SpringerBriefs in Optimization

Alexander J. Zaslavski

Optimization in Banach Spaces

Springer

SpringerBriefs in Optimization

SpringerBriefs in Optimization showcases algorithmic and theoretical techniques, case studies, and applications within the broad-based field of optimization. Manuscripts related to the ever-growing applications of optimization in applied mathematics, engineering, medicine, economics, and other applied sciences are encouraged.

Alexander J. Zaslavski

Optimization in Banach Spaces

Springer

Alexander J. Zaslavski
Department of Mathematics
The Technion – Israel Institute of
Technology
Haifa, Israel

ISSN 2190-8354 ISSN 2191-575X (electronic)
SpringerBriefs in Optimization
ISBN 978-3-031-12643-7 ISBN 978-3-031-12644-4 (eBook)
https://doi.org/10.1007/978-3-031-12644-4

Mathematics Subject Classification: 65K10, 90C25, 90C26, 90C48

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

Preface

The book is devoted to the study of constrained minimization problems on closed and convex sets in Banach spaces with a Frechet differentiable objective function. Such problems are well studied in a finite-dimensional space and in an infinite-dimensional Hilbert space equipped with an inner product which induces a complete norm. When the space is Hilbert, there are many algorithms for solving optimization problems, including the gradient projection algorithm, which is one of the most important tools in the optimization theory, nonlinear analysis, and their applications.

An optimization problem is described by an objective function and a set of feasible points. For the gradient projection algorithm, each iteration consists of two steps. The first step is a calculation of a gradient of the objective function, while in the second one, we calculate a projection on the feasible set. In each of these two steps, there is a computational error. In general, these two computational errors are different. In our recent research [48, 53, 54], we show that the gradient projection algorithm generates a good approximate solution, if all the computational errors are bounded from above by a small positive constant. Moreover, if we know computational errors for the two steps of the algorithm, we find out what an approximate solution can be obtained and how many iterates one needs for this.

It should be mentioned that in all these works, the properties of a Hilbert space play an important role. When we consider an optimization problem in a general Banach space, the situation becomes more difficult and less understood. On the other hand, such problems arise in the approximation theory.

In this book, our goal is to obtain a good approximate solution of the constrained optimization problem in a general Banach space under the presence of computational errors. It is shown that the algorithm generates a good approximate solution, if the sequence of computational errors is bounded from above by a small constant. The book consists of four chapters. In the first, we discuss several algorithms which are studied in the book and prove a convergence result for an unconstrained problem which is a prototype of our results for the constrained problem. In Chap. 2, we analyze convex optimization problems. Nonconvex optimization problems are studied in Chap. 3. In Chap. 4, we study continuous algorithms for minimization problems under the presence of computational errors.

The author believes that this book will be useful for researchers interested in the optimization theory and its applications.

Haifa, Israel
February 28, 2022

Alexander J. Zaslavski

Contents

1

Introduction

In this book, we study algorithms for constrained minimization problems in a general Banach space. Our goal is to obtain a good approximate solution of the problem in the presence of computational errors. It is shown that the algorithm generates a good approximate solution, if the sequence of computational errors is bounded from above by a small constant. In this section, we discuss several algorithms that are studied in the book. We also prove a convergence result for an unconstrained problem that is a prototype of our results for the constrained problem.

1.1 Notation

In this section, we collect the notation that will be used in the book.

Let $(X, \|\cdot\|)$ be a Banach space equipped with the norm $\|\cdot\|$ that induces the topology in X. We denote by X^* its dual space with the norm $\|\cdot\|_*$. For $x \in X$ and $l \in X^*$, we set $l(x) = \langle l, x\rangle$. The symbol $\langle \cdot, \cdot \rangle$ is referred to as the duality pairing between X^* and X. For each $x \in X$ and each $r > 0$, set

$$B(x, r) = \{y \in X :\ \rho(x, y) \le r\}.$$

For each function $f : Y \to R^1$, where Y is nonempty, we set

$$\inf(f, Y) = \inf\{f(y) :\ y \in Y\}.$$

If a real-valued function f is defined in a neighborhood a point x in a Banach space X, then by $f'(x)$ we denote a Frechet derivative of f at x if it exists.

We denote by $\mathrm{mes}(\Omega)$ the Lebesgue measure of a Lebesgue measurable set $\Omega \subset R^1$ and define

$$\chi_\Omega(x) = 1 \text{ for all } x \in \Omega,\ \chi_\Omega(x) = 0 \text{ for all } x \in R^1 \setminus \Omega.$$

A. J. Zaslavski, *Optimization in Banach Spaces*, SpringerBriefs in Optimization, https://doi.org/10.1007/978-3-031-12644-4_1

In the sequel, we denote by Card(D) the cardinality of a set D; suppose that the sum over an empty set is zero and that the infimum over an empty set is ∞.

