Applied Mathematical Sciences

Asen L. Dontchev

Lectures on Variational Analysis

Springer

Applied Mathematical Sciences

Volume 205

The mathematization of all sciences, the fading of traditional scientific boundaries, the impact of computer technology, the growing importance of computer modeling and the necessity of scientific planning all create the need both in education and research for books that are introductory to and abreast of these developments. The purpose of this series is to provide such books, suitable for the user of mathematics, the mathematician interested in applications, and the student scientist. In particular, this series will provide an outlet for topics of immediate interest because of the novelty of its treatment of an application or of mathematics being applied or lying close to applications. These books should be accessible to readers versed in mathematics or science and engineering, and will feature a lively tutorial style, a focus on topics of current interest, and present clear exposition of broad appeal. A compliment to the Applied Mathematical Sciences series is the Texts in Applied Mathematics series, which publishes textbooks suitable for advanced undergraduate and beginning graduate courses.

More information about this series at http://www.springer.com/series/34

Asen L. Dontchev

Lectures on Variational Analysis

Springer

Asen L. Dontchev
University of Michigan–Ann Arbor
Ann Arbor
MI, USA

ISSN 0066-5452 ISSN 2196-968X (electronic)
Applied Mathematical Sciences
ISBN 978-3-030-79910-6 ISBN 978-3-030-79911-3 (eBook)
https://doi.org/10.1007/978-3-030-79911-3

Mathematics Subject Classification: 49-02, 90-02, 49J52, 49J53, 49K40, 90C31

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Terry Rockafellar,
mathematician, mentor, and friend

Preface

Variational analysis can be briefly described as the mathematics of optimization and control. The name became commonly accepted after the publication of the book *Variational Analysis* by R. T. Rockafellar and R. J-B. Wets in 1998. However, basic concepts, ideas, and results that are now regarded as being in the heart of the area have been around much longer. Variational analysis unifies theories and techniques that have been developed in *calculus of variations*, *mathematical programming*, and *optimal control* and covers areas such as *convex analysis, nonlinear analysis, nonsmooth analysis,* and *set-valued analysis*. In a separate direction, it has been defined to include topics in differential geometry and functional analysis. It is a good example of mathematics with deep roots, rich theory, and a variety of applications, whose rapid growth in the last decades has been driven not only by the curiosity of researchers but also by very practical problems such as financial planning or steering a flying object.

The aim of this book is to present an introduction to variational analysis for graduate students, researchers and practitioners in the broad area of mathematical sciences and engineering, including operations research, economics, and finances. The focus is on problems with *constraints*, the analysis of which involves set-valued mappings and functions that are not differentiable, at least in the usual sense. A typical reader of the book should be familiar with multivariable calculus and linear algebra. Some basic knowledge in optimization, control, and elementary functional analysis is desirable but not necessary—all needed background material is included in the book.

The book is structured in 20 lectures, each one of which is devoted to a particular topic. The choice of topics reflects the author's views and interests and by no means covers all aspects of the field. The material is presented in a theorem-proof format, but there are also examples, exercises, and discussions. In each lecture, the proofs that are essential for understanding the concepts and techniques involved are given in the text, and reading these proofs is highly recommended. More technical or lengthy

e.g., as in the book by M. Morse, *Variational analysis: critical extremals and Sturmian extensions,* Interscience Publishers, 1973, which is mainly about calculus of variations at large.

proofs are given at the end of the lecture and could be skipped at a first reading. Most of the exercises are supplied with hints or detailed guides. Some classical results in analysis such as, e.g., (a version of) the Baire category theorem, are given without proofs in the lecture where they are first used. Each lecture has its own numbering of formulas; citing formulas from other lectures is avoided.

The book starts with a preparatory part presenting in a condensed form notations and terminology, along with some basic concepts and facts from functional analysis. Lecture 1 introduces standard optimization problems and in particular shows how to derive necessary optimality conditions in the format of generalized equations. Lecture 2 is about continuity of set-valued mappings. In addition, conditions are presented for continuity of the optimal value and optimal solution mappings of a general optimization problem depending on a parameter. Lipschitz continuity of set-valued mappings is also introduced in that lecture but explored more broadly in the following Lecture 3, which deals in particular with polyhedral mappings with application to linear programming, as well as outer Lipschitz continuity of piecewise polyhedral mappings.

The next several lectures are devoted to regularity properties of set-valued mappings with applications to nonlinear optimization. *Metric regularity* of mapping acting in metric spaces, together with the companion properties of Aubin continuity and linear openness, are introduced in Lecture 4. Lecture 5 is devoted to stability of metric regularity under perturbations as defined by the Lyusternik-Graves theorem. Lecture 6 presents the Robinson-Ursescu theorem, which gives a characterization of metric regularity of mappings with closed and convex graphs. Lecture 7 deals with derivative criteria for metric regularity that utilize generalized differentiation of set-valued mappings. *Strong regularity* is introduced in Lecture 8 together with the Robinson theorem, a far reaching extension of the classical implicit function theorem. The next one, Lecture 9, is devoted to two basic implicit function theorems for nondifferentiable functions, one due to F. Clarke and the other to B. Kummer. Lecture 10 characterizes strong regularity of mappings appearing in quadratic optimization problems in a Hilbert space setting, and in a basic nonlinear programming problem. *Strong subregularity* is introduced and discussed in Lecture 11. Lecture 12 presents a version of the Bartle-Graves theorem for set-valued mappings, which is concerned with the existence of *continuous selections* of the inverses of metrically regular mappings. Lecture 13 deals with so-called radius theorems associated with regularity properties. Lectures 7, 12, and 13 are technically involved and can be skipped at a first reading of the book; they are not used in the further lectures. Lecture 14 broadens the spectrum of applications of the regularity theory; it shows convergence of the Newton method applied to generalized equations involving mappings having the regularity properties given in the preceding lectures. A version of the Newton method for a class of nondifferentiable functions, the so-called semismooth functions, is presented in Lecture 15.

