

**WILEY-INTERSCIENCE SERIES IN DISCRETE  
MATHEMATICS**

**ADVISORY EDITORS:** Ronald L. Graham  
Bell Laboratories  
Murray Hill, New Jersey

Dr Jan Karel Lenstra  
Stichting Mathematisch Centrum  
Amsterdam, The Netherlands

Graham, Rothschild and Spencer · RAMSEY THEORY

Tucker · APPLIED COMBINATORICS

Pless · INTRODUCTION TO THE THEORY OF  
ERROR-CORRECTING CODES

Nemirovsky and Yudin · PROBLEM COMPLEXITY AND  
METHOD EFFICIENCY IN  
OPTIMIZATION

Goulden and Jackson · COMBINATORIAL ENUMERATION

Gondran and Minoux · GRAPHS AND  
ALGORITHMS

**PROBLEM COMPLEXITY AND  
METHOD EFFICIENCY IN  
OPTIMIZATION**

**A. S. NEMIROVSKY**

Senior Scientific Fellow, State University of Moscow, USSR.

**D. B. YUDIN**

Professor of Mathematical Methods, Faculty of Economics,  
State University of Moscow, USSR

Translated by

**E. R. DAWSON,**

University of Dundee

A Wiley-Interscience Publication

**JOHN WILEY & SONS**

Chichester · New York · Brisbane · Toronto · Singapore

Translated from SLOZHNOST' ZADACH I EFFEKTIVNOST'  
METODOV OPTIMIZATSII by E. R. Dawson

© Glavnaya redaktsiya fiziko-matematicheskoi literatury,  
Izdatelstva "Nauka", 1979

Copyright © 1983 by John Wiley & Sons, Ltd.

All rights reserved.

No part of this book may be reproduced by any means, nor  
transmitted, nor translated into a machine language without the  
written permission of the publisher.

**Library of Congress Cataloging in Publication Data:**

Nemirovsky, Arkadii Semenovich.  
Problem complexity and method efficiency in optimization.  
(Wiley series in discrete mathematics)  
'A Wiley-Interscience publication.'  
Translation of: Slozhnost' zadach i effektivnost'  
metodov optimizatsii.  
Bibliography: p.  
Includes index.  
1. Programming (Mathematics) I. Yudin, David  
Borisovich. II. Title. III. Series: Wiley-Interscience  
series in discrete mathematics.  
QA402.5.N45413 1983 519.7 82-11065  
ISBN 0 471 10345 4

**British Library Cataloguing in Publication Data:**

Nemirovsky, A. S.  
Problem complexity and method efficiency in  
optimization—(Wiley series in discrete  
mathematics)  
1. Mathematical optimization  
I. Title. II. Yudin, D. B. III. Slozhnost'  
zadach i effektivnost' metodov optimizatsii.  
*English*  
515 QA402.5  
ISBN 0 471 10345 4

Typeset by Macmillan India Ltd., Bangalore

PAGE  
BROS Printed in Great Britain by  
Page Bros (Norwich) Ltd

QA  
402.5  
N45413  
1983  
3-7-86  
PAC

# Contents

## 1 INTRODUCTION

1.1	Formulation of the problem of estimating the complexity of optimization problems, and the main results of the work. An informal description	1
1.2	Families of mathematical programming problems. Approximate solutions and their error	10
1.3	Numerical methods of solving mathematical program- ming problems	16
1.4	Characteristics of methods on a problem and on a class. Complexity	23
1.5	Some propositions about mathematical programming methods	29
1.6	On the complexity of classes of multi-extremal problems	35
1.7	Proof of Theorem 1.3.4	38

## 2 CONVEX PROGRAMMING. LINEARLY CONVERGENT METHODS FOR CLASSES OF GENERAL CONVEX PROBLEMS

2.1	Convex sets and convex problems	42
2.2	Classes of convex extremal problems	50
2.3	The centres-of-gravity method for solving general convex problems	57
2.4	Special versions of the MCG	65
2.5	A realizable version of the MCG	72

<b>3 METHODS OF MIRROR DESCENT</b>	<b>83</b>
3.1 The idea of the methods	83
3.2 Regular spaces	89
3.3 MD-methods on classes of Lipschitz-convex problems	93
3.4 MD-methods on classes of general convex problems	107
3.5 Some additional proofs	120
<b>4 COMPLEXITY OF CLASSES OF GENERAL CONVEX PROBLEMS (EXACT FIRST-ORDER ORACLE)</b>	<b>125</b>
4.1 Complexity of classes of general convex problems	125
4.2 Complexity of classes of Lipschitz-convex problems	137
4.3 Recommendations about the use of methods of solving general convex or Lipschitz-convex problems	142
4.4 Proof of lower bounds for the complexity. I	151
4.5 Proof of lower bounds for the complexity. II	164
4.6 Proof of lower bounds for the complexity. III	174
<b>5 PROBLEMS OF CONVEX STOCHASTIC PROGRAMMING</b>	<b>181</b>
5.1 Classes of convex stochastic programming problems	182
5.2 Unconstrained optimization	188
5.3 The complexity of stochastic programming problems	197
<b>6 SOLUTION OF CONVEX–CONCAVE GAMES AND CONSTRAINED STOCHASTIC PROGRAMMING PROBLEMS</b>	<b>206</b>
6.1 Classes of convex–concave game problems	206
6.2 MD-methods of solving games: the case of a deterministic oracle	210
6.3 MD-methods of solving games: the case of a stochastic oracle	219

<b>6.4 Solution of convex operator inequalities</b>	<b>225</b>
6.5 Solution of extremal problems with operator constraints	233
6.6 Solution of constrained stochastic problems	241
6.7 ‘Complex’ stochastic programming problems	248
<b>7 STRONGLY CONVEX PROBLEMS</b>	<b>254</b>
7.1 Classes of smooth and strongly convex extremal problems	254
7.2 Quadratic programming and lower bounds of the complexity of strongly convex classes	261
7.3 Strongly convex programming; unconstrained problems	272
7.4 Minimax problems	277
7.5 Solution of constrained strongly convex and smooth problems	285
7.6 Proof of Theorem 7.4.3	294
<b>8 EFFICIENCY OF STANDARD METHODS OF STRONGLY CONVEX PROGRAMMING</b>	<b>307</b>
8.1 On standard methods of strongly convex programming	307
8.2 The Fletcher–Reeves method	313
8.3 The Polak–Ribi��re method	319
8.4 The Zoutendijk projection method	327
8.5 Remarks on the Davidon–Fletcher–Powell method	330
<b>9 CONVEX PROGRAMMING METHODS OF ZEROTH ORDER</b>	<b>333</b>
9.1 Methods of zeroth order: deterministic oracle. I	334
9.2 Methods of zeroth order: deterministic oracle. II	341
9.3 Methods of zeroth order: stochastic oracle. I	349
9.4 Methods of zeroth order: stochastic oracle. II	354

<b>APPENDIX: MATHEMATICAL SUPPLEMENT</b>	<b>362</b>
A.1 Polish spaces, Borel functions, measures	362
A.2 Banach spaces	369
<b>PRINCIPAL NOTATION</b>	<b>376</b>
<b>REFERENCES</b>	<b>380</b>
<b>INDEX</b>	<b>383</b>

## Preface

There is an extensive literature devoted to numerical methods of solving extremal problems. The central position in it is occupied by studies relating to actual algorithms for optimization—their description, range of applicability, rate of convergence. There are also a number of papers on empirical comparison (based on examples) of the effectiveness of algorithms, with a view to selecting the ‘best’ method from the existing arsenal of such methods.

Far less attention has been paid to questions such as the following. What, in general, can be expected from numerical methods of solving problems of a given type? What are the potentially attainable limits of these methods? How complicated are problems of one sort or another, not with respect to a given concrete method, but in relation to ‘all (in general)’ methods of solution? It is this natural question of the ‘potentially attainable effectiveness of numerical methods applied to problems of a given type’ which forms the subject of investigation in this monograph.

A typical question which we shall consider is of this form. Given a family of optimization problems together with a source of information, accessible to the methods, about each solvable problem of this family, what are the potential lower bounds of laboriousness of all possible methods which solve all problems of the family with a given accuracy? Which method realizes this potential lower limit and is therefore the best one? Clearly, a precise formulation of such a question requires a formalization of the concepts of ‘method’, ‘laboriousness of a method’, etc. We fix a definite formalization of this kind (in our opinion, the most convenient formalization for studying the ‘continuous’ mathematical programming problems with which we shall be concerned), and we then investigate the question posed above as applied to a number of the standard non-linear programming problems: smooth multi-extremal problems, ‘all convex’ problems, strongly convex problems, and so on. In most cases a sufficiently conclusive answer is successfully obtained to the question in which we are interested.

We do not claim that the formalization we have adopted of the problem of selecting the best possible numerical method of optimization is entirely adequate to represent the full content of the original formulation. There is no

need to dwell now on this aspect of the matter, since a full motivation of the approach we have adopted and a discussion of its merits and limitations are given in Section 1.1. We remark only that, in our opinion, the formalization we have adopted enables rather useful, if rough, information (orders of magnitude only) to be obtained about the potential possibilities of numerical methods of solving extremal problems of the standard types. It is up to the reader to share or repudiate this opinion.

Another question closely related to that of finding potential lower bounds for the laboriousness of methods which solve problems of a given class with a given error, i.e. with the problem (as we have defined it) of computing the 'complexity' of a class, is the question of estimating the effectiveness of one or other method of solving problems of this class. It is natural to define this effectiveness as the inverse of the ratio of the laboriousness of the method in question to a 'standard laboriousness', i.e. to the complexity of the class. The effectiveness of a method shows to what extent it can be improved as regards laboriousness, i.e. to what extent it is non-optimal.

We touch on only a small part of the problem of estimating the effectiveness of traditional numerical methods of optimization. This is understandable; to solve this problem it is not sufficient merely to have available a standard of effectiveness (basically our efforts are concentrated precisely on obtaining such a standard); it is further necessary to have estimates of the laboriousness and error of the method in which we are interested on the class of problems under consideration. The profusion of standard numerical methods obviously precludes the possibility of the authors' being able to estimate these characteristics for some arbitrary representative of the group of methods.

In this book the effectiveness of some of the most natural and simple methods is evaluated. For reasons which will become clear later, all these methods are the methods of convex programming. Of the methods of non-smooth convex programming, we evaluate the gradient method and the Kelly method (these might be said to exhaust the list of traditional algorithms for non-smooth convex optimization). The extensive field of algorithms for the minimization of smooth and (strongly) convex problems is examined to a far less extent; here we restrict our attention to the gradient method with minimization in the anti-gradient direction and to some simple versions of the method of conjugate directions. The methods considered for strongly convex programming problems turn out to be inefficient; they are unnecessarily sensitive to the degree of conditionality of the problem under consideration, and their effectiveness tends to zero as the conditionality deteriorates. We remark that negative results of this kind also enable certain conclusions to be drawn regarding the effectiveness of a number of traditional methods which are not explicitly considered in this book.

Let us give an example. An extensive family of methods of feasible directions for solving constrained convex problems is known. The rate of

convergence of these methods is generally estimated under the hypothesis of strong convexity of the problem under solution. Thus it is natural to study these methods on the class of strongly convex problems. In Chapter 7 it is shown that the complexity of this class is determined essentially only by the required accuracy, the conditionality of the problem, and its dimension, but not by the number of constraints. On the other hand, if there are no constraints, then most versions of the method of feasible directions turn into the method of gradient descent with minimization in the direction of descent. So the effectiveness of the methods of feasible directions cannot be essentially greater than that of the gradient method, and therefore it too tends to zero as the conditionality of the problem deteriorates.

The limited size of the book does not allow us to dwell on consequences of this sort; they will certainly be self-evident to the reader. We remark further that the judgement expressed previously about the method of feasible directions being ineffective (like similar statements in the main text) is a judgement made on the basis of the definition of laboriousness which has been adopted and which turned out to be not quite adequate for the intuitively understood computational complexity of a method. It should not therefore be interpreted as a call for unconditional discrimination against the corresponding methods; categorical verdicts of this kind are scarcely admissible generally.

We mention some differences of the approaches adopted in this monograph, and of the results obtained, from the traditional treatments in optimization theory.

The traditional approach to estimates of the rate of convergence of numerical methods of optimization is usually of an asymptotic character; the type of asymptotic behaviour of the laboriousness of a method for a required accuracy is elucidated. The question of when the 'exit' on to this asymptotic behaviour takes place is investigated comparatively rarely, as is, incidentally, the important question of the effect of other parameters, apart from accuracy, of the class of problems (parameters such as, for example, the dimension of the problems under solution).

In this monograph, on the other hand, all the estimates given for the laboriousness of numerical methods of optimization (as also, incidentally, for most of the estimates of complexity) are of a non-asymptotic character. We write down in explicit form the upper bounds for the methods under examination, as a function of the required accuracy (measured in a sensible way) and of the parameters which distinguish the class of problems to be solved (such as the dimensionality of the problem, the number of constraints, the characteristics of the geometry of the domain of the problem, etc.).

In the literature on methods of optimization the rate of convergence (the rate, not the fact of convergence itself) is established generally only as applied

to classes of sufficiently 'good' problems. In the present monograph optimization methods are studied as a rule on wider classes of problems. In most parts of the book nothing more is required of the problems to be solved other than their convexity, unless perhaps the boundedness of the domain of the problem and the continuity of the functionals appearing in it (sometimes these functionals have to satisfy a Lipschitz condition). Even the classes of strongly convex problems considered in Chapter 7 are still somewhat wider than classes of 'good' functions on which it is traditionally accepted that optimization methods should be evaluated. Of course, evaluating methods on wider families of problems than usual leads to a worsening of the potentially attainable guarantees of their work. It turns out, however, that in many cases this worsening is not too considerable, and it represents an acceptable payment for the extension of the store of problems over which the new guarantees are extended.

Having sketched in general terms the purpose of the monograph, we shall briefly characterize the contents of the work by sections (a more detailed survey of the results is given in Section 1.1). The first chapter is of an introductory character: here we form the language in which later we formulate and solve the problems concerning the potential effectiveness of numerical methods. The separation of the later material into chapters is dictated by the necessity to examine separately the various classes of extremal problems. In the last section of Chapter 1, smooth (but not necessarily convex) problems are examined. The results in this section, some well-known, some new, are of a negative character (it turns out that this class of problem is, in the general case, hopelessly too complicated for solution). Convex programming, to which the main attention is paid in the book, presents a much more optimistic picture. In Chapters 2, 3, and 4 we examine classes of 'all (in general)' convex problems (including non-smooth ones) which can be solved by first-order methods when there is exact calculation of the values and derivatives of the components in the problem. In Chapters 5 and 6 we study convex problems in which the values and derivatives of the components are observed mixed with noise, i.e. problems of convex stochastic programming. In Chapter 7 the classes of smooth convex problems and strongly convex problems are considered, and in Chapter 9 we deal with problems solvable by zeroth-order methods, i.e. methods working with the values but not with the derivatives of the components in the problem. Chapter 8 stands somewhat by itself; it deals with the estimation of the laboriousness of some popular methods of solving strongly convex problems. The appendix contains a résumé of a number of classical mathematical concepts and theorems which are used in the book and which are not always familiar to applied mathematicians. In general it should be mentioned that the treatment in this book is such that it should be accessible to a reader having the normal training of a numerical analyst interested in optimization theory. More complicated mathematical apparatus

is brought in only in the 'formal niceties' (of secondary importance) in some of the proofs.

The exposition of material relating to concrete classes of extremal problems usually comprises the following main steps:

- (1) a description of the class of problems in question;
- (2) a description of some methods of solving problems of the given class;
- (3) a lower bound (over all conceivable methods of solving the problems of a given class) for the potentially attainable laboriousness of these methods for a given error.

As a rule the choice of actual methods in (2) leads to estimates for the laboriousness which essentially are the same as the potential lower bounds in (3). As a result we obtain, on the one hand, a sufficiently complete idea of the 'objective complexity' of the given class of problems, and on the other hand, a basis for the theoretical recommendations on the use of the methods in (2) which cannot essentially be improved as regards laboriousness. We mention that the 'sub-optimal' methods which we adduce are, in a number of cases, in substantial measure new.

