ and $\boldsymbol{\xi}_2 = m\boldsymbol{\xi}$.

Example 2.1.4 Consider the special case of the two-well function,

$$F(v_1, v_2) = \begin{cases} 0 & \text{if } \ v_1^2 + v_2^2 = 0, \\ 1 + v_1^2 + v_2^2 & \text{if } \ v_1^2 + v_2^2 \neq 0. \end{cases} \tag{2.9}$$

The convex envelope of F is equal to

$$\mathcal{C}F(v_1, v_2) = \begin{cases} 2\sqrt{v_1^2 + v_2^2} & \text{if } \ v_1^2 + v_2^2 \leq 1, \\ 1 + v_1^2 + v_2^2 & \text{if } \ v_1^2 + v_2^2 > 1. \end{cases} \tag{2.10}$$

Here the envelope is a cone if it does not coincide with F and a paraboloid if it coincides with F.

Indeed, the graph of the function $F(v_1, v_2)$ is rotationally symmetric in the plane $v_1,\ v_2$; therefore, the convex envelope is symmetric as well: $\mathcal{C}F(v_1, v_2) = f(\sqrt{v_1^2 + v_2^2})$. The convex envelope $\mathcal{C}F(\boldsymbol{v})$ is supported by the point $\boldsymbol{v} - (1-m)\boldsymbol{\xi} = \mathbf{0}$ and by a point $\boldsymbol{v} + m\boldsymbol{\xi} = \boldsymbol{v}^0$ on the paraboloid $\phi(\boldsymbol{v}) = 1 + v_1^2 + v_2^2$. We have

$$\boldsymbol{v}^0 = \frac{1}{1-m}\boldsymbol{v}$$

and

$$\mathcal{C}F(\boldsymbol{v}) = \min_m \left\{ (1-m)\phi\left(\frac{1}{1-m}\boldsymbol{v}\right)\right\}. \tag{2.11}$$

The calculation of the minimum gives (2.10).

Example 2.1.5 Consider the nonconvex function $F(v)$ used in Example **??**:

$$F(v) = \min\{(v-1)^2, (v+1)^2\}.$$

It is easy to see that the convex envelope $\mathcal{C}F$ is

$$\mathcal{C}F(v) = \begin{cases} (v+1)^2 & \text{if } \; v \le -1, \\ 0 & \text{if } \; v \in (-1,1), \\ (v-1)^2 & \text{if } \; v \ge 1. \end{cases}$$

Example 2.1.6 Compute convex envelope for a more general two-well function:

$$F(v) = \min\{(av)^2, (bv+1)^2\}.$$

The envelope $\mathcal{C}F_n(v)$ coincides with either the graph of the original function or the linear function $l(v) = A\,v + B$ that touches the original graph in two points (as it is predicted by the Carathéodory theorem; in this example $n = 1$). This function can be found as the common tangent $l(v)$ to both convex branches (wells) of $F(v)$:

$$\begin{cases} l(v) = av_1^2 + 2av_1(v - v_1), \\ l(v) = (bv_2^2 + 1) + 2bv_2(v - v_2), \end{cases} \tag{2.12}$$

where v_1 and v_2 belong to the corresponding branches of F_p:

$$\begin{cases} l(v_1) = av_1^2, \\ l(v_2) = bv_2^2 + 1. \end{cases} \tag{2.13}$$

Solving this system for $v,\ v_1,\ v_2$ we find the coordinates of the supporting points

$$v_1 = \sqrt{\frac{b}{a(a-b)}}, \quad v_2 = \sqrt{\frac{a}{b(a-b)}}, \tag{2.14}$$

and we calculate the convex envelope:

$$\mathcal{C}F(v) = \begin{cases} av^2 & \text{if } \; |v| < v_1, \\ 2v\sqrt{\frac{ab}{a-b}} - \frac{b}{a-b} & \text{if } \; v \in [v_1,\ v_2], \\ 1 + bv^2 & \text{if } \; |v| < v_2 \end{cases} \tag{2.15}$$

that linearly depends on v in the region of nonconvexity of F.

Hessian of Convex Envelope We mention here an algebraic property of the convex envelope that we will use later. If the convex envelope $\mathcal{C}F(\boldsymbol{v})$ does not coincide with $F(\boldsymbol{v})$ for some $\boldsymbol{v} = \boldsymbol{v}_n$, then the graph of $\mathcal{C}F(\boldsymbol{v}_n)$ is convex, but not strongly convex. At these points the Hessian $He(F) = \frac{\partial^2}{\partial v_i \partial v_j} F(\boldsymbol{v})$ is semi-positive; it satisfies the relations

$$He(\mathcal{C}F(\boldsymbol{v})) \geq 0, \quad \det H(\mathcal{C}F(\boldsymbol{v})) = 0 \quad \text{if} \;\; \mathcal{C}F < F, \tag{2.16}$$

which say that $He(\mathcal{C}F)$ is a nonnegative degenerate matrix. These relations can be used to compute $\mathcal{C}F(\boldsymbol{v})$. For example, compute the Hessian of the convex envelope $\mathcal{C}F(v_1, v_2) = \sqrt{v_1^2 + v_2^2}$ obtained in Example 2.1.4. The Hessian is

$$He\left(\sqrt{v_1^2 + v_2^2}\right) = \frac{1}{(v_1^2 + v_2^2)^{\frac{3}{2}}} \begin{pmatrix} v_1^2 & v_1 v_2 \\ v_1 v_2 & v_2^2 \end{pmatrix}$$

and its determinant is clearly zero.

Comparing the minimization problems

$$I = \min_{x \in R^n} F(x) \quad \text{and} \quad I_c = \min_{x \in R^n} \mathcal{F}(x)$$

we observe that (i) $I = I_c$ – the minimum of a function coincides with the minimum of its convex envelope, and (ii) the convex envelope of a function does not have local minima but only not global one.

Remark 2.1.2 (Convex envelope as second conjugate) We may as well compute convex envelope in more regular way as a second conjugate of the original function as described later in Section ??.

Convex envelope are used below in the next Section to address ill-posed variational problems.

2.2 Infinitely oscillatory solutions: Relaxation

2.2.1 Nonconvex Variational Problems.

Consider the variational problem

$$\inf_u J(u), \quad J(u) = \inf_u \int_0^1 F(x, u, u')dx, \quad u(0) = a_0, \; u(1) = a_1 \tag{2.17}$$

with Lagrangian $F(x, \boldsymbol{y}, \boldsymbol{z})$ and assume that the Lagrangian is nonconvex with respect to $\boldsymbol{z}$, for some values of z, $z \in \mathcal{Z}_\mathrm{f}$.

Definition 2.2.1 We call the *forbidden region* $\mathcal{Z}_\mathrm{f}$ the set of $\boldsymbol{z}$ for which $F(x, \boldsymbol{y}, \boldsymbol{z})$ is not convex with respect to $\boldsymbol{z}$.

The Weierstrass test requires that the derivative u' of an extremal never assume values in the set $\mathcal{Z}_{\rm f}$,

$$u' \notin \mathcal{Z}_{\rm f}. \tag{2.18}$$

On the other hand, a stationary trajectory u may be required by Euler equations to pass through this set. Such trajectories fail the Weierstrass test and must be rejected. We conclude that the true minimizer (the limit of a minimizing sequence) is not a classical differentiable curve, otherwise it would satisfy both the Euler equation and the Weierstrass test.

We will demonstrate that a minimizing sequence tends to a "generalized curve." It consists of infinitely many infinitesimal zigzags. The derivative of the minimizer "jumps over" the forbidden set, and does it infinitely often. Because of these jumps, the derivative of a minimizer stays outside of the forbidden interval but its average can take any value within or outside the forbidden region. The limiting curve – the minimizer – has a dense set of points of discontinuity of the derivative.

Example of a nonconvex problem Consider a simple variational problem that yields to an irregular solution [**?**]:

$$\inf_u I(u) = \inf_u \int_0^1 G(u, u')dx, \quad u(0) = u(1) = 0 \tag{2.19}$$

where

$$G(u, v) = u^2 + \begin{cases} (v-1)^2, & \text{if } \; v \geq \frac{1}{2} \qquad\quad \text{Regime 1} \\ \frac{1}{2} - v^2 & \text{if } \; -\frac{1}{2} \leq v \leq \frac{1}{2} \quad \text{Regime 2} \\ (v+1)^2 & \text{if } \; v \leq -\frac{1}{2} \qquad \text{Regime 3} \end{cases} . \tag{2.20}$$

The graph of the function $G(., v)$ is presented in **??**B; it is a nonconvex differentiable function of v of superlinear growth.

The Lagrangian G penalizes the trajectory u for having the speed $|u'|$ different from ± 1 and penalizes the deflection of the trajectory u from zero. These contradictory requirements cannot be resolved in the class of classical trajectories.

Indeed, a differentiable minimizer satisfies the Euler equation (**??**) that takes the form

$$\begin{array}{ll} u'' - u = 0 & \text{if} \quad |u'| \geq \frac{1}{2} \\ u'' + u = 0 & \text{if} \quad |u'| \leq \frac{1}{2}. \end{array} \tag{2.21}$$

The Weierstrass test additionally requires convexity of $G(u, v)$ with respect to v; the Lagrangian $G(u, v)$ is nonconvex in the interval $v \in (-1,\ 1)$ (see **??**). The Weierstrass test requires the extremal (2.21) to be supplemented by the constraint (recall that $v = u'$)

$$u' \notin (-1,\ 1) \quad \text{at the optimal trajectory.} \tag{2.22}$$

The second regime in (2.21) is never optimal because it is realized inside of the forbidden interval. It is not clear how to satisfy both the Euler equations and

Weierstrass test because the Euler equation does not have a freedom to change the trajectory to avoid the *forbidden interval*.

We can check that the stationary trajectory can be broken at any point. The Weierstrass-Erdman condition (**??**) (continuity of $\frac{\partial L}{\partial u'}$) must be satisfied at a point of the breakage. This condition permits switching between the first ($u' > 1/2$) and third ($u' < -1/2$) regimes in (2.20) when

$$\left[\frac{\partial L}{\partial u'}\right]_-^+ = 2\left(u'_{(1)} - 1\right) - 2\left(u'_{(3)} + 1\right) = 0$$

or when

$$u'_{(1)} = 1, \quad u'_{(3)} = -1$$

which means the switching from one end of the forbidden interval $(-1, 1)$ to another.

Remark 2.2.1 Observe, that the easier verifiable Legendre condition $\frac{\partial^2 F}{\partial (u')^2} \geq 0$ gives a twice smaller forbidden region $|u'| \leq \frac{1}{2}$ and is not in the agreement with Weierstrass-Erdman condition. One should always use stronger conditions!

Minimizing sequence The minimizing sequence for problem (2.19) can be immediately constructed. Indeed, the infimum of (2.19) obviously is nonnegative, $\inf_u I(u) \geq 0$. Therefore, any sequence u^s with the property

$$\lim_{s\to\infty} I(u^s) = 0 \tag{2.23}$$

is a minimizing sequence.