The final five lectures are devoted to applications of variational analysis to optimal control problems. The main focus is on the linear-quadratic optimal control problem with control constraints, which is relatively simple but still rich enough to allow demonstrating applications of basic ideas and techniques of variational analysis, without going deep into differential equations and functional analysis. Lecture 16 derives optimality conditions for the linear-quadratic problem. Lecture 17 considers a more general nonlinear optimal control problem and shows how to apply the theorems of Lyusternik-Graves, Robinson-Ursescu, and Robinson in order to obtain conditions for metric regularity and strong regularity of feasibility and optimality mappings for that problem. Lecture 18 is devoted to a discrete approximation of the linear-quadratic and a nonlinear optimal control problem, for which error estimates are obtained by using regularity properties of the mappings involved. In Lecture 19, optimal feedback control is introduced and discussed, and the existence of such a feedback is proven for the control-constrained linear-quadratic problem. The final Lecture 20 is devoted to model predictive control, which is presented as an approach to approximate optimal feedback; the main result presented is about the accuracy of such an approximation.

Parts of this book can be used in various graduate courses in the general area of optimization and control, depending on the audience and the specific goals of the lecturer. For example, combining the material of lectures 1–3, 8–11, and 15 with some adjustments and additions may become a one-semester course on modern optimization theory. Lectures 1–8, and 13–17 provide a basis for a course on variational analysis for mathematically mature students. Lectures 4–6, 12, and 15–20 may be used in a course on control, addressing in particular optimality, regularity, as well as more specific topics such as discrete approximations and model predictive control.

The idea to write this book was born in the spring of 2020, when I gave a one-semester graduate course on variational analysis in the Faculty of Mathematics and Informatics, Sofia University, Bulgaria. I was not able to find a suitable book that could be used as a textbook for the course. Then I prepared lecture notes which evolved into this book. While working on the book, new topics have been added to obtain a broader coverage of nonsmooth analysis and regularity of mappings, as well as topics in optimal control. The choice of the material in the book was also inspired by the decades long research collaboration of the author with students, postdocs, and faculty from the University of Michigan in Ann Arbor.

I am indebted to a number of people who have helped me with this book. During the preparation of the manuscript, I benefited from extensive discussions with V. Veliov, in particular concerning the lectures on control. Special thanks to R. Goebel, C. Josz, M. Krastanov, D. Liao-McPherson, J. Leung, Y. Stoev, and N. Zlateva, who had read preliminary drafts and made valuable suggestions, and especially to R. Rozenov, who also helped with the figures. I am also thankful to R. Bot, D. Drusvyatskiy, D. Klatte, I. Kolmanovsky, S. Robinson, and T. Zolezzi for their advice and encouragement. Many thanks to the anonymous referees for their valuable remarks. I obtained additional feedback from students who attended my course at Sofia University, and in particular from G. Angelov, S. Apostolov, M. Konstantinov, M. Nikolova, B. Stefanov, M. Tasheva, and I. Vasilev. The support from AFOSR under the grant FA9550-20-1-0385 is greatly appreciated.

This book would have never been written without the moral support of my family. In particular, I am thankful to my wife Dora for her patience and understanding, to my daughter Mira for her help with typesetting, and to my son Kiko for the many questions he asked me about this project.

Ann Arbor, MI, USA Asen L. Dontchev
June 2021

In Memoriam. In finishing the Preface above, Asen Dontchev knew he had a terminal illness. However, in putting September 2021 beneath his name he had no inkling that, already on the 16th of that month, his life would come to its end.

Through his original research and many publications, he contributed hugely to the mathematics of optimization and control. This volume of lectures, brought forth in troublesome final circumstances, will help to spread the understanding of that subject and its evolving applications. It shows his remarkable qualities as a thinker and writer, being able to combine theoretical with practical while revealing the vital heart of every topic. He always stressed the importance of explaining things to newcomers to variational analysis, building on their diverse motivations from other areas and trying to ease their way around technicalities. For all this, he will be sadly missed and lastingly remembered.