We point out that a reader can, if he wishes, restrict his attention to the sections which deal with some particular class of problems.

As regards the nature of the exposition it is worth mentioning the following. We have tried to distinguish as clearly as possible the ideas which form the basis of the constructions here presented, and to describe the numerical methods precisely. The formal proofs are kept separate; some of them are given in separate sections. In a first reading these proofs could, if so desired, be omitted, and this would not hinder the reader from using the methods described, although it would make detailed understanding of their mechanism more difficult.

Many of the results are formulated as exercises, inviting the reader to prove some proposition which has only been formulated. If some not entirely trivial fact is concerned, then often a proof is given (enclosed in angular brackets  $\langle \dots \rangle$ ). We stress that the reader should acquaint himself with the propositions enunciated in the exercises, whether or not he actually carries out the exercises themselves.

Regarding terminology: apart from the standard terminology, which we use without special explanation, we have to use extensively a number of specific concepts and special notation. On first encountering non-standard notation and terms appearing without commentary, the reader should consult the list of notation at the end of the book, or the subject index, where he will find a reference to the section in which the corresponding object is first defined. An exception occurs with certain secondary concepts and notation used only in some one chapter. Accordingly, to understand a particular section of a

chapter, the reader will generally need to be acquainted with all the preceding sections of that chapter.

A few words about the bibliography and references. The list includes only those works which are directly referred to in the main text; the list is short and makes no claim to cite all the works which deal with the theme of the monograph. When using the results of others, the authors have *tried* to mention the fact by pointing out the source of their information (without setting themselves the task of identifying the original source without fail). In speaking of results which have become part of the chrestomathy, so to say, we have replaced direct citation by phrases such as 'it is well known that . . .'. It is quite possible that some of the results regarded by the authors as original may in actual fact be re-discoveries of already known facts (that is why we used the word 'tried' in a previous sentence); in that case we beg in advance the pardon of the first discoverers.

In conclusion we regard it as our pleasant duty to thank E. G. Goldshtain and B. T. Polyak for stimulating discussions of the results of the work.

A. S. NEMIROVSKY, D. B. YUDIN

June 1978

## Preface to the English Edition

This book is one of the series 'Theory and methods of systems analysis' published under the guidance of an editorial board of economists and cyberneticists headed by D. M. Gvishiani, a Soviet philosopher and son-in-law of the former Soviet prime minister, Alexei N. Kosygin. Of the authors of the present work, Professor D. B. Yudin holds the chair of mathematical methods in the faculty of economics in Moscow university, and Dr. A. S. Nemirovsky, a senior scientific fellow at the same university, is a disciple of the late distinguished mathematician G. E. Shilov. The book is based on, and is an extension of, a series of papers by these authors published mainly in the journal *Economics and Mathematical Methods*.

The authors set up their own mathematical model in order to investigate questions concerning the complexity of optimization problems and efficiency of methods of solving them. They obtain bounds for the potential efficiency of methods of solving standard classes of optimization problems, and propose new methods which largely realize these potential bounds. They apply their apparatus to draw perhaps surprising conclusions about a number of popular methods of optimization. But, as with all mathematics, the reader must remember that the technical terms have precisely the meaning assigned to them in the definition of the concepts; this is particularly necessary when every-day words such as 'method' and 'complexity' are being used as technical terms. In particular, as the authors themselves point out, their analysis does not deal at all with such practically important aspects of methods as the simplicity of their computational organization and computational stability.

When not being strictly formal, as in the definitions and statements and proofs of theorems, the authors write in a lively and informal style, which may at times even be humorous. They have a habit of frequently putting words into quotation marks, presumably to point up and lend immediacy to their exposition; this practice has been followed in the translation.

F. R. Dawson  
Department of Mathematical Sciences, University of Dundee

# 1

## Introduction

### 1.1 FORMULATION OF THE PROBLEM OF ESTIMATING THE COMPLEXITY OF OPTIMIZATION PROBLEMS, AND THE MAIN RESULTS OF THE WORK. AN INFORMAL DESCRIPTION

In this section we describe informally the set of problems we shall be concerned with and the direction of our investigations. Our aim is to show that the approach adopted in this work for evaluating the complexity of problems and the effectiveness of methods is a natural one.

#### 1.1.1

We shall study the potentialities of numerical methods in solving mathematical programming problems. We need hardly mention how important such methods are in the application of mathematics to practical problems. The widening field of applications and the power of computers is leading firstly to a sharp growth in the complexity of the optimization problems which have to be 'worked out to the answer', and secondly to a continuous reinforcement of the arsenal of methods used for this purpose. In this situation there is naturally a growing tendency to take a hard look at the theory of these methods themselves. By contemporary standards, its mere convergence gives no method the right to exist; 'decency' requires us also to estimate its laboriousness.

The next stage is that one wants to find the potentially attainable lower limits for the amount of labour needed to solve a given type of problem, and to construct methods which attain these limits, i.e. methods—in some sense optimal—which ensure solutions of the required quality for all the problems in question, with the least possible amount of labour. These are precisely the problems to which the present work is devoted. Our target is a theoretical analysis of the potentialities of numerical methods.

A strict formulation of the problems arising in this connection requires a formalization of ideas such as 'a class of problems of a given type', 'the

laboriousness of a method', etc. There are many formalizations of this kind, which seem natural in one way or another.

### 1.1.2

The most general approach to the analysis of computations is developed within the framework of the general theory of algorithms and is related to the articles [35, 15]. A reading of the short but exceptionally deep paper by Levin [19] may be recommended to the reader wishing to familiarize himself with the spirit of this approach. Its essence, stated in an informal and general form, is as follows. It is assumed that a computational method is an algorithm (according to any of the formal definitions adopted in mathematical logic, for example, a Turing machine). Any such algorithm, as we know, processes words in some alphabet (e.g. binary) into certain other words. In order to apply an algorithm to some problem formulated perhaps in a non-formalized language (in the mathematico-logical sense), let us say, in the language of classical mathematics, it is necessary to associate it with a word in the input alphabet of the algorithm—the code of the problem. The algorithm processes the code of the problem into the output word, which can then be interpreted as the code of the solution. This code is translated according to definite rules into the original language in which the problem was stated. The laboriousness of the solution of a problem is measured by the number of times the algorithm works on the code.

Other measures of laboriousness which are related to the amount of memory storage used, etc., can also be considered. A number of general and extremely deep results have been obtained on these lines. There is no doubt that the apparatus developed in the general theory of algorithms is the one best fitted for studying the general problems of computation. At the same time it is clear that, with this approach, the problem of translating a problem from the language (or 'jargon', as logicians would say) of classical mathematics into a formalized language—the problem of coding the problem!—remains outside the theory. For problems with a clear, precise algebraic structure (e.g. for integer programming problems, to which, in particular, combinatorial problems reduce) the coding problem is easily resolved: the methods for a natural coding are in this case obvious. Although there are many natural methods of coding, they can all be transformed one into the other by simple algorithms, and so essentially the results do not depend on the method of coding.

### 1.1.3

The situation changes when we pass from discrete problems (or problems with a simple algebraic structure) to continuous problems like integrating or minimizing an 'arbitrary' function (a smooth function, say) defined, for example, on an interval of the real line. The coding problem here is by no

means so simple. Of course, for its practical solution, any problem necessarily is coded (in our examples, by an algorithm which computes the values of the function at an 'arbitrary' point) and it is discretized, so that formally one is always talking about solving discrete problems, and a method of solution is a certain algorithm for processing words. However, with such a formalized (in the mathematico-logical sense) approach, it is difficult to discuss 'continuous', 'classical' properties of the problem (in our examples, the properties of smoothness, convexity, etc.), and the description and, *a fortiori*, the optimization of the methods become exceptionally difficult (not to say hopelessly complicated) matters.

It is natural to describe methods of solving 'continuous' problems in the usual classical language, without realizing that, strictly speaking, it is impossible to multiply real numbers or calculate a sine by means of an algorithm. But even the description itself of the methods of solving non-algebraic problems, in terms of an idealized computer which carries out purely arithmetical operations, comparisons, computations of standard functions, and so on, is still 'too detailed'. It seems that a study in these terms of, let us say, the Newton–Raphson method of minimizing a function of several variables will 'founder' in trying to elucidate the question of the minimal number of operations needed for the inversion of a matrix (this question, incidentally, is up to now unresolved). A construction in this manner of a theory of numerical methods of solving problems of some definite (but arbitrarily wide) class (let us say, extremal problems) would necessarily have to include within itself, roughly speaking, a theory of 'all numerical methods in general'. But then it would not be adapted to the precise specific character of a given class of problems.

Thus, as well as the most general, universal theory of computations, there must also exist special investigations relating to special classes of computational problems. The language of these investigations (i.e. primarily the formalization of the concepts 'problem' and 'method') must, without pretending to be universal, be adapted to the specific character of the class under consideration. Investigations of this type would enable essential (although special) information to be obtained, which could hardly be derived by a universal theory.

There is nothing surprising in the variety of approaches; it would scarcely be possible to rely on one unique formalization of the concepts relating to the analysis of numerical methods which would be satisfactory in all respects without exception.

### 1.1.4

The approach adopted in this book (in our opinion the most convenient one for studying methods of 'continuous optimization') dates back to a paper by

N. S. Bakhvalov [3]. According to this approach, a numerical method is a set of rules for accumulating information about the problem to be solved, to enable its solution subsequently to be formed. The source of this information is fixed in advance. The process of obtaining information is governed by the method in the following sense: in advance there is fixed a set of 'questions'  $X$  which can be put to the source of information, the 'oracle'. The work of the method proceeds in steps. At the  $i$ th step one question  $x_i \in X$  can be put to the oracle and an answer obtained to it; this answer is a point of a certain set  $I$  (the information space, which is the set of possible answers of the oracle). Successive questions may be freely chosen within the limits of  $X$ . After a certain number of such questions, on the basis of the accumulated information (the collection of answers from the oracle) the result of applying the method to the problem is formed. The method of solution thus consists in a set of rules for forming successive questions, moments of pause, and results which depend on the information accumulated up to the given moment. No restrictions as regards computability are imposed on these rules. They can be arbitrary functions of the corresponding arguments ('arbitrary', that is, up to formal reservations which are entirely non-restrictive in practice).

The laboriousness of solving a given problem by a given method is defined as the number of steps in the work of the method on that problem. Apart from its laboriousness, a method applied to a given problem is characterized by its error (the measure of the inexactness of the result of its application to the problem, regarded as the approximate solution of the latter).

We remark that, in the approach just described, the restrictions on the possibilities of methods are of a purely informational nature; a method 'does not know' in advance exactly what problem it has to solve. Essentially its objective consists precisely in 'identifying' the problem to such a degree that it is then possible to form the solution with the required accuracy.

We point out a contrast. In the general theory of computations, complete information about the problem (its code) proves to be the entrance for the method, and the possibilities for the method are limited only by the means which it is decided to apply to process the code of the problem into the code of the answer (the method must be an algorithm in the mathematico-logical sense). But in the approach just described, the picture is the reverse: the means which can be applied are in no way limited, but on the other hand the initial information about the problem is incomplete, and its acquisition has to be paid for.

It is now clear that it is meaningless to pose the question of choosing the best method of solving a given completely concrete problem. The method would consist simply in a description of the problem's solution, and it has zero laboriousness. It should not be thought that this remark demonstrates that the approach outlined is void of meaning. The phenomenon just mentioned is not pathological. If it is necessary to solve just one problem, then the question of

method does not arise. By the very meaning of the word, the concept 'method' relates not to an individual problem but to a mass problem, i.e. to a class of problems of a given type. Thus the discussion must be not about 'a method of solving an actual problem' but about 'a method of solving problems of a given class' (i.e. problems from some previously fixed set).

We remark, incidentally, that it is scarcely possible sensibly to pose the question of 'the best method, as regards minimal laboriousness, for solving an individual problem' no matter how the concept of 'method' is defined. Thus, in terms of the general theory of algorithms, such a 'minimal' laboriousness is essentially the length of the word which codes the solution. There is not much more meaning in such a result than there is in the assertion that this laboriousness is zero.

### 1.1.5

The study of numerical methods on the basis of the approach just described thus presupposes the fixation of the following objects:

- (i) classes of problems which are to be solved by the methods under examination (in our case these will always be classes of extremal problems);
- (ii) sources of information about the problems to be solved, i.e. the oracles;
- (iii) methods of defining the error of the results regarded as approximate solutions of the problems under consideration.

Having fixed these objects, we distinguish clearly the class of methods corresponding to them, and for each method we determine the characteristics, i.e. the laboriousness and error, when it is applied to each problem of the class. This is still not enough to state the problem of optimizing the methods. For, generally speaking, of two methods one will be better on some problems, and worse on other problems, than the other; and so it is still necessary to specify the way in which methods are to be compared on the whole class of problems and not just on particular members of that class. For this purpose we define the characteristics of a method on a class of problems—by minimax considerations.

The discussion of the legitimacy of this minimax approach is postponed to Section 1.4.5.

Making precise the concept of 'the characteristics of a method on a class of problems' also enables us to state precisely the problem of studying the potentialities of methods of solving problems of a given class.

### 1.1.6

Let us consider the approach we have adopted. It is universal in the sense that a suitable fixation of the concepts 'class', 'oracle', etc., enables us to apply it to

any class of problems. However, this universality is only apparent. The fact is that the definition of laboriousness which corresponds to it is in no way adequate for the intuitively grasped concept of 'complexity of solution'. Our definition concerns information only, and it does not take into account in any way the expenditure in computational means for processing the information obtained.

If it is possible after a small number of questions,  $k$  say, to distinguish the problem from its class (this is typical for 'algebraic' problems), then the optimal laboriousness of solving problems of this class is small—it is not greater than  $k$ . In such a situation our approach may give empty results. It is nevertheless clear that the laboriousness of a method according to the definition adopted already gives automatically a lower bound for the 'real' laboriousness. In a number of cases these estimates are useful. What is more, 'optimal' methods (in our sense) turn out often enough to be, apparently, completely acceptable from the practical point of view as well. In short, the approach adopted is in no way a panacea for all difficulties and it does not give an exhaustive analysis of all the sets of characteristics of computational methods (for example, the important questions of computational stability lie, in general, beyond its compass), but we consider (and hope that the substance of this book will bring others to this opinion) that the results given by this approach provide a rather large amount of useful information about numerical methods of optimization. Like any other means, this approach is effective provided, of course, that it is used sensibly without demanding too much from it.

Another 'limiting factor' is that proceeding in the way mentioned requires a new entity to be brought in, namely, an oracle (in the general theory of computations this problem arises only in special cases). The corresponding limitation is not very important from the standpoint of possible applications. An oracle for solving 'continuous' optimization problems—mathematical programming problems—is fortunately traditional (it calculates the values—or the values and derivatives—of the functionals of the problem at a point indicated to it). Nevertheless the authors do not understand why there should be a restriction to just this oracle. We can motivate the use of this tradition in all the concrete results of this book only by citing the tradition mentioned. However, the wide prevalence of a blemish cannot justify it.

### 1.1.7

We now give a short résumé of our results. As already mentioned, we consider concrete classes of problems (together with a fixed oracle and a fixed method of measuring the error of approximate solutions) and the class of methods corresponding to these objects. Each of these methods applied to the class of problems considered is characterized by its laboriousness and error, i.e. by upper bounds (over the problems of the class) for the number of steps in its

work on the problem and by the error of the result. We further define the complexity  $N(v)$  of a given class of problems as a function of the error  $v$ ;  $N(v)$  is equal to the least possible laboriousness of a method which solves every problem of the class with an error not exceeding  $v$ . Our object is, firstly, to estimate the 'potential' lower bound of laboriousness for  $N(v)$ , and, secondly, to construct methods which 'really do' realize this lower bound. Before stating the results achieved in this direction, we remark that from now on the error involved is necessarily defined to be the 'relative' error. The only non-trivial case is when  $v < 1$ , and it is this case which is discussed from now on.