Consider a set of functions $\tilde{u}^s(x)$ with the derivatives equal to ± 1 at each point,

$$\tilde{u}'(x) = \pm 1 \quad \forall x.$$

These functions belong to the boundary of the *forbidden interval* of the nonconvexity of $G(., v)$; they make the second term in the Lagrangian (2.20) vanish, $G(\tilde{u}, \tilde{u}') = u^2$, and the problem becomes

$$I(\tilde{u}^s, (\tilde{u}^s)') = \min_{\tilde{u}} \int_0^1 (\tilde{u}^s)^2 dx. \tag{2.24}$$

The sequence $\tilde{u}^s$ oscillates near zero if the derivative $(\tilde{u}^s)'$ changes its sign on intervals of equal length. The cost $I(\tilde{u}^s)$ depends on the density of switching points and tends to zero when the number of these points increases (see **??**). Therefore, the minimizing sequence consists of the saw-tooth functions $\tilde{u}^s$; the heights of the teeth tend to zero and their number tends to infinity as $s \to \infty$.

Note that the minimizing sequence $\{\tilde{u}^s\}$ does not converge to any classical function. This minimizer $\tilde{u}^s(x)$ satisfies the contradictory requirements, namely, the derivative must keep the absolute value equal to one, but the function itself must be arbitrarily close to zero:

$$|(\tilde{u}^s)'| = 1 \quad \forall x \in [0, 1], \qquad \max_{x\in[0,1]} \tilde{u}^s \to 0 \quad \text{as } s \to \infty. \tag{2.25}$$

The limiting curve u_0 has zero norm in $C_0[0,\ 1]$ but a finite norm in $C_1[0,\ 1]$.

Remark 2.2.2 Below, we consider this problem with arbitrary boundary values; the solution corresponds partly to the classical extremal (2.21), (2.22), and partly to the saw-tooth curve; in the last case u' belongs to the boundary of the forbidden interval $|u'| = 1$.

Regularization and relaxation We may apply regularization to discourage the solution to oscillate infinitely often. For example, we may penalize for the discontinuity of the u' adding the stabilizing term $\epsilon(u'')^2$ to the Lagrangian. Doing this, we pass to the problem

$$\min_u \int_0^1 (\epsilon^2(u'')^2 + G(u, u'))dx$$

that corresponds to Euler equation:

$$\begin{array}{ll} \epsilon^2 u^{IV} - u'' + u = 0 \ \text{ if } & |u'| \geq \frac{1}{2} \\ \epsilon^2 u^{IV} + u'' + u = 0 \ \text{ if } & |u'| \leq \frac{1}{2}. \end{array} \tag{2.26}$$

The Weierstrass condition this time requires the convexity of the Lagrangian with respect to u''; this condition is satisfied.

One can see that the solution of equation (2.26) is oscillatory; the period of oscillations is of the order of $\epsilon \ll 1$: The solution still tends to an infinitely often oscillating distribution. When ϵ is positive but small, the solution has finite but large number of wiggles. The computation of such solutions is difficult and often unnecessary: It strongly depends on an artificial parameter ϵ, which is difficult to justify physically. Although formally the solution of regularized problem exists, the questions remain. The problem is still computationally difficult and the difficulty grows when $\epsilon \to 0$ because the finite frequency of the oscillation of the solution tends to infinity.

Below we describe the relaxation of a nonconvex variational problem. The idea of relaxation is in a sense opposite to regularization. Instead of penalization for fast oscillations, we admit oscillating functions as legitime minimizers enlarging set of minimizers. The main problem is to find an adequate description of infinitely often switching controls in terms of smooth functions. It turns out that the limits of oscillating minimizers allows for a parametrization and can be effectively described by a several smooth functions: the values of alternating limits for u' and the average time that minimizer spends on each limit. The relaxed problem has the following two basic properties:

- The relaxed problem has a classical solution.
- The infimum of the functional (the cost of the problem) in the initial problem coincides with the cost of the relaxed problem.

Here we will demonstrate two approaches to relaxation based on necessary and sufficient conditions. Each of them yields to the same construction but uses different arguments to achieve it. In the next chapters we will see similar procedures applied to variational problems with multiple integrals; sometimes they also yield the same construction, but generally they result in different relaxations.

2.2.2 Minimal Extension

We introduce the idea of relaxation of a variational problem. Consider the class of Lagrangians $\mathcal{N}F(x,y,z)$ that are smaller than $F(x,y,z)$ and satisfy the Weierstrass test $\mathcal{W}(\mathcal{N}F(x,y,z)) \geq 0$:

$$\begin{cases} \mathcal{N}F(x,y,z) - F(x,y,z) \leq 0, \\ \mathcal{W}(\mathcal{N}F(x,y,z)) \geq 0 \end{cases} \quad \forall\ x,y,z. \tag{2.27}$$

Let us take the maximum on $\mathcal{N}F(x,y,z)$ and call it $\mathcal{S}F$. Clearly, $\mathcal{S}F$ corresponds to turning one of these inequalities into an equality:

$$\begin{array}{llll} \mathcal{S}F(x,y,z) = F(x,y,z), & \mathcal{W}(\mathcal{S}F(x,y,z)) \geq 0 & \text{if} & z \notin \mathcal{Z}_{\mathrm{f}}, \\ \mathcal{S}F(x,y,z) \leq F(x,y,z), & \mathcal{W}(\mathcal{S}F(x,y,z)) = 0 & \text{if} & z \in \mathcal{Z}_{\mathrm{f}}. \end{array} \tag{2.28}$$

This variational inequality describes the extension of the Lagrangian of an unstable variational problem. Notice that

1. The first equality holds in the region of convexity of F and the extension coincides with F in that region.
2. In the region where F is not convex, the Weierstrass test of the extended Lagrangian is satisfied as an equality; this equality serves to determine the extension.

These conditions imply that $\mathcal{S}F$ is convex everywhere. Also, $\mathcal{S}F$ is the maximum over all convex functions that do not exceed F. Again, $\mathcal{S}F$ is equal to the convex envelope of F:

$$\mathcal{S}F(x,y,z) = \mathcal{C}_z F(x,y,z). \tag{2.29}$$

The cost of the problem remains the same because the convex envelope corresponds to a minimizing sequence of the original problem.

Remark 2.2.3 Note that the geometrical property of convexity never explicitly appears here. We simply satisfy the Weierstrass necessary condition everywhere. Hence, this relaxation procedure can be extended to more complicated multidimensional problems for which the Weierstrass condition and convexity do not coincide.

Recall that the derivative of the minimizer never takes values in the region $\mathcal{Z}_{\mathrm{f}}$ of nonconvexity of F. Therefore, a solution to a nonconvex problem stays the same if its Lagrangian $F(x,\boldsymbol{y},\boldsymbol{z})$ is replaced by any Lagrangian $\mathcal{N}F(x,\boldsymbol{y},\boldsymbol{z})$ that satisfies the restrictions

$$\begin{array}{l} \mathcal{N}F(x,\boldsymbol{y},\boldsymbol{z}) = F(x,\boldsymbol{y},\boldsymbol{z}) \quad \forall\ z \notin \mathcal{Z}_{\mathrm{f}}, \\ \mathcal{N}F(x,\boldsymbol{y},\boldsymbol{z}) > \mathcal{C}F(x,\boldsymbol{y},\boldsymbol{z})\ \forall\ z \in \mathcal{Z}_{\mathrm{f}}. \end{array} \tag{2.30}$$

Indeed, the two Lagrangians $F(x, \boldsymbol{y}, \boldsymbol{z})$ and $\mathcal{N}F(x, \boldsymbol{y}, \boldsymbol{z})$ coincide in the region of convexity of F. Therefore, the solutions to the variational problem also coincide in this region. Neither Lagrangian satisfies the Weierstrass test in the forbidden region of nonconvexity. Therefore, no minimizer can distinguish between these two problems: It never takes values in Z_{f}. The behavior of the Lagrangian in the forbidden region is simply of no importance. In this interval, the Lagrangian cannot be computed back from the minimizer.

Minimizing Sequences Let us prove that the considered extension preserves the value of the objective functional. Consider the extremal problem (2.17) of superlinear growth and the corresponding stationary solution $u(x)$ that may not satisfy the Weierstrass test. Let us perturb the trajectory u by a differentiable function $\omega(x)$ with the properties:

$$\max_x |\omega(x)| \leq \varepsilon, \quad \omega(x_k) = 0 \ k = 1 \ldots N \tag{2.31}$$

where the points x_k uniformly cover the interval (a, b). The perturbed trajectory wiggles around the stationary one, crossing it at N uniformly distributed points x_k; the derivative of the perturbation is not bounded.

The integral $J(u, \omega)$

$$J(u, \omega) = \int_0^1 F(x, u + \omega, u' + \omega')dx$$

on the perturbed trajectory is estimated as

$$J(u, \omega) = \int_0^1 F(x, u, u' + \omega')dx + o(\varepsilon).$$

because of the smallness of ω (see (2.31)). The derivative $\omega'(x) = v(x)$ is a new minimizer constrained by N conditions (see (2.31))

$$\int_{\frac{k}{N}}^{\frac{k+1}{N}} v(x)dx = 0, \quad k = 0, \ldots N - 1; \tag{2.32}$$

correspondingly, the variational problem can be rewritten as

$$J(u, \omega) = \sum_{k=1}^{N-1} \int_{\frac{k}{N}}^{\frac{k+1}{N}} F(x, u, u' + \omega')dx + o\left(\frac{1}{N}\right).$$

Perform minimization of a term of the above sum with respect of v, treating u as a fixed variable:

$$I_k(u) = \min_{v(x)} \int_{\frac{k}{N}}^{\frac{k+1}{N}} F(x, u, u' + v)dx \quad \text{subject to} \quad \int_{\frac{k}{N}}^{\frac{k+1}{N}} v(x)dx = 0$$

This is exactly the problem (1.1) of the convex envelope with respect to v.

By referring to the Carathéodory theorem (2.5) we conclude that the minimizer $v(x)$ is a piece-wise constant function in $(\frac{k}{N}, \frac{k+1}{N})$ that takes at most $n+1$ values $v_1, \dots v_{n+1}$ at each interval. These values are subject to the constraints (see (2.32))

$$m_i(x) \geq 0, \quad \sum_{i=1}^{n} m_i = 1, \quad \sum_{i=1}^{p} m_i \boldsymbol{v}_i = 0. \tag{2.33}$$

This minimum coincides with the convex envelope of the original Lagrangian with respect to its last argument (see (2.5)):

$$I_k = \min_{m_i, \boldsymbol{v}_i \in (2.33)} \frac{1}{N} \left(\sum_{i=1}^{p} m_i F(x, \boldsymbol{u}, u' + \boldsymbol{v}_i) \right) \tag{2.34}$$

Summing I_k and passing to the limit $N \to \infty$, we obtain the relaxed variational problem:

$$I = \min_{\boldsymbol{u}} \int_0^1 \mathcal{C}_{\boldsymbol{u}'} F\left(x, \boldsymbol{u}(x), \boldsymbol{u}'(x)\right) dx. \tag{2.35}$$

Note that $n+1$ constraints (2.33) leave the freedom to choose $2n+2$ inner parameters m_i and $\boldsymbol{v}_i$ to minimize the function $\sum_{i=1}^{p} m_i F(u, \boldsymbol{v}_i)$ and thus to minimize the cost of the variational problem (see (2.34)). If the Lagrangian is convex, $\boldsymbol{v}_i = 0$ and the problem keeps its form: The wiggle trajectories do not minimize convex problems.

