Chapter 1

Preliminary Remarks

1.1 Origins of extremal problems

Optimization The desire for optimality (perfection) is inherent in humans. The search for extremes inspires mountaineers, scientists, mathematicians, and the rest of the human race. The development of Calculus of Variation is driven by this noble desire. Simultaneously with Calculus technique of finding minima of functions. A mathematical technique of minimization of functionals was developed in the eighteen century to describe the best possible geometric objects: The minimal surface, the shortest distance, or the trajectory of fastest travel.

In the twentieth century, control theory emerged to address the extremal problems in science, engineering, and decision-making. These problems study the influence on the objective of the available free-chosen time-dependent functions called controls. Optimal design theory addresses space-dependent analog of control problems focusing on multivariable control. Minimax problems address optimization in conflict situations or in an undetermined environment. A special branch of the theory uses minimization principles to create effective numerical algorithms such as finite element method for computing the numerical solutions.

Description of fundamental laws of Nature For centuries, philosophers and scientists tried to prove that the Universe is rational, symmetric, or optimal in some sense. Attempts were made to formulate laws of natural sciences as extreme problems (variational principles) and to use the variational calculus as a scientific instrument to derive and investigate the motion and equilibria in Nature (Fermat, Lagrange, Gauss, Hamilton, Gibbs..). It was observed by Fermat that light always takes the trajectory that minimizes the time of travel. Equilibria correspond to the local minima of the energy, motion of mechanical systems corresponds to stationarity of a functional called *action*, etc. In turn, the variational principles link together conservation laws and symmetries.

Does the actual trajectory minimize the action? This question motivated great researchers starting from Leibnitz and Fermat to develop variational methods to justify the Nature's "desire" to choose the most economic way to move, and it caused much heated discussions that involved philosophy and theology. The general principle by Maupertuis proclaims: "If there occur some changes in nature, the amount of action necessary for this change must be as small as possible." In a sense, this principle would prove that our world is "the best of all worlds" – the conclusion defended by Fermat, Leibnitz, Maupertuis, and Euler but later ridiculed by Voltaire. In mid-nineteen century Jacobi showed that the action is minimized by short trajectories, but for long ones. This mathematical fact was disappointing for philosophical speculations, "a beautiful conjecture is ruined by an ugly fact." Still later, the relativity and the notion of the world lines in a sense returns the principle of minimization of a quantity at the real trajectory over all other trajectories.

Concise description of the state of an object No matter if the actual trajectories minimize the action or not, the variational methods in physics become an important tool for investigation of motions and equilibria. First, the variational formulation is convenient and economic: Instead of formulation of equations it is enough to write down a single functional that must be optimized at the actual configuration. The equations of the state of the system follow from the optimality requirement. Second, variational approach allows for accounting of symmetries, for invariants of the configuration, and (through duality) for different differential equations that describe the same configuration in different terms.

There are several ways to describe a shape or a motion. The most explicit way is to describe positions of all points: Sphere is described by the functions $-\sqrt{1-x^2-y^2} \leq z(x,y) \leq \sqrt{1-x^2-y^2}$. The more implicit way is to formulate a differential equation which produces these positions as a solution: The curvature tensor is constant everywhere in a sphere. An even more implicit way is to formulate a variational problem: Sphere is a body with given volume that minimizes its surface area. The minimization of a single quantity produces the "most economic" shape in each point. Such implicit description goes back to Platonic ideals and is opposite to the Aristotelian principle to explicit description/classification of factual events (here, the explicit functions)

1.2 Remarks on history

For the rich history of Calculus of variation we refer to such books as [Kline, Boyer].. Here we make several short remarks about the ideas of its development. Calculus of variations is a rare disciplile that has a birthdate. The story started with the challenge called the *brachistochrone problem*:

> Given two points A and B in a vertical plane, what is the curve traced out by a point acted on only by gravity, which starts at A and reaches B in the shortest time.

The brachistochrone problem was posed by Johann Bernoulli in Acta Eruditorum in June 1696[1]. He introduced the problem with the following preambula:

> *I, Johann Bernoulli, address the most brilliant mathematicians in the world. Nothing is more attractive to intelligent people than an honest, challenging problem, whose possible solution will bestow fame and remain as a lasting monument. Following the example set by Pascal, Fermat, etc., I hope to gain the gratitude of the whole scientific community by placing before the finest mathematicians of our time a problem which will test their methods and the strength of their intellect. If someone communicates to me the solution of the proposed problem, I shall publicly declare him worthy of praise.*

Johann Bernoulli's solution divides the plane into strips and he assumes that the particle follows a straight line in each strip. The path is then piecewise linear. The problem is to determine the angle of the straight line segment in each strip and to do this he appeals to Fermat's principle, namely that light always follows the shortest possible time of travel. If v is the velocity in one strip at angle a to the vertical and u in the velocity in the next strip at angle b to the vertical then, according to the usual sine law v/sin a = u/sin b.