The first (pessimistic) results relate to classes of smooth (multi-extremal) problems. It turns out (section 1.6) that for the complexity  $N(v)$  of the class of all extremal problems with  $k$ -times continuously differentiable functionals on a compact field  $G$  in  $E^n$  the lower bound

$$N(v) \geq c(k, G)(1/v)^{n/k}$$

holds both for the ordinary (deterministic) methods of solution and for random-search methods. The catastrophic growth of  $N(v)$  as  $v \rightarrow 0$  and especially as  $n \rightarrow \infty$  shows that it is meaningless to pose the question of constructing universal methods of solving smooth problems of any appreciable dimensionality 'generally'. It is interesting to notice that the same estimates hold even for unconditional problems generated by uni-extremal (but not convex) functions.

The irregular behaviour of the complexity of classes of non-convex problems compels us to focus our attention in future on the far more rewarding object of convex programming.

We begin with classes of 'all (in general)' convex extremal problems on a convex compact set  $G$  of dimension  $n$ ; the oracle supplies the values and subgradients of the components in the problem. It turns out that the complexity of this class admits the estimate

$$c_1 \leq \frac{N(v)}{1 + n \ln 1/v} \leq c_2,$$

(here and always the  $c_i$  are positive absolute constants). The right-hand inequality holds for all  $v < 1$ , and the left-hand one holds asymptotically when  $v \rightarrow 0$  (Chapters 2, 4). The 'moment when the asymptotic behaviour is established' depends on the affine properties of  $G$ ; this dependence is the subject of a special investigation, the results of which are stated in the next paragraph. A method of error  $v$  and laboriousness  $O(n \ln(1/v))$  which 'essentially realizes' the complexity of the class when  $v \rightarrow 0$  (the idea of such a method was suggested by A. Yu. Levin in [18]) unfortunately cannot be utilized in practice when  $n > 3$  because of the extreme complexity of the step. We describe another method, which can be realized in practice, of laboriousness  $O(n^2 \ln(n/v))$  with a number  $O(n^4 \ln(n/v))$  of elementary operations on the organization and with

a memory store  $O(n^2)$  (see Section 2.5). We emphasize that the applicability and guarantee of accuracy and laboriousness of the methods mentioned are not associated with the presence (or number) of constraints of the problem, nor with its smoothness and degree of conditionality, but assume only that the problem is convex.

Thus, the asymptotic formula for the complexity of the class considered is  $O(n \ln(1/\nu))$  as  $\nu \rightarrow 0$ . The time has now come to explain how the complexity behaves as the dimension  $n$  of the problem increases. It turns out that everything depends on the affine properties of  $G$ . If  $G$  is a parallelepiped, then

$$N(\nu) \sim n \ln(1/\nu) \quad \text{for all } \nu < \frac{1}{4} \text{ and for all } n;$$

so the complexity increases linearly with  $n$ . But if  $G$  is an ellipsoid, then, as  $n$  increases,  $N(\nu)$  is stabilized at a level  $\sim 1/\nu^2$ , i.e. the complexity of ‘general’ convex problems on ellipsoids is bounded above by a quantity which is  $O(1/\nu^2)$  independent of the dimension, and which, for a given  $\nu$ , is  $O(1/\nu^2)$  for all sufficiently large  $n$ . An upper bound  $O(1/\nu^2)$  for the complexity can be obtained by a standard gradient method, and thus it cannot be ‘essentially improved’ for problems of high dimension. What then determines the principal differences, so far as we are now concerned, between a parallelepiped and an ellipsoid? It turns out (Chapter 3) that an ellipsoid is ‘uniformly convex’ and a parallelepiped is not.

The study of the question in Chapter 3 leads to a new view of the nature of the gradient method. It is well known that, applied to non-smooth, convex problems, the efficiency of the gradient method is connected, not with the fact that it works with a local linear approximation (a point of view inherited from the case of smooth problems), but with the fact that, for a displacement along the anti-gradient, we approach the minimum point (with an accuracy up to the square of the magnitude of the displacement). This fact has not appeared (to the authors, at least) to be a miracle specially performed by the Almighty for the benefit of convex programming. It turns out that there is no miracle; there is a simple and natural construction which associates methods of solving non-smooth convex problems with a wide class of functions (we have called these ‘methods of mirror descent’, MD-methods). This construction enables us to ‘mass-produce the miracles’. The standard gradient method is the simplest case where this construction is used.

We construct MD-methods for solving convex problems on convex fields  $G$  of the type of balls in  $L_p$  ( $1 \leq p < \infty$ ). If  $G$  is such a ball, then an estimate for the laboriousness of the corresponding method is  $c(p)(1/\nu)^{\max(2, p)}$  when  $p > 1$  (when  $p = 1$  we have  $c(1)(1/\nu)^2 \ln n$ ; this case  $p = 1$  is evidently not without interest for applications). It turns out that in their asymptotic behaviour with respect to dimension the methods constructed cannot be ‘essentially improved’.

Having examined the question of the class of ‘all (in general)’ convex problems with an explicit oracle of the first order, we then consider the case where the oracle is stochastic, i.e. the observations of component values in the problem and of their support functionals are distorted by stochastic noise (Chapters 5, 6). In this situation, too, we have been able to apply MD-methods successfully. It turns out that their laboriousness, when the error  $\nu$  (averaged over the noise) is guaranteed to be  $< 1$ , and for the case where  $G$  is a convex field like an  $L_p$ -ball, is

$$c(p)(1/\nu)^{\max(2, p)} \ln^2(1/\nu) \ln(m+2), \quad 1 < p < \infty,$$

(where  $m$  is the number of constraints in the problem; the hypotheses regarding the noise are the standard ones for problems of stochastic approximation; when  $m = 0$ , the factor  $O(\ln^2 1/\nu)$  can be omitted). It turns out that this estimate is, under broad hypotheses, essentially the same as the complexity of the corresponding class, and therefore in principle it is ‘almost’ unimprovable (i.e. with accuracy up to the logarithmic factor  $O(\ln^2(1/\nu))$ ). *En route* to these results we shall construct MD-methods (not without their own independent interest) of solving problems in game-theory. It should be pointed out that, even in the simplest situation ( $p = 2, m = 0$ ), the proposed method does not fully coincide with the traditional methods of stochastic approximation.

Having examined in the first approximation the complexity of the class of ‘all (in general)’ convex problems, we then study in Chapter 7 its most important subclass, viz. problems with smooth, strongly convex components; the minimal eigenvalues of the matrices of second derivatives are distinct from 0 and the maximum are bounded above; the ratio  $Q$  of the corresponding upper and lower bounds for the eigenvalues—the modulus of strong convexity of the problem—characterizes its degree of conditionality. The simplest gradient method of solving problems of this class has a laboriousness  $O(Q \ln(1/\nu))$ . It turns out that it is ‘too sensitive’ to the degree of conditionality. We construct another, less sensitive, method with an estimate  $O(\sqrt{Q} \ln^2 Q \ln(1/\nu))$  for the laboriousness (independent of the number of constraints in the problem). This last estimate cannot be improved essentially, since it is possible to show that the complexity of the class in question (at least asymptotically with respect to the dimension) admits the lower bound

$$N(\nu) \geq c_3 \sqrt{Q} \ln(1/\nu).$$

In Chapter 8 an attempt is made to establish which of the standard methods of solving strongly convex problems, which are regarded as the most effective (the methods of conjugate gradients), realize this potential limit. The gradient method, the Polak–Ribi  re method, the Fletcher–Reeves method, and the

Zoutendijk method are considered. For these methods negative results are obtained; within the criteria adopted for the estimate none is better than the gradient method.

In Chapter 9 we examine zeroth-order methods of solving convex problems (that is, only the values—exact or with stochastic noise—are observed, not the derivatives of the functionals of the problem). No conclusive results are obtained here; methods with the estimates for the laboriousness  $O(P(n) \ln(n/v))$ , where  $P(n)$  is a polynomial, and the information is deterministic, and  $O(P(n) \ln^2(1/v) \ln(m+2) \cdot 1/v^2)$  when the information is distorted by stochastic noise, are obtained. The character of the dependence of these estimates on  $v$  cannot, in view of what has been said earlier, be improved even in the class of first-order methods.

## 1.2 FAMILIES OF MATHEMATICAL PROGRAMMING PROBLEMS. APPROXIMATE SOLUTIONS AND THEIR ERROR

In this section and the next we describe the basic concepts with which we shall operate; for example, ‘a mathematical programming problem’, ‘a class of such problems’, ‘a method of solving problems of a given class’, and so on. Our purpose is to give formal definitions by means of these concepts. Without a formalization of this kind we cannot formulate precisely the problem of estimating the potentialities of numerical methods of optimization.

### 1.2.1

We shall be concerned fundamentally with numerical methods of solving mathematical programming problems. Such a problem is written in the form

$$f_0(x) \rightarrow \min \mid x \in G, f_j(x) \leq 0, 1 \leq j \leq m, \quad (2.1)$$

where  $G$  is a subset (always closed) of some real Banach space  $E$ , and the  $f_j(x)$ ,  $0 \leq j \leq m$ , are real (always continuous) functions defined on  $G$  at least. For technical reasons it is convenient to take the space  $E$  to be separable (from the practical point of view this is scarcely a serious restriction). From now on, unless the contrary is stated,  $E$  will be regarded as separable. The function  $f_0$  is called the *objective functional* of the problem, and the functions  $f_j$ ,  $1 \leq j \leq m$ , the *functional constraints* (or, simply, the *constraints*);  $G$  is the *domain* of the problem.

A point  $x \in E$  is called a *feasible point* of the problem (2.1) if  $x \in G$  and  $f_j(x) \leq 0$ ,  $1 \leq j \leq m$  (i.e.  $x$  lies in the domain and satisfies the constraints). The problem itself consists in finding a feasible point which gives the least possible value of the objective functional  $f_0$ . Accordingly, any feasible point  $\bar{x}$  such that, if  $x'$  is any other feasible point, then  $f_0(\bar{x}) \leq f_0(x')$ , is called a *solution* of the problem.

A problem which has feasible points is said to be *compatible* (otherwise it is *incompatible*). A problem which has a solution is said to be *solvable*. A solvable problem is necessarily compatible.

From what has been said, it is clear that a mathematical programming problem with  $m$  constraints is, from the formal standpoint, a set of objects

$$(f = (f_0, \dots, f_m); G; E),$$

where  $G$  is the domain of the problem,  $E$  the corresponding Banach space, and  $f = (f_0, \dots, f_m)$  is a continuous,  $(m+1)$ -dimensional vector-function.

*Remark.* Generally speaking, in formalizing the concept of a mathematical programming problem there is no need to provide  $G$  with a topology, and still less to fix the ‘Banach embedding’  $G \subset E$ . However, ‘continuous’ problems (and we shall study only these) are conveniently defined precisely in this way.

As a rule we shall be dealing with families of problems having  $m$ ,  $G$ , and  $E$  in common. In such a situation it is convenient to identify the problems themselves with the corresponding functions  $f$ ; this we shall always do in future. Thus instead of the phrase ‘the problem (2.1) defined by the function  $f$ ’ we shall write ‘the problem  $f$ ’, it will always be clear from the context what  $G$  and  $E$  are being considered.

The optimal value of the objective functional of the problem  $f$  is denoted by  $f_*$ :

$$f_* = \begin{cases} +\infty, & f \text{ incompatible}, \\ \inf \{f_0(x) | x \text{ is a feasible point of } f\}, & f \text{ compatible}. \end{cases}$$

### 1.2.2

Any family of mathematical programming problems (2.1) having  $m$ ,  $G$ ,  $E$  in common is called a *field of problems*. Every such family can be identified with a certain set  $\mathcal{F}$  of  $(m+1)$ -dimensional vector-functions on  $G$ . The field of problems generated by the objects  $G$ ,  $E$ ,  $m$ , and  $\mathcal{F}$  is denoted by  $\mathfrak{U}(\mathcal{F}, G, m, E)$ .

Suppose, for example,  $G$  is a convex, closed, bounded set in  $R^n = E$ , and  $\mathcal{F}$  is formed by all component-wise continuous, convex functions  $f$  on  $G$ . Then the corresponding field is the set of all convex problems with  $m$  constraints and domain  $G \subset R^n$ .

### 1.2.3 Approximate solutions and their error

#### 1.2.3.1

Numerical methods, as a rule, cannot ensure an exact solution of an arbitrary problem from some wide field. The results of applying these methods to problems are approximate solutions of the latter. In order to determine the

characteristics of methods we have to learn how to measure the error of the result of the work of a method, this result being regarded as an approximate solution of the problem in question. This result may be a point of the set  $G$  or a statement that the problem is incompatible (we shall denote the latter result by the symbol  $*$ ). Thus the set of possible results of applying a method to a problem  $f \in \mathfrak{U}(\mathcal{F}, G, m, E)$  is  $G_* = G \cup \{*\}$ . With each point  $x \in G$  and each point  $f \in \mathfrak{U}$  we must associate a number—the error of the point  $x$  regarded as an approximate solution of  $f$ .

### 1.2.3.2

There are two main ways of defining the error. The first way is to measure, in the metric of  $G$ , the closeness of the candidate solution  $x$  to the true solution  $\bar{x}$ . The second way is to examine the deviations of the values of  $f_j(x)$  from the nominal values, i.e. those demanded from  $\bar{x}$ . In this book we shall use only the second of these two ways.

Let us say briefly why we have made this choice. First of all, if a problem really is an extremal problem, then approximation of the solution 'by functional' corresponds to the heart of the matter, and one cannot talk about approximation 'by solution'. But this does not quite settle the question. Problems which are non-extremal in content are often put into extremal form (for example, a system of linear equations is reduced to a quadratic optimization problem). In such cases, a good approximation to the solution 'by functional' does not in itself necessarily guarantee a satisfactory solution to the original formulation of the problem. But there is another, the main, reason for our choice. The point is that in all the standard situations (at least, so far as this book is concerned), the problem of approximating the solution 'by functional' is well-posed (with a proper formulation), whereas in a number of important cases the problem of approximating 'by solution' is not well-posed and cannot be solved 'with a guarantee of accuracy' after any finite time (cf. Section 4.3, Exercise 5). Thus if an approximation 'by solution' is required, then special assumptions about the problem must be made in order to make this objective attainable. But the standard assumptions of this kind (e.g. strong convexity) are such that approximation 'by functional' implies the approximation 'by solution', and the corresponding errors can be effectively expressed one in terms of the other. So in this case too we can restrict ourselves to approximation of the solution 'by functional'.

### 1.2.3.3

Correspondingly to what has been said, we introduce two measures of the error of a point  $x \in G_*$  regarded as an approximate solution of a problem  $f$ : a vector measure  $\varepsilon(x, f)$  and a scalar measure  $\varepsilon_*(x, f)$ .

The vector measure of the error is defined as

$$\begin{aligned}\varepsilon(x, f) &= (\varepsilon_0(x, f), \dots, \varepsilon_m(x, f)) \\ &= \begin{cases} (+\infty, \dots, +\infty) & \text{if } x = * \text{ and } f \text{ is compatible,} \\ (0, \dots, 0) & \text{if } x = * \text{ and } f \text{ is incompatible,} \\ \{[f_0(x) - f_*]_+, [f_1(x)]_+, \dots, [f_m(x)]_+\} & \text{in the remaining cases, i.e. for } x \in G.\end{cases} \quad (2.2)\end{aligned}$$

Here and elsewhere,  $[t]_+$  (where  $t$  is a scalar) denotes  $\max\{0, t\}$ . The content of the definition (2.2) is clear:  $\varepsilon(x, f)$  shows by how much the values of the functionals of the problem  $f$  at the point  $x$  exceed their nominal values, i.e. the values required for an exact solution. The nominal value of the objective function is  $f_*$ , and the nominal values of the functional constraints are equal to 0. Further, the announcement that a compatible problem is incompatible is punished by a penalty of  $+\infty$ . We point out that smallness of  $\varepsilon(x, f)$  still does not imply that  $x$  is a feasible point of the problem: in estimating the error of the solution by means of the vector  $\varepsilon(x, f)$  we do not insist that the constraints be satisfied exactly, but we fix the corresponding residuals. Smallness of  $\varepsilon(x, f)$  implies, in particular, that these residuals are small:  $x$  is an 'almost optimal almost feasible point'.