The cost of the reformulated (relaxed) problem (2.35) corresponds to the cost of the problem (2.17) on the minimizing sequence (**??**). Therefore, the cost of the relaxed problem is equal to the cost of the original problem (2.17). The extension of the Lagrangian that preserves the cost of the problem is called the *minimal extension*. The minimal extension enlarges the set of classical minimizers by including generalized curves in it.

2.2.3 Examples

Relaxation of nonconvex problem in Example ?? We revisit Example **??**. Let us solve this problem by building the convex envelope of the Lagrangian $G(u, v)$:

$$\begin{aligned} \mathcal{C}_v G(u, v) &= \min_{m_1, m_2} \min_{v_1, v_2} \left\{ u^2 + m_1(v_1 - 1)^2 + m_2(v_2 + 1)^2 \right\}, \\ v &= m_1 v_1 + m_2 v_2, \quad m_1 + m_2 = 1, \quad m_i \geq 0. \end{aligned} \tag{2.36}$$

The form of the minimum depends on the value of $v = u'$. The convex envelope $\mathcal{C}G(u, v)$ coincides with either $G(u, v)$ if $v \notin [0,\ 1]$ or $\mathcal{C}G(u, v) = u^2$ if $v \in [0,\ 1]$; see Example 2.1.5. Optimal values $v_1^0,\ v_2^0,\ m_1^0\ m_2^0$ of the minimizers and the convex envelope $\mathcal{C}G$ are shown in Table 2.1. The relaxed form of the problem with zero boundary conditions

$$\min_u \int_0^1 \mathcal{C}G(u,\ u') dx, \quad u(0) = u(1) = 0, \tag{2.37}$$

Average derivative	Pointwise derivatives	Optimal concentrations	Convex envelope $\mathcal{C}G(u,v)$
$v < -1$	$v_1^0 = v_2^0 = v$	$m_1^0 = 1,\ m_2^0 = 0$	$u^2 + (v-1)^2$
$\|v\| < 1$	$v_1^0 = 1, v_2^0 = -1$	$m_1^0 = m_2^0 = \frac{1}{2}$	u^2
$v > 1$	$v_1^0 = v_2^0 = v$	$m_1^0 = 0,\ m_2^0 = 1$	$u^2 + (v+1)^2$

Table 2.1: Characteristics of an optimal solution in Example **??**.

has an obvious solution,

$$u(x) = u'(x) = 0, \tag{2.38}$$

that yields the minimal (zero) value of the functional. It corresponds to the constant optimal value $m_{\rm opt}$ of $m(x)$:

$$m_{\rm opt}(x) = \frac{1}{2}\ \forall x \in [0,1]$$

.

The relaxed Lagrangian is minimized over four functions u, m_1, v_1, v_2 bounded by one equality, $u' = m_1 v_1 + (1-m_1)v_2$ and the inequalities $0 \le m \le 1$, while the original Lagrangian is minimized over one function u. In contrast to the initial problem, the relaxed one has a differentiable solution in terms of these four controls.

Inhomogeneous boundary conditions Let us slightly modify this example. Assume that boundary conditions are

$$u(0) = V\ \ (0 < V < 1), \quad u(1) = 0$$

In this case, an optimal trajectory of the relaxed problem consists of two parts,

$$u' < -1 \quad \text{if } x \in [0, x_0), \quad u = u' = 0 \quad \text{if } x \in [x_0, 1]$$

At the first part of the trajectory, the Euler equation $u'' - u = 0$ holds; the extremal is

$$u = \begin{cases} Ae^x + Be^{-x} & \text{if } x \in [0, x_0) \\ 0 & \text{if } x \in [x_0, 1] \end{cases}$$

Since the contribution of the second part of the trajectory is zero, the problem becomes

$$I = \min_{u, x_0} \int_O^{x_0} \mathcal{C}_v G(u, u') dx$$

To find unknown parameters A, B and x_0 we use the conditions

$$u(0) = V, \quad u(x_0) = 0, \quad u' = -1$$

The last condition expresses the optimality of x_0, it is obtained from the condition (see (**??**))

$$\mathcal{C}_v G(u, u')|_{x = x_0} = 0.$$

We compute

$$A + B = V, \quad Ae^{x_0} + Be^{-x_0} = 0, \quad Ae^{x} - Be^{-x} = 1$$

which leads to

$$\begin{aligned} u(x) &= \begin{cases} \sinh(x - x_0) & \text{if } x < x_0, \\ 0 & \text{if } x > x_0, \end{cases} \\ x_0 &= \sinh^{-1}(V) \end{aligned}$$

The optimal trajectory of the relaxed problem decreases from V to zero and then stays equal zero. The optimal trajectory of the actual problem decays to zero and then become infinite oscillatory with zero average.

Relaxation of a two-wells Lagrangian We turn to a more general example of the relaxation of an ill-posed nonconvex variational problem. This example highlights more properties of relaxation. Consider the minimization problem

$$\min_{u(x)} \int_0^z F_p(x, u, u')dx, \quad u(0) = 0, \; u'(z) = 0 \tag{2.39}$$

with a Lagrangian

$$F_p = (u - \alpha x^2)^2 + F_n(u'), \tag{2.40}$$

where

$$F_n(v) = \min\{a\, v^2, \; b\, v^2 + 1\}, \quad 0 < a < b, \; \alpha > 0.$$

Note that the second term F_n of the Lagrangian F_p is a nonconvex function of u'.

The first term $(u - \alpha x^2)^2$ of the Lagrangian forces the minimizer u and its derivative u' to increase with x, until u' at some point reaches the interval of nonconvexity of $F_n(u')$, after which it starts oscillating by alternation of the values of the ends of this interval, because u' must vary outside of this forbidden interval at every instance. (see **??**)

To find the convex envelope $\mathcal{C}F$ we must transform $F_n(u')$ (in this example, the first term of F_p (see (2.40)) is independent of u' and it does not change after the convexification). The convex envelope $\mathcal{C}F_p$ is equal to

$$\mathcal{C}F_p = (u - \alpha x^2)^2 + \mathcal{C}F_n(u'). \tag{2.41}$$

The convex envelope $\mathcal{C}F_n(u')$ is computed in Example 2.1.6 (where we use the notation $v = u'$). The relaxed problem has the form

$$\min_u \int \mathcal{C}F_p(x, u, u')dx, \tag{2.42}$$

where

$$\mathcal{C}F_p(x, u, u') = \begin{cases} (u - \alpha x^2)^2 + a(u')^2 & \text{if } \; |u'| \le v_1, \\ (u - \alpha x^2)^2 + 2u'\sqrt{\frac{ab}{a-b}} - \frac{b}{a-b} & \text{if } \; v_1 \le |u'| \le v_2, \\ (u - \alpha x^2)^2 + b(u')^2 + 1 & \text{if } \; |u'| \ge v_2. \end{cases}$$

Note that the variables u, v in the relaxed problem are the averages of the original variables; they coincide with those variables everywhere when $\mathcal{C}F = F$. The Euler equation of the relaxed problem is

$$\begin{aligned} au'' - (u - \alpha x^2) &= 0 \quad \text{if} \quad |u'| \le v_1, \\ (u - \alpha x^2) &= 0 \quad \text{if} \quad v_1 \le |u'| \le v_2, \\ bu'' - (u - \alpha x^2) &= 0 \quad \text{if} \quad |u'| \ge v_2. \end{aligned} \tag{2.43}$$

The Euler equation is integrated with the boundary conditions shown in (2.39). Notice that the Euler equation degenerates into an algebraic equation in the interval of convexification. The solution u and the variable $\frac{\partial}{\partial u'}\mathcal{C}F$ of the relaxed problem are both continuous everywhere.

Integrating the Euler equations, we sequentially meet the three regimes when both the minimizer and its derivative monotonically increase with x (see **??**). If the length z of the interval of integration is chosen sufficiently large, one can be sure that the optimal solution contains all three regimes; otherwise, the solution may degenerate into a two-zone solution if $u'(x) \le v_2 \ \forall x$ or into a one-zone solution if $u'(x) \le v_1 \ \forall x$ (in the last case the relaxation is not needed; the solution is a classical one).

Let us describe minimizing sequences that form the solution to the relaxed problem. Recall that the actual optimal solution is a generalized curve in the region of nonconvexity; this curve consists of infinitely often alternating parts with the derivatives v_1 and v_2 and the relative fractions $m(x)$ and $(1 - m(x))$:

$$v = \langle u'(x) \rangle = m(x) v_1 + (1 - m(x)) v_2, \quad u' \in [v_1, v_2], \tag{2.44}$$

where $\langle \ \rangle$ denotes the average, u is the solution to the original problem, and $\langle u \rangle$ is the solution to the homogenized (relaxed) problem.

The Euler equation degenerates in the second region into an algebraic one $\langle u \rangle = \alpha x^2$ because of the linear dependence of the Lagrangian on $\langle u \rangle'$ in this region. The first term of the Euler equation,

$$\frac{d}{dx} \frac{\partial F}{\partial \langle u \rangle'} \equiv 0 \quad \text{ if } \ v_1 \le | \langle u \rangle' | \le v_2, \tag{2.45}$$

vanishes at the optimal solution.

The variable m of the generalized curve is nonzero in the second regime. This variable can be found by differentiation of the optimal solution:

$$(\langle u \rangle - \alpha x^2)' = 0 \quad \Longrightarrow \quad \langle u \rangle' = 2\alpha x. \tag{2.46}$$

This equality, together with (2.44), implies that

$$m = \begin{cases} 0 & \text{if} \ \ |u'| \le v_1, \\ \frac{2\alpha}{v_1 - v_2} x - \frac{v_2}{v_1 - v_2} & \text{if} \ \ v_1 \le |u'| \le v_2, \\ 1 & \text{if} \ \ |u'| \ge v_2. \end{cases} \tag{2.47}$$

Variable m linearly increases within the second region (see **??**). Note that the derivative u' of the minimizing generalized curve at each point x lies on the boundaries v_1 or v_2 of the forbidden interval of nonconvexity of F; the average derivative varies only due to varying of the fraction $m(x)$ (see **??**).

Chapter 3

Localization and Discontinuous minimizers

3.1 Solutions with unbounded derivative. Regularization

3.1.1 Lagrangians of linear growth

A minimizing sequence may tend to a discontinuous function if the Lagrangian growth slowly with the increase of u'. Here we investigate discontinuous solutions of Lagrangians of linear growth. Assume that the Lagrangian F satisfies the limiting equality

$$\lim_{|u'|\to\infty} \frac{F(x,u,u')}{|u'|} \leq \beta u \tag{3.1}$$

where β is a nonnegative constant.

Considering the scalar case (u is a scalar function), we assume that the minimizing sequence tends to a finite discontinuity (jump) and calculate the impact of it for the objective functional. Let a miniming sequence u^ϵ of differentiable functions tend to a discontinuous at the point x_0 function, as follows

$$\begin{aligned} u^\epsilon(x) &= \phi(x) + \psi^\epsilon(x) \\ \psi^\epsilon(x) &\rightharpoondown \alpha H(x - x_0), \quad \beta \neq 0 \end{aligned}$$

where ϕ is a differentiable function with the bounded everywhere derivative, and H is the Heaviside function.

