m 5500 Calculus of Variations. Spring 2013.

Andrej Cherkaev

Introduction

1 Subjects of Calculus of variations

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 was driven by this noble desire. A mathematical technique of minimization of curves 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 function 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.

Description of fundamental laws of Nature For centuries, philosophers and scientists tried to prove that the Universe is rational, symmetric, or optimal in another 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 "chooses" the trajectory that minimizes the time of travel, equilibria correspond to the local minimum of the system's energy, motion of mechanical systems correspond to stationarity of a functional called the *action*, etc. In turn, the variational principles link together conservation laws and symmetries.

Lagrange reformulated variational mechanics as a variational problem, making calculus of variations a universal tool for the applications. The methods were developed and new test for optimality were suggested by by Jacobi and Weierstrass, and a number of *universal variational principles* were proposed by Gauss (principle of least constraint), Hertz (principle of least curvature), Hamilton, and others. Today, aspecial branch of the theory uses minimization principles to create effective numerical algorithms such as finite element method for computing the equilibria.

Does the actual trajectory minimize the action? This question motivated great researcher 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 and ridiculed by Voltaire. It turns out that the action is minimized on sufficiently short trajectories, but correspond to only stationary value of action for long ones, because of violation of so-called Jacobi conditions. This mathematical fact was disappointing for philosophical speculations, "a beautiful conjunction is ruined by an ugly fact." However, the relativity and the notion of the world lines 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 real trajectories minimize the action or not, the variational methods in physics become an important tool for investigation of motions and equilibria. Firstly, the variational formulation is convenient and economic: Instead of formulation of all the 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. Secondly, variational approach allows for accounting of symmetries, invariants of the configuration, and derive (through the duality) several 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.

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, Γ-limits and other fundamental concepts of modern analysis.

Many computational methods as motivated by optimization problems and use of the technique 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 concern with the question: How to find or approximate the minimizer?

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

Origin 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. The story started with the challenge:

> 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. [1]

The brachistochrone problem was posed by Johann Bernoulli in Acta Eruditorum in June 1696. He introduced the problem as follows:

> *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*

[1] Johann Bernoulli was not the first to consider the brachistochrone problem. Galileo in 1638 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 circle.

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.

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 solutions to the brachistochrone problem by Leibniz, Johann Bernoulli , Jacob Bernoulli, and Newton. The solution by de L'Hôpital was discovered only in 1988 Jeanne Peiffer presented it in [].

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.

The optimal trajectory turns out to be a cycloid (see Section **??** for the derivation). Cycloid was a well investigated curve in seventeen century. 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.*

The methods which Bernoulli developed were put in a general setting by Euler in his 1744 work *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes sive solutio problematis isoperimetrici latissimo sensu accepti.*[2] In this work, Euler found the differential equation (Euler equation or Euler-Lagrange equation) for a minimizer.

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. In the introduction

[2] *Method for finding plane curves that show some property of maxima and minima.*

to the paper Lagrange gives the historical development of the ideas which we have described above but it seems appropriate to end this article by giving what is in effect a summary of the developments in Lagrange's words:- *The first problem of this type [calculus of variations] which mathematicians solved was that of the brachistochrone, or the curve of fastest descent, which Johann Bernoulli proposed towards the end of the last century.*

2 Properties of the extremals

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

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

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_0

$$I_0 = I(x_0) = \min_{x \in \mathcal{X}} I(x)$$

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
   $$\min_{x \in \mathcal{X}_1} F(x) \leq \min_{x \in \mathcal{X}_2} F(x)$$

   Restricting the class of admitted functions or parametrizing, we find an upper estimate of the minimum, and increasing that class, e.i. lifting some restrictions, we come to the lower estimate of it.

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. Generalizing the previous property, we consider a multiplication of the minimizing functional by a real number c:

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

4. Minimum of sum is not smaller than the sum of minima of additives.

$$\min_x[f(x) + g(x)] \geq \min_x f(x) + \min_x g(x)$$

Splitting the minimizer into two and evaluating the parts separately, we find a lower estimate of it.