The scalar measure of the error is defined as

$$\begin{aligned}\varepsilon_*(x, f) &= \\ &= \begin{cases} +\infty, & x = * \text{ and } f \text{ compatible,} \\ +\infty, & x \in G \text{ but is not a feasible point of } f, \\ 0, & x = * \text{ and } f \text{ is incompatible,} \\ f_0(x) - f_*, & \text{in the remaining cases (i.e. when } x \in G \text{ and is a feasible point of } f).\end{cases} \quad (2.3)\end{aligned}$$

The error  $\varepsilon_*(x, f)$  in fact eliminates from consideration any approximate solution  $x \in G$  which is not a feasible point of  $f$  (because the error of such a solution is equal to  $+\infty$ ). It is worth using this measure in cases where, from intrinsic considerations, only those approximate solutions which satisfy the constraints exactly are admissible.

We shall have to do mainly with the vector measure of error (the reason again is that the problem of finding feasible points is not well-posed unless additional assumptions are made about the problem itself (see Section 4.1, equation 1.17)).

We shall call the measure  $\varepsilon(x, f)$  the *absolute error* of the point  $x$  regarded as an approximate solution of the problem  $f$ , and the measure  $\varepsilon_*(x, f)$  the *\*-absolute error*.

## 1.2.3.4

It is not always convenient to use absolute errors, particularly in cases where one is speaking about potential bounds for the laboriousness of methods which guarantee a given error in solving all problems of a given field. For example, it is meaningless to pose the question of the laboriousness of a method which solves all, let us say convex, problems

$$f(x) \rightarrow \min |x \in \mathbb{R}, |x| \leq 1 \quad (2.4)$$

with an absolute error of 1. For, a method of this sort would, in fact, necessarily allow us to solve all the problems indicated with an *arbitrary* error however small; because, if we want to solve a problem  $f$  with an accuracy  $\varepsilon$ , we need only apply the method under investigation to the problem  $f(x) = (1/\varepsilon)f(x)$ . It is clear that a solution of  $f$  with absolute error 1 will be a solution of  $f$  with absolute error  $\varepsilon$ . Similarly it is meaningless to ask a question about the laboriousness of such a method on the whole field of problems under consideration; it will, of course, be infinite.

The way out of this situation is to change over from measuring the residuals in 'absolute' units to measuring them in a suitable relative scale. In other words, instead of the absolute error vector  $\epsilon(x, f)$  we shall consider the *relative error* vector

$$\nu(x, f) = (\nu_0(x, f), \dots, \nu_m(x, f)) = \left( \frac{\epsilon_0(x, f)}{r_0(f)}, \dots, \frac{\epsilon_m(x, f)}{r_m(f)} \right),$$

where the  $r_j(f)$  are the units in which the error in the  $j$ th functional is measured. The choice of these units is dictated by considerations of mathematical convenience and the striving for the clearest results. It is clear, incidentally, that this choice does not play any essential part. Consider, for example, the problem (2.4). Here the most natural choice for  $r(f)$  is the variation of  $f$  on the interval  $\Delta = \{|x| \leq 1\}$ , i.e. the number

$$\sup_{|x| \leq 1} f(x) - \inf_{|x| \leq 1} f(x).$$

With this choice  $r(f)$  represents essentially the maximal possible absolute error of a point  $x \in \Delta$  regarded as an approximate solution of the problem  $f$ , and so the clause ' $x \in \Delta$ ' is an approximate solution of the problem  $f$  with the relative error  $\nu$  now has a clear, precise meaning. It asserts that  $x$  is a  $1/\nu$  times better approximate solution of  $f$  than the worst possible solution which could be obtained by a 'trivial search' (see page 52).

The choice of the numbers  $r_j(f)$ —we shall call them the *normalizing factors*—will be described explicitly later for each of the classes of problems subsequently considered. We shall always choose them to be non-negative. As a rule the normalizing factors will be chosen so that the relative error of an arbitrary point (or of some point known *a priori*)  $x \in G$  regarded as a solution

of any problem  $f \in \mathfrak{U}$  will not exceed 1. A field of problems  $\mathfrak{U}$  together with a rule for forming the normalizing factors, i.e. the *normalizing map*

$$f \rightarrow r(f) = (r_0(f), \dots, r_m(f)), \quad f \in \mathfrak{U},$$

will be called a *weighted field of problems*.

## 1.2.3.5

The vector of relative errors introduced above is still not a very convenient measure of error for the analysis of methods. It would be desirable to measure error by a scalar, not a vector quantity (otherwise it is difficult to compare methods precisely). There are various 'natural' ways in which a scalar measure of error could be derived from the vector measure. The simplest is to take the maximum component of the relative-error vector, and that is how we shall actually proceed. At a first glance, in such a definition all the components of a problem seem to be equivalent as regards the influence of their 'residuals' on the error of  $x$  as a solution of a problem  $f$ , but in fact this is not always true. However, this is only an apparent drawback: any *a priori* 'non-equivalence' of the residuals among the various components of the problem can always be taken care of by a suitable choice of the normalizing factors.

Finally we adopt two definitions of the *relative error* of a point  $x \in G_*$  regarded as an approximate solution of a problem  $f \in \mathfrak{U}$ :

$$\begin{aligned} \nu(x, f) &= \begin{cases} 1, & f \text{ compatible, } x = *, \\ 0, & f \text{ incompatible, } x = *, \\ \max \left\{ \frac{f_0(x) - f_*}{r_0(f)}, \frac{f_1(x)}{r_1(f)}, \dots, \frac{f_m(x)}{r_m(f)} \right\}, & x \neq *; \end{cases} \\ \nu_*(x, f) &= \begin{cases} 1, & (f \text{ compatible, } x = *) \text{ or } (x \in G \text{ is not a feasible point of } f)_1 \\ 0, & f \text{ incompatible, } x = * \\ \frac{f_0(x) - f_*}{r_0(f)} & \text{otherwise.} \end{cases} \end{aligned} \quad (2.5)\dagger$$

Here the  $r_j(f) \geq 0$ ,  $0 \leq j \leq m$ , are the normalizing factors; it will be shown how to choose them in each of the cases later considered.

The possibility that some denominators in (2.5) may be zero has to be admitted, in order to avoid making trivial reservations, and in this connection let us agree once and for all on the convention that

$$\frac{a}{0} = \begin{cases} +\infty, & a > 0, \\ 0, & a = 0, \\ -\infty, & a < 0. \end{cases}$$

<sup>†</sup> The symbol  $\dagger$  before the number of a formula indicates that that number refers to the group of formulae which is preceded by the symbol  $\lceil$  and followed by the  $\rfloor$ .

We note that the penalty  $+\infty$  in (2.2), (2.3) has been replaced in (2.5) by the penalty 1. The reasons for this change will be given in Section 1.4.1.

We hope that the motivation for the introduction of the measures of error given in (2.2), (2.3), (2.5) is sufficiently convincing. So far as the formal side of the matter is concerned, the definitions adopted satisfy the natural requirements for error measures: the quantities  $\varepsilon(x, f)$ ,  $\varepsilon_*(x, f)$ ,  $v(x, f)$ ,  $v_*(x, f)$  are non-negative numbers (vectors) and are equal to zero if and only if  $x$  is either an exact solution of  $f$  or is the correct answer as regards the incompatibility of  $f$ .

### 1.3 NUMERICAL METHODS OF SOLVING MATHEMATICAL PROGRAMMING PROBLEMS

In this section we introduce a formal definition of the fundamental object of our investigations, a numerical method of optimization. We begin by describing the information basis of the methods, i.e. *oracles*.

#### 1.3.1 Oracles

As already mentioned, a method of solving problems of a certain family does not, at the start of its work, 'know precisely what problem it has to solve'. All its *a priori* knowledge is the fact that the problem presented belongs to some earlier known field of problems  $\mathfrak{U}$ . The method stores the further information about the problem needed for its solution by turning to the source of information, by asking questions of the oracle. The oracle's answers, in general, are distorted by noise.

We formalize the concept of an oracle in the following sufficiently general form. The *oracle*  $\mathcal{O}$  for a field of problems  $\mathfrak{U}(F, G, m, E)$  is a set of several objects, *viz.*,

- a space*  $\Omega$  of 'oracle noises' with a probability distribution  $F_\omega$  defined on  $\Omega$  ( $\Omega$  is assumed to be a Polish space, and  $F_\omega$  a regular Borel measure, complete with respect to Lebesgue measure) (see Section A.1);
- an observation function  $\psi(x, f, \omega)$ :  $G \times F \times \Omega \rightarrow I$  taking values in some set  $I$  (the *information space*).

The definition includes the requirement that  $\psi$  be Borel with respect to  $\omega$ ,  $x$ ;  $I$  is assumed to be a Polish space.

If the observation function does not depend on  $\omega$  (i.e. if the oracle's answer is uniquely determined by the ordered pair  $(x, f)$ ), then the oracle is said to be *deterministic*. In this case, of course,  $\Omega$  can be regarded as a 1-point set. An oracle which is not deterministic is called a *stochastic oracle*. The oracle determined by the objects  $\Omega$ ,  $F_\omega$ ,  $I$ ,  $\psi(x, f, \omega)$  will be denoted by  $\mathcal{O}((\Omega, F_\omega), I, \psi(x, f, \omega))$ .

Let  $\omega_i$  be independent random variables with values in  $\Omega$  and with a distribution function  $F_\omega$ . We shall suppose that at the  $i$ th step the method can put a question about the problem  $f$  under solution at any point  $x_i \in G$ . The oracle's answer is a point  $\psi(x_i, f, \omega_i) \in I$ . The random variable  $\omega_i$  is the oracle noise occurring at the  $i$ th cycle. All the information accumulated by the method in the first  $i$  steps of its work on the problem  $f$  is a sequence  $\{x_1, \psi(x_1, f, \omega_1); x_2, \psi(x_2, f, \omega_2); \dots; x_i, \psi(x_i, f, \omega_i)\}$  of the questions posed and the answers to them.

#### 1.3.1.1

We give a simple example. Let  $\Omega$  be a 1-point set (so that the oracle is in fact free from noise), let  $I = R^{m+1}$ , and let  $\psi(x, f) = f(x)$ . The oracle so defined is capable of giving at each point of  $G$  the values of all the functionals of the problem.

More 'informative' oracles can provide not only the values of the functionals of the problem but also the values of their derivatives up to some order. Oracles of this type have a certain property of being localized, which in the general case can be defined as follows. An oracle  $\mathcal{O} = \mathcal{O}((\Omega, F_\omega), I, \psi(x, f, \omega))$  for a field of problems  $\mathfrak{U}(F, G, m, E)$  is said to be *local* if for every  $x, f, f'(x \in G, f, f' \in F)$ , such that  $f \equiv f'$  in some neighbourhood of  $x$  in  $G$ , we have  $\psi(x, f, \omega) = \psi(x, f', \omega)$  for all  $\omega \in \Omega$ .

We emphasize that the neighbourhood which appears in this definition may depend on  $x, f, f'$ . A local oracle 'feels' only the local structure of the functionals of the problem at the 'interrogated' point. All the 'usual' oracles in mathematical programming are local, and it is precisely with such oracles that we shall be working later.

#### 1.3.2

We are now in a position to define the basic object of our studies—a method of solving problems of a given class. Let  $\mathfrak{U}(F, G, m, E)$  be a weighted field of problems and let  $\mathcal{O}((\Omega, F_\omega), I, \psi(x, f, \omega))$  be the corresponding oracle. What is meant by 'a method of solving problems in  $\mathfrak{U}$  using the oracle  $\mathcal{O}$ ' is clear. In fact, the process of solution consists in putting a series of questions to the oracle and forming, on the basis of the answers received to the successive questions, the result at a given moment of time. The method itself then must be a set of rules governing the process of solution. Thus, a *method of solving problems in  $\mathfrak{U}$ , using the oracle  $\mathcal{O}$*  (for brevity, an  $(\mathfrak{U}, \mathcal{O})$ -method), is a set of rules for forming the successive questions, moments of pause, and presentation of the result, and also of this result itself as a function of the information accumulated up to the given moment about the problem being solved. These rules may be either deterministic or randomized, leading respectively to deterministic or stochastic methods of mathematical programming.

Thus the work of a method on a problem  $f$  consists in forming a certain sequence  $x_1, \dots, x_{N_f}$  of points of  $G$  (a sequence of questions put to the oracle in solving this problem) and of an additional point  $\tilde{x} \in G \cup \{*\}$ —the result of applying the method to  $f$ . The number  $N_f$  of steps of the method on the problem  $f$  may, in the general case, be formed by the method itself (and therefore it depends on the problem). The sequence  $x_1, \dots, x_{N_f}, \tilde{x}$  (the lower subscript  $f$  will often be omitted) could be called the *trajectory of the method* on the given problem. Formally, however, it is more convenient to adjoin to  $G$ , as well as the element  $*$ , another element  $\emptyset$  (a symbol of ‘inaction’), and to regard the trajectory of the method on  $f$  as the infinite sequence  $x_1, \dots, x_{N_f}, \tilde{x}, \emptyset, \emptyset, \dots, \emptyset, \dots$ .

It is reasonable to start the formal definition of the concept ‘method’ with a formal definition of a trajectory.

### 1.3.2.1

**DEFINITION.** Let  $\mathfrak{U}(F, G, m, E)$  be a weighted field of problems and let  $\mathcal{O} = \mathcal{O}(\Omega, F_\omega, I, \psi)$  be the corresponding oracle; let  $\bar{G} = G \cup \{*\} \cup \{\emptyset\}$ .

1. A *G-trajectory* is any sequence  $x^\infty = (x_1, x_2, \dots)$ ,  $x_i \in \bar{G}$ .
2. The *laboriousness* of a trajectory  $x^\infty$  is the number  $l(x^\infty)$  of members distinct from  $\emptyset$  of the sequence  $x^\infty$ . (The laboriousness of a trajectory can take non-negative integer values and the value  $+\infty$ .)
3. A trajectory  $x^\infty$  is said to be *resultative* if the set  $\{i | x_i \neq \emptyset\}$  is not void and is finite (i.e. if  $0 < l(x^\infty) < +\infty$ ). The *result*  $\tilde{x}(x^\infty)$  of a resultative trajectory is the last element distinct from  $\emptyset$  of the sequence  $x^\infty$ .
4. The *error* of a trajectory  $x^\infty$  on a problem  $f \in \mathfrak{U}$  is the number

$$v(x^\infty, f) = \begin{cases} +\infty & \text{if } x^\infty \text{ is not resultative,} \\ v(\tilde{x}(x^\infty), f) & \text{otherwise.} \end{cases}$$

Thus the error of a resultative trajectory is the error of its result regarded as an approximate solution of the problem  $f$ . The error of a ‘non-resultative’ trajectory is  $+\infty$  by definition.

Similarly the  $*$ -*error* of a trajectory on a problem  $f$  is defined to be

$$v_*(x^\infty, f) = \begin{cases} +\infty & \text{if } x^\infty \text{ is not resultative,} \\ v_*(\tilde{x}(x^\infty), f) & \text{otherwise.} \end{cases}$$

### 1.3.3

It is now clear what an ‘ $(\mathfrak{U}, \mathcal{O})$ -method’ is. Every such method is a set of recurrence rules for forming a trajectory as a function of the information available at a given moment.

We give a precise definition. Let

$$J = \bar{G} \times I, \quad J^i = \underbrace{J \times \dots \times J}_{i \text{ times}} \quad (J^0 \text{ is a 1-point set}).$$

We shall denote points of  $J$  as  $\eta = (x, \xi)$ ,  $x \in \bar{G}$ ,  $\xi \in I$ , and points of  $J^i$  as

$$\eta^i = (\eta_1, \dots, \eta_i), \quad \eta_s \in J.$$

#### 1.3.3.1

We define first a deterministic  $(\mathfrak{U}, \mathcal{O})$ -method.

**DEFINITION.** A *deterministic  $(\mathfrak{U}, \mathcal{O})$ -method* is a set  $\mathscr{B} = \{\tilde{x}_i(\eta^{i-1})\}_{i=1}^\infty$  of Borel functions  $\tilde{x}_i(\cdot): J^{i-1} \rightarrow \bar{G}$ .