Assume that functions ψ^ϵ that approximate the jump at the point x_0 are piece-wise linear,

$$\psi^\epsilon(x) = \begin{cases} 0 & \text{if } x < x_0 - \epsilon \\ \frac{\alpha}{\epsilon}(x - x_0 + \epsilon) & \text{if } x_0 - \epsilon \leq x \leq x_0 \\ \alpha & \text{if } x > x_0. \end{cases}$$

The derivative $(\psi^\epsilon)'$ is zero outside of the interval $[x_0-\epsilon, x_0]$ where it is equal to a constant,

$$\psi' = \begin{cases} 0 & \text{if } x \notin [x_0-\epsilon, x_0] \\ \frac{\alpha}{\epsilon} & \text{if } x \in [x_0-\epsilon, x_0] \end{cases}$$

The Lagrangian is computed as

$$F(x,u,u') = \begin{cases} F(x,\phi,\phi') & \text{if } x \notin [x_0-\epsilon, x_0] \\ F\left(x,\phi+\psi^\epsilon,\phi'+\frac{\alpha}{\epsilon}\right) = \frac{\alpha\beta}{\epsilon} + o\left(\frac{1}{\epsilon}\right) & \text{if } x \in [x_0-\epsilon, x_0] \end{cases}$$

Here, we use the condition (3.1) of linear growth of F.

The variation of the objective functional is

$$\int_a^b F(x,u,u')dx \leq \int_a^b F(x,\phi,\phi')dx + \alpha\beta.$$

We observe that the contribution $\alpha\beta$ due to the discontinuity of the minimizer is finite when the magnitude $|\alpha|$ of the jump is finite. Therefore, discontinuous solutions are tolerated in the problems with Lagrangian of linear growth: They do not lead to infinitely large values of the objective functionals. To the contrary, the problems with Lagrangians of superlinear growth $\beta = \infty$ do not allow for discontinuous solution because the penalty is infinitely large.

Remark 3.1.1 The problems of Geometric optics and minimal surface are or linear growth because the length $\sqrt{1+u'^2}$ linearly depends on the derivative u'. To the contrary, problems of Lagrange mechanics are of quadratic (superlinear) growth because kinetic energy depends of the speed $\dot{q}$ quadratically.

3.1.2 Examples of discontinuous solutions

Example 3.1.1 (Discontinuities in problems of geometrical optics) We have already seen in Section ?? that the minimal surface problem

$$I_0 = \min_{u(x)} I(u), \quad I(u) = \int_o^L u\sqrt{1+(u')^2}dx, \quad u(-1)=1, \quad u(1)=1, \tag{3.2}$$

can lead to a discontinuous solution (Goldschmidt solution)

$$u = -H(x+1) + H(x-1)$$

if L is larger than a threshold.

Particularly, the Goldschmidt solution corresponds to zero smooth component $u(x)=0$, $x=(a,b)$ and two jumps M_1 and M_2 of the magnitudes $u(a)$ and $u(b)$, respectively. The smooth component gives zero contribution, and the contributions of the jumps are

$$I = \frac{1}{2}\left(u^2(a) + u^2(b)\right)$$

The next example (Gelfand & Fomin) shows that the solution may exhibit discontinuity if the superlinear growth condition is violated even at a single point.

Example 3.1.2 (Discontinuous extremal and viscosity-type regularization) Consider the minimization problem

$$I_0 = \min_{u(x)} I(u), \quad I(u) = \int_{-1}^{1} x^2 u'^2 dx, \quad u(-1) = -1, \quad u(1) = 1, \tag{3.3}$$

We observe that $I(u) \geq 0 \ \forall u$, and therefore $I_0 \geq 0$. The Lagrangian is convex function of u', and the third condition is satisfied. However, the second condition is violated in $x = 0$:

$$\lim_{|u'|\to\infty} \left. \frac{x^2 u'^2}{|u'|} \right|_{x=0} = \lim_{|u'|\to\infty} x^2 |u'| \Big|_{x=0} = 0$$

The functional is of sublinear growth at only one point $x = 0$.

Let us show that the solution is discontinuous. Assume the contrary, that the solution satisfies the Euler equation $(x^2 u')' = 0$ everywhere. The equation admits the integral

$$\frac{\partial L}{\partial u'} = 2x^2 u' = C.$$

If $C \neq 0$, the value of $I(u)$ is infinity, because then $u' = \frac{C}{2x^2}$, the Lagrangian becomes

$$x^2 u'^2 = \frac{C^2}{x^2} \quad \text{if } C \neq 0.$$

and the integral of Lagrangian diverges. A finite value of the objective corresponds to $C = 0$ which implies that $u_0'(x) = 0$ if $x \neq 0$. Accounting for the boundary conditions, we find

$$u_0(x) = \begin{cases} -1 & \text{if } x < 0 \\ \ \ 1 & \text{if } x > 0 \end{cases}$$

and $u_0(0)$ is not defined.

We arrived at the unexpected result that violates the assumptions used when the Euler equation is derived: $u_0(x)$ is discontinuous at $x = 0$ and u_0' exists only in the sense of distributions:

$$u_0(x) = -1 + 2H(x), \quad u_0'(x) = 2\delta(x)$$

This solution delivers absolute minimum $(I_0 = 0)$ to the functional, is not differentiable and satisfies the Euler equation in the sense of distributions,

$$\int_{-1}^{1} \frac{d}{dx} \left. \frac{\partial L}{\partial u'} \right|_{u=u_0(x)} \phi(x) dx = 0 \quad \forall \phi \in L_\infty[-1, 1]$$

Regularization A slightly perturb the problem (regularization) yields to the problem that has a classical solution and this solution is close to the discontinuous solution of the original problem. This time, regularization is performed by adding to the Lagrangian a stabilizer, a strictly convex function $\epsilon\rho(u')$ of superlinear growth.

Consider the perturbed problem for the Example 3.3:

$$I_\epsilon = \min_{u(x)} I_\epsilon(u), \quad I_\epsilon(u) = \int_{-1}^{1} \left(x^2 u'^2 + \epsilon^2 u'^2\right) dx, \quad u(-1) = -1, \quad u(1) = 1, \tag{3.4}$$

Here, the perturbation $\epsilon^2 u'$ is added to the original Lagrangian $\epsilon^2 u'$; the perturbed Lagrangian is of superlinear growth everywhere.

The first integral of the Euler equation for the perturbed problem becomes

$$(x^2 + \epsilon^2)u' = C, \quad \text{or } du = C\frac{dx}{x^2 + \epsilon^2}$$

Integrating and accounting for the boundary conditions, we obtain

$$u_\epsilon(x) = \left(\arctan\frac{1}{\epsilon}\right)^{-1} \arctan\frac{x}{\epsilon}.$$

When $\epsilon \to 0$, the solution $u_\epsilon(x)$ converges to $u_0(x)$ although the convergence is not uniform at $x = 0$.

Unbounded solutions in constrained problems The discontinuous solution often occurs in the problem where the derivative satisfies additional inequalities $u' \geq c$, but is unbounded. In such problems, the stationary condition must be satisfied everywhere where derivative is not at the constrain, $u' > c$. The next example shows, that the measure of such interval can be infinitesimal.

Example 3.1.3 (Euler equation is meaningless) Consider the variational problem with an inequality constraint

$$\max_{u(x)} \int_0^\pi u' \sin(x) dx, \quad u(0) = 0, \; u(\pi) = 1, \; u'(x) \geq 0 \; \forall x.$$

The minimizer should either corresponds to the limiting value $u' = 0$ of the derivative or satisfy the stationary conditions, if $u' > 0$. Let $[\alpha_i, \beta_i]$ be a sequence of subintervals where $u' = 0$. The stationary conditions must be satisfied in the complementary set of intervals $(\beta_i, \alpha_{i+1}])$ located between the intervals of constancy. The derivative cannot be zero everywhere, because this would correspond to a constant solution $u(x)$ and would violate the boundary conditions.

However, the minimizer cannot correspond to the solution of Euler equation at any interval. Indeed, the Lagrangian L depends only on x and u'. The first integral $\frac{\partial L}{\partial u'} = C$ of the Euler equation yields to an absurd result

$$\sin(x) = \text{ constant} \quad \forall x \in [\beta_i, \alpha_{i+1}]$$

The Euler equation does not produce the minimizer. Something is wrong!

The objective can be immediately bounded by the inequality

$$\int_0^\pi f(x)g(x)dx \le \left(\max_{x\in[0,\pi]} g(x)\right)\int_0^\pi |f(x)|dx.$$

that is valid for all functions f and g if the involved integrals exist. We set $g(x) = \sin(x)$ and $f(x) = |f(x)| = u'$ (because u' is nonnegative), account for the constraints

$$\int_0^\pi |f(x)|dx = u(\pi) - u(0) = 1 \quad \text{and} \quad \max_{x\in[0,\pi]} \sin(x) = 1,$$

and obtain the upper bound

$$I(u) = \int_0^\pi u' \sin(x)dx \le 1 \quad \forall u.$$

This bound corresponds to the minimizing sequence u_n that tends to a Heaviside function $u_n(x) \to H(x - \pi/2)$. The derivative of such sequence tends to the δ-function, $u'(x) = \delta(x - \pi/2)$. Indeed, immediately check that the bound is realizable, substituting the limit of u_n into the problem

$$\int_0^\pi \delta\left(x - \frac{\pi}{2}\right)\sin(x)dx = \sin\left(\frac{\pi}{2}\right) = 1.$$

The reason for the absence of a stationary solution is the openness of the set of differentiable function. This problem also can be regularized. Here, we show another way to regularization, by imposing an additional pointwise inequality $u'(x) \le \frac{1}{\gamma}\ \forall x$ (Lipschitz constraint). Because the intermediate values of u' are never optimal, optimal u' alternates the limiting values:

$$u'_\gamma(x) = \begin{cases} 0 & \text{if } x \notin \left[\frac{\pi}{2} - \gamma, \frac{\pi}{2} + \gamma\right], \\ \frac{1}{2\gamma} & \text{if } x \in \left[\frac{\pi}{2} - \gamma, \frac{\pi}{2} + \gamma\right], \end{cases}$$

The objective functional is equal to

$$I(u_\gamma) = \frac{1}{2\gamma}\int_{\frac{\pi}{2}-\gamma}^{\frac{\pi}{2}+\gamma} \sin(x)dx = \frac{1}{\gamma}\sin(\gamma)$$

When γ tends to zero, I_M goas to its limit

$$\lim_{\gamma\to 0} I_\gamma = 1,$$

the length γ of the interval where $u' = \frac{1}{2\gamma}$ goes to zero so that $u'_\gamma(t)$ weakly converges to the δ-function for u', $u'_\gamma(t) \rightharpoonup \delta\left(x - \frac{\pi}{2}\right)$.

This example clearly demonstrates the source of irregularity: The absence of the upper bound for the derivative u'. The constrained variational problems are studied in the control theory; they are are discussed later in Section 5.1.

3.1.3 Regularization by penalization

Regularization as smooth approximation The smoothing out feature of regularization is easy demonstrated on the following example of a quadratic approximation of a function by a smoother one.