Terry Rockafellar

Contents

0 Notation, Terminology, and Some Functional Analysis 1

1 Basics in Optimization 11

2 Continuity of Set-Valued Mappings 23

3 Lipschitz Continuity of Polyhedral Mappings 29

4 Metric Regularity 37

5 Lyusternik–Graves Theorem 47

6 Mappings with Convex Graphs 53

7 Derivative Criteria for Metric Regularity 61

8 Strong Regularity 77

9 Variational Inequalities over Polyhedral Sets 87

10 Nonsmooth Inverse Function Theorems 97

11 Lipschitz Stability in Optimization 103

12 Strong Subregularity 113

13 Continuous Selections 125

14 Radius of Regularity 139

15 Newton Method for Generalized Equations 147

16 The Constrained Linear-Quadratic Optimal Control Problem 157

17 Regularity in Nonlinear Control 167

18 Discrete Approximations 177

19 Optimal Feedback Control 189

20 Model Predictive Control 197

Bibliographical Remarks and Further Reading 205

References 209

List of Symbols 215

Index 217

Lecture 0
Notation, Terminology, and Some Functional Analysis

We start with a brief introduction to metric spaces. Let X be a set, $I\!R$ be the set of reals and $I\!R_+$ be the set of nonnegative reals. A *metric* ρ in X is a mapping $\rho : X \times X \to I\!R_+$ which has the following properties:

– $\rho(x, y) = 0 \quad\Longrightarrow\quad x = y$;

– $\rho(x, y) = \rho(y, x)$;

– $\rho(x, z) \le \rho(x, y) + \rho(y, z)$ (the triangle inequality).

The set X equipped with a metric ρ is called a *metric space* and denoted (X, ρ). In such a space the closed ball with center $x \in X$ and radius $r \in I\!R_+$ is the set

$$I\!B_r(x) = \left\{ y \in X \,\middle|\, \rho(y, x) \le r \right\}.$$

The open ball with center x and radius r is defined in the same way with $\le$ replaced by $<$ and denoted $\operatorname{int} I\!B_r(x)$.

A metric space X is said to be *complete* when every Cauchy sequence $\{x_n\}$ in X converges. That is to say: if $\rho(x_n, x_m) \to 0$ as both n and m independently go to infinity, then there is $y \in X$ such that $\rho(x_n, y) \to 0$.

A *linear* or *vector* space is a set supplied with two operations: addition and multiplication by a scalar, which obey standard rules such as commutativity, associativity, and distributivity. All linear spaces considered are over the reals. A metric ρ acting in a linear space X is said to be *shift-invariant* when

$$\rho(y + z, y' + z) = \rho(y, y') \ \text{ for all } \ y, y', z \in X.$$

A linear space X is said to be *normed* when there is a mapping $\|\cdot\| : X \to I\!R$ defined on X and called *norm* that has the following properties:

– $\|x\| \ge 0$ for all $x \in X$ and $\|x\| = 0 \quad\Longleftrightarrow\quad x = 0$;

– $\|\alpha x\| = |\alpha|\|x\|$ for all $x \in X$ and $\alpha \in I\!R$, where $|\alpha|$ denotes the absolute value of a real α: $|\alpha| = \max\{0, \alpha\}$;

– $\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in X$.

A. L. Dontchev, *Lectures on Variational Analysis*, Applied Mathematical Sciences 205, https://doi.org/10.1007/978-3-030-79911-3_0

A normed space is a metric space with a metric $\rho(x, y) = \|x - y\|$. In such a space the closed unit ball $I\!B_1(0)$ is denoted by $I\!B$ and the open unit ball is then $\operatorname{int} I\!B$. A space that is linear, normed and complete is called a *Banach space* .

Given a linear space X, an *inner* or *scalar* product on X is a function $\langle \cdot, \cdot \rangle : X \times X \to I\!R$ with the following properties:

– $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in X$;
– $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle$ for all $x_1, x_2, y \in X$;
– $\langle \lambda x, y \rangle = \lambda \langle y, x \rangle$ for all $x, y \in X$ and $\lambda \in I\!R$;
– $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \iff x = 0$.

Then $\sqrt{\langle x, x \rangle}$ is a norm of x. A linear space X equipped with a scalar product and an associated norm which is complete with respect to that norm is said to be a *Hilbert space*. A standard fact for such a space is the Cauchy-Schwarz inequality

$$\langle u, v \rangle \leq \|u\| \|v\|.$$

The space consisting of all linear and continuous real-valued functions on a Banach space X is another Banach space which is *dual* or *adjoint* to X, denoted by X^*. The value that an $x^* \in X^*$ assigns to an $x \in X$ is called the duality mapping. The dual of the Banach space X^* is the bidual X^{**} of X. When every function $x^{**} \in X^{**}$ on X^* can be represented as $x^* \mapsto \langle x^*, x \rangle$ for some $x \in X$, the space X is called *reflexive*. In particular, when X is a Hilbert space with inner product $\langle x, y \rangle$, each $x^* \in X^*$ corresponds to a function $x \mapsto \langle x, y \rangle$ for some $y \in X$, so that X^* can be identified with X itself.

The n-dimensional Euclidean space, denoted $I\!R^n$, is a linear space of vectors

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

equipped with the Euclidean norm

$$\|x\| = \sqrt{\langle x, x \rangle} = \left[\sum_{j=1}^n x_j^2 \right]^{1/2}$$

which is associated with the canonical inner product

$$\langle x, y \rangle = \sum_{j=1}^n x_j y_j.$$

The space $I\!R$ is the set of reals, the real line, whose norm is the absolute value $| \cdot |$.

Given a metric space (X, ρ), a point $x \in X$ and a set $C \subset X$, the quantity

$$d(x, C) = \inf_{y \in C} \rho(x, y)$$

is called the *distance* from x to C. We adopt the convention that the distance from any point to the empty set is $+\infty$. Any point y of C which is closest to x in the sense of achieving this distance is called a *projection* of x on C. The set of projections is denoted by $P_C(x)$.