(Note: The topology on  $\bar{G}$  coincides on  $G$  with the original topology of  $G$ ;  $*$  and  $\emptyset$  are isolated elements.)

Let

$$\omega^\infty = (\omega_1, \dots, \omega_k, \dots) \in \Omega \times \Omega \times \dots \times \Omega \times \dots = \Omega^\infty.$$

A *trajectory of a deterministic method  $\mathscr{B}$*  on a problem  $f \in \mathfrak{U}$  with noise  $\omega^\infty$  is a sequence  $x^\infty(\omega^\infty, f, \mathscr{B})$  defined by the recurrence rules

$$x_i = \tilde{x}_i(x_1, \psi(x_1, f, \omega_1); \dots; x_{i-1}, \psi(x_{i-1}, f, \omega_{i-1})).$$

(Note: The function  $\psi(x, f, \omega)$  was not defined for  $x = *$  or for  $x = \emptyset$ ; we shall take  $\psi(*, f, \omega) = \psi(\emptyset, f, \omega) = a$ , where  $a$  is a fixed point of  $I$ .)

It can be shown (see Section A.1.2) that  $x^\infty(\omega^\infty, f, \mathscr{B})$  is, as a function of  $\omega^\infty$ , a Borel mapping of  $\Omega^\infty$  into  $\bar{G}^\infty = \bar{G} \times \dots \times \bar{G} \times \dots$  (the topology of direct products is always the Tikhonov topology). Therefore one can speak of the distribution of values of  $x^\infty(\omega^\infty, f, \mathscr{B})$ , induced by the measure  $F_{\omega^\infty} = F_{\omega_1} \times \dots \times F_{\omega_k} \times \dots$  on  $\Omega^\infty$ . This distribution, which we denote by  $\Phi_{x^\infty}(f, \mathscr{B})$ , is called the *distribution of trajectories of the method  $\mathscr{B}$  on the problem  $f$* . One can also consider the joint distribution of  $(x^\infty, \omega^\infty)$ , which we denote by  $\Phi_{x^\infty, \omega^\infty}(f, \mathscr{B})$ , and call it the *joint distribution of the trajectories and noise of the method  $\mathscr{B}$  on the problem  $f$* .

**Remark.** The requirement that the search rules  $\tilde{x}_i(\eta^{i-1})$  and the observation functions be Borel functions, and also the requirements that  $E$  be separable and that  $I$  be a Polish space, are essential only when stochastic oracles are being considered. These hypotheses ensure that the methods’ trajectories are measurable, and are required only so that the language of probability theory can be applied to the analysis of methods. But if the oracle is deterministic and if we are studying deterministic methods, then the trajectories of the method on a problem, the result of the work of the method, and also the characteristics (defined later) of the method on a problem or on a class of problems, can be

given a natural definition without assumptions of the type that the functions of observation of the oracle be Borel, that the search rules be Borel, that  $E$  and  $J$  be separable, etc. Accordingly, in future, when describing deterministic methods, of one sort or another, of solving optimization problems when the oracle is deterministic, we shall not bother about the separability of  $E$  or the 'Borelness' of the search rules of the methods described. Of course, only spaces  $E$  which are separable are, in fact, of any interest, and in this case the methods described later will automatically have Borel search rules, but we shall not specially watch over this.

### 1.3.3.2

We shall now define a stochastic  $(\mathfrak{U}, \mathcal{O})$ -method. We begin with the most natural definition, which is, however, not the most convenient one.

**DEFINITION.** A stochastic  $(\mathfrak{U}, \mathcal{O})$ -method is a set  $\{\Phi_{x_i|\eta^{i-1}}\}_{i=1}^{\infty} = \tilde{\mathcal{B}}$  of probability distributions on  $G$  which depend in a Borel manner on a parameter  $\eta^{i-1} \in J^{i-1}$ .

(Note: In connection with Borel families of measures and operations on measures which we utilize, see Section A.1.)

This means that  $\Phi_{x_i|\eta^{i-1}}$  is the distribution according to which the  $i$ th successive element of the trajectory of the method  $\tilde{\mathcal{B}}$  is selected, it being understood that the information available at this step is  $\eta^{i-1}$ .

We now define the distribution of trajectories of a method  $\tilde{\mathcal{B}}$  on a problem  $f$  in the following natural way. First we define recurrently the distributions  $\Phi_{x^i|\omega^\infty}(\tilde{\mathcal{B}}, f)$  of fragments  $x^i = (x_1, \dots, x_i)$  of trajectories with a fixed noise  $\omega^\infty = (\omega_1, \omega_2, \dots)$  in the oracle:

$$\Phi_{x^i|\omega^\infty}(\tilde{\mathcal{B}}, f) = \Phi_{x_i|\psi^{i-1}(x^{i-1}, f, \omega^{i-1})} \otimes \Phi_{x^{i-1}|\omega^\infty}(f, \tilde{\mathcal{B}}).$$

Here

$$\psi^i(x^i, f, \omega^i) = (x_1, \psi(x_1, f, \omega_1); \dots; x_i, \psi(x_i, f, \omega_i)).$$

It is clear that the family of measures  $\{\Phi_{x^i|\omega^\infty}(f, \tilde{\mathcal{B}})\}_{i=1}^{\infty}$  is concordant, and therefore it defines a measure  $\Phi_{x^\infty|\omega^\infty}(f, \tilde{\mathcal{B}})$  on  $G^\infty$  which, together with all the measures  $\Phi_{x^i|\omega^\infty}(f, \tilde{\mathcal{B}})$ , is Borel with respect to  $\omega^\infty$ . The measure  $\Phi_{x^\infty|\omega^\infty}(f, \tilde{\mathcal{B}}) = F_{\omega^\infty} \otimes \Phi_{x^1|\omega^\infty}(f, \tilde{\mathcal{B}})$  is called the joint distribution with noise of the trajectories of  $\tilde{\mathcal{B}}$  on the problem  $f$ , and its projection  $\Phi_{x^\infty}(f, \tilde{\mathcal{B}})$  on  $G^\infty$  is called the distribution of the trajectories of  $\tilde{\mathcal{B}}$  on  $f$ .

### 1.3.4 Mixtures, and the structure of stochastic methods

In the previous section the definition of a stochastic  $(\mathfrak{U}, \mathcal{O})$ -method is presented in an entirely natural way, a way which seems to be the 'most general' one for

including the possibility of a randomization of search. But there is another method of randomization, which, at first sight, seems much more trivial; that is, formation of a mixture of deterministic methods. Let  $T$  be a space with a measure  $F_t$ , and let  $\tilde{\mathcal{B}} = \{\mathcal{B}^t = \{\tilde{x}_i(\eta^{i-1}, t)\}_{i=1}^{\infty}\}_{t \in T}$  be a family of deterministic  $(\mathfrak{U}, \mathcal{O})$ -methods, indexed by the elements  $t$  of the set  $T$ . We suppose that this family is Borel with respect to  $t$  (i.e. all the  $\tilde{x}_i(\eta^{i-1}, t)$  are Borel with respect to the set  $\eta^{i-1}, t$ ).

We now imagine the following process of solving problems in  $\mathfrak{U}$ . Before the solution begins, an index  $t \in T$  is selected at random (according to the probability distribution  $F_t$ ), and then throughout the course of solution the method  $\mathcal{B}^t$  is always used. This last sentence defines (informally) a certain randomized method of solving solutions of the class  $\mathfrak{U}$ ; we shall call it a mixture of methods  $\tilde{\mathcal{B}}$  and denote it by  $\tilde{\mathcal{B}}$ :

$$\tilde{\mathcal{B}} = \int_T \mathcal{B}^t dF_t.$$

In a sense, a mixture is a mixed strategy of solution (the pure strategies are, of course, identified with deterministic methods of solution).

It is easy to give a formal definition of a mixture. An  $(\mathfrak{U}, \mathcal{O})$ -mixture is an ordered pair consisting of a measure space  $(T, F_t)$  and families  $\{\mathcal{B}^t\}$ , Borel with respect to  $t$ , of deterministic  $(\mathfrak{U}, \mathcal{O})$ -methods. It is also easy to define the basic characteristic of a mixture  $\tilde{\mathcal{B}} = \int_T \mathcal{B}^t dF_t$ , on a problem  $f \in \mathfrak{U}$ ; this is the joint distribution with noise,  $\Phi_{x^\infty|\omega^\infty}(f, \tilde{\mathcal{B}})$ , of its trajectories on the problem  $f$ . To be precise, let  $x^\infty(\omega^\infty, t, f)$  be a trajectory of  $\mathcal{B}^t$  on the problem  $f$  with oracle noise  $\omega^\infty$ . Since the family  $\mathcal{B}^t$  is Borel with respect to  $t$ , it is easily deduced that the map

$$(\omega^\infty, x^\infty(\omega^\infty, t, f)) : \Omega^\infty \times T \rightarrow \Omega^\infty \times G^\infty \quad (f \text{ is a parameter})$$

is Borel with respect to  $(\omega^\infty, t)$ . The distribution of values of this mapping induced by the measure  $F_{\omega^\infty} \times F_t$  on  $\Omega^\infty \times T$  is, by definition,  $\Phi_{x^\infty|\omega^\infty}(f, \tilde{\mathcal{B}})$ . This definition corresponds exactly to the original informal description of the idea of a 'mixture'. In terms of  $\Phi_{x^\infty|\omega^\infty}(f, \tilde{\mathcal{B}})$  we define the distribution  $\Phi_{x^\infty}(f, \tilde{\mathcal{B}})$  of the trajectories of  $\tilde{\mathcal{B}}$  on the problem  $f$  in the same way as in Section 1.3.3.2.

It turns out that, in a certain precise sense, the possibilities of mixtures are in no way more restricted than those of stochastic methods as defined earlier.

**Theorem.** For every stochastic  $(\mathfrak{U}, \mathcal{O})$ -method  $\tilde{\mathcal{B}}$  there is a mixture

$$\tilde{\mathcal{B}} = \int_0^1 \mathcal{B}^t dt$$

of deterministic  $(\mathfrak{U}, \mathcal{O})$ -methods which is equivalent to  $\tilde{\mathcal{B}}$  in the following

precise sense: the joint distributions with noise of the trajectories of  $\mathcal{B}$  or of  $\tilde{\mathcal{B}}$  are the same on every problem in  $\mathfrak{U}$ .

The proof of this theorem is postponed to Section 1.7. Let us discuss the result itself. At first sight it may seem surprising, since the definition of a mixture is more trivial than the original definition of a stochastic method. We see that to realize any such method it is sufficient to turn to a random-number generator, of the normal sort, and to do this once only. It is, by the way, easy to give an intuitively convincing justification of this assertion. For how can a stochastic method be realized on an electronic computer, say? Clearly, by means of a program in which, at each cycle of its operation, recourse is had to a random-number generator working quite independently. But then it would be possible, in advance of running the program, to use the random-number generator the requisite number of times and to write into the memory store the whole set of random numbers. During the work the memory store can be consulted, not the random-number generator. The resulting program of work is the program of a deterministic method. A mixture of such methods is indeed the original stochastic method.

Of course, the assertion just formulated does not imply that, in practice, all methods of random search should be organized just as a mixture. The equivalence established by the theorem is an equivalence of behaviour on problems and certainly not an equivalence of internal organization.

However, within the framework of our approach, which ignores the internal structure of methods, methods which are equivalent according to the theorem are indistinguishable.

The significance of the theorem is not a practical, but a formal, one. The theorem allows us to interpret stochastic methods as certain simple conglomerates of deterministic methods. As a technical tool, this result is the basis of the whole analysis of randomized methods.

We remark that so far we have still not asserted that every mixture is equivalent to some stochastic method. This is, of course, indeed true (under extremely general hypotheses about the properties of the objects which appear in the definition), but we shall not specify these properties nor prove the relevant theorem. Instead of that, we shall adopt the definition of a mixture as the only definition operative in this book of a stochastic  $(\mathfrak{U}, \mathcal{O})$ -method, and cancel the definition given in Section 1.3.3.2, which has now served its purpose. We shall denote the set of all stochastic  $(\mathfrak{U}, \mathcal{O})$ -methods by  $\mathcal{B}(\mathfrak{U}, \mathcal{O})$ . It is clear that deterministic  $(\mathfrak{U}, \mathcal{O})$ -methods are particular cases of mixtures; or speaking rather more strictly, deterministic methods are equivalent to suitable mixtures in the sense of Theorem 1.3.4; and so the set of these methods is identified with a certain subset  $\mathcal{B}(\mathfrak{U}, \mathcal{O})$  in  $\mathcal{B}(\mathfrak{U}, \mathcal{O})$ . In view of this inclusion, it is sufficient to give subsequent definitions only for stochastic methods.

#### 1.4 CHARACTERISTICS OF METHODS ON A PROBLEM AND ON A CLASS. COMPLEXITY

We continue the formation of the language in which we shall describe the properties of mathematical programming methods. Let  $\mathfrak{U}(\mathcal{F}, G, m, E)$  be a fixed weighted field of problems and let  $\mathcal{O}((\Omega, F_\omega), I, \psi)$  be the corresponding oracle. Our ultimate purpose is to examine the best  $(\mathfrak{U}, \mathcal{O})$ -methods. In order to progress in this direction, we must first study how to compare methods one with another, first on each problem in  $\mathfrak{U}$ , and then on  $\mathfrak{U}$  as a whole. We start with the characteristics of methods on a problem.

##### 1.4.1

As already mentioned, the approach we have adopted ignores properties of methods which relate to their internal organization, and it is limited to a description of methods in terms of 'quality (error) of the solution *versus* laboriousness'.

From this point of view the behaviour of a method  $\mathcal{B}$  on a problem is exhaustively described by the distribution of its trajectories  $\Phi_{x^\infty}(f, \mathcal{B})$  and even by the distributions 'laboriousness *versus* error' derived therefrom, i.e. by the joint distribution  $\Phi_{v,l}(f, \mathcal{B})$ , induced by  $\Phi_{x^\infty}$ , of the quantities  $(v(x^\infty, f), l(x^\infty))$ .

In some cases we shall also examine the distribution 'laboriousness *versus* \*-error'  $\Phi_{v,l}^*(f, \mathcal{B})$ , i.e. the joint distribution of the ordered pairs  $(v_*(x^\infty, f), l(x^\infty))$ . Both these distributions are measures on  $[0, \infty] \times [0, \infty]$ .

Description of methods in terms of their 'laboriousness *versus* error' distributions on problems does not make direct comparison of methods possible. To make such a comparison possible, one has to 'coarsen' the description by changing over to characterization of methods on problems by numbers instead of by distributions. There are many 'natural' ways of associating a number with a distribution; the simplest is averaging, and this is the one we adopt as basis. That is, we define the (*mean*) laboriousness, error, and \*-error of a method  $\mathcal{B}$  on a problem  $f \in \mathfrak{U}$  by the relations

$$\begin{aligned} l(\mathcal{B}, f) &= \int l d\Phi_l(f, \mathcal{B}), & v(\mathcal{B}, f) &= \int v d\Phi_v(f, \mathcal{B}), \\ v_*(\mathcal{B}, f) &= \int v d\Phi_{v,l}^*(f, \mathcal{B}), \end{aligned}$$

where  $\Phi_l$ ,  $\Phi_v$ ,  $\Phi_{v,l}^*$  are the projections of the measures  $\Phi_{v,l}^*$ ,  $\Phi_{v,l}^*$  on the corresponding lines of the components of  $[0, \infty] \times [0, \infty]$ .

More 'stringent' characteristics are the *strong laboriousness*, the *strong error*, and the *strong \*-error*:

$$\begin{aligned} \bar{l}(\mathcal{B}, f) &= \text{ess sup}_{\Phi_l(f, \mathcal{B})} l, & \bar{v}(\mathcal{B}, f) &= \text{ess sup}_{\Phi_v(f, \mathcal{B})} v, \\ \bar{v}_*(\mathcal{B}, f) &= \text{ess sup}_{\Phi_{v,l}^*(f, \mathcal{B})} v_*, \end{aligned}$$

with these a method is characterized by the properties of its trajectories.