Approximate a function $f(x)$ where $x \in \mathcal{R}$, by the function $u(x)$, adding a quadratic stabilizer; this problem takes the form

$$\min_u \int_{-\infty}^{\infty} [\epsilon^2 (u')^2 + (u-f)^2] dx$$

The Euler equation

$$\epsilon^2 u'' - u = -f \tag{3.5}$$

can be easily solved using the Green function

$$G(x,y) = \frac{1}{2\epsilon} \exp\left(-\frac{|x-y|}{\epsilon}\right)$$

of the operator in the left-hand side of (3.5). We have

$$u(x) = \frac{1}{2\epsilon} \int_{-\infty}^{\infty} \exp\left(-\frac{|x-y|}{\epsilon}\right) f(y) dy$$

that is the expression of the averaged f. The smaller is ϵ the closer is the average to f.

Quadratic stabilizers Besides the stabilizer $\varepsilon u'^2$, other stabilizers can be considered: The added term εu^2 penalizes for large values of the minimizer, $\varepsilon (u'')^2$ penalizes for the curvature of the minimizer and is insensitive to linearly growing solutions. The stabilizers can be inhomogeneous like $\varepsilon(u - u_{\text{target}})^2$; they force the solution stay close to a target value. The choice of a specific stabilizer depends on the physical arguments (see Tikhonov).

For example, solve the problem with the Lagrangian

$$F = \epsilon^4 (u'')^2 + (u - f(x)^2$$

Show that $u = f(x)$ if $f(x)$ is any polynomial of the order not higher than three. Find an integral representation for $u(f)$ if the function $f(x)$ is defined at the interval $|x| \leq 1$ and at the axis $x \in R$.

Regularization of a finite-dimensional problem

As the most of variational methods, the regularization has a finite-dimensional analog. It is applicable to the minimization problem of a convex but not strongly convex function which may have infinitely many solutions. The idea of regularization is to slightly perturb the function by small but a strictly convex term; the perturbed problem has a unique solution to matter how small the perturbation is. The numerical advantage of the regularization is the convergence of minimizing sequences.

Let us illustrate ideas of regularization by studying a finite dimensional problem. Consider a linear system

$$Ax = b \tag{3.6}$$

where A is a square $n \times b$ matrix and b is a known n-vector.

We know from linear algebra that the Fredholm Alternative holds:

- If $\det A \neq 0$, the problem has a unique solution:

$$x = A^{-1}b \quad \text{if } \det A \neq 0 \tag{3.7}$$

- If $\det A = 0$ and $Ab \neq 0$, the problem has no solutions.

- If $\det A = 0$ and $Ab = 0$, the problem has infinitely many solutions.

In practice, we also deal with an additional difficulty: The determinant $\det A$ may be a "very small" number and one cannot be sure whether its value is a result of rounding of digits or it has a "physical meaning." In any case, the errors of using the formula (3.7) can be arbitrary large and the norm of the solution is not bounded.

To address this difficulties, it is helpful to restate linear problem (3.6) as an extremal problem:

$$\min_{x \in R^n} (Ax - b)^2 \tag{3.8}$$

This problem does have at least one solution, no matter what the matrix A is. This solution coincides with the solution of the original problem (3.6) when this problem has a unique solution; in this case the cost of the minimization problem (3.8) is zero. Otherwise, the minimization problem provides "the best approximation" of the non-existing solution.

If the problem (3.6) has infinitely many solutions, so does problem (3.8). Corresponding minimizing sequences $\{x^s\}$ can be unbounded, $\|x^s\| \to \infty$ when $s \to \infty$.

In this case, we may select a solution with minimal norm. We use the *regularization*, passing to the perturbed problem

$$\min_{x \in R^n} (Ax - b)^2 + \epsilon x^2$$

The solution of the last problem exists and is unique. Indeed, we have by differentiation

$$(A^T A + \epsilon I)x - A^T b = 0$$

and

$$x = (A^T A + \epsilon I)^{-1} A^T b$$

We mention that

1. The inverse exists since the matrix $A^T A$ is nonnegative defined, and ϵ is positively defined. The eigenvalues of the matrix $(A^T A + \epsilon I)^{-1}$ are not smaller than ϵ^{-1}

2. Suppose that we are dealing with a well-posed problem (3.6), that is the matrix A is not degenerate. If $\epsilon \ll 1$, the solution approximately is $x = A^{-1}b - \epsilon(A^2A^T)^{-1}b$ When $\epsilon \to 0$, the solution becomes the solution (3.7) of the unperturbed problem, $x \to A^{-1}b$.

3. If the problem (3.6) is ill-posed, the norm of the solution of the perturbed problem is still bounded:
$$\|x\| \le \frac{1}{\epsilon}\|b\|$$

Remark 3.1.2 Instead of the regularizing term ϵx^2, we may use any positively define quadratic $\epsilon(x^T Px + p^T x)$ where matrix P is positively defined, $P > 0$, or other strongly convex function of x.

3.2 Lagrangians of sublinear growth

Discontinuous extremals Some applications, such as an equilibrium in organic or breakable materials, deal with Lagrangians of sublinear growth. If the Lagrangian $F_{\text{sub}}(x, u, u')$ growths slower that $|u'|$,

$$\lim_{|z|\to\infty} \frac{F_{\text{sub}}(x, y, z)}{|z|} = 0 \quad \forall x, y$$

then the discontinuous trajectories are expected because the functional is insensitive to finite jumps of the trajectory.

The Lagrangian is obviously a nonconvex function of u', The convex envelope of a bounded from below function $F_{\text{sub}}(x, y, z)$ of a sublinear with respect to z growth is independent of z.

$$\mathcal{C}F_{\text{sub}}(x, y, z) = \min_z F_{\text{sub}}(x, y, z) = F_{\text{conv}}(x, y)$$

In the problems of sublinear growth, the minimum $f(x)$ of the Lagrangian correspond to pointwise condition

$$f(x) = \min_u \min_v F(x, u, v)$$

instead of Euler equation. The second and the third argument become independent of each other. The condition $v' = u$ is satisfied (as an average) by fast growth of derivatives on the set of dense set of interval of arbitrary small the summary measure. Because of sublinear growth of the Lagrangian, the contribution of this growth to the objective functional is infinitesimal.

Namely, at each infinitesimal interval of the trajectory $x_0, x_0 + \varepsilon$ the minimizer is a broken curve with the derivative

$$u'(x) = \begin{cases} v_0 & \text{if } x \in [x_0, x_0 + \gamma\varepsilon] \\ v_0 & \text{if } x \in [x_0 + \gamma\varepsilon, x_0 + \varepsilon] \end{cases}$$

where $v_0 = \arg\min_z F(x, y, z)$, $1 - \gamma \ll 1$, and v_1 is found from the equation

$$u'(x) \approx \frac{u(x+\varepsilon) - u(x)}{\varepsilon} = \frac{v_1 \gamma \varepsilon + v_2(1-\gamma)\varepsilon}{\varepsilon}$$

to approximate the derivative u'. When $\gamma \to 1$, the contribution of the second interval becomes infinitesimal even if $v_2 \to \infty$.

The solution $u(x)$ can jump near the boundary point, therefore the main boundary conditions are irrelevant. The optimal trajectory will always satisfy natural boundary conditions that correspond to the minimum of the functional, and jump at the boundary points to meet the main conditions.

Example 3.2.1 (Jump at the boundary)

$$F = \log^2(u + u') \quad u(0) = u(1) = 10$$

The minimizing sequence converges to a function from the family

$$u(x) = A\exp(-x) + 1 \quad x \in (0, 1)$$

(A is any real number) and is discontinuous on the boundaries.

A problem with everywhere unbounded derivative This example shows an instructive minimizing sequence in a problem of sublinear growth. Consider the problem with the Lagrangian

$$J(u) = \int_0^1 F(x, u, u')dx, \quad F = (ax - u)^2 + \sqrt{|u'|}$$

This is an approximation problem: we approximate a linear function $f(x) = ax$ on the interval $[0, 1]$ by a function $u(x)$ using function $\sqrt{|u'|}$ as a penalty. We show that the minimizer is a distribution that perfectly approximate $f(x)$, is constant almost everywhere, and is nondifferentiable everywhere.

We mention two facts first: (i) The cost of the problem is nonnegative,

$$J(u) \geq 0 \quad \forall u,$$

and (ii) when the approximating function simply follows $f(x)$, $u_{trial} = ax$, the cost J of the problem is $J = \sqrt{a} > 0$ because of the penalty term.

Minimizing sequence Let us construct a minimizing sequence $u^k(x)$ with the property:

$$J(u^k) \to 0 \quad \text{if } s \to \infty$$

Partition the interval $[0, 1]$ into N equal subintervals and request that approximation $u(x)$ be equal to $f(x) = ax$ at the ends $x_k = \frac{k}{N}$ of the subintervals, and that the approximation is similar in all subintervals of partition,

$$\begin{aligned} u(x) &= u_0\left(x - \frac{k}{N}\right) + a\frac{k}{N} \quad \text{if } x \in \left[\frac{k}{N}, \frac{k+1}{N}\right], \\ u_0(0) &= 0, \quad u_0\left(\frac{1}{N}\right) = \frac{a}{N} \end{aligned}$$

Because of self-similarity, he cost J of the problem becomes

$$J = N \int_0^{\frac{1}{N}} \left((ax - u_0)^2 + \sqrt{|u_0'|} \right) dx \tag{3.9}$$

The minimizer $u_0(x)$ in a small interval $x \in \left[0, \frac{1}{N}\right]$ is constructed as follows

$$u_0(x) = \begin{cases} 0 & \text{if } x \in [0, \epsilon] \\ a\frac{1+\delta}{\delta}(x - \epsilon) & \text{if } x \in [\epsilon, \epsilon(1+\delta)] \end{cases}$$

Here, ϵ and δ are two small positive parameters, linked by the condition $\epsilon(1 + \delta) = \frac{1}{N}$. The minimizer stays constant in the interval $x \in [0, \epsilon]$ and then linearly growths on the supplementary interval $x \in [\epsilon, \epsilon(1 + \delta)]$. We also check that

$$u_0\left(\frac{1}{N}\right) = u_0(\epsilon + \delta\epsilon) = \frac{a}{N}$$

Derivative $u_0'(x)$ equals

$$u_0'(x) = \begin{cases} 0 & \text{if } x \in [0, \epsilon] \\ a\frac{1+\delta}{\delta} & \text{if } x \in [\epsilon, \epsilon(1+\delta)] \end{cases}$$

Computing the functional (3.9) of the suggested function u_0,

$$J = N \left(\int_0^{\epsilon} ((ax)^2 dx + \int_{\epsilon}^{\epsilon+\delta} \left[\left(ax - a\frac{1+\delta}{\delta}(x - \epsilon) \right)^2 + \sqrt{a\frac{1+\delta}{\delta}} \right] dx \right)$$

we obtain, after obvious simplifications,

$$J = N \left(\frac{a^2\epsilon^3}{3}(1+\delta) + \epsilon\sqrt{a(1+\delta)\delta} \right)$$

Excluding $\epsilon = \frac{1}{N(1+\delta)}$ we finally compute

$$J = \frac{a^2}{3N^2(1+\delta)^2} + \sqrt{\frac{a\delta}{1+\delta}}$$

Increasing N, $N \to \infty$ and decreasing δ, $\delta \to 0$ we can bring the cost functional arbitrary close to zero.