In any metric space, a *neighborhood* of x is any set U for which there exists a positive number r such that $\mathbb{B}_r(x) \subset U$. We recall that the interior of a set $C \subset \mathbb{R}^n$ consists of all points x such that C is a neighborhood of x, whereas the closure of C consists of all points x such that the *complement* of C is *not* a neighborhood of x; C is *open* if it coincides with its interior and *closed* if it coincides with its closure. A nonempty set $C \subset X$ is closed if and only if every $x \in X$ with $d_C(x) = 0$ belongs to C. The interior is denoted by $\operatorname{int} C$. The union of *any* number of open sets is open, while the intersection of a *finite* number of open sets is open.

A set C is said to be *locally closed* at a point $x \in C$ when there exists a neighborhood U of x such that the intersection $C \cap U$ is a closed set. A set C is said to be *dense* in a closed set D when the closure $\operatorname{cl} C$ of C coincides with D, or equivalently, when for any $x \in D$ any neighborhood U of x contains elements of C. A set C is said to be *compact* when every open cover of it has a finite subcover. That is, C is compact if for every collection $\mathcal{T}$ of open sets $U \subset C$ such that $C \subset \cup_{\mathcal{T}} U$, there is a finite subset $\mathcal{F}$ of $\mathcal{T}$ such that $C \subset \cup_{\mathcal{F}} U$. In metric spaces this is the same as saying that every sequence $\{x_n\}$ with $x_n \in C$ has a subsequence which is convergent to an element of C. If $C \subset \mathbb{R}^n$ then compactness of C is equivalent to C being both bounded and closed.

A set C in a linear space X is said to be *convex* if for every $x, x' \in C$ and every $\lambda \in [0, 1]$, the point $\lambda x + (1 - \lambda)x'$ is in C. A function $f : X \to \mathbb{R}$ with domain containing a convex set C is said to be *convex* on C, or over C, or relative to C, if for every $x, x' \in C$ and every $\lambda \in (0, 1)$,

$$f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda) f(x').$$

If $\leq$ is replaced by $<$ for $x \neq x'$, the function f is said to be *strictly convex* on C. A function $f : X \to \mathbb{R}$ defined in a Banach space X is said to be *strongly convex* on C if there exists a constant $\alpha > 0$ such that the function $f(x) - \alpha\|x\|^2$ is convex. A function $f : X \to \mathbb{R}$ is *concave* when $-f$ is convex.

The *convex hull* of a set $C \subset \mathbb{R}^n$, which will be denoted by $\operatorname{co} C$, is the smallest convex set that includes C. It can be identified as the intersection of all convex sets that include C, but also can be described as consisting of all linear combinations $\lambda_0 x_0 + \lambda_1 x_1 + \cdots + \lambda_n x_n$ with $x_i \in C$, $\lambda_i \geq 0$, and $\lambda_0 + \lambda_1 + \cdots + \lambda_n = 1$; this is the Carathéodory theorem. The *closed convex hull* of C is the closure of the convex hull of C and denoted $\operatorname{cl} \operatorname{co} C$; it is the smallest closed convex set that contains C.

A sequence x_k with elements in a metric space (X, ρ) is said to strongly converge to x, or simply to converge to x, when for every $\varepsilon > 0$ there exists a natural number K such that for all $k \geq K$ we have $\rho(x_k, x) \leq \varepsilon$. A sequence x_k in a Banach space X with a dual X^* is weakly convergent to x when for every $x^* \in X^*$ the sequence $\langle x^*, x_k \rangle$ converges to $\langle x^*, x \rangle$. For a Hilbert space X weak convergence means that for every $y \in X$ the sequence of reals $\langle y, x_k \rangle$ is convergent to $\langle y, x \rangle$. Weak convergence determines weak compactness, i.e., a set C in a Hilbert space X is weakly compact when every sequence has a weakly convergent subsequence. A basic result in that context says that any bounded, closed and convex set in a Hilbert space is weakly compact.

In further lines we introduce notations and terminology for mappings. A mapping F acting from a set X to a set Y is generally denoted as

$$X \ni x \mapsto F(x) \subset Y,$$

where $x \in X$ and $F(x)$ is the image of x which in general is a subset of Y. The *domain* of a mapping F acting between X and Y is

$$\mathrm{dom}\ F = \{x \in X \mid F(x) \neq \emptyset\},$$

the *graph* of F is

$$\mathrm{gph}\ F = \{(x, y) \in X \times Y \mid y \in F(x)\},$$

while the *range* of F is

$$\mathrm{rge}\ F = \{y \in Y \mid \text{ there exists } x \in \mathrm{dom}\ F \text{ with } y \in F(x)\}.$$

A mapping $F : X \to Y$ is said to be a *function* when for every $x \in X$ the image $F(x) \subset Y$ is just one point or the empty set. We will also consider mappings for which the image $F(x)$ of a point x may consist of more than one point and call such mappings *set-valued mappings*. In the notation for set-valued mappings we use capital letters and double arrows, e.g., $F : X \rightrightarrows Y$, versus small letters and single arrows for functions, e.g., $f : X \to Y$. Every function may be viewed as a set-valued mapping. A set-valued mapping which is not a function, that is, having multiple values at certain points in its domain, is said to be a *multivalued mapping*, in contrast to a function, which is a *single-valued mapping* in its domain.