We remark further that in the important cases where the oracle is free from noise and the methods are deterministic, all the ways mentioned of characterizing methods coincide.

*Remark.* We can now give the motivation for changing the penalty in (2.5) to 1 instead of  $+\infty$  as it was in (2.2) and (2.3). The point is that when there is noise present it is impossible to count on a reliable solution of the question whether the problem is compatible (and to guarantee that the problem's constraints are exactly satisfied as a result of the work of the method). Under such circumstances infinite penalties for non-fulfilment of the requirements indicated would lead to infinite mean (and strong) errors of the methods on problems.

#### 1.4.2

It is now opportune to interrupt the stream of definitions and to analyse the interrelations of the concepts introduced. Our object was to formalize the concepts of 'numerical methods of solving optimization problems of a given type' and 'characteristics of a given method'. Let us look at the way in which we have reached this objective.

The starting point was the normal concept of a

##### (1) mathematical programming problem

(and the directly related concept of a field of such problems, meaning the set of all problems of a given type).

To formalize the concept of 'a method of solving problems of a given type' we had to introduce a further concept

##### (2) an oracle for a given field of problems.

By fixing the objects 'a field of problems' and 'an oracle' we were able to define the concept of a

##### (3) method,

which formalizes the meaning of the concept of 'a method of solving problems of a given type using a given source of information about the problem'.

We emphasize that, according to the definition given, the store of  $(\mathfrak{U}, \mathcal{O})$ -methods does not in fact depend on the objects  $\mathfrak{U}$  and  $\mathcal{O}$  themselves, but on their components, viz., the domain  $G$  of problems of the field  $\mathfrak{U}$  and the information space  $I$  of the oracle  $\mathcal{O}$ . The ordered pair  $(G, I)$  uniquely determines the set of  $(\mathfrak{U}, \mathcal{O})$ -methods. But the operation of a given method  $\mathfrak{B}$  on a given problem  $f \in \mathfrak{U}$  (i.e. the joint distribution with noise of the trajectories of  $\mathfrak{B}$  on  $f$ ) is determined not only by the objects  $f, G$ , and  $I$ , but also by the oracle system  $\mathcal{O}$ . However, this distribution still does not depend on the

definition of the measures of the errors of points of  $G_*$  regarded as approximate solutions of  $f$ , i.e. on

##### (4) the normalizing map

provided for the field  $\mathfrak{U}$ . We have to fix this mapping in order to define the characteristics (laboriousness and error) of a method on a problem. When we have available not only the concept of a method but also a way of characterizing it on problems in  $\mathfrak{U}$ , we are then in a position to pose the question of how to select the best method of solving problems in  $\mathfrak{U}$ .

Thus the starting-point for the approach under discussion to the analysis of the possibilities of numerical methods of optimization is the fixation of a set of the following three objects:

- (I) a field of problems  $\mathfrak{U}$ ;
- (II) a normalizing mapping  $f \rightarrow r(f)$  ( $f \in \mathfrak{U}$ ); and
- (III) an oracle  $\mathcal{O}$  for the field of problems  $\mathfrak{U}$ .

The objects (I) and (III) define the store of methods under examination, and (II) enables their work on problems in  $\mathfrak{U}$  to be characterized.

We shall call the objects (I)–(III) the *class of mathematical programming problems* (it will be denoted by  $\mathfrak{A}$ ). As already stated, this concept serves as the starting-point for the analysis of methods, the purpose of which is to estimate the potential possibilities of methods of solving problems of a given class and to find the best (in a certain sense) of these methods.

#### 1.4.3

Thus we have characterized a method  $\mathfrak{B}$  on each problem  $f \in \mathfrak{U}$  by a pair of numbers, the laboriousness and the error. But again this is insufficient for a comparison of one method with another. Of two methods one will be better on some problems in  $\mathfrak{U}$  and worse on others. We are again compelled to coarsen the characterizations of methods, to change over from characteristic functions of problems of a class to characteristic numbers. For this purpose we define the *laboriousness*, the *error*, and the *\*-error of an  $(\mathfrak{U}, \mathcal{O})$ -method  $\mathfrak{B}$  on the class  $\mathfrak{U}$*  as

$$\begin{aligned} l(\mathfrak{B}, \mathfrak{U}) &= \sup_{f \in \mathfrak{U}} l(\mathfrak{B}, f), \quad v(\mathfrak{B}, \mathfrak{U}) = \sup_{f \in \mathfrak{U}} v(\mathfrak{B}, f), \\ v_*(\mathfrak{B}, \mathfrak{U}) &= \sup_{f \in \mathfrak{U}} v_*(\mathfrak{B}, f). \end{aligned}$$

Similarly, by taking the upper bounds  $\bar{l}(\mathfrak{B}, f)$ ,  $\bar{v}(\mathfrak{B}, f)$ , and  $\bar{v}_*(\mathfrak{B}, f)$  over the problems of the class, we define the *strong characteristics  $\bar{l}(\mathfrak{B}, \mathfrak{U})$ ,  $\bar{v}(\mathfrak{B}, \mathfrak{U})$ ,  $\bar{v}_*(\mathfrak{B}, \mathfrak{U})$  of a method on the class*.

If the ordinary error, strong error, resp. \*-error of a method  $\mathfrak{B}$  on the class  $\mathfrak{U}$  does not exceed  $v$ , then we shall also say that the method  $\mathfrak{B}$  has on  $\mathfrak{U}$  an *ordinary accuracy*  $v$ , *strong accuracy*  $v$ , resp. *\*-accuracy*  $v$ .

*Exercise 1.* Let  $\mathfrak{A}$  be no more than countable, and let  $\mathcal{B}$  be an  $(\mathfrak{A}, \mathcal{O})$ -method. Show that there is a deterministic  $(\mathfrak{A}, \mathcal{O})$ -method  $\tilde{\mathcal{B}}$  such that  $\bar{l}(\tilde{\mathcal{B}}, \mathfrak{A}) \leq \bar{l}(\mathcal{B}, \mathfrak{A})$ ,  $\bar{v}(\tilde{\mathcal{B}}, \mathfrak{A}) \leq \bar{v}(\mathcal{B}, \mathfrak{A})$ , and  $\bar{v}_*(\tilde{\mathcal{B}}, \mathfrak{A}) \leq \bar{v}_*(\mathcal{B}, \mathfrak{A})$ . Thus under 'strong' characterization of methods, randomization of search becomes superfluous.

*Hint.* See Theorem 1.3.4.

#### 1.4.4

We are now in a position to define the potential bound for the laboriousness of solving problems of a given class, that is, the complexity of the class of problems  $\mathfrak{A}$ . Roughly speaking, the complexity is defined as a function of accuracy, equal to the minimal possible laboriousness of a method which still ensures solution, with the necessary accuracy, of all the problems of the class. Depending on how the accuracy and laboriousness are defined (and we have given several definitions of these concepts) and on what restrictions are imposed on the methods used, we obtain not one definition of complexity, but several. We restrict ourselves to two—the 'weakest' and the 'strongest'. Namely, let

$$\begin{aligned}\bar{N}(v) &= \inf \{l \mid \exists \mathcal{B} \in \mathcal{B}(\mathfrak{A}, \mathcal{O}): l(\mathcal{B}, \mathfrak{A}) \leq l, v(\mathcal{B}, \mathfrak{A}) \leq v\}, \\ N(v) &= \inf \{l \mid \exists \mathcal{B} \in \mathcal{B}(\mathfrak{A}, \mathcal{O}): \bar{l}(\mathcal{B}, \mathfrak{A}) \leq l, v(\mathcal{B}, \mathfrak{A}) \leq v\},\end{aligned}$$

(Note: we always take  $\inf \{t \mid t \in \emptyset\} = +\infty$ .) The function  $\bar{N}(v)$  is called the *stochastic complexity*, and  $N(v)$  the *strong deterministic complexity*, of the class  $\mathfrak{A} = (\mathfrak{A}, \mathcal{O}, r(\cdot))$ .

Clearly,  $\bar{N}(v) \leq N(v)$ . If in the definitions of the complexity functions the error is replaced by the  $*$ -error, then the definition of the  $*$ -complexities  $\bar{N}_*(v)$  and  $N_*(v)$  is obtained.

The main content of this book consists in the calculation of estimates of the complexity of the standard classes of mathematical programming problems and in constructing methods of solving the problems of these classes which realize these estimates.

#### 1.4.5

The motivation for our choice of the minimax method for characterizing methods on classes 'according to the worst case' undoubtedly requires some explanation. There is a widespread belief that the 'minimax approach' is too pessimistic; it is considered more sensible to average the characteristics of the methods on individual problems of the class according to some *a priori* distribution. Such a 'Bayesian' approach postulates that, 'in life', problems of a

given type are distributed in a definite way. We point out, however, that for arbitrary, broad classes of problems (for example, for all the problems considered below), there is no way, justified in any degree, of giving such an *a priori* distribution over the class of problems. Hopes of an 'experimental determination' of such a distribution are unfounded; if the class of problems is parametric, with 50 parameters say, then any reliable 'direct' construction of their joint distribution would require a selection of a fantastic size, which is certainly unrealizable. So, even in the simplest linear problems, an empirical approach to the construction of the *a priori* distribution is hopeless.

Thus the 'Bayesian' approach to the study of methods of solving arbitrarily wide classes of problems has no future in practice: the recommended methods would have to work well with an *arbitrary a priori* distribution; but then they would also be good in the minimax sense.

#### 1.4.6

Let us go further into a difference between the definition given above of the concept of 'method' and the traditional idea of numerical methods. From the point of view adopted here, a method is a process with a finite number of steps which, after a finite time, forms a result of the required accuracy, and then stops. The traditional idea is that a method works for an endless time (there are no 'finite' methods, of the linear programming type, for general classes of problems) and forms a sequence of approximations which converges in some sense or other to the exact solution. The higher the rate of convergence, the better an 'infinite' method is. The scheme we have adopted for applying a method is like this: a required accuracy of solution  $v$  is given, and from this  $v$  a method of solution  $\mathcal{B}$ , is chosen. The laboriousness of the best of such methods is taken to be precisely the complexity of the class.

To all appearances, this scheme has no defects on the theoretical plane. However, in practice people often proceed differently: they 'allow' in principle a method with an infinite number of steps, and set it to work for so much time as the available computational resources allow. Whatever accuracy turns up, that is what turns up. No explicit guarantees in this respect are given as a rule. Notwithstanding that such an approach is unsatisfactory theoretically, in practice it is simpler.

Speaking more formally, let  $\mathcal{B}^\infty$  be an infinite-step method of solving problems of the class  $\mathfrak{A}$  using an oracle  $\mathcal{O}$ , and let  $v(N)$  be the maximal (over the class of problems) error given by it at the  $N$ th step of approximation. (For the sake of definiteness, let the error of this approximation on a problem  $f$  be  $v(\mathcal{B}^N, f)$ , where  $\mathcal{B}^N$  is the obvious  $N$ -step 'truncation' of  $\mathcal{B}^\infty$ . It would be possible also in a similar way to consider other methods of measuring the error.) Let

$$N_{\mathcal{B}^\infty}(v) = \min \{N \mid v(M) \leq v, M \geq N\}.$$

The maximal guarantees relating to the use of  $\mathcal{B}^\infty$  consist in the assertion that for every  $v$  the method  $\mathcal{B}^\infty$  will, after working for a time  $N_{\mathcal{B}^\infty}(v)$ , attain the accuracy  $v$ , whichever the problem  $f \in \mathfrak{U}$  may be. (As already mentioned, traditionally methods are compared by their 'rate of convergence', the latter being estimated, as a rule, rather coarsely—by the order of the dependence on  $N$  of the accuracy attained on a given problem after  $N$  steps. It is clear that such estimates provide no special guarantees.)

Now let  $\tilde{\mathcal{B}}^l(\mathfrak{U})$  be a class of finite-step  $\mathfrak{U}$ -methods 'stopping at the step  $l$ ' (the solution is independent of whatever may be assumed at the  $l$ th step; we shall call such methods *l-step methods*), and let  $N_{\mathfrak{U}}(v)$  be the complexity of  $\mathfrak{U}$  in relation to the family of classes of methods  $\{\tilde{\mathcal{B}}^l\}_{l=1}^\infty$ :

$$N_{\mathfrak{U}}(v) = \min \{l \mid \exists \tilde{\mathcal{B}} \in \tilde{\mathcal{B}}^l(\mathfrak{U}): v(\tilde{\mathcal{B}}, \mathfrak{U}) \leq v\}.$$

For each  $v$  specified in advance it is possible to pick out from the class  $\tilde{\mathcal{B}}^l = \cup_i \tilde{\mathcal{B}}^i$  a method which ensures this accuracy more rapidly than  $\mathcal{B}^\infty$ . It is clear from the definition that  $N_{\mathfrak{U}}(v) - 1 \leq N_{\mathcal{B}^\infty}(v)$  (recall that a method in  $\mathcal{B}^\infty$  forms the result based on  $l - 1$  steps). However, a 'good' method of this sort is special for each particular  $v$ , and it is not clear whether one can 'construct' from them an equally good infinite-step method for all  $v$ . *A priori* it is possible, therefore, that the function  $N_{\mathfrak{U}}(v)$  (which is what we shall actually study) is 'essentially less' (let us say, asymptotically as  $v \rightarrow 0$ ) than any of the functions  $N_{\mathcal{B}^\infty}(v)$  and hence it is not an effective lower bound for the laboriousness of the traditional infinite-step methods.

It might therefore be supposed that the approach adopted here does not permit the estimation of the potential effectiveness of such methods, and therefore does not take in the scheme, widely adopted in practice, of applying numerical methods.

Fortunately, everything turns out happily:  $N_{\mathfrak{U}}(v)$  is an effective bound for the laboriousness of 'convergent' methods. In order to prove this, we shall indicate a way of obtaining a 'good' method of this kind from 'good' finite-step methods. Let  $N_{\mathfrak{U}}(v)$ ,  $0 \leq v \leq \bar{v}$ , be the function defined above of the complexity of the class  $\mathfrak{U}$  (equipped with an oracle  $\mathcal{O}$ ), and let  $\varphi(v) \geq 1$  be any continuous non-increasing function with values in  $[0, \infty]$  which majorizes  $N_{\mathfrak{U}}(v) - 1$  (such a function can be chosen to coincide with  $N_{\mathfrak{U}}(v) - 1$  outside an arbitrarily small neighbourhood of the countable set of points of discontinuity of  $N_{\mathfrak{U}}(v)$  and of the set on which  $N_{\mathfrak{U}}(v) = 1$ ).

**Theorem.** Under the above hypotheses there is an infinite-step  $(\mathfrak{U}, \mathcal{O})$ -method  $\mathcal{B}^\infty$  such that  $N_{\mathcal{B}^\infty}(v) \leq 6\varphi(v)$ ,  $0 \leq v \leq \bar{v}$ .

(Note: the reader is invited to define the concept of an infinite-step  $(\mathfrak{U}, \mathcal{O})$ -method for himself.)

**Proof.** The case  $\varphi \equiv +\infty$  is trivial, so let  $\varphi \not\equiv +\infty$  and let  $y = \inf\{v \mid \varphi(v) < \infty\}$ . We put  $\bar{v} = v_1$ , and suppose that  $v_1, \dots, v_s$  have already been defined; then  $v_{s+1}$  is defined to be  $\max\{v \mid \varphi(v) \geq 2\varphi(v_s)\}$ .