Within a year, five solutions were obtained, Newton, Jacob Bernoulli, Leibniz and de L'Hôpital solving the problem in addition to Johann Bernoulli. The May 1697 publication of Acta Eruditorum contained Leibniz's solution to the brachistochrone problem on page 205, Johann Bernoulli's solution on pages 206 to 211, Jacob Bernoulli's solution on pages 211 to 214, and a Latin translation of Newton's solution on page 223. The solution by de L'Hôpital was not published untilnearly 300 years later, in 1988 (Jeanne Peiffer). Thus, the calculus of variation appeared simultaneously with the Calculus, and was created by the same brilliant group.

The optimal trajectory turns out to be a cycloid (see Section **??** for the derivation). Cycloid was an investigated in seventeen century parametric curve. Huygens had shown in 1659, prompted by Pascal's challenge, that the cycloid is the tautochrone of isochrone: The curve for which the time taken by a particle sliding down the curve under uniform gravity to its lowest point is independent of its starting point. Johann Bernoulli ended his solution with the remark: *Before I end I must voice once more the admiration I feel for the unexpected identity of Huygens' tautochrone and my brachistochrone. ... Nature always tends to act in the simplest way, and so it here lets one curve serve two different functions, while under any other hypothesis we should need two curves.*

For a halh of a century, the methods for funding the best curves remain a collection of examples. Then, Leonhard Euler unified them in his 1744 work *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes sive solutio problematis isoperimetrici latissimo sensu accepti. (Method for finding*

[1] Johann Bernoulli was not the first to consider the brachistochrone problem. Galileo had studied the problem in 1638 in his famous work Discourse on Two New Sciences. He correctly concluded that the straight path is not the fastest one, but made an error concluding that an optimal trajectory is a part of a circumference.

plane curves that show some property of maxima and minima.) In that work, Euler suggested a general approach to variational problems. He demonstrated that extremal curve satisfies a differential equation that he derived and that since was named after him (Euler equation or Euler-Lagrange equation),

Lagrange, in 1760, published Essay on a new method of determining the maxima and minima of indefinite integral formulas. It gave an analytic method to attach calculus of variations type problems and applied it to classical mechanics: He announced: "The admirers of the Analysis will be pleased to learn that Mechanics became one of its new branches" and "The reader will find no figures in this work. The methods which I set forth do not require either constructions or geometrical or mechanical reasonings: but only algebraic operations, subject to a regular and uniform rule of procedure." (Lagrange, *Mcanique analytique*. Preface)

New mathematical concepts Working on optimization problems, mathematicians met paradoxes related to absence of optimal solution or its weird behavior; resolving these was useful for the theory itself and resulted in new mathematical development such as weak solutions of differential equations and related functional spaces (Hilbert and Sobolev spaces), various types of convergence of functional sequences, distributions and other limits of function's sequences, and other fundamentals of modern analysis.

Many computational methods are motivated by optimization problems and use techniques of minimization. Methods of search, finite elements, iterative schemes are part of optimization theory. The classical calculus of variation answers the question: What conditions must the minimizer satisfy? while the computational techniques are concerned with the question: How to find or approximate the minimizer?

The list of main contributors to the calculus of variations includes the most distinguished mathematicians of the last three centuries such as Leibnitz, Newton, Bernoulli, Euler, Lagrange, Gauss, Jacobi, Hamilton, Hilbert.

Today, the most attention is paid to the extremal problem without classical solution. In these problem, the minimizing sequence of differentiable minimizers may turn either to a discontinuous function, or to a distribution, of to infinitely often oscillating function. Each of these cases requires the rethinking of the concept of "solution", "function" and leads to new formulations of the problem.

1.3 Properties of the extrema

Every optimization problem contains several necessary components. It deals with a set $\mathcal{X}$ of admissible elements x, that can be real or complex numbers, differentiable curves, integrable functions, shapes, people in the town, or ants in the colony. A real-valued function $I(x)$ called objective is put into correspondence to each admissible element. The objective could be an absolute value of a complex number, value of the function at a certain point, value of the integral of a function over an interval, weight of a town inhabitant, or length of an ant.

The goal is to find or characterize the element x_0 called minimizer, such that

$$I(x_0) \le I(x), \quad \forall x \in \mathcal{X}$$

We denote this element as

$$x_0 = \arg\min_{x \in \mathcal{X}} I(x)$$

and we denote the value $I(x_0$ as

$$I(x_0) = \min_{x \in \mathcal{X}} I(x)$$

Next, we list the basic properties of any extreme problem that are based on the definition of the minimizer.