We define the *inverse* of a set-valued mapping $F : X \rightrightarrows Y$ as

$$Y \ni y \mapsto F^{-1}(y) = \{x \in X \mid y \in F(x)\}.$$

According to this definition, every mapping has an inverse. In particular, the inverse of a function always exists, but it may be multivalued, that is, not a function. A simple example is given in Fig. 0.1.

We introduce next a concept which identifies the case when a set-valued mapping is locally a function . Let X and Y be metric spaces.

Single-Valued Graphical Localization of a Set-Valued Mapping. *For $F : X \rightrightarrows Y$ and a pair $(\bar{x}, \bar{y}) \in \mathrm{gph}\ F$, a function s is said to be a* single-valued graphical localization *of F around $\bar{x}$ for $\bar{y}$ if there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that $U \subset \mathrm{dom}\ s$ and*

$$\mathrm{gph}\ s = (U \times V) \cap \mathrm{gph}\ F,$$

so that

$$s : x \mapsto \begin{cases} F(x) \cap V & \text{when } x \in U, \\ \emptyset & \text{otherwise.} \end{cases}$$

In the example displayed in Fig. 0.1, the inverse of the function $x \mapsto x^2$ has a single-valued localization around any $x > 0$ for $y = \sqrt{x}$ and another one around any $x > 0$ for $y = -\sqrt{x}$; moreover, it has no single-valued localization around 0 for 0 and is empty valued for any $x < 0$.

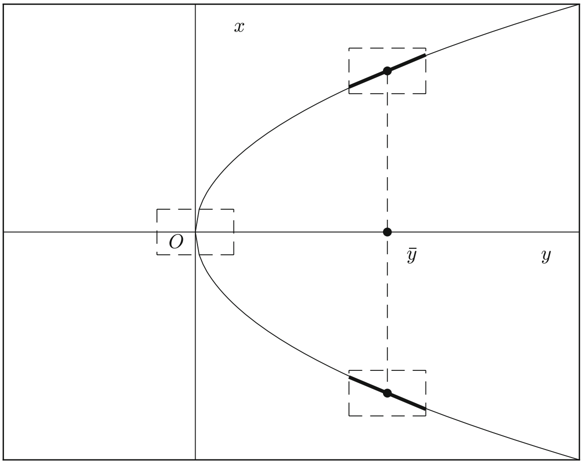

Fig. 0.1: The inverse of the function $x \mapsto x^2$

Let X and Y be Banach spaces. A *linear* mapping A acting from X to Y is a function with dom $A = X$ which obeys the rule for linearity:

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay \quad \text{for all } x,\ y \in X \text{ and all scalars } \alpha, \beta \in I\!R.$$

A linear mapping $A : X \to Y$ acting between Banach spaces X and Y is *bounded* when there exists a constant $\alpha \geq 0$ such that

$$\|Ax\| \leq \alpha \|x\| \quad \text{for all } x \in X.$$

The space of linear bounded mappings acting from a Banach space X to a Banach space Y is denoted by $\mathcal{L}(X,Y)$. It is a Banach space when equipped with the operator norm $\sup_{x \in I\!B} \|Ax\|$. A mapping $A \in \mathcal{L}(X,Y)$ is said to be *surjective* or *onto* when for every $y \in Y$ there is a $x \in X$ with $Ax = y$; this is denoted as $AX = Y$. We also use the notation $\ker A = \{x \in X \mid Ax = 0\}$.

Let A be a linear and bounded mapping A acting from Hilbert space X into itself. Then the mapping A^* defined as

$$\langle A^*x, y\rangle = \langle Ay, x\rangle \quad \text{for all } x, y \in X$$

is linear and bounded and is called the *adjoint mapping* associated with A. If $A = A^*$ the mapping A is said to be *selfadjoint*.

In $\mathbb{R}^n$ we make a distinction between a linear mapping and its matrix. Specifically, a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ is represented by a matrix, which, with some abuse of notation, we denote again by A; here the matrix A is with m rows, n columns, and components $a_{i,j}$. The transpose of A, denoted by A^{T}, represents the adjoint of the mapping A. If the matrix A is nonsingular, which requires $m = n$, then the inverse A^{-1} of A is also a linear mapping represented by the inverse of the matrix A. More generally, if $m \leq n$ and the rows of the matrix A are linearly independent, then the rank of the matrix A is m and the mapping A is surjective. In this case the matrix AA^{T} is nonsingular. On the other hand, if $m \geq n$ and the columns of A are linearly independent then $A^{\mathsf{T}}A$ is nonsingular. Both the identity mapping and its matrix will be denoted by I, regardless of dimensionality.

