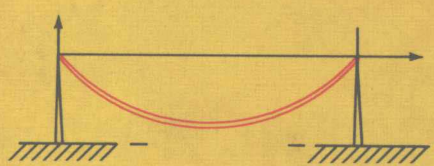
Undergraduate Texts in Mathematics

John L. Troutman

Variational Calculus with Elementary Convexity



$$J(y+v) - J(y) \geqslant \delta J(y;v)$$

Springer-Verlag New York • Heidelberg • Berlin

John L. Troutman

With the assistance of W. Hrusa

Variational Calculus with Elementary Convexity

With 73 Illustrations

Springer-Verlag
New York Heidelberg Berlin

John L. Troutman
Department of Mathematics
Syracuse University
200 Carnegie
Syracuse, NY 13210
U.S.A.

William Hrusa
Department of Mathematics
Carnegie-Mellon University
Pittsburgh, PA 15213
U.S.A.

AMS Subject Classifications (1980): 26-01, 49-01

Library of Congress Cataloging in Publication Data
Troutman, John L.
Variational calculus with elementary convexity.
(Undergraduate texts in mathematics)
Bibliography: p
Includes index.
1. Calculus of variations. 2. Convex functions.
I. Hrusa, W. (William) II. Title. III. Series.
QA315.T725 1983 515′.64 82-19257

Typeset by Composition House Ltd., Salisbury, England.
Printed and bound by R. R. Donnelley & Sons, Harrisonburg, VA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-90771-8 Springer-Verlag New York Heidelberg Berlin
ISBN 3-540-90771-8 Springer-Verlag Berlin Heidelberg New York

Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics

Apostol: Introduction to Analytic Number Theory.
1976. xii, 338 pages. 24 illus.

Armstrong: Basic Topology.
1983. xii, 251 pages. 132 illus.

Bak/Newman: Complex Analysis.
1982. x, 224 pages. 69 illus.

Banchoff/Wermer: Linear Algebra Through Geometry.
1983. x, 257 pages. 81 illus.

Childs: A Concrete Introduction to Higher Algebra.
1979. xiv, 338 pages. 8 illus.

Chung: Elementary Probability Theory with Stochastic Processes.
1975. xvi, 325 pages. 36 illus.

Croom: Basic Concepts of Algebraic Topology.
1978. x, 177 pages. 46 illus.

Fischer: Intermediate Real Analysis.
1983. xiv, 770 pages. 100 illus.

Fleming: Functions of Several Variables. Second edition.
1977. xi, 411 pages. 96 illus.

Foulds: Optimization Techniques: An Introduction.
1981. xii, 502 pages. 72 illus.

Franklin: Methods of Mathematical Economics. Linear and Nonlinear Programming. Fixed Point Theorems.
1980. x, 297 pages. 38 illus.

Halmos: Finite-Dimensional Vector Spaces. Second edition.
1974. viii, 200 pages.

Halmos: Naive Set Theory.
1974, vii, 104 pages.

Iooss/Joseph: Elementary Stability and Bifurcation Theory.
1980. xv, 286 pages. 47 illus.

Kemeny/Snell: Finite Markov Chains.
1976. ix, 224 pages. 11 illus.

Lax/Burstein/Lax: Calculus with Applications and Computing, Volume 1.
1976. xi, 513 pages. 170 illus.

LeCuyer: College Mathematics with A Programming Language.
1978. xii, 420 pages. 144 illus.

Macki/Strauss: Introduction to Optimal Control Theory.
1981. xiii, 168 pages. 68 illus.

Malitz: Introduction to Mathematical Logic: Set Theory - Computable Functions - Model Theory.
1979. xii, 198 pages. 2 illus.

Martin: The Foundations of Geometry and the Non-Euclidean Plane.
1975. xvi, 509 pages. 263 illus.

Martin: Transformation Geometry: An Introduction to Symmetry.
1982. xii, 237 pages. 209 illus.

Millman/Parker: Geometry: A Metric Approach with Models.
1981. viii, 355 pages. 259 illus.