5. Superposition. Consider the 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:

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

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. Minimax theorem

$$\max_y \min_x f(x,y) \leq \min_x \max_y f(x,y)$$

providing that the quantities in the left and righthand side exist.

The next several colloraries are often used in the applications:

7. Linearity: If b and $c > 0$ are real numbers, than

$$\min_x\,(c\,f(x) + b) = c\left(\min_x f(x)\right) + b$$

8. The minimizer is invariant to the superposition of the objective with any monotonic function. Namely, The minimizer

$$x_0 = \arg\min_x f(x)$$

where $f:\ X \to Y \subset R_1$ is also the minimizer of the problem

$$x_0 = \arg\min_{x \in X} g\left(f(x)\right)$$

where $g:\ Y \to R_1$ is monotone everywhere on Y.

9. Maximum of several minima is not larger than minimum of several maxima:

$$\max\left\{\min_x f_1(x), \dots \min_x f_N(x)\right\} \le \min_x f_{\max}(x)$$

where

$$f_{\max}(x) = \max\{f_1(x), \dots f_N(x)\}$$

The listed properties can be proved by the straightforward use of the definition of the minimizer. We leave the prove to the reader.

3 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 differentiable function 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 $u(x)$, and its derivative $u'(x)$ as well as on the independent variable x. The examples of variational problems are: The shortest path on a surface, the surface of minimal area, the best approximation by a smooth curve of the experimental data, the most economical strategy, etc.

The classical variational problem is formulated as follows: Find

$$I(u_0) = \min_{u(x) \in \mathcal{U}_b} J(u) \quad \mathcal{U}_b = \{u:\ u \in C_1(a, b),\ u(a) = \alpha,\ u(b) = \beta\} \qquad (1)$$

where $x \in [a, b]$, $u_0(x)$ is an unknown function called the *minimizer*, the boundary values of u are fixed, $J(u)$ is the functional of the type

$$J(u) = \int_a^b F(x, u(x), u'(x))dx. \qquad (2)$$

F is a function of three arguments, $x, u(x), u'(x)$, called *Lagrangian*, and it is assumed that the integral in (**??**) 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

$$I(u_0) \leq I(u) \quad \forall u \in \mathcal{U} \tag{3}$$

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 the ratio of two integrals.
- The integration can be performed over a spacial 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 $I(u_1, \ldots u_n)$ if a function(s) $\phi(u_1, \ldots u_n)$ 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.

Outline of the methods There are several groups of methods aimed to find the minimizer of an extremal problem.

1. **Methods of necessary conditions (variational methods).** Using these methods, we establish 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 two compared curves 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 yield to only necessary conditions of optimality; they detect *locally optimal* curves. On the other hand, variational 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.

2. **Methods of 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, the methods of convexity are commonly used. The method often requires a guess of the global minimizer u_0 and is applicable to relatively simple extremal problems.

   For example the polynom $F(x) = 6 + 13x^2 - 10x + 2x^4 - 8x^3$ has its global minimum at the point $x = 1$ because it can be presented in the form $F(x) = 3 + (x-1)^2 + 2(x-1)^4$; the second and the third term reach minimum at $x = 1$.

3. **Sequential optimization algorithms** These 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 a project with an improved one and chose the best of two. Of course, additional assumptions help to optimize the search but it can be conducted without these. 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.

	Global methods	Variational methods	Algorithmic search
Objectives	Search for the global minimum	Search for a local minimum	An improvement of existing solution
Means	Sufficient conditions	Necessary conditions	Algorithms of sequential improvement
Tools	Inequalities, Fixed point methods	Analysis of features of optimal trajectories	Gradient-type search
Existence of solution	Guaranteed	Not guaranteed	Not discussed
Applicability	Special problems	Large class of problems	Universal

Table 1: Approaches to variational problems