Let (X, ρ) and (Y, ρ) be metric spaces. A function $f : X \to Y$ is said to be *continuous* at $\bar{x} \in \operatorname{dom} f$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho(f(x), f(\bar{x})) \leq \varepsilon$ whenever $\rho(x, \bar{x}) \leq \delta$. A function f is said to be continuous on, or over, or relative to, a set $D \subset \operatorname{dom} f$ if it is continuous at every $x \in D$. A function $f : X \to Y$ is said to be *open* at $\bar{x} \in \operatorname{dom} f$ when for any neighborhood U of $\bar{x}$ the set $f(U)$ is a neighborhood of $f(\bar{x})$. A function $f : X \to \mathbb{R}$ is said to be *lower semicontinuous* on a set C if for every point $x \in C$ and every sequence $x_k \in C$ convergent to x one has $\liminf_{k\to\infty} f(x_k) \geq f(x)$. Symmetrically, a function $f : X \to \mathbb{R}$ is *upper semicontinuous* on a set C if for every point $x \in C$ and every sequence $x_k \in C$ convergent to x one has $\limsup_{k\to\infty} f(x_k) \leq f(x)$. Recall the classical Weierstrass theorem:

Weierstrass Theorem. *Let C be a nonempty compact set in a metric space X. Then every lower semicontinuous function on C attains its minimum on C and every upper semicontinuous function on C attains its maximum on C. Hence, any continuous on C function attains both its minimum and maximum on C.*

A function $f : X \to \mathbb{R}$ acting on a Hilbert space X is said to be weakly lower semicontinuous at a point x if it is lower semicontinuous with respect to the weak convergence in X. The Weierstrass theorem extends to the weak versions of compactness and lower semicontinuity: any weakly lower semicontinuous function f on Hilbert space X attains its minimum on a weakly compact set and, in particular, on a closed and convex set. A continuous function which is convex, is weakly lower semicontinuous.

Let X and Y be Banach spaces. A function $f : X \to Y$ is said to be *Fréchet differentiable* at a point x when $x \in \operatorname{int} \operatorname{dom} f$ and there is a linear and bounded mapping $A : X \to Y$ with the property that for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|f(x+h) - f(x) - Ah\| \leq \varepsilon\|h\| \quad \text{for every } h \in X \text{ with } \|h\| < \delta.$$

If such a mapping A exists, it is unique; it is denoted by $Df(x)$ and called the *Fréchet derivative* of f at x. A function $f : X \to Y$ is said to be *twice Fréchet differentiable* at a point $x \in \operatorname{int} \operatorname{dom} f$ when it is Fréchet differentiable at x and there is a bilinear mapping $N : X \times X \to Y$ with the property that for every $\varepsilon > 0$ there exists $\delta > 0$

with

$$\|f(x+h) - f(x) - Df(x)h - N(h,h)\| \le \varepsilon\|h\|^2 \quad \text{for every } \; h \in X \; \text{ with } \; \|h\| < \delta.$$

If such a mapping N exists, it is unique and is called the *second Fréchet derivative* of f at x, denoted by $D^2 f(x)$. Higher-order derivatives can be defined accordingly. (For functions acting in finite dimensions these are the usual derivatives from calculus; then we omit Fréchet.) When the Fréchet derivative mapping $x \mapsto Df(x)$ exists and is continuous (with respect to the operator norm) on an open set $C \subset X$, then we say that the function f is *continuously differentiable* on C; we also call such a function *smooth* on C. Analogously, for an integer k we define k times continuously differentiable functions. The set of such functions is denoted by C^k.

For a function $f : P \times X \to Y$ and a pair $(p,x) \in \operatorname{int} \operatorname{dom} f$, the *partial Fréchet derivative* mapping $D_x f(p,x) : X \to Y$ of f with respect to x at (p,x) is the Fréchet derivative of the function $g(y) = f(p,y)$ at x. If the partial derivative mapping is a continuous function in a neighborhood of (p,x), then f is said to be continuously differentiable with respect to x around (p,x).

For a function $f : I\!\!R^n \to I\!\!R^m$ we distinguish between the derivative as a linear mapping and its matrix. The $m \times n$ matrix that represents the derivative $Df(x)$ at x is called the *Jacobian* of f at x and is denoted by $\nabla f(x)$. The second derivative is denoted by $\nabla^2 f(x)$ and so on. In the notation $x = (x_1, \ldots, x_n)$ and $f = (f_1, \ldots, f_m)$, the components of $\nabla f(x)$ are the partial derivatives of the component functions f_i:

$$\nabla f(x) = \left(\frac{\partial f_i}{\partial x_j}(x) \right)_{i,j=1}^{m,n}.$$

For $f : I\!\!R^d \times I\!\!R^n \to I\!\!R^m$ the partial derivative $\nabla_x f(p,x)$ is represented by an $m \times n$ matrix, denoted $\nabla_x f(p,x)$ and called the partial Jacobian. It's a standard fact from calculus that if the function $(p,x) \mapsto f(p,x)$ is differentiable with respect to both p and x around $(\bar{p},\bar{x})$ and the partial Jacobian mappings $(p,x) \mapsto \nabla_x f(p,x)$ and $(p,x) \mapsto \nabla_p f(p,x)$ are continuous around $(\bar{p},\bar{x})$, then f is continuously differentiable around $(\bar{p},\bar{x})$.

In this book we will also employ the following weaker notion of derivative; for simplicity we stick with finite dimensions. For $f : I\!\!R^n \to I\!\!R^m$, a point $\bar{x} \in \operatorname{dom} f$ and a vector $d \in I\!\!R^n$, the limit

$$f'(\bar{x}; d) = \lim_{t \searrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t},$$

when it exists, is called the *directional derivative* of f at $\bar{x}$ for d. If this directional derivative exists for every d, f is said to be *directionally differentiable* at $\bar{x}$.