Prenowitz/Jantosciak: Join Geometries: A Theory of Convex Set and Linear Geometry.
1979. xxii, 534 pages. 404 illus.

Priestly: Calculus: An Historical Approach.
1979, xvii, 448 pages. 335 illus.

Protter/Morrey: A First Course in Real Analysis.
1977. xii, 507 pages. 135 illus.

(continued on page 365)

This book is dedicated to
my parents, and to
RGB, the artist.

Preface

The calculus of variations, whose origins can be traced to the works of Aristotle and Zenodoros, is now a vast repository supplying fundamental tools of exploration not only to the mathematician, but—as evidenced by current literature—also to those in most branches of science in which mathematics is applied. (Indeed, the macroscopic statements afforded by variational principles may provide the only valid mathematical formulation of many physical laws.) As such, it retains the spirit of natural philosophy common to most mathematical investigations prior to this century. However, it is a discipline in which a single symbol (δ) has at times been assigned almost mystical powers of operation and discernment, not readily subsumed into the formal structures of modern mathematics. And it is a field for which it is generally supposed that most questions motivating interest in the subject will probably not be answerable at the introductory level of their formulation.

In earlier articles,[1,2] it was shown through several examples that a complete characterization of the solution of optimization problems may be available by elementary methods, and it is the purpose of this work to explore further the convexity which underlay these individual successes in the context of a full introductory treatment of the theory of the variational calculus. The required convexity is that determined through Gâteaux variations, which can be defined in any real linear space and which provide an unambiguous foundation for the theory. In applications, this convexity for integral functions is assured by a partial convexity of their integrands.

[1] (With W. Hrusa) Elementary characterization of classical minima. *MAA Monthly*, May 1981 (321–327).

[2] Partially convex functions in the variational calculus. *Real Analysis Exchange*, **7**, 1981–1982 (89–92).

This book is intended as an introduction to the use of variational methods in the formulation and solution of optimization problems of both mathematical and physical interest. It is an outgrowth of lectures presented in the Mathematics Department of Syracuse University during the past several years. In preliminary form, the material from Chapters 0–6 comprised a standard one-semester course of three hours per week given to upper-level students in mathematics, physics, computer science, and engineering, and that from the remainder of the book was offered on request during a succeeding semester.

When pursued systematically, it is designed to carry those who have completed a standard course in multidimensional calculus—with some exposure to differential equations—through the rudiments of rigorous analysis in (normed) linear spaces. For it is in the presence of the linear spaces of continuously differentiable functions that a theory for the integral inequalities which generated the calculus of variations can first be clarified. In this setting, the fundamentals of the subject are explored, including the natural role of the differential equations of Euler–Lagrange in relating the behavior of integral inequalities to the physical principles of stationarity. (The actual plan of the text is presented in §1.5.)

However, this book departs from previous introductory texts by placing initial emphasis on (global) sufficiency considerations as the basis for motivation and development, thereby postponing (until Chapter 5) the more subtle mathematical questions of (local) necessity. Instead, it is shown (in Chapter 3) that after suitable formulation, many minimization problems of interest exhibit a natural convexity which can be readily discerned, and which leads directly to their (unique) solution through the equations of Euler–Lagrange. Within this framework, generalization to vector valued and multidimensional problems is straightforward, and even the method of Lagrangian multipliers becomes not a mystery, but a simple observation. Thereby made available at an introductory level are some of the benefits which convexity considerations supply to modern functional analytic treatments of this subject, such as [E–T], [I–T], [K–S], [R] and [V].

Surprisingly, most of the standard problems in the subject yield at least partial solution by this method. Moreover, convexity may supply insight and direction for analysis of the remaining problems, and it offers, for example, a mechanism by which Hamilton's principle of *stationary* action can be reduced to Bernoulli's principle of *minimum* potential energy. Finally, when the general sufficiency arguments for a minimum are presented (in Chapter 9) it will be recognized that those integrand functions amenable to attack by the (field theory) methods there presented also exhibit a *partial* convexity, albeit to a lesser degree than those studied previously. Effective notation is employed to facilitate the recognition and manipulation of partially convex functions and to reduce the complexity of formal appearance which plagues any presentation of the variational calculus.