1. Minimum over a larger set is equal or smaller than minimum of the smaller set

   If $\mathcal{X}_1 \supseteq \mathcal{X}_2$, then
   $$\min_{x \in \mathcal{X}_1} F(x) \le \min_{x \in \mathcal{X}_2} F(x)$$

2. Minimum of a function $F(x)$ is equal to the negative of maximum of $-F(x)$,
   $$\min_{x \in \mathcal{X}} F(x) = -\max_{x \in \mathcal{X}} (-F(x))$$

   This property allows us not to distinguish between minimization and maximization problems: We always can reformulate the maximization problem as the minimization one.

3. Generalization the previous property has the form:

   $$\min_{x \in \mathcal{X}} [cF(x)] = \begin{cases} c \min_{x \in \mathcal{X}} F(x) & \text{if } c \ge 0 \\ -c \max_{x \in \mathcal{X}} (-F(x)) & \text{if } c \le 0 \end{cases}$$

4. Minimum of a sum of functions is not smaller than the sum of minima of additives.
   $$\min_x [f(x) + g(x)] \ge \min_x f(x) + \min_x g(x)$$

5. The minimizer is invariant to the superposition of the objective with a monotonically increasing function. Consider functions $F : X \subset R^n \to Y \subset R^1$ and $G : Y \to Z \subset R^1$ and assume that G monotonically increases:

   $$G(y_1) - G(y_2) \le 0 \quad \text{ if } y_1 \le y_2$$

   Then minima of $F(x)$ and of $G(F(x))$ are reached at the same minimizer,

   $$x_0 = \arg\min F(x) = \arg\min G(F(x))$$

6. Maximum of several minima is not larger than minimum of several maxima:
$$\max_x f_{\min}(x) \le \min_x f_{\max}(x)$$
where
$$f_{\max}(x) = \max\{f_1(x), \dots f_N(x)\}$$
$$f_{\min}(x) = \min\{f_1(x), \dots f_N(x)\}$$

7. Minimax theorem
$$\max_y \min_x f(x,y) \le \min_x \max_y f(x,y)$$

The listed properties can be proved by the straightforward use of the definition of the minimizer. We show the prove of the minimax theorem, leaving other to the reader.

Consider functions
$$\phi(x) = \max_y f(x,y), \quad \psi(y) = \min_x f(x,y)$$
that satisfy the obvious inequalities
$$\psi(y) \le f(x,y) \le \phi(x), \quad \forall x, y$$
Therefore, the extreme value of these functions satisfy the same inequalities:
$$\max_y \psi(y) \le \min_x \phi(x), \quad \forall x, y$$
which is the statement of the minimax theorem.

1.4 Variational problem

The extremal (variational) problem requires to find an optimal function $u_0(x)$ which can be visualized as a curve (or a surface). Function $u_0(x)$ belongs to a set of admissible functions $\mathcal{U}$: $u \in \mathcal{U}$; it is assumed that $\mathcal{U}$ is a set of continuously differentiable functions on the interval $[a,b]$ that is denoted as $C_1[a,b]$. To measure the optimality of a curve, we define a *functional* (a real number) $I(u)$ which may depend on $u(x)$, its derivative $u'(x)$, and on independent variable x. Examples of functionals are: the length of the curve, the area of the contour, weight of a construction, its strength, etc.

The examples of variational problems are: The shortest path on a surface, the surface of minimal area, the best approximation of experimental data by a smooth curve, the construction of given strength and minimal weight, etc.

Simplest variational problem The classical (simplest) variational problem is formulated as follows: Consider an integral functional of the type

$$J(u) = \int_a^b F(x, u(x), u'(x))dx. \tag{1.1}$$

where F is a known function of three arguments, $x, u(x), u'(x)$, called *Lagrangian*. For example, the length of a curve $u(x)$ corresponds to $F = \sqrt{1 + (u')^2}$.

Find the differentiable function $u_0(x)$ that corresponds to minimum of $J(u)$ assuming that the boundary values of u are fixed:

$$I(u_0) = \min_{u(x) \in \mathcal{U}_b} J(u) \quad \mathcal{U}_b = \{u : \ u \in C_1(a, b), \ u(a) = \alpha, \ u(b) = \beta\} \tag{1.2}$$

Function $u_0(x)$ is called the *minimizer*. For example, a curve of *minimal length* corresponds to a straight line,

$$u' = \text{ constant}, \quad u(x) = u_a + \frac{u_b - u_a}{b - a}(x - a)$$

Minimizing sequence It is assumed that the set of minimizers $\mathcal{U}_b$ is chosen so that the integral in (1.1) exists. However, there is no guarantee of existence of the C_1 minimizer. Indeed, C_1 set is open and a sequence of C_1-functions may lead to a discontinuous function, or to a distribution.