Let X and Y be metric spaces where both metrics are denoted by ρ but may be different. A function $f : X \to Y$ is said to be *Lipschitz continuous* relative to a set C, or on a set C, if $C \subset \operatorname{dom} f$ and there is a constant $\kappa \ge 0$ such that

$$\rho(f(x'), f(x'')) \le \kappa\rho(x', x'') \quad \text{for all } \; x', x'' \in C. \tag{1}$$

If f is Lipschitz continuous relative to a neighborhood of a point $x \in \operatorname{int} \operatorname{dom} f$, then f is said to be Lipschitz continuous *around* x. We say further, in the case of an open set C, that f is locally Lipschitz continuous on C if it is a Lipschitz continuous function around every point x of C.

For a function $f : X \to Y$ and a point $x \in \operatorname{int} \operatorname{dom} f$, the *Lipschitz modulus* of f at x, denoted $\operatorname{lip}(f; x)$, is the infimum of the set of values of κ for which there exists a neighborhood C of x such that (1) holds. Equivalently,

$$\operatorname{lip}(f; x) = \limsup_{\substack{x', x'' \to x, \\ x' \neq x''}} \frac{\rho(f(x'), f(x''))}{\rho(x', x'')}.$$

A function f is Lipschitz continuous around x if and only if $\operatorname{lip}(f; x) < \infty$. For an open set C, a function f is *locally Lipschitz continuous* on C exactly when $\operatorname{lip}(f; x) < \infty$ for every $x \in C$.

Let P, X and Y be metric spaces. A function $f : P \times X \to Y$ is said to be Lipschitz continuous with respect to x around $(p, x) \in \operatorname{int} \operatorname{dom} f$ when the function $y \mapsto f(p, y)$ is Lipschitz continuous around x; the associated Lipschitz modulus of f with respect to x is denoted by $\operatorname{lip}_x(f; (p, x))$. We say f is Lipschitz continuous with respect to x *uniformly in* p around $(p, x) \in \operatorname{int} \operatorname{dom} f$ when there are neighborhoods Q of p and U of x along with a constant κ and such that

$$\rho(f(p', x''), f(p', x')) \leq \kappa\rho(x'', x') \quad \text{for all } x'', x' \in U \text{ and } p' \in Q.$$

Accordingly, the partial uniform Lipschitz modulus with respect to x has the form

$$\widehat{\operatorname{lip}}_x(f; (p, x)) := \limsup_{\substack{x'', x' \to x, p' \to p, \\ x'' \neq x'}} \frac{\rho(f(p', x''), f(p', x'))}{\rho(x'', x')}.$$

A one-point version of the Lipschitz continuity is a property called *calmness*. A function $f : X \to Y$ is said to be *calm* at x relative to a set D in X if $x \in D \cap \operatorname{dom} f$ and there exists a constant $\kappa \geq 0$ such that

$$\rho(f(x'), f(x)) \leq \kappa\rho(x', x) \quad \text{for all } x' \in D \cap \operatorname{dom} f. \tag{2}$$

For a function $f : X \to Y$ and a point $x \in \operatorname{dom} f$, the calmness modulus of f at x, denoted $\operatorname{clm}(f; x)$, is the infimum of the set of values $\kappa \geq 0$ for which there exists a neighborhood D of x such that (2) holds. The definition of the partial uniform calmness modulus is completely analogous to that of the partial uniform Lipschitz modulus.

Let X and Y be Banach spaces. Having the concept of calmness, we can interpret the Fréchet differentiability of a function $f : X \to Y$ at a point $x \in \operatorname{int} \operatorname{dom} f$ as the existence of a linear mapping $Df(x) : X \to Y$ such that

$$\operatorname{clm}(e; x) = 0 \quad \text{for } e(x') = f(x') - [f(x) + Df(x)(x' - x)].$$

Furthermore, we have that $\operatorname{clm}(f;x) = \|Df(x)\|$. A sharper concept of derivative is tied up with the Lipschitz modulus. A function $f : X \to Y$ is said to be *strictly* Fréchet differentiable at a point x if there is a linear and bounded mapping $A : X \to Y$ such that

$$\operatorname{lip}(e;x) = 0 \quad \text{for} \quad e(x') = f(x') - [f(x) + A(x' - x)].$$

Specifically, in this case we have that for each $\varepsilon > 0$ there exists a neighborhood U of x such that

$$\|f(x'') - [f(x') + Df(x)(x'' - x')]\| \le \varepsilon\|x'' - x'\| \quad \text{for every} \quad x'', x' \in U.$$

Clearly, the strictly Fréchet differentiable functions are Fréchet differentiable and the continuously Fréchet differentiable functions are strictly Fréchet differentiable functions, with $\operatorname{lip}(f;x) = \|Df(x)\|$. Sometimes, when clear from the context, we omit "Fréchet."

We introduce next the notion of semidifferentiability. First, we need the following definition: A function $\varphi : X \to Y$ is said to be *positively homogeneous* if $0 = \varphi(0)$ and $\varphi(\lambda w) = \lambda\varphi(w)$ for all $w \in \operatorname{dom} \varphi$ and $\lambda > 0$. This mean geometrically that the graph of φ is a cone in $X \times Y$. Any linear function is positively homogeneous in particular.