The methods considered herein are directed toward obtaining exact solutions for the problems encountered—usually in the form of functions defined explicitly or parametrically—although recognition is taken of the possibility and value of approximations. Each attack by exact methods requires finding a solution of the Euler–Lagrange equation(s) with given boundary conditions (which need not be achievable). If extremal values are sought, then the nature of the solution must be further characterized. Hence, the supply of completely workable problems is limited, and further progress demands that the reader be drawn into the fabric of the theoretical development. Partly for this purpose, and partly to reserve the text proper for a presentation of essentials, I have intended that some of the problems at the end of each chapter be considered as integral to the exposition. Alternatively, these problems, frequently starred and cited at the appropriate points in the text, could be used as a basis for supplementary lectures. An effort has been made to distribute the difficulty of each problem thoughout a partitioned statement from which partial assignment may be made within context of the whole. Answers are provided for selected problems.

A significant number of problems for each chapter have been contributed by William Hrusa, who has maintained the interest in this subject, begun here at Syracuse, throughout his graduate career at Brown University. Moreover, he has been associated with this book since its inception and has supplied assistance and cogent comment during its preparation.

Although some effort has been made to acknowledge original contributors to this subject, I have embraced (with relief) the customary practice which permits drawing from unspecified sources that which might benefit this text. Most of these works are listed in the bibliography; to their authors, I am indebted.

I wish also to express my appreciation to those who first made me aware of the elegance and power of variational methods—Daniel Frederick at V.P.I., M. M. Schiffer at Stanford, and C. Lanczos as author. Special gratitude is offered to my colleagues at Syracuse, Philip Church and Wolfgang Jurkat who used some of this material in preliminary form in class and supplied valuable suggestions. In addition, useful comment was received from J. Erdman, and from my colleague Daniel Waterman, whose concerned encouragement has sustained my determination to produce this book. It is essential to recognize the many contributions from my students over the years, whose willingness to work with sets of notes in a state of daily transition—and to ask for more—convinced me of the value of the effort. They are too numerous to name here, but they will be remembered.

Finally, I wish to recognize the efforts of those responsible for the transformation of this work, from manuscript to printed page—of Louise Capra and Esther Clark, the principal typists; of Jody Bush, who endured innumerable revisions; and the staff of Springer-Verlag, who combined expertise with understanding during its production.

Syracuse
New York

JOHN L. TROUTMAN

Contents

CHAPTER 0

Review of Optimization in $\mathbb{R}^d$

This chapter presents a brief summary of the standard terminology and basic results related to characterizing the maximal and minimal values of a real valued function f defined on a set D in Euclidean space. With the possible exception of the remarks concerning convexity ((0.8) and (0.9)), this material is covered in texts on multidimensional calculus; the notation is explained in §1.5.

For $d = 1, 2, 3, \ldots$, let $\mathbb{R}^d$ denote d-dimensional real Euclidean space where a typical point or vector $X = (x_1, x_2, \ldots, x_d)$ has the length $|X| = (\sum_{j=1}^{d} |x_j|^2)^{1/2}$ which is positive unless $X = \mathcal{O} = (0, 0, 0, \ldots, 0)$.

On $\mathbb{R}^d$, with $Y = (y_1, y_2, \ldots, y_d)$, we have the vector space operations of componentwise addition

$$X + Y \overset{\text{def}}{=} (x_1 + y_1, x_2 + y_2, \ldots, x_d + y_d),$$

and scalar multiplication:

$$aX \overset{\text{def}}{=} (ax_1, ax_2, \ldots, ax_d), \qquad \forall\, a \in \mathbb{R}.$$