The minimizing sequence consists of the functions that are constant almost everywhere and contain a dense set of intervals of rapid growth. It tends to a nowhere differentiable function of the type of Cantor's "devils steps." The derivative is unbounded on a dense in $[0, 1]$ set. Because of slow growth of F,

$$\lim_{|u'| \to \infty} \frac{F(x, u, u')}{|u'|} \to 0$$

the functional is not sensitive to large values of u', if the growth occurs at the interval of infinitesimal measure. The last term of the Lagrangian does not contribute at all to the cost.

Regularization and relaxation To make the solution regular, we may go in two different directions. The first way is to forbid the wiggles by adding a penalization term $\epsilon(u' - a)^2$ to the Lagrangian which is transformed to:

$$F_\epsilon = (u - ax)^2 + \sqrt{|u'|} + \epsilon(u' - a)^2$$

The solution would become smooth, but the cost of the problem would significantly increase because the term $\sqrt{|u'|}$ contributes to it and the cost $J\epsilon = J(F_\epsilon)$ would depend on ϵ and will rapidly grow to be close to $\sqrt{a}$. Until the cost grows to this value, the solution remain nonsmooth.

Alternatively, we may "relax" the problem, replacing it with another one that preserves its cost and has a classical solution that approximates our nonregular minimizing sequence. To perform the relaxation, we simply ignore the term $\sqrt{|u'|}$ and pass to the Lagrangian

$$F_{\text{relax}} = (u - ax)^2$$

which corresponds the same cost as the original problem and a classical solution $u_{\text{class}} = ax$ that in a sense approximate the true minimizer, but not its derivative; it is not differentiable at all.

3.3 Nonuniqueness and improper cost

Unbounded cost functional An often source of ill-posedness (the nonexistence of the minimizer) is the convergence to minimizing functional to $-\infty$ or the maximizing functional to $+\infty$. To illustrate this point, consider the opposite of the brachistochrone problem: Maximize the travel time between two points. Obviously, this time can be made arbitrary large by different means: For example, consider the trajectory that has a very small slop in the beginning and then rapidly goes down. The travel time in the first part of the trajectory can be made arbitrary large (Do the calculations!). Another possibility is to consider a very long trajectory that goes down and then up; the larger is the loop the more time is needed to path it. In both cases, the maximizing functional goes to infinity. The sequences of maximizing trajectories either tend to a discontinuous curve or is unbounded and diverges. The sequences do not convergence to a finite differentiable curve.

Generally, the problem with an improper cost does not correspond to a classical solution: a finite differentiable curve on a finite interval. Such problems have minimizing sequences that approach either non-smooth or unbounded curve or do not approach anything at all. One may either accept this "exotic solution," or assume additional constraints and reformulate the problem. In applications, the improper cost often means that something essential is missing in the formulation of the problem.

Nonuniqueness Another source of irregular solutions is nonuniqueness. If the problem has families of many extremal trajectories, the alternating of them can occur in infinitely many ways. The problem could possess either classical or nonclassical solution. To detect such problem, we investigate the Weierstrass-Erdman conditions which show the possibilities of broken extremals.

An example of nonuniqueness, nonconvex Lagrangian As a first example, consider the problem

$$I(v) = \min_u \int_0^1 \left(1 - (u')^2\right)^2 dx, \quad u(0) = 0, \ u(1) = v \tag{3.10}$$

The Euler equation admits the first integral, because the Lagrangian depends only on u',

$$\left(1 - (u')^2\right)(1 - 2u') = C;$$

the optimal slope is constant everywhere and is equal to V.

When $-1 \leq v \leq 1$, the constant C is zero and the value of I is zero as well. The solution is not unique. Indeed, in this case one can joint the initial and the final points by the curve with the slope equal to either one or negative one in all points. The Weierstrass-Erdman condition

$$\left[\left(1 - (u')^2\right)(1 - 2u')\right]_-^+ = 0$$

is satisfied if $u' = \pm 1$ to the left and to the right of the point of break. There are infinitely many extremals with arbitrary number of breaks that all join the end points and minimize the functional making it equal to zero. Notice that Lagrangian is not convex function of u'.

Similarly to the finite-dimensional case, regularization of variational problems with nonunique solutions can be done by adding a penalty $\epsilon(u')^2$, or $\epsilon(u'')^2$ to the minimizer. Penalty would force the minimizer to prefer some trajectories. Particularly, the penalty term may force the solution to become infinitely oscillatory at a part of trajectory.

Another example of nonuniqueness, convex Lagrangian Work on the problem

$$I(v) = \min_u \int_0^1 (1 - u')^2 \sin^2(mu) dx, \quad u(0) = 0, \ u(1) = v \tag{3.11}$$

As in the previous problem, here there are two kinds of "free passes" (the trajectories that correspond to zero Lagrangian that is always nonnegative): horizontal ($u = \pi k/m$, $u' = 0$) and inclined ($u = c + x$, $u' = 1$). The Weierstrass-Erdman condition

$$[\sin(mu)^2(1 - u')]_-^+ = 0$$

allows to switch these trajectories in infinitely many ways.

Unlike the previous case, the number of possible switches is finite; it is controlled by parameter m. The optimal trajectory is monotonic; it becomes unique if $v \geq 1$ or $v \leq 0$, and if $|m| < \frac{1}{\pi}$.

3.4 Conclusion and Problems

We have observed the following:

- A one-dimensional variational problem has the fine-scale oscillatory minimizer if its Lagrangian $F(x, u, u')$ is a nonconvex function of its third argument.
- Homogenization leads to the relaxed form of the problem that has a classical solution and preserves the cost of the original problem.
- The relaxed problem is obtained by replacing the Lagrangian of the initial problem by its convex envelope. It can be computed as the second conjugate to F.
- The dependence of the Lagrangian on its third argument in the region of nonconvexity does not effect the relaxed problem.

To relax a variational problem we have used two ideas. First, we replaced the Lagrangian with its convex envelope and obtained a stable variational problem of the problem. Second, we proved that the cost of variational problem with the transformed Lagrangian is equal to the cost of the problem with the original Lagrangian if its argument $\boldsymbol{u}$ is a zigzag-like curve.

Problems

1. Formulate the Weierstrass test for the extremal problem
$$\min_u \int_0^1 F(x, u, u', u'')$$
that depends on the second derivative u''.

2. Find the relaxed formulation of the problem
$$\min_{u_1, u_2} \int_0^1 \left(u_1^2 + u_2^2 + F(u_1', u_2')\right),$$
$$u_1(0) = u_2(0) = 0, \quad u_1(1) = a, \quad u_2(1) = b,$$
where $F(v_1, v_2)$ is defined by (2.9). Formulate the Euler equations for the relaxed problems and find minimizing sequences.

3. Find the relaxed formulation of the problem
$$\min_u \int_0^1 \left(u^2 + \min\left\{|u' - 1|,\ |u' + 1| + 0.5\right\}\right),$$
$$u(0) = 0, \quad u(1) = a.$$
Formulate the Euler equation for the relaxed problems and find minimizing sequences.

4. Find the conjugate and second conjugate to the function

$$F(x) = \min\left\{x^2,\ 1 + ax^2\right\}, \quad 0 < a < 1.$$

   Show that the second conjugate coincides with the convex envelope $\mathcal{C}F$ of F.

5. Let $x(t) > 0$, $y(t)$ be two scalar variables and $f(x, y) = x\, y^2$. Demonstrate that

$$f(\langle x\rangle, \langle y\rangle) \geq \langle y\rangle^2 \left\langle \frac{1}{x} \right\rangle^{-1}.$$

   When is the equality sign achieved in this relation?

   Hint: Examine the convexity of a function of two scalar arguments,

$$g(y, z) = \frac{y^2}{z}, \quad z > 0.$$

Chapter 4

Hamilton-Jacobi Theory

So far we discussed properties of a single minimizing trajectories. Now we look at the family of the minimizers that differ by their destination points. Each curve of this family –the minimizer – can be treated as a coordinate curve for some curvilinear system of coordinates. We define the orthogonal curvilinear coordinate below as a solution of the problem of canonic transformation.

4.1 Geometric Optics. Eikonal

Let us start with an example. Consider again the problem of geometric optics or a problem of maximally quick passage from the starting point x_0 to a final point x. A logical difficulty of such formulation is the necessity to fix the destination point x which the light will reach at a future instance and choose the present trajectory depending on the future destination position. To resolve this difficulty, we consider a *family* of rays (trajectories) that aim to all directions from x_0 and later choose the right trajectory. In the simplest case of constant speed v of propagation the minimizers (rays) are straight lines originated at the point x_0.

Let $t = W(x)$ be a minimal time required for the light to reach a point x; call it the *Eikonal* or the optical length. The surface $W(x)$ is a propagation front for the family of rays that propagate from x_0 in different directions. Surface $W(x)$ is orthogonal to the rays in the family. In the simplest case of constant speed, $W(x)$ is a sphere centered at x_0 with the radius equal to the time needed to reach a point on it, that is

$$t = \frac{\|x\|}{v} = W(x)$$

We may describe the propagation of the light in terms of evolution of the surface $W(x)$ rather than in terms of the rays.

To describe the evolution of this surface, we consider an infinitesimal step dt in its propagation. We follow Huygens principle. Namely, we assume that the surface $W(x)$ radiate a light as if it would be new light source. In other words, the optimal trajectory from x_0 to a point x that lasts the time equal to $t + dt$ is a superposition of an optimal trajectory from x_0 to a point on the surface $W(x)$ which requires the time t to reach and an infinitesimal and also optimal trajectory from that point to the point of destination x.

The source at each point on the surface $W(x)$ highlights a spherical (or, in anisotropic case, elliptical) neighborhood; the common boundary (envelope) of these neighborhoods determines the surface $W(x, t+dt)$. A point of the surface $W(x + dx)$ that is the closest to a fixed point x at the surface $W(x)$ lies at the normal to this surface. The light propagated along the normal $N = \frac{\nabla W}{|\nabla W|}$ to the surface $W(x)$ with the speed $v(x)$; therefore

$$dx = N v\, dt = \frac{\nabla W}{|\nabla W|} v\, dt. \tag{4.1}$$

We arrive at the equation

$$W\,(x + dx) = W(x) + dt \tag{4.2}$$

Its left-hand side is transformed by linearization

$$W\,(x + dx) = W(x) + \nabla W\, dx$$

and by using (4.1),

$$W\,(x + dx) = W(x) + \frac{\nabla W^2}{|\nabla W|} v\, dt = W(x) + v|\nabla W| dt \tag{4.3}$$

and the equation (4.2) of Eikonal becomes

$$(\nabla W)^2 = \frac{1}{v^2}.$$

Solution of this nonlinear first order partial differential equation represents the moving boundary of the lighted part of the space, and the rays are the characteristics of the equation. The particle move along the (generally curved) rays with the speed $v(x)$.

Example 4.1.1 The Eikonal for the homogeneous medium ($v =$ constant)

$$(\nabla W)^2 = \left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 = \frac{1}{v^2}. \tag{4.4}$$

where x, y, z are Cartesian coordinates, has a solution

$$W = \frac{1}{v}\sqrt{x^2 + y^2 + z^2} \tag{4.5}$$

which can be checked by the differentiation.

The Eikonal equation is a nonlinear first-order partial differential equation. It can be integrated with various boundary condition

$$W(x_0) = 0 \quad \forall x_0 \in S_0$$

that introduces the shape S_0 of the original radiating body.