We shall break off the process at the first step for which  $\varphi(v) < 2\varphi(v_s)$  for all  $v < v_s$  (this can happen only if  $\varphi(0) < +\infty$ ). As a result, for some  $\bar{s} \leq +\infty$  and for all natural numbers  $i < \bar{s}$ ,  $\bar{s} > 1$ , there will be defined numbers  $v_i \in [v, \bar{v}]$  such that  $v_1 > v_2 > \dots > v_j > \dots, j \leq \bar{s} - 1$ , and  $+\infty > \varphi(v_j) = 2\varphi(v_{j-1})$ . If  $\bar{s} < \infty$ , then  $\varphi(0) < 2\varphi(v_{\bar{s}-1})$ . Since  $\varphi(\cdot)$  is a majorant of  $N_{\mathfrak{U}}(\cdot) - 1$ , we can, for each  $i \leq \bar{s} - 1$ , find an  $N_i$ -step  $(\mathfrak{U}, \mathcal{O})$ -method  $\mathcal{B}^i$  with  $N_i = N_{\mathfrak{U}}(v_i) \leq \varphi(v_i) + 1$  ( $\mathcal{B}^i$  is an element of  $\tilde{\mathcal{B}}^{N_i}(\mathfrak{U}, \mathcal{O})$ ) which has error  $v_i$  on the class  $\mathfrak{U}$ . If  $\bar{s} < \infty$ , then let  $\mathcal{B}^{\bar{s}}$  be an  $N_{\bar{s}}$ -step  $(\mathfrak{U}, \mathcal{O})$ -method with error 0 on  $\mathfrak{U}$ ,  $N_{\bar{s}} = N_{\mathfrak{U}}(0)$ .

We construct a method  $\mathcal{B}^\infty$  in the following way. The method is made up of stages, of which the  $i$ th stage ( $i \leq \bar{s}$ ) contains  $N_i - 1$  steps and reduces to applying the method  $\mathcal{B}^i$  to solve the problem under solution. The successive approximations, determined by  $\mathcal{B}^\infty$  on the  $i$ th stage, coincide with the result of applying the method  $\mathcal{B}^{i-1}$  to the problem under solution (when  $i = 1$  this approximation is an arbitrary point of  $G$ ). If  $\bar{s} < \infty$ , then at the final stage (i.e. the  $(\bar{s} + 1)$ -th stage) no optimization takes place, and the successive approximations are the result of applying the method  $\mathcal{B}^{\bar{s}}$ .

Let us verify that  $\mathcal{B}^\infty$  is the desired method. Let  $v \in [0, \bar{v}]$ . It may be that  $\varphi(v) = +\infty$ ; in that case the assertion is trivial. Now suppose that  $\varphi(v) < +\infty$ , and so  $v \geq y$ . Then, for some  $i_0 < \bar{s}$ , we have  $v_{i_0} \geq v \geq v_{i_0+1}$  (for  $\bar{s} < \infty$  we take  $v_{\bar{s}} = 0$ ). Hence  $\varphi(v_{i_0}) \leq \varphi(v) \leq \varphi(v_{i_0+1})$ . Since  $v_{i_0+1} \leq v$ , it is clear from the construction that, beginning with the  $(i_0 + 2)$ -th step, the error of  $\mathcal{B}^\infty$  on any problem is not greater than  $v$ . Hence

$$\begin{aligned} N_{\mathcal{B}^\infty}(v) &\leq N_1 + \dots + N_{i_0+1} - (i_0 + 1) \leq \varphi(v_1) + \dots + \varphi(v_{i_0+1}) \\ &\leq 3\varphi(v_{i_0+1}) \end{aligned}$$

(we have taken into account that  $\varphi(v_{i+1}) = 2\varphi(v_i)$  when  $i + 1 < \bar{s}$  and that always  $\varphi(v_{i+1}) \geq \varphi(v_i)$ ). But at the same time  $\varphi(v_{i_0+1}) \leq 2\varphi(v_{i_0})$ , i.e.  $N_{\mathcal{B}^\infty}(v) \leq 6\varphi(v_{i_0}) \leq 6\varphi(v)$ . The theorem has been proved.

**Remark.** The assertion of the theorem is clearly also true in the case where we restrict our attention to deterministic finite-step and infinite-step methods. In such cases, of course, the definition of the convexity function  $N_{\mathfrak{U}}(v)$  has to be altered in an obvious way.

## 1.5 SOME PROPOSITIONS ABOUT MATHEMATICAL PROGRAMMING METHODS

In this section certain general propositions, extensively used later, about methods of solving extremal problems and about complexity functions are brought together. Let  $\mathfrak{U} = (\mathfrak{U}(\mathcal{F}, G, m, E), \mathcal{O}((\Omega, F_\omega), I, \psi), r(\cdot))$  be the class of problems.

### 1.5.1 Canonical forms of methods

According to the definition, an  $\mathfrak{U}$ -method can form as the trajectories *any* sequences of elements of  $G$ , including even those in which the ‘symbol of inaction’ is encountered right up to the result. Trajectories of this sort have little natural meaning, and some discussions of methods are carried out more conveniently if it is supposed that a method does not form such trajectories. On the other hand, in proving general theorems about methods (theorems about the structure of a random search, for instance), the definition given earlier is convenient (and that is exactly why we choose it). It turns out—and this is intuitively obvious—that every method can be ‘reduced to a canonical form’, i.e. can be changed immaterially to ensure that it forms only ‘natural’ trajectories.

**DEFINITION.** A deterministic  $\mathfrak{U}$ -method  $\mathcal{B}$  is said to be *regular* if  $\mathcal{B} = \{\bar{x}_i(\xi^{i-1})\}$  (i.e. the arguments in the rules are just the set of answers of the oracle and do not include the set of questions given at the preceding steps), and if

- (1)  $\bar{x}_i(\xi^{i-1}) = \emptyset$  implies  $\bar{x}_{i+1}(\xi^{i-1}, \xi_i) = \emptyset$  for all  $\xi_i \in I$ ;
- (2)  $\bar{x}_i(\xi^{i-1}) = *$  implies  $\bar{x}_{i+1}(\xi^{i-1}, \xi_i) = \emptyset$  for all  $\xi_i \in I$ .

It is clear from the definition that a regular method forms trajectories of one of the three types:

$$\begin{aligned} & \emptyset, \emptyset, \emptyset, \dots; \\ & x_1, \dots, x_n, \quad \emptyset, \emptyset, \dots, x_1, \dots, x_{n-1} \in G, \quad x_n \in G_*; \\ & x_1, \dots, x_n, \dots, x_i \in G. \end{aligned}$$

A *regular  $\mathfrak{U}$ -mixture* is a mixture of regular  $\mathfrak{U}$ -methods. It turns out that the study of ‘all’ methods can be reduced to the study of regular methods.

**DEFINITION.** Two  $\mathfrak{U}$ -methods  $\mathcal{B}$  and  $\mathcal{B}'$  are said to be *strongly equivalent* (resp. *equivalent*) if the joint distributions with noise of their trajectories (resp. their ‘laboriousness–error’ distributions) coincide with one another on every problem of the class.

We are interested only in those properties of methods which are uniquely determined by their ‘laboriousness–error’ distribution. Therefore equivalent, and *a fortiori* strongly equivalent, methods are indistinguishable within the framework of the approach adopted.

**Theorem.** Every deterministic  $\mathfrak{U}$ -method  $\mathcal{B}$  is equivalent to a regular deterministic  $\mathfrak{U}$ -method. Every  $\mathfrak{U}$ -mixture is equivalent to a regular  $\mathfrak{U}$ -mixture.

We omit the proof of this intuitively obvious theorem.

*Example of the use of the theorem.* In Section 1.4 we introduced the function  $N_n(v)$ , a certain characteristic of the complexity of the class  $\mathfrak{U}$ , and we

demonstrated the rôle of this characteristic as a lower bound of the effectiveness of ‘infinite-step’ methods. Let us prove that  $N_n(v)$  satisfies the inequalities

$$N(v) \geq N_n(v) \geq \tilde{N}(v).$$

These inequalities show that information about the basic objects  $N(v)$  and  $\tilde{N}(v)$  studied here is also information about  $N_n(v)$ .

The first inequality is obvious. To prove the second, we observe that

$$N(v) = \inf \{ l \mid \exists \mathcal{B} \in \mathfrak{B}(\mathfrak{U}): l(\mathcal{B}, \mathfrak{U}) \leq l, v(\mathcal{B}, \mathfrak{U})v \}.$$

By Theorem 1.5.1 the methods  $\mathcal{B}$  appearing in the definition of  $N(\cdot)$  can be regarded as regular. But every deterministic *regular* method  $\mathcal{B}$  with a strong laboriousness  $l$  on the class  $\mathfrak{U}$  produces a result on any problem not later than the  $[l]$ -th step with probability 1 (and then this can be regarded as ‘certainly’). (Note:  $[ \cdot ]$  means, as usual, the integer part of a number.) Therefore any such method is a deterministic method from  $\mathfrak{B}^{[l]}(\mathfrak{U})$  (see Section 1.4.6). But then the inequality  $N(v) \geq N_n(v)$  follows from the definition of  $N_n(v)$ .

The above theorem enables us in future to work only with regular deterministic methods and mixtures of them (and from this point onwards we shall avail ourselves of this possibility, regarding all the methods under consideration, once and for all, as regular methods, without any further special mention of the fact). We further point out that if a deterministic  $\mathfrak{U}$ -method  $\mathcal{B}$  is capable of forming the trajectory  $\emptyset, \emptyset, \emptyset, \dots$ , then it will form this trajectory on any problem and with any noise (why?). Therefore the error of such a method is identically equal to  $+\infty$  and there is no point in considering it. Accordingly, in future ‘methods’ will always be considered to be either deterministic methods which do not form trajectories  $\emptyset, \emptyset, \emptyset, \dots$ , or mixtures of such deterministic methods.

### 1.5.2 Mixtures of problems

Let  $\mathfrak{U}_0$  be a not more than countable subset of  $\mathfrak{U}$ , and let  $\mathbf{P}_f$  be a probability distribution on  $\mathfrak{U}_0$ . We shall call the ordered pair  $\mathfrak{U}_0 = (\mathfrak{U}_0, \mathbf{P}_f)$  an  $\mathfrak{U}$ -mixture of problems. The  $\mathfrak{U}$ -methods  $\mathcal{B}$  on the mixture  $\mathfrak{U}_0$  will be characterized by the *means* (over the distribution  $\mathbf{P}_f$ ) of the laboriousness and error (resp.  $*$ -error), i.e. by the quantities

$$l(\mathcal{B}, \mathfrak{U}_0) = \int_{\mathfrak{U}_0} l(\mathcal{B}, f) d\mathbf{P}_f,$$

$$\tilde{v}(\mathcal{B}, \mathfrak{U}_0) = \int_{\mathfrak{U}_0} v(\mathcal{B}, f) d\mathbf{P}_f,$$

$$\tilde{v}_*(\mathcal{B}, \mathfrak{U}_0) = \int_{\mathfrak{U}_0} v_*(\mathcal{B}, f) d\mathbf{P}_f.$$

It turns out that with such a ‘Bayesian’ characterization, the possibilities of randomized methods are approximately the same as those of deterministic methods.

**Theorem.** Let  $\mathcal{B}$  be any  $\mathfrak{U}$ -method, and let  $\mathfrak{U}_0$  be an  $\mathfrak{U}$ -mixture of problems. There is a deterministic  $\mathfrak{U}$ -method  $\mathcal{B}'$  such that

$$\mathcal{I}(\mathcal{B}, \mathfrak{U}_0) \leq 2\mathcal{I}(\mathcal{B}', \mathfrak{U}_0) \quad \text{and} \quad \bar{v}(\mathcal{B}, \mathfrak{U}_0) \leq 2\bar{v}(\mathcal{B}', \mathfrak{U}_0)$$

(and there is also a deterministic method  $\mathcal{B}$  such that

$$\mathcal{I}(\mathcal{B}, \mathfrak{U}_0) \leq 2\mathcal{I}_*(\mathcal{B}, \mathfrak{U}_0) \quad \text{and} \quad \bar{v}_*(\mathcal{B}, \mathfrak{U}_0) \leq 2\bar{v}_*(\mathcal{B}, \mathfrak{U}_0).$$

**Exercise 1.** Prove this theorem.

< It suffices to prove the first assertion; the second is proved in exactly the same way. We can suppose that  $\mathcal{B}$  is a mixture  $\int_0^1 \mathcal{B}' dt$ . Then, clearly

$$\mathcal{I}(\mathcal{B}, \mathfrak{U}_0) = \int_{\mathfrak{U}_0 \times [0, 1]} \mathcal{I}(\mathcal{B}', f) dP_f dt = \int_0^1 \mathcal{I}(\mathcal{B}', \mathfrak{U}_0) dt$$

and similarly

$$\bar{v}(\mathcal{B}, \mathfrak{U}_0) = \int_0^1 \bar{v}(\mathcal{B}', \mathfrak{U}_0) dt.$$

Since  $\mathcal{I}(\cdot)$  is non-negative, there is a Borel set  $T \subset \Delta = [0, 1]$  with  $\text{mes } T \geq \frac{1}{2}$  such that  $\mathcal{I}(\mathcal{B}', \mathfrak{U}_0) \leq 2\mathcal{I}(\mathcal{B}', \mathfrak{U}_0)$  for  $t \in T$ . Further, we have  $\int_T \bar{v}(\mathcal{B}', \mathfrak{U}_0) dt \leq \bar{v}(\mathcal{B}, \mathfrak{U}_0)$ , and, since  $\text{mes } T \geq \frac{1}{2}$ , this inequality implies that, for some  $t_0 \in T$ , we have  $\bar{v}(\mathcal{B}'^{t_0}, \mathfrak{U}_0) \leq 2\bar{v}(\mathcal{B}, \mathfrak{U}_0)$ . Clearly,  $\mathcal{B}'^{t_0}$  is the desired method. >

This simple theorem generalizes, to the case of ‘vector’ characterization of the resolvent rules, the well-known proposition in the theory of estimates: in a Bayesian approach, randomization of the rules is not necessary. This fact will be extremely useful in obtaining lower bounds for the ‘stochastic complexity’  $\tilde{N}(v)$ ; for, if we define in the natural way the ‘mean’ complexity  $\tilde{N}_{\mathfrak{U}_0}(v)$  of an  $\mathfrak{U}$ -mixture  $\mathfrak{U}_0$  (starting from the means over  $\mathfrak{U}_0$  of the characteristics of the methods), then the obvious inequality  $\tilde{N}_{\mathfrak{U}_0}(v) \geq N_{\mathfrak{U}_0}(v)$  will hold. This result shows that to estimate  $\tilde{N}_{\mathfrak{U}_0}(v)$  it is in fact sufficient to restrict attention to deterministic methods, and this, as a rule, can be carried out more simply than examining all stochastic methods.

### 1.5.3 Complexities related to local oracles

The lower bounds for the complexity which are calculated in subsequent chapters will, as a rule, be invariant in relation to the choice of the oracle with which the class of problems under examination is provided. It is important only that this oracle should be local (see Section 1.3.1.1). Let us outline the

method of obtaining these estimates. As a rule, in obtaining lower bounds for the complexity we shall reduce matters to the study of the behaviour of deterministic methods using deterministic oracles. The basis for this procedure is the following obvious, but extremely useful, fact.

**Lemma on indistinguishability.** Let  $\mathfrak{U} = (\mathcal{F}, G, m, E)$ ;  $\mathcal{O}(\Omega, F_\omega, I, \psi)$ ;  $r(\cdot)$  be a class of problems equipped with a deterministic oracle  $\mathcal{O}$ , and let  $\mathcal{B}$  be a deterministic  $(\mathfrak{U}, \mathcal{O})$ -method. Then the following assertions are true:

- (i) if  $f, f' \in \mathfrak{U}$ , and if  $x^\infty = (x_1, \dots, x_i, \dots) \in \bar{G}^\infty$  is such that  $x^\infty$  is a trajectory of  $\mathcal{B}$  on  $f$ , and if  $\psi(x_i, f) = \psi(x_i, f')$  for all  $i$  such that  $x_i \in G$ , then  $x^\infty$  is a trajectory of  $\mathcal{B}$  on  $f'$ ;
- (ii) let  $N$  be a natural number or  $+\infty$ , and let  $f_i \in \mathfrak{U}$ ,  $0 \leq i < N$ . Further, suppose, for  $0 \leq i < N$ , that  $x_{i+1}$  is the  $(i+1)$ -th point of the trajectory of  $\mathcal{B}$  on  $f_i$ . Suppose that, for all  $i < N$  such that  $x_{i+1} \in G$ , we have  $\psi(x_{i+1}, f_i) = \psi(x_{i+1}, f_{i+1})$  for all  $j$  such that  $N > j \geq i+1$ . Then, for all  $i < N$ , the points  $(x_1, \dots, x_{i+1})$  are the first  $i+1$  points of the trajectory of  $\mathcal{B}$  on  $f_i$ .