Semidifferentiability. *A function $f : X \to Y$ is said to be semidifferentiable at x if there exists a continuous and positively homogeneous function $\varphi : X \to Y$ such that*

$$\operatorname{clm}(e;x) = 0 \quad \textit{for} \quad e(x') = f(x') + \varphi(x' - x).$$

If the stronger condition holds that

$$\operatorname{lip}(e;x) = 0 \quad \textit{for} \quad e(x') = f(x') + \varphi(x' - x),$$

then f is said to be strictly semidifferentiable at x. Either way, the function φ, necessarily unique, is called the semiderivative of f at x and denoted by $Df(x)$. In the literature, this kind of derivative is also called Bouligand or B-derivative.

For a function f and a point $x \in \operatorname{dom} f$, semidifferentiability of f at x is equivalent to the existence of the limit

$$\lim_{t \searrow 0, w' \to w} \frac{f(x + tw') - f(x)}{t} \quad \text{for every } w \in X \ .$$

If a function f is semidifferentiable at x, then f is in particular directionally differentiable at x and has

$$f'(x;w) = Df(x)(w) \quad \text{for all} \quad w.$$

When $\operatorname{lip}(f;x) < \infty$, directional differentiability at x in turn implies semidifferentiability at x.

When the semiderivative $Df(x)$ is linear, semidifferentiability turns into differentiability, and strict semidifferentiability turns into strict differentiability. The connections known between $Df(x)$ and the calmness modulus and Lipschitz modulus of f at x under differentiability can be extended to semidifferentiability by adopting the definition that

$$\|\varphi\| = \sup_{\|x\|\le 1} \|\varphi(x)\| \quad \text{for a positively homogeneous function } \varphi.$$

We then have $\operatorname{clm}(Df(x);0) = \|Df(x)\|$ and consequently $\operatorname{clm}(f;x) = \|Df(x)\|$, which in the case of strict semidifferentiability becomes $\operatorname{lip}(f;x) = \|Df(x)\|$.

Examples.

(1) The function $f(x) = e^{|x|}$ for $x \in I\!R$ is not differentiable at 0, but it is semidifferentiable there and its semiderivative is given by $Df(0) : w \mapsto |w|$. This is actually a case of strict semidifferentiability. Away from 0, f is of course continuously differentiable (hence strictly differentiable).

(2) The function $f(x_1, x_2) = \min\{x_1, x_2\}$ on $I\!R^2$ is continuously differentiable at every point away from the line where $x_1 = x_2$. On that line, f is strictly semidifferentiable with

$$Df(x_1, x_2)(w_1, w_2) = \min\{w_1, w_2\}.$$

(3) A function of the form $f(x) = \max\{f_1(x), f_2(x)\}$, with f_1 and f_2 continuously differentiable from $I\!R^n$ to $I\!R$, is strictly differentiable at all points x where $f_1(x) \neq f_2(x)$ and semidifferentiable where $f_1(x) = f_2(x)$, the semiderivative being given there by

$$Df(x)(w) = \max\{Df_1(x)(w), Df_2(x)(w)\}.$$

However, f might not be strictly semidifferentiable at such points.

The semiderivative obeys standard calculus rules, such as semidifferentiation of a sum, product and ratio, and, most importantly, the chain rule. Here are some further properties of the semiderivative: Let f be semidifferentiable at x and let g be Lipschitz continuous and semidifferentiable at $y := f(x)$. Then the composition $g\circ f$ is semidifferentiable at x and

$$D(g\circ f)(x) = Dg(y)\circ Df(x).$$

Let f be strictly semidifferentiable at x and g be strictly differentiable at $f(x)$. Then $g\circ f$ is strictly semidifferentiable at x.

Lecture 1
Basics in Optimization

An *optimization problem* is typically a problem of finding minimum or maximum of a real-valued function f relative to a set C. The function f is called the *objective* or *cost* function, while the set C over which the minimization or maximization takes place is called the *feasible set* usually given by *constraints*. The problem of minimizing f over C consists of finding an element $\bar{x}$ in C such that

$$f(\bar{x}) \le f(x) \quad \text{for all} \quad x \in C.$$

This problem is written as

$$\min f(x) \ \text{ subject to } \ x \in C \qquad \text{or} \qquad \min_{x \in C} f(x).$$

Stated in that way, this is a problem of finding a *global* minimum of f over C. A point $\bar{x}$ is a *local* minimum of f over C if there exists a neighborhood U of $\bar{x}$ such that $f(\bar{x}) \le f(x)$ for all $x \in C \cap U$. Clearly, every global minimum is also a local minimum but the converse is not true. A maximization problem consists of finding $\bar{x} \in C$ such that $f(\bar{x}) \ge f(x)$ for all $x \in C$ and can be stated equivalently as a minimization problem, inasmuch as

$$\max_{x \in C} f(x) = -\min_{x \in C}(-f(x)).$$

In this book we will consider mainly two kinds of optimization problems. First come *mathematical programming* problems, where the feasible set is a subset of an Euclidean space usually given by equalities and inequalities. The name "programming" most likely stems from the time when optimization problems were solved on early computers; it propagated to problem classes such as linear programming, quadratic programming, convex programming, nonlinear programming, etc., where linear, quadratic, convex, and nonlinear correspond to the type of functions involved in the objective function and the constraints. Then we will focus on *optimal control* problems, where the feasible set is a set of functions in an infinite-dimensional

A. L. Dontchev, *Lectures on Variational Analysis*, Applied Mathematical Sciences 205, https://doi.org/10.1007/978-3-030-79911-3_1