Example 4.1.2 One checks immediately, that the radiating line $S_0: \ x = y = 0$ produces the Eikonal

$$W = \frac{1}{v}\sqrt{x^2 + y^2}.$$

and the radiation plane $z = 0$ corresponds to the Eikonal

$$W = \frac{1}{v}|z|.$$

These Eikonals satisfy the equation (4.4)

Generally, the Eikonal in an homogeneous isotropic media ($v = \ \mathrm{constant}(x)$) describes the manifold that consists of points equally distant from the initially given surface. Combining the obtained solutions for a radiating point, line, and plane one can obtain the Eikonal for a radiation polyhedron. For example, the Eikonal surface for an illuminating cube is piece-wise analytic and contains the images of radiating sides, edges and corners. The surface consists of six squares separated by twelve cylindrical quarters and eight spherical octants.

The Eikonal equations can be solved in the appropriate coordinate system by separation of variables.

Example 4.1.3 Consider the Eikonal corresponding to the radiating sphere of the radius R. Introduce the spherical coordinates r, ϕ, θ in which Eikonal equation becomes

$$\left(\frac{\partial W}{\partial r}\right)^2 + \frac{1}{r^2 \sin^2\theta}\left(\frac{\partial W}{\partial \phi}\right) + \frac{1}{r^2}\left(\frac{\partial W}{\partial \theta}\right) = \frac{1}{v^2}$$

To solve, assume that $W = W(r)$. The above equation simplifies to

$$\frac{dW}{dr} = \pm\frac{1}{v}$$

and the initial condition becomes $W(R) = 0$. The problem has two solutions that correspond to different signs of the right-hand side. The first solution

$$W_1 = \frac{r - R}{v}$$

describes the light propagating away from the sphere. It exist for all $r \geq R$. The second solution

$$W_1 = \frac{R - r}{v}$$

describes the light propagating inside the sphere from its surface. Because of the natural constraints $r \geq 0$ and $W \geq 0$, it exist for r smaller than $0 \leq r \leq R$.

Example: Radiating ellipsoid in the homogeneous medium.
The solution of the Eikonal can have singularities.
Example: Refraction on a wedge.

4.2 Hamilton-Jacobi equation

Hamilton (1834) applied the approach to mechanical systems as well, using Lagrange principle instead of Fermat principle and respectively, Lagrangian L instead of Eikonal.

$$S(t,x) = \int_\gamma L(x,u,u')dt$$

where γ is the trajectory between the fixed initial point t_0, x_0 and the current point t, x.

Jacobi (1837) suggested to consider the function S for an arbitrary variational problem. All extremals originated at t_0, x_0 cross the surface level of the action transversally, that is in a manner similar to (4.4) for the rays of light. This implies the representation

$$dS = p^T du - H dt$$

where $p = \frac{\partial L}{\partial \dot{u}}$ is the impulse and $H = \dot{u}^T \frac{\partial L}{\partial \dot{u}} - L$ is the Hamiltonian. The last representation leads to the following equation called the *Hamilton-Jacobi* equation for the function $S(t,u)$:

$$\frac{\partial S}{\partial t} + H\left(x, u, \frac{\partial S}{\partial u}\right) = 0 \tag{4.6}$$

where H is the Hamiltonian of the system, $H = u'p - L$. The Hamilton-Jacobi equation is equivalent to the Euler or Hamilton equations of an extremal.

In a conservative system, Hamiltonian $H(u,p)$ is constant along the trajectory, $H(u,p) = E$, where E is the energy. This suggest a convenient representation of Hamilton-Jacobi function S,

$$S(t,q) = -Et + S^*(u)$$

where S^* is the *modified Jacobi function*. In terms of S^*, the equation (4.6) for the conservative system takes the form

$$H\left(x, u, \frac{\partial S^*}{\partial u}\right) = E. \tag{4.7}$$

The Hamilton-Jacobi equation contains the information of a family of extremals directed toward various end points; it can be used to investigate the variation of the objective functional caused by the variation of the position of the end point, etc. The Hamilton-Jacobi equation is a nonlinear partial differential equation of the first order which solution may exist only in a proximity of initial surface or not exist at all. In spite of its appearing complexity, the investigation of the Hamilton-Jacobi equation is very fruitful in Optics, Mechanics, and Geometry; in control theory, this equation leads to Bellman's dynamic programming, see Chapter 5.1.

Example 4.2.1 (Hamilton-Jacobi equation for a free particle) Consider motion of one particle in constant gravitational field $f = mg$ directed along Z axis. The kinetic energy T of the mass is

$$T = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2 \right)$$

and the potential energy V is $V = gz$. The impulses are

$$p_x = \frac{\partial (T - V)}{\partial \dot{x}} = m\dot{x}, \quad p_y = \frac{\partial (T - V)}{\partial \dot{y}} = m\dot{y}, \quad p_z = \frac{\partial (T - V)}{\partial \dot{z}} = m\dot{z}$$

and the Hamiltonian H is

$$H(x, y, z, p_x, p_y, p_z) = gz + \frac{m}{2} \left(p_x^2 + p_y^2 + p_z^2 \right)$$

The Hamilton-Jacobi equation is obtained when we formally replace the impulses with the corresponding derivatives of S^*, as follows

$$gz + \frac{m}{2} \left(\frac{\partial S^*}{\partial x}^2 + \frac{\partial S^*}{\partial y}^2 + \frac{\partial S^*}{\partial z}^2 \right) = E$$

Separating the variables, we look for the solution in the form

$$S^*(x, y, z) = X(x) + Y(y) + Z(z);$$

the previous equation then becomes

$$X'(x)^2 + Y'(y)^2 + \left(Z'(z)^2 + \frac{2g}{m} z \right) = \frac{2E}{m}$$

The three terms in the left-hand side depends only on x, y, and z, respectively, and therefore each of them is constant. We obtain

$$X(x) = a_1 x + b_1, \quad Y(y) = a_2 y + b_2, \quad Z(z) = \frac{(-2mgz + m^2 a_3^2)^{\frac{3}{2}}}{3m^2 g}$$

where a_i and b_i, $i = 1, 2, 3$ are constants,

$$a_1^2 + a_2^2 + a_3^2 = 2\frac{E}{m}$$

that are determined from the initial position and speed of the particle. Function S becomes

$$a_1 x + a_2 y + \frac{(-2mgz + m^2 a_3^2)^{\frac{3}{2}}}{3m^2 g} + b = Et$$

where b is a constant. It determines the time-dependent family of the surfaces that are perpendicular to the trajectory of a particle.

4.3 Canonic transformation and Jacobi theorem

We may look at the Hamilton-Jacobi equations from a geometric perspective. The extremals (rays) and the orthogonal to them front surfaces W form a curved coordinate system in the space. The Hamilton-Jacobi equation describes these curved coordinates; more precisely, it describes the mapping from a initial cartesian coordinates to the coordinates defined by the extremals. For example, a light front from a point-like source in a homogeneous space form a family of concentrated spheres centered around the source point, and the rays are the radii. In spherical coordinates ϕ, θ, r, these are lines $\phi =$ constant, $\theta =$ constant and a flat surfaces $r =$ constant, respectively.

In the proper curved coordinates, the front propagates straightforward and with the constant speed; the problem therefore can be reformulated to the problem of finding a proper mapping. We start with the general representation of the canonic transformation that keeps the form of Hamiltonian.

Canonic transformations Consider a system

$$q_k = \frac{\partial H}{\partial p_k}, \quad p_k = -\frac{\partial H}{\partial q_k}, \quad k = 1, \ldots, n$$

where $H(q, p)$ is the Hamiltonian. Let us make the transform

$$Q_k = Q_k(q, p), \quad P_k = P_k(q, p)$$

such that

1. There exists a function $K = K(Q, P)$ such that

$$\dot{Q}_k = \frac{\partial K}{\partial P_k}, \quad \dot{P}_k = -\frac{\partial K}{\partial Q_k}, \quad k = 1, \ldots, n$$

2.

3.

Such transformation is called *canonical transformation*. The transform that we want to find is a canonical transformation with K =constant. We observe:

$$\dot{Q}_k = \dot{P}_k = 0, \quad \text{or } Q(t) = Q(0) \text{ and } P(t) = P(0) \quad \text{if } K = \text{ constant}$$

In the properly transformed coordinates, all paths are straight and all velocities are constant. The question becomes: How to find this canonical transformation?

Chapter 5

Control theory

5.1 Optimal control

The theory of optimal control was developed starting from 1950s to meet the needs of designed the automatic control systems. Optimal control problem is essentially the variational problem with an additional feature: The pointwise constraints of the set of admissible controls.

5.1.1 Formulation

Variational and control problems Let us begin with reformulation of a classical variational problem using the optimal control notations. Consider the simplest variational problem

$$I = \min_{z} \int_0^1 F(x, z, z')dx \quad z(0) = z_0, \ z(1) = z_1. \tag{5.1}$$

The minimizer z is a differentiable function, but its derivative, z' is free of any pointwise constraints. We rewrite (5.1) introducing a new variable $u = z'$ called *control*. The control can be arbitrarily assigned at each point of trajectory. When the control is fixed, the trajectory z is uniquely defined.

The problem becomes a constrained one:

$$I = \min_{z,u} \int_0^1 F(x, z, u)dx \quad z(0) = z_0, \ z(1) = z_1, \quad z' = u. \tag{5.2}$$

We solve this problem using the standard Lagrange multipliers method. The extended functional is

$$I = \min_{z,u} \int_0^1 G(t, z, u, \mu)dt \quad z(0) = z_0, \ z(1) = z_1, \quad G = F + \mu(u - z'). \tag{5.3}$$

The Euler equations are obtained by variation of z and u:

$$\delta z : \ \frac{\partial F}{\partial z} + \mu' = 0, \quad \delta u : \ \frac{\partial F}{\partial u} + \mu = 0 \tag{5.4}$$

We obtained the two first order equations for z and μ and an algebraic relation for the control u:

$$u' = z, \quad \mu' = \frac{\partial F}{\partial z}, \quad \mu = -\frac{\partial F}{\partial u}$$

supplemented with two boundary conditions $z(0) = z_0,\ z(1) = z_1$. Excluding μ, we obtain the classical Euler equation.

$$\frac{d}{dx}\frac{\partial F}{\partial z} - \frac{\partial F}{\partial u} = 0$$

Therefore, a variational problem can always be viewed as an optimal control problem; z can be either a scalar or a vector. The inverse statement is, however, incorrect: The control problem can handle a more general type of extremal problem, accounting for conditions of the type $u(t) \in \mathcal{U}, \forall t$; for example $|u| > 1$..

A general control problem Formulating a control problem, the variables are separated into two groups

1. Controls $u = [u_1 \ldots, u_k]$ that are arbitrary assigned at each point and may be subject to some pointwise algebraic constraints

$$u \in \mathcal{U}(u, z, t)$$

   so that they are bounded in $L_\infty[0, 1]$.

2. Phase variables $z = [z_1, \ldots, z_n]$ that are (i) differentiable and (ii) entirely defined by the controls through differential constraints that are represented in the Cauchy form:

$$z_i' = f_i(x, z, u) \tag{5.5}$$

   Some boundary conditions are applied

$$b(z(0), z(1)) = 0 \tag{5.6}$$

The minimizing functional may consist of an integral term and a boundary term:

$$I = \int_{t_0}^{t_1} \Phi(x, z, u)dx + \phi(z(t_1)).$$

Example 5.1.1 (Control of a boat) Consider a control of a motor boat moving in a river. The differential equation os the motion is

$$m\ddot{x} + \gamma(\dot{x} - s) = \psi(u) \tag{5.7}$$

where x is the coordinate of the boat, m is its mass, γ is the viscosity of the water, s is the speed of the stream, $f(u)$ is the driving force applied from the propeller, and u is the speed of fuel consumption.