We may also express $|X| = (X \cdot X)^{1/2}$, utilizing the scalar or dot product

$$X \cdot Y \overset{\text{def}}{=} \sum_{j=1}^{d} x_j y_j,$$

which is subject to the *Cauchy* inequality

$$|(X \cdot Y)| \leq |X|\,|Y|. \tag{1}$$

The Cauchy inequality (1) is used to prove the so-called triangle inequality

$$|X + Y| \le |X| + |Y|, \tag{2a}$$

an alternate form of which is

$$||X| - |Y|| \le |X - Y|, \tag{2b}$$

where

$$X - Y \stackrel{\text{def}}{=} X + (-1)Y; \quad \text{(Problem 0.1).}$$

$|X - Y|$ defines the *Euclidean distance* between X and Y.

When $X_0 \in \mathbb{R}^d$, then for $\delta > 0$, the "sphere"

$$S_\delta(X_0) \stackrel{\text{def}}{=} \{X \in \mathbb{R}^d : |X - X_0| < \delta\}$$

is called an (open) *neighborhood* of X_0, and X_0 is said to be an *interior point* of each set D which contains this neighborhood for some $\delta > 0$. D is *open* when it consists entirely of interior points. An open set D is a *domain* when each pair of its points may be connected by a (polygonal) curve which lies entirely in D. Each open sphere is a domain, as is each open "box"

$$B = \{X \in \mathbb{R}^d : a_j < x_j < b_j, j = 1, 2, \ldots, d\},$$

but the union of disjoint open sets is *not* a domain, although it remains open.

A point *not* in the interior of a set S, and *not* interior to its complement, $\mathbb{R}^d \sim S$, is called a *boundary point* of S. The set of such points, denoted ∂S, is called the *boundary* of S. For example, if $S = \{(X \in \mathbb{R}^d : |X| \le 1\}$, then $\partial S = B = \{X \in \mathbb{R}^d : |X| = 1\}$; also $\partial B = B$.

We suppose that we are given a real valued function f defined on a set $D \subseteq \mathbb{R}^d$ for which we wish to find *extremal* values. That is, we wish to find points in D (called *extremal* points) at which f assumes maximum or minimum values. With such optimization problems we should note the following facts:

(0.0) *f need not have extremal values on D.*

For example, when $D = \mathbb{R}^1$, then the function $f(X) = x_1$ is unbounded in both directions on D. Moreover, when $D = (-1, 1) \subseteq \mathbb{R}^1$, this same function, although bounded, takes on values as near -1 or 1 as we please but does *not* assume the values ± 1 on $(-1, 1)$. On the closed interval, $D = [-1, 1]$, this function does assume both maximum and minimum values, but the function

$$f(X) = \frac{1}{x_1}, \qquad x_1 \ne 0,$$

$$f(\mathcal{O}) = 0,$$

is again unbounded.

(0.1) *f may assume only one extremal value on D.*

For example, on $D = (-1, 1]$ the function $f(X) = x_1$ assumes a maximum value $(+1)$, but not a minimum value, while on $(-1, 1)$ the function $f(X) = x_1^2$ assumes a minimum value (0) but not a maximum value.

(0.2) *f may assume an extremal value at more than one point.*

On $D = [-1, 1]$, $f(X) = x_1^2$ assumes a maximum value (1) at $x_1 = \pm 1$, while on $D = \mathbb{R}^2$, $f(X) = x_1^2$ assumes its minimum value (0) at every point located on the x_2 axis.

The only reasonable conditions which guarantee the existence of extremal values are contained in the following theorem whose proof is deferred. (See Proposition 5.3.)

(0.3) **Theorem.** *If $D \subseteq \mathbb{R}^d$ is compact and $f: D \to \mathbb{R}$ is continuous, then f assumes both maximum and minimum values on D.*

In $\mathbb{R}^d$, a *compact* set is a bounded set which is *closed* in that it contains each of its boundary points. In particular, each "box" of the form

$$\bar{B} = \{X \in \mathbb{R}^d: a_j \le x_j \le b_j, j = 1, 2, \ldots, d\}$$

for given real numbers $a_j \le b_j$, $j = 1, 2, \ldots, d$ is compact. However, the interval $(-1, +1)$ is *not* compact. (See §A.0.)