1.2 Constrained Optimization

In our book, we consider a minimization problem

$$f(x) \to \min \qquad (P)$$

$$x \in C,$$

where $C \subset U$ is a convex closed set in a Banach space $(X, \|\cdot\|)$, $f : U \to R^1$ is a convex Frechet differentiable function that is Lipschitz on bounded sets, and U is a convex open set in X. This problem is well studied when the space X is finite-dimensional and when X is an infinite-dimensional Hilbert space equipped with an inner product denoted by $\langle \cdot, \cdot \rangle$ that induces a complete norm $\|\cdot\|$.

When the space X is Hilbert, the problem is solved by using the subgradient projection algorithm that is one of the most important tools in the optimization theory, nonlinear analysis, and their applications. See, for example, [1–3, 5–7, 9–15, 17, 20–26, 28, 31–37, 39–41, 43–47, 49–55]. For this algorithm, we do not need to assume that the function f is Frechet differentiable. Instead of the Frechet derivative, subgradients of a convex function are used [29, 30].

Our problem is described by the objective function f and the set of feasible points C. For the subgradient projection algorithm, each iteration consists of two steps. The first step is a calculation of a subgradient of the objective function, while in the second one, we calculate a projection on the feasible set. In each of these two steps, there is a computational error. In general, these two computational errors are different. In our recent research [48, 53, 54], we show that the subgradient projection algorithm generates a good approximate solution, if all the computational errors are bounded from above by a small positive constant. Moreover, if we know computational errors for the two steps of the algorithm, we find out what an approximate solution can be obtained and how many iterates one needs for this.

It should be mentioned that in all these works, the properties of a Hilbert space play an important role. When X is a general Banach space, the situation becomes more difficult and less understood. Probably the first result for the minimization problem (P) in a Banach space was obtained in [16]. Recently, problem (P) in a Banach space was studied in [18, 19, 42] using greedy algorithms.

Now we describe our algorithm. Let $\lambda \in (0, 1]$ and $c_* \geq 1$ be fixed.

Assume that $x_t \in C$ is a current iteration vector where $t \geq 0$ is an integer. In our first step, we calculate the Frechet derivative $f'(x_t)$. Since we take into account computational errors produced by our computer system instead of $f'(x_t)$, we get its approximation $g_t \in X^*$ satisfying

$$\|g_t - f'(x_t)\| \leq \delta,$$

where $\delta > 0$ is a computational error.

In our second step, we solve the following auxiliary minimization problem:

$$\langle g_t, \eta \rangle \to \min \qquad (P_a)$$

$$\eta \in B(0, c_*) \cap (C - x_t).$$

Since X is a general Banach space, the existence of a solution of this problem is not guaranteed. Taking into account this fact and also the presence of computational errors at the second step, we get a vector

$$l_t \in C - x$$

such that there exists

$$\xi_t \in B(0, c_*) \cap B(l_t, \delta),$$

which satisfies

$$\langle g_t, \xi_t \rangle \leq \lambda \inf\{\langle g_t, \eta \rangle : \ \eta \in B(0, c_*) \cap (C - x_t)\} + \Delta.$$

Here, $\delta, \Delta > 0$ are computational errors produced by our computer system that occur when we solve the auxiliary problem (P_a).

In the third step, we calculate the next iteration $x_{t+1} = x_t + \alpha_t l_t$, where $\alpha_t \in [0, 1]$ is a step size.

In our book, we deal with three options of choosing the step size α_t. In algorithm $(\mathcal{A}1)$, α_t is an approximate solution of the auxiliary optimization problem

$$f(x_t + \alpha l_t) \to \min, \ \alpha \in [0, 1].$$

In algorithm $(\mathcal{A}2)$, it is given a sequence $\{\alpha_t\}_{t=0}^{\infty}$ satisfying

$$\lim_{t \to \infty} \alpha_t = 0 \text{ and } \sum_{t=0}^{\infty} \alpha_t = \infty.$$

In algorithm $(\mathcal{A}3)$, $\alpha_t \in [\beta_1, \beta_0]$, where $0 < \beta_1 < \beta_0$ are given constants. (It is possible that $\beta_1 = \beta_0$, of course.)

Since we take into account the computation errors, the inequality $f(x_t + \alpha_t l_t) \leq f(x_t)$ cannot be guaranteed. As a result, each of algorithms $(\mathcal{A}1)$–$(\mathcal{A}3)$ has two subcases: in the first case, we always define $x_{t+1} = x_t + \alpha_t l_t$, while in the second subcase, we define x_{t+1} by the equality above only if

$$f(x_t + \alpha_t l_t) \leq f(x_t);$$

otherwise, $x_{t+1} = x_t$. So actually we have six different algorithms.

The discussion above leads us to the following definition that allows us to describe the second steps of our algorithms shortly.