Assume that we want to minimize the total fuel spend in an hour

$$I = \int_0^1 u\,dx$$

and we need to reach a certain point A. The boundary conditions become

$$x(0) = \dot{x}(0) = 0, \quad x(1) = A$$

To rewrite the problem as an optimal control one, we transform the second-order equation (5.7) to the Cauchy form: Call $x = z_1$ and obtain the system

$$z_1' = z_2, \quad z_2' = \frac{1}{m}(-\gamma z_2 - s + f(u))$$

with the boundary conditions

$$z_1(0) = 0, \quad z_2(0) = 0, \quad z_1(1) = A$$

(notice that we have three boundary conditions here, the last one expresses the requirement of reaching a prescribed point A). The variables can be calculated if the control u is known, provided that the control is chosen to satisfy the extra boundary condition.

Example 5.1.2 (Canonic equations for rocket motion) Original equations are:

$$(mx')' - \gamma x' = f, \quad m' = \alpha f^\beta, \quad f \in [0, A]$$

The first one is the equation of the motion in a viscous media (air, γ is the viscosity coefficient) of a mass m under the reactive force f, the second shows that the force is proportional to the rate of burning of the mass, the third says that the force is positive and restricted.

We rewrite the system as

$$(mx' - \gamma x)' = f, \quad m' = \alpha f^\beta, \quad f \in [0, A]$$

Canonic variables are:

$$z_1 = x, \quad z_3 = m, \quad z_2 = mx' - \gamma x, \quad f = u$$

and the canonic system becomes

$$\begin{aligned} z_1' &= \frac{z_2 + \gamma z_1}{z_3} \\ z_2' &= u \\ z_3' &= \alpha u^\beta \end{aligned}$$

5.1.2 Adjoint system

Accounting for differential constraints, the minimization problem can be rewritten in the form

$$\min_{z,u\in\mathcal{U}} \int_{t_0}^{t_1} (\Phi(x,z,u) + \mu[-z' + f(x,z,u)])\, dx + \phi(z(t_1)) \quad u(x) \in \mathcal{U} \tag{5.8}$$

The stationary conditions (Euler equation) are:

$$\delta z: \ \frac{\partial \Phi}{\partial z} + \mu^T \frac{\partial f}{\partial z} + \mu' = 0 \tag{5.9}$$

the supplementary boundary conditions

$$-\mu_i(0) + \gamma_0 \frac{\partial b}{\partial z_i(0)} = 0, \quad -\mu_i(1) + \gamma_1 \frac{\partial b}{\partial z_i(1)} = 0 \tag{5.10}$$

where γ_0 and γ_1 are Lagrange multipliers.

The conditions for the optimal control (if u belongs to the interior of $\mathcal{U}$) are

$$\delta u: \ \frac{\partial \Phi}{\partial u} + \mu^T \frac{\partial f}{\partial u} = 0, \quad u \in Int\mathcal{U}. \tag{5.11}$$

We postpone discussion of the conditions that holds if u belongs to the boundary of $\mathcal{U}$.

The system of n differential equations (5.9) defines the vector of *adjoint variables* (a.k.a. Lagrange multipliers) μ. The boundary conditions for μ can be found from the system:

$$\mu^T \cdot \delta z|_{t=0} = 0, \quad \left(\frac{\partial \phi}{\partial z} + \mu^T\right) \cdot \delta z|_{t=1} = 0 \tag{5.12}$$

where δz are subject to constraints originated from the given initial or boundary conditions. Notice that the total order of the system for differential constraints and Lagrange functions is $2n$, and there are $2n$ boundary conditions for it.

Example 5.1.3 (continue) The Lagrange functions are

$$\mu_1' = 0 \quad \mu_2' = -\mu_1 + \frac{\gamma}{m}\mu_2$$

The variation of $z_2(1)$ is undetermined, therefore, $\mu_2(1) = 0$. The variations of other boundary terms are zero, and no other boundary conditions for μ arise. The order of the system for differential constraints and Lagrange functions is four, and there are four boundary conditions.

Hamiltonian The system of differential equations is further rewritten as

$$\mu' = -\frac{\partial H}{\partial z}, \quad z' = \frac{\partial H}{\partial \mu}, \quad \text{where } H = F + \mu^T f \tag{5.13}$$

Here $H(t,z,\mu,u)$ is called *Hamiltonian*; it contains all the information of the extremal problem and it is an algebraic function of its argument.

5.1.3 Pontryagin's maximum principle

Needle variation Because of the constrained set of controls, the Weierstrass-type variation is not always possible. Indeed, this variation would require such a perturbation of u that its average value is zero, because u plays the role of derivative in the classical variation problem. If u in on the boundary (say in a corner point) of the set $\mathcal{U}$, this variation is not possible. At the other hand, the consideration of closed sets of control is the main objective of the control theory.

Instead of the Weierstrass variation, we use the needle-type variation

$$\Delta u(x) = 0, \text{ if } x \notin [x_0, x_0 + \epsilon], \quad u(x) = U \in \mathcal{U}, \text{ if } x \notin [x_0, x_0 + \epsilon].$$

This variation investigates the sign of the increment of functional if a control switch its value at a small time interval. It is applicable every time if only the set $\mathcal{U}$ contains more than one element. The admissible controls do not need to be continuous; the set $\mathcal{U}$ may even consists of several isolated points.

Unlike the perturbation by the Weierstrass variation, the perturbation of the trajectory δz caused by the needle variation is not zero outside of the interval $[x_0, x_0 + \epsilon]$ of variation, but it is of the order of ϵ everywhere. The main term of the increment is of the order of ϵ and consists of two parts:

$$\delta I = \epsilon([H(x_0, , ., u, .) - H(x_0, .., U, .)] + \int_0^1 \mu \left(\delta z' - \frac{\partial H}{\partial z} \delta z \right) dt + o(\epsilon)$$

The first term is the main term of the expansion of the integral

$$\int_{x_0}^{x_0+\epsilon} [H(x, z(u), \mu(u), u) - H(x, z(U), \mu(U), U)] dx =$$
$$\epsilon [H(x_0, z, \mu, u) - H(x_0, z, \mu, U)] + o(\epsilon).$$

Here, the continuity of z and μ is used. In the perturbed system, these quantities differs from the optimal values not more than by the term of the order of ϵ, see (5.5), (5.9), and therefore can be replaced with the optimal values in the approximation.

The second term

$$\int_0^1 \mu \left(\delta z' - \frac{\partial H}{\partial z} \delta z \right) dt = -\int_0^1 \left(\mu' + \frac{\partial H}{\partial z} \right) \delta z \, dt + \mu' z |_0^1$$

contains terms of the order of ϵ. This term, however, disappears due to the choice of the adjoint variables (by virtue of (**??**)).

The remaining first term leads to the condition

$$u_{opt} = arg \left\{ \min_{u \in \mathcal{U}} H(z, \mu, u) \right\} \tag{5.14}$$

where z and μ are computed along the optimal trajectory. This condition is called the *Pontryagin's maximum principle*

Remark 5.1.1 Traditionally, problem of the control theory is to maximize (not minimize) the functional; wherefore the name *maximum principle* is originated.

Now we have the complete set of equations to determine u, z, μ: N first-order differential equations (5.5) for the differential constraints, N first-order differential equations (5.9) for the adjoint system, and p equations (5.14) for an optimal control. The system is supplemented by $2N$ boundary conditions (5.12) and (**??**).

Control inside $\mathcal{U}$ Particularly, if u is inside $\mathcal{U}$, one can make $U - u$ infinitely small and obtain the conditions

$$\frac{\partial H}{\partial u} = 0, \quad \frac{\partial^2 H}{\partial u^2} \geq 0 \text{ if } x \in Int\,(\mathcal{U}) \tag{5.15}$$

These equalities serve to find u

Control on the boundary of $\mathcal{U}$ If $u \in \partial\mathcal{U}$, u is on the boundary of $\mathcal{U}$, the condition depend on the type of the boundary point (a point on a smooth boundary, the corner point, isolated point, etc.) These conditions are differently expressed in these cases. For example, if the constrain on a scalar control restricts its values, $u_- \leq u(t) \leq u_+$ the conditions for the small variations are

$$\begin{aligned} u &= u_- \quad \text{if } \frac{\partial H}{\partial u} > 0, \\ u &= u_+ \quad \text{if } \frac{\partial H}{\partial u} < 0, \\ \frac{\partial H}{\partial u} &= 0, \ \frac{\partial^2 H}{\partial u^2} \geq 0, \quad \text{if } u_- \leq u(t) \leq u_+. \end{aligned}$$

In all cases, we have one equality to find the optimal control and one inequality to check. The condition (5.14) is obviously stronger then these last conditions, but its verification is also more difficult.

Example 5.1.4 (boat; continue) Revisiting the example, we find that

$$H = u + \mu_1 z_1 + \mu_2 z_2 - \mu_1 z_2 - \mu_2 \frac{1}{m}(-\gamma z_2 - s + f(u))$$

or

$$H = u - \mu_2 \frac{1}{m} f(u) + \text{terms independent of } u$$

Assuming the constraints $0 \leq u \leq 1$, we find the last condition for u

$$U = arg \left(\max_{u \in (0,1)} \left[u - \mu_2 \frac{1}{m} f(u) \right] \right)$$

which leads to conditions for control inside the interval

$$u \in (0,1) \quad \text{if } m - \mu_2 f'(u) = 0, \quad \mu_2 f''(u) \geq 0,$$

(the equation serves to determine u), and on the boundaries

$$u = 0 \quad \text{if } m - \mu_2 f'(0) \geq 0;$$
$$u = 1 \quad \text{if } m - \mu_2 f'(1) \leq 0.$$

5.2 Developments and examples

5.2.1 Various types of constraints

The control theory formulation is flexible enough to incorporate additional isoperimetric constraints, uncertain interval, etc.

- The integral constraint
$$\int_0^1 G(z,u)dt = A$$
can be rewritten in the control theory setting as following. A new variable z_{n+1} is introduced by the equation
$$z'_{n+1} = G(z,u), \quad z_{n+1}(0) = 0, \quad z_{n+1}(1) = A.$$
This equation is added to the system (5.5).

- Similarly, one can always maximize (or minimize) the boundary value of a variable instead of an integral. The minimization of the integral of $\Phi(z,u)$ is equivalent to minimization of $z_{n+1}(t_1)$ where
$$z'_{n+1} = \Phi(z,u), \quad z_{n+1}(0) = 0$$

- If the final time T of the process is to be minimized, we may change the independent variable:
$$t = \theta T, \quad \theta \in [0,1]$$
and
$$\frac{d}{dt} = \frac{1}{T}\frac{d}{d\theta}$$
and minimize T subject to the constraints
$$\frac{d}{d\theta} z = T f(t,z,u)$$

- Similarly, algebraic constraints $G(t,u,z) = 0$ can be incorporated using Lagrange multiplier's technique.