$f: D \to \mathbb{R}$ is *continuous* at $X_0 \in D$ *iff* for each $\varepsilon > 0$, $\exists\, \delta > 0$, such that when $X \in D$ and $|X - X_0| < \delta$, then $|f(X) - f(X_0)| < \varepsilon$; and f is continuous on D *iff* it is continuous at each point $X_0 \in D$.

The previous examples show that neither compactness nor continuity can alone assure the existence of extremal values.

(0.4) *The maximum value of f is the minimum value of $-f$ and vice versa.*

Thus it suffices to characterize the *minimum* points, those $X_0 \in D$ for which

$$f(X) \ge f(X_0), \qquad \forall\, X \in D. \tag{3}$$

As we have seen, such points may be present even on a noncompact set.

(0.5) *When D contains a neighborhood of X_0, an extremal point of f, in which f has continuous partial derivatives $f_{x_j} = \partial f/\partial x_j$, $j = 1, 2, \ldots, d$, then for each vector $U \in \mathbb{R}^d$ of unit length, the (two-sided) directional derivative:*

$$\partial_U f(X_0) \stackrel{\text{def}}{=} \lim_{\varepsilon \to 0} \left[\frac{f(X_0 + \varepsilon U) - f(X_0)}{\varepsilon}\right] = \frac{\partial f}{\partial \varepsilon}(X_0 + \varepsilon U)\bigg|_{\varepsilon = 0}$$
$$= 0.$$

[The bracketed quotient reverses sign as the sign of ε is changed. The existence and continuity of the partial derivatives ensures the existence of the limit which must therefore be zero.]

Introducing the *gradient vector* $\nabla f \stackrel{\text{def}}{=} (f_{x_1}, f_{x_2}, \ldots, f_{x_d})$, we may also express $\partial_U f(X_0) = \nabla f(X_0) \cdot U$, and conclude that at such an interior extremal point X_0,

$$\nabla f(X_0) = \mathcal{O}. \tag{4}$$

(0.6) *The points X_0 at which* (4) *holds, called stationary points* (*or critical points*) *of f, need not give either a maximum or a minimum value of f.*

For example, on $D = [-1, 1]$, the function $f(X) = x_1^3$ has $x_1 = 0$ as its only stationary point, but its maximum and minimum values occur at the end points 1 and -1 respectively.

On $D = \mathbb{R}^2$, the function $f(X) = x_2^2 - x_1^2$ has $X_0 = (0, 0)$ as its only critical point; it has there maximal behavior in the x_1 direction ($x_2 = 0$) and minimal behavior in the x_2 direction ($x_1 = 0$).

(In such cases, X_0 is said to be a *saddle point* of f.)

(0.7) *A stationary point X_0 may be* (*only*) *a local extremal point for f*; *i.e., one for which $f(X) \geq f(X_0)$* (*or $f(X) \leq f(X_0)$*) *for all $X \in D$ which are sufficiently near X_0.*

For example, the polynomial $f(X) = x_1^3 - 3x_1$ has on $D = [-3, 3]$, stationary points at $x_1 = -1, 1$; the first is (only) a local maximum point while the second is (only) a local minimum point for f. (See Figure 0.1.)

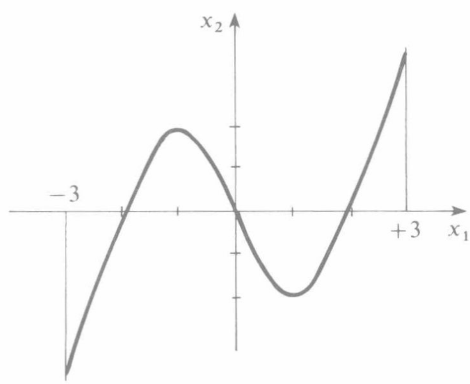


Figure 0.1

(0.8) *When f is a convex function on D then it assumes a minimum value at each stationary point in D.*