To the contrary, the minimizing sequence always exists. The value of the objective functional $I(u)$ (also called the cost functional) is a real number. Since real numbers are ordered, one can compare functionals $J(u_1), J(u_2), \ldots$ of different admissible functions $u_1, u_2, \ldots$, and build *minimizing sequences* of functions

$$u_1, \ u_2, \ldots, u_n, \ldots$$

with the property:

$$I(u_1) \geq I(u_2) \geq \ldots \geq I(u_n) \ldots$$

The limit u_0 of a minimizing sequence (if it exists) is called the *minimizer*; it delivers the minimum of I

$$I(u_0) \leq I(u) \quad \forall u \in \mathcal{U} \tag{1.3}$$

The minimizing sequence can always be built independently of the existence of the minimizer.

Generalization The formulated problem can be generalized in several ways.

- The minimizer and an admissible function can be a vector-function; the functional may depend of higher derivatives, and be of a more general form such as a ratio of two integrals.
- The integration can be performed over a spatial domain instead of the interval $[a, b]$; this domain may be completely or partly unknown and should be determined together with the minimizer.

- The problem may be constrained in several ways: The isoperimetric problem asks for the minimum of $I(u)$ if the value of another functional $I_r(u)$ is fixed. Example: find a domain of maximal area enclosed by a curve of a fixed length. The constrained problem asks for the minimum of $I(u_1, \dots u_n)$ if a function(s) $\phi(u_1, \dots u_n)$ is fixed everywhere. Example: The problem of geodesics: the shortest distance between two points on a surface. In this problem, the path must belong to the surface everywhere.

1.5 Outline of the methods

There are several groups of methods aimed to find a minimizer of an extremal problem of different compexity.

Global optimization (sufficient conditions). These rigorous methods directly establish the inequality $I(u_0) \leq I(u), \forall u \in \mathcal{U}$. They are applicable to a limited variety of problems, and the results are logically perfect. To establish the above inequality, methods based on convexity or stable point theorem are commonly used. The methods usually require a guess of the global minimizer u_0 and therefore are applicable to relatively simple extremal problems.

For example, the variational problem (1.2), (1.1) with the Lagrangian

$$F(x,u) = \frac{u(x)^2 + \sin^2(x)}{2} - u(x)\,\sin(x)$$

reaches its minimum, zero, if $u(x) = \sin(x)$. This follows from the inequality between arithmetic and harmonic means of $u(x)^2$ and $\sin^2(x)$ that is valid for all functions $u(x)$.

Methods of necessary conditions (variational methods). Using these methods, we establish and analyze necessary conditions for $u(x)$ to provide a local minimum. In other words, the conditions tell that there is no other curve $u + \delta u$ that is (i) sufficiently close to the chosen curve u (that is assuming $\|\delta u\|$ is infinitesimal), (ii) satisfies the same boundary or other posed conditions, and (iii) corresponds to a smaller value $I(u + \delta u) < I(u)$ of the objective functional. The closeness of compared curves $u(x)$ and $u(x) + \delta u(x)$ allows for a relative simple form of the resulting *variational conditions* of optimality. On the other hand, it restricts the generality of the obtained conditions.

Variational methods that are discussed below yield to only necessary conditions of optimality; they detect *locally optimal* curves. These methods are regular and robust; they are applicable to a great variety of extremal problems called *variational problems*. Necessary conditions are the true workhorses of extremal problem theory, while exact sufficient conditions are rare and remarkable exceptions.

Direct optimization methods These methods are aimed to building the minimizing sequence $\{u^s\}$ and provide a sequence of better solutions. Generally, the convergence to the true minimizer may not be required, but it is guaranteed that the solutions are improved on each step of the procedure: $I(u^s) \leq I(u^{s-1})$ for all s. These methods require no a priori assumption of the dependence of functional on the minimizer only the possibility to compare two values of the aim functional and chose the better of two. Of course, additional assumptions help to optimize the search and a smart iterative strategy takes advantage of variational necessary conditions in forming of a minimizing sequence. However, the search can be conducted without these assumptions. As an extreme example, one can iteratively find the oldest person from the alphabetic telephone directory calling at random, asking for the age of the responder and comparing the age with the maximum from the already obtained answers.

There is also no guarantee that the obtained solution is truly optimal.Nevertheless, such methods are used for practical applications when the model is too complicated or the data are too unreliable for more refined methods.

The next table summarizes the discussion.

	Global optimization	Variational methods	Direct methods
Objectives	Search for the global minimum	Search for a local minimum	An improvement of existing solution
Means	Sufficient conditions	Necessary conditions	Sequential improvement
Tools	Inequalities, Convex analysis, Fixed point methods	Analysis of features of optimal trajectories	Gradient-type search
Existence of solution	Guaranteed	Not guaranteed	Not discussed

Table 1.1: Approaches to variational problems