For each $x \in C$ and each pair of numbers $\delta, \Delta \in (0, 1]$, denote by $E(x, \delta, \Delta)$ the set of all $l \in C - x$ for which there exist $g \in X^*$ such that

$$\|g - f'(x)\|_* \leq \delta$$

and

$$\xi \in B(0, c_*) \cap B(l, \delta)$$

such that

$$\langle g, \xi \rangle \leq \lambda \inf\{\langle g, \eta \rangle : \ \eta \in B(0, c_*) \cap (C - x)\} + \Delta.$$

With this definition, the second step of our algorithms is described as the choice of

$$l_t \in E(x_t, \delta, \Delta).$$

Section 2.3 contains Theorems 2.2–2.10 that show the behavior of the algorithms. In these results, we show that for a given $\epsilon > 0$, there exist a sufficiently small error $\delta, \Delta > 0$ such that after a certain number of iterations we obtain a point $x_s \in K$ satisfying $f(x_s) \leq \inf(f, C) + \epsilon$. Of course, it is interesting to obtain an explicit estimation for δ, Δ. It is done in Theorem 2.11 (see Sect. 2.13) under some additional assumptions on the objective function f. Namely, we assume that its Frechet derivative is Holder continuous on bounded sets and that the corresponding constants are known. In this case, we obtain an explicit dependence δ, Δ and a number of iterations on ϵ. This dependence allows us easily to solve an inverse problem: if we know δ, Δ what ϵ can be obtained?

Note that in [16] it was considered an exact version of algorithm $\mathcal{A}1$ when the set C is bounded. Implicitly, it was assumed that the functional $f'(x_t)$ has a minimizer on C. This is true when C is weakly compact. In [18, 42], also algorithm $\mathcal{A}1$ is studied. In both of these works [18, 42] at the second step, the direction

$$l_t \in C - x$$

satisfies

$$\langle f'(x_t), l_t \rangle \leq \lambda \inf\{\langle f'(x_t), \eta \rangle : \ \eta \in C - x_t\}.$$

In [42], the step size α_t is an exact solution on the auxiliary minimization problem, while in [18], it is an approximate one.

In the next section of this chapter, we prove a convergence result in the case of unconstrained problems ($C = X$) that is a prototype of our results for constrained problems. It should be mentioned that unconstrained problems were studied in Chapter 8 of [38] using algorithms induced by regular vector fields.

In Chap. 3 of the book, we study nonconvex minimization problems. Continuous versions of our algorithms are studied in Chap. 4.

1.3 Unconstrained Optimization

Assume that $(X, \|\cdot\|)$ is a Banach space and that $f : X \to R^1$ is a convex function that is Lipschitz on all bounded sets in X. We assume that f is Frechet differentiable at any point $x \in X$ such that $f(x) > \inf(f, X)$. If $x \in X$ satisfies $f(x) = \inf(f, X)$, then we set $f'(x) = 0$.

We assume that the following two assumptions hold:

(A1) There exists a nonempty bounded set $X_0 \subset X$ such that $\inf(f, X_0) = \inf(f, X)$.

(A2) For every pair $M, \epsilon > 0$, the Frechet derivative $f'(\cdot)$ is uniformly continuous on the set

$$\{x \in B(0, M) :\ f(x) > \inf(f, X) + \epsilon\}.$$

In this section, we consider the minimization problem

$$f(x) \to \min,\ x \in X.$$

In order to solve this problem, we use two iterative processes. Let $c_* > 1,\ \lambda \in (0, 1]$ and $\{\alpha_t\}_{t=0}^{\infty},\ \{\epsilon_t\}_{t=0}^{\infty} \subset (0, 1]$ satisfy

$$\lim_{t\to\infty} \epsilon_t = 0,\ \lim_{t\to\infty} \alpha_t = 0,\ \sum_{t=0}^{\infty} \alpha_t = \infty.$$

In both processes, we select an arbitrary $x_0 \in X$. Given a current iteration vector $x_t \in X$, we calculate $f'(x_t)$. If $f'(x_t) = 0$, then x_t is a solution of our minimization problem. Assume that $f'(x_t) \neq 0$. We find $l_t \in X$ satisfying

$$\|l_t\| \le c_*$$

and

$$\langle f'(x_t), l_t\rangle \le -\lambda\|f'(x_t)\|_* + \epsilon_t.$$

In the first process, we find $x_{t+1} \in \{x_t + \alpha l_t :\ \alpha \in [0, 1]\}$ such that

$$f(x_{t+1}) \le f(x_t + \alpha l_t),\ \alpha \in [0, 1].$$

In the second process, if $f(x_t + \alpha_t l_t) < f(x_t)$, then we set $x_{t+1} = x_t + \alpha_t l_t$; otherwise $x_{t+1} = x_t$.

In Sect. 1.5, we prove the following result.

Theorem 1.1. *Let* $c_* > 1,\ \lambda \in (0, 1]$ *and* $\{\epsilon_t\}_{t=0}^{\infty} \subset (0, 1]$ *satisfy*

$$\lim_{\to\infty} \epsilon_t = 0. \tag{1.1}$$

Assume that $\{x_t\}_{t=0}^{\infty} \subset C,\ \{l_t\}_{t=0}^{\infty} \subset X$ *and that for each integer* $t \ge 0$*,*

$$\|l_t\| \le c_*,\ \langle f'(x_t), l_t\rangle \le -\lambda\|f'(x_t)\|_* + \epsilon_t. \tag{1.2}$$