Suppose, in particular, that the oracle  $\mathcal{O}$  is local as well as deterministic. Then for (i) to hold it suffices that  $f = f'$  in a neighbourhood of each of the points lying within  $G$  of the trajectory of  $\mathcal{B}$  on  $f$ . For (ii) to hold it suffices that  $f_{i+1} = f_i$  in a neighbourhood of  $x_{i+1}$  for all  $i, j$  such that  $x_{i+1} \in G$  and that  $N > i+1 \geq 1, j \geq i+1$ . (In the last two statements the neighbourhoods are supposed to be within  $G$ .)

The lemma on indistinguishability, for all its obviousness, is extremely useful in obtaining lower bounds for the complexity, and this is easy to understand: the assertions of the lemma reflect the *unique* restriction imposed on the possibilities of methods, the restriction of ‘informational realizability’, consisting in the fact that the trajectory of a method is recursively determined by the information accumulated during the running of the method.

**Exercise 1.** Prove the lemma.

< We restrict ourselves to the proof of assertion (ii); it is by induction over  $i$ .

The case  $i = 0$  is trivial, because the initial point of the trajectory of the method  $\mathcal{B}$  is formed by it in the absence of all information about the problem to be solved, and so this is one and the same point for all problems. For  $i = 1$  assertion (ii) is also trivial: by definition  $x_2$  is the second point of the trajectory of  $\mathcal{B}$  on  $f_1$  (and we have already seen that  $x_1$  is the first point of this trajectory).

**The inductive step.** Suppose we already know that for all  $s$ ,  $1 \leq s \leq k$ , the assertion

$$I(s): \quad x_1, \dots, x_s \text{ is the trajectory of } \mathcal{B} \text{ on } f_{s-1}$$

is true. We shall deduce from this that  $I(k+1)$  is also true, provided that  $k+1 \leq N$ . This will complete the proof. We already know that  $I(2)$  is true, and so we shall suppose that  $k+1 > 2$ .

Since  $\mathcal{B}$  is a regular method and  $x_1, \dots, x_k$  are the first  $k$  points of the trajectory of  $\mathcal{B}$  on  $f_{k-1}$ , so for some  $\bar{k}$  with  $\bar{k} \leq k$  we have  $x_i \in G, i \leq \bar{k}$ , whereas  $x_i \in \{\emptyset, *\}$  when  $\bar{k} < i \leq k$ . Suppose it is already known that

$$\psi(x_i, f_{k-1}) = \psi(x_i, f_k), \quad 1 \leq i \leq k-1,$$

In view of the facts that the method  $\mathcal{B}$  is deterministic and that the rules for forming trajectories are of a recursive kind, it obviously follows from this that  $x_1, \dots, x_k$  are the first  $k$  points of the trajectory of  $\mathcal{B}$  on  $f_k$ . By definition  $x_{k+1}$  is the  $(k+1)$ -th point of this trajectory, and so  $I(k+1)$  is true. So it remains to verify that

$$\psi(x_i, f_{k-1}) = \psi(x_i, f_k)$$

is indeed true. When  $x_i \in \{\emptyset, *\}$ , the function  $\psi(x_i, f)$  does not depend on  $f$ , and so it suffices to verify the formulated assertion for  $i \leq \min\{\bar{k}, k-1\}$ . Let  $i \leq k-1$  be such that  $1 \leq i \leq \bar{k}$ . Then by hypothesis  $\psi(x_i, f_j) = \psi(x_i, f_i)$  when  $j \geq i$ , i.e. in particular,  $\psi(x_i, f_{k-1}) = \psi(x_i, f_i) = \psi(x_i, f_k)$ , as required. The lemma is proved.  $\rangle$

*Exercise 2.* Using the lemma, prove the following (intuitively obvious) result. Let  $G$  consist of  $N$  points, and let the class of problems be formed by all problems of the form  $f_0(x) \rightarrow \min |x \in G$ , let the oracle provide the value of  $f_0$  at the point  $x$  'consulted', and let  $r_0(f) \equiv 1$ . Prove that, for all  $v < 1$ , the strong deterministic complexity of the class  $\mathfrak{U}$  is equal to  $N$ .

#### 1.5.4

In conclusion we make the following remark. Up to now we have supposed that the vector-function  $f$  generating the problem (2.1) is defined only on  $G$ . Similarly the oracle, too, was supposed to give information only at points in  $G$ . But sometimes it is known *a priori* that the function  $f$  is defined in a domain  $G_i$  larger than  $G$  (e.g. on the whole of  $E$ ), and that the oracle is capable of giving information about  $f$  over the whole of this larger domain. It is clear that from the informational point of view, such an 'exit beyond the limits of  $G$ ' may prove useful, and so, in such a situation, it is expedient to allow the methods to put questions about the problem under solution even outside  $G$  (within the limits of  $G_i$ ). The scheme already given for describing classes of problems, methods of solving them, and characteristics of methods, extends in an obvious way to this more general case, and the propositions above (suitably reformulated) all still remain valid.

We shall not, of course, rehearse the new versions of the previous definitions and propositions. We point out only that, in the situation now being

considered,  $G_i$ , like  $G$  previously, must come into the list of objects which specify the class of problems. Moreover, we must now be able to define the error of a point  $x$  regarded as an approximate solution for points  $x \in G_i \setminus G$  as well; we shall take it to be equal to  $+\infty$  (or to a vector with co-ordinates  $+\infty$ , if we are dealing with vector measures of error). After this remark the reader himself will be able without difficulty to extend the scheme described above to the case where  $G_i \supset G$  (and for this reason we refrained from burdening the exposition with yet another object,  $G_i$ ).

## 1.6 ON THE COMPLEXITY OF CLASSES OF MULTI-EXTREMAL PROBLEMS

The most easily verified and the most often encountered property of mathematical programming problems is that of smoothness, which consists in the continuous differentiability for a certain number of times of the functionals in the problem. Correspondingly, methods of solving classes of smooth problems could be considered over the widest applications. Unfortunately, 'universal' methods of this type cannot fail to have catastrophically large and entirely unacceptable laboriousness. To convince ourselves of this, we adduce the corresponding lower bounds for the complexity. First we describe the classes of problems under consideration.

### 1.6.1.

Let  $G \subset E^n$  be a bounded closed set, and let  $k$  be a natural number. We shall consider problems of the form

$$f_0(x) \rightarrow \min |f_i(x) \leq 0, \quad 1 \leq i \leq m, \quad x \in G, \quad (6.1)$$

generated by  $k$ -smooth functions  $f = (f_0, \dots, f_m)$ , i.e. functions  $f$  defined over the whole of  $E^n$  and continuously differentiable  $k$  times. A degree of smoothness contributes little, obviously, to the obtaining of constructive results. It is further necessary to limit the rate of variation of the functions  $f_i$  and their derivatives. The most convenient way of doing this is as follows. Let  $L_0, \dots, L_m$  be given positive numbers. Let  $S^k(L_0, \dots, L_m)$  denote the set of all  $k$ -times continuously differentiable vector-functions  $f = (f_0, \dots, f_m)$  such that the  $k$ th derivative of  $f_i$  in any direction does not exceed  $L_i$ :

$$\left| \frac{d^k}{dt^k} f_i(x + th) \right| \leq L_i \|h\|^k, \quad x \in E^n, h \in E^n, \quad 0 \leq t \leq 1. \quad (6.2)$$

Further, let  $\mathcal{O}$  be an arbitrary local deterministic oracle for the set  $S^k(L_0, \dots, L_m)$  (questions can be put over the whole of  $E^n$ ). Let  $S^k(G; L_0, \dots, L_m)$  denote the class of all problems of the form (6.1) generated by the functions  $f \in S^k(L_0, \dots, L_m)$ . It is convenient to take the normalizing

factors for the class to be

$$r_j(f) \equiv r_j = \frac{\rho^k(G) L_j}{k!},$$

where  $\rho(G)$  is the radius of  $G$ . The significance of  $r_j$  is that it is the natural estimate of the maximal (in  $G$ ) difference between the function  $f_j$  and its Taylor polynomial of order  $(k-1)$  constructed at the 'centre' of the set  $G$ .

### 1.6.2

The following theorem determines the lower bound of the complexity of the class  $S_c^k(G; L_0, \dots, L_m)$ .

**Theorem.** Let the interior of  $G$  be non-empty, so that the asphericity  $\alpha_{\parallel\parallel}(G)$  of the body  $G$  is finite. Then the stochastic complexity of the class of problems  $S_c^k(G; L_0, \dots, L_m)$  admits the lower bound

$$\tilde{N}(v) \geq d^n(k) \alpha_{\parallel\parallel}^{-n}(G) \left(\frac{1}{v}\right)^{n/k} \equiv \Phi_{G,k}(v). \quad (6.3)$$

Here  $d(k) > 0$ .

(For the proof see the authors' paper [34].)

Let us discuss the estimate (6.3). First of all we point out that a similar estimate for the deterministic complexity of the class  $S_c^k(G; L_0, \dots, L_m)$  is well-known; according to the authors' information it was first obtained in [13]. Let us consider some consequences of (6.3). It is clear that, from the practical point of view, the nature of the dependence of  $\Phi_{G,k}(v)$  on  $v$  and  $n$  leads to a catastrophic growth of the complexity of the class when  $v \rightarrow 0$  and when  $n \rightarrow \infty$ , provided that  $k$  is fixed and  $\alpha_{\parallel\parallel}(G)$  remains constant (i.e. unless problems of high dimension are considered on 'strongly condensed' bodies). By increasing the smoothness,  $\Phi_{G,k}(v)$  can be somewhat reduced; but in practice  $k$  is usually not very big ( $k=1$  or  $2$ ). Of course, the point here is not that the actual problems are not sufficiently smooth; it is simply that we are not in a position to estimate the highest derivatives of the components in the problem and to know into precisely which class  $S_c^k(G; \dots, \dots)$  they fall. Moreover, the derivatives 'usually' increase rapidly with growth of their order  $k$ , and the effect of increasing the smoothness is 'eaten up' by the fact that, to ensure the requisite absolute accuracy, one has to employ methods which, although orientated towards greater smoothness, are, on the other hand, constructed on an ever higher relative accuracy.

Of course, from the practical point of view, matters might turn out not too badly, if the quantity  $d(k)$  were sufficiently small. But this is not so. Even the extremely rough (lowered) lower bounds  $\tilde{N}(v)$  determined by the proof of Theorem 1.6.2 (see [33]) show that when  $\alpha_{\parallel\parallel}(G) = 1$  ( $G$  is a ball) and  $k=1$  or  $2$ ,

a guaranteed solution of all multi-extremal problems with a relative error of  $10^{-3}$ , say, is an absolutely hopeless matter when the dimension  $n \sim 20$ .

We remark further that the lower bound for complexity (6.3) is exact (asymptotically as  $v \rightarrow 0$ ) in the following precise sense: with a suitable  $\emptyset$  there are methods of solving problems of the class  $S_c^k(G; L_0, \dots, L_m)$  with an error  $\leq v$  and with an upper bound for the laboriousness of the form  $c(G, k)(1/v)^{n/k}$  (see [13]).

We emphasize that the estimate (6.3) was obtained for the stochastic complexity of the class. Thus the use of methods of random search for solving multi-extremal problems of high dimensionality is just as hopeless as the use of deterministic methods (we are speaking, of course, of methods which give certain guarantees: on individual problems any method 'may be lucky'). This being so, we cannot see the sources of the optimism displayed by devotees of random search as regards its applicability to multi-extremal problems.

Roughly speaking, not only multi-extremal smooth problems, but even 3-extremal smooth problems are not accessible to solution by methods with guaranteed (and applicable in appreciable dimensionalities) estimates of laboriousness: in fact, it can be shown that the estimate (6.3) obtains on problems with  $m=0$  and a not very large number (3 altogether) of critical points throughout the whole of  $E^n$  (i.e. points  $x$  at which  $f'_0(x)=0$ ).

Even more can be asserted. Consider the subclass  $S_c^k(G, 1)$  of the class  $S_c^k(G, 1)$  (corresponding to the unit ball  $G$  of the space  $E^n$ ), consisting of all possible 1-extremal problems  $f_0 \in S^k(1)$  such that  $f(x) = p_f x^2$  when  $\|x\| \geq 1$ ,  $p_f > 0$ ,  $\inf_G f_0 = \inf_{E^n} f_0$ . We emphasize that, when  $f \in S_c^k(G, 1)$ , the equation  $\nabla f(x) = 0$  has, by definition of the class  $S_c^k(G, 1)$ , just one solution in the whole of  $E^n$ . It can be shown that the deterministic complexity  $N(v)$  of this class satisfies an estimate of the form

$$N(v) \geq c(n, k)(1/v)^{(n-1)/k}. \quad (6.4)$$

The estimate (6.4) is not much better than (6.3). Thus even 1-extremal (non-convex) problems do not admit a guaranteed method of solution with an acceptable bound for the laboriousness.

### 1.6.3

Let us discuss some deductions from the results obtained. We have seen that not only all smooth problems, but even 1-extremal smooth problems, constitute a class which is too complex to admit methods, acceptable as regards laboriousness, of solving all problems of the class. This deduction is based on the low (informational) estimates of the actual computational complexity of the methods, and so, from the latter point of view, matters are still worse. On the other hand, from the point of view of applications, it is precisely such sorts of classes of problems which appear most natural. Of course, in practice,

methods are usually applied without thinking too much about what guarantees they can provide, and people are satisfied with the results so obtained, which are at least no worse (and sometimes they may be considerably better) than the original (zeroth) approximations to the solution. Whatever success may be credited to this purely empirical approach to a problem, it cannot be regarded as capable of replacing a theoretical basis for recommendations about the use of this or that method. We see that a sufficiently general positive theory of this kind cannot relate to the class of all optimization problems (or even the class of smooth problems). A way out from this situation would be to single out those classes of optimization problems which are, on the one hand, sufficiently general and natural (that is, in the final analysis, those which embrace sufficiently many practical problems) and which, on the other hand, have a complexity (albeit an informational complexity) of acceptable magnitude. So far as the authors know, only one class of non-linear problems which satisfies these requirements has been singled out, namely, the class of convex programming problems. Convex programming, it would seem, has a sufficiently wide field of applications, and at the same time this class is also acceptable as regards complexity. For this reason we shall go over to the study of classes of convex extremal problems.

### 1.7 PROOF OF THEOREM 1.3.4

See Section A.1 for the concepts used in this proof. The result was obtained in the authors' paper [33].

**Lemma.** Let  $\Phi_{x|y}$  be a family of probability distributions (regular, Borel, and Lebesgue-complete) on the Polish space  $X$ , which depends in a Borel manner on a parameter  $y \in Y$ , where  $Y$  is again a Polish space. There is a function, Borel with respect to  $s$  and  $y$ ,

$$\varphi(s, y): [0, 1] \times Y \rightarrow X$$

such that the distribution of values of  $\varphi$  as a function of  $s \in \Delta = [0, 1]$  ( $\Delta$  is equipped with Lebesgue measure) is, for each  $y \in Y$ ,  $\Phi_{x|y}$ .

#### Derivation of theorem 1.3.4 from this lemma.

Let  $\{\Phi_{x|\eta^{i-1}}\}_{i=1}^{\infty}$  be a set of distributions corresponding to the stochastic method  $\tilde{\mathcal{B}}$  under consideration. Let

$$\Delta^{\infty} = \prod_{i=1}^{\infty} \Delta = \{t^{\infty} = (t_1, \dots, t_k, \dots), t_k \in \Delta\}$$

(the product is given the Tikhonov topology and a measure  $F_{t^{\infty}}$ , which is the product of the Lebesgue measures on the factors). By the lemma, there is a Borel function  $\bar{x}_i(\eta^{i-1}, t_i): \Delta^{\infty} \times J^{i-1} \rightarrow G$  such that, for each  $\eta^{i-1}$ , the distribution of its values as functions of  $t^{\infty}$  on  $\Delta^{\infty}$  (the distribution being induced by the measure  $F_{t^{\infty}}$ ) is  $\Phi_{x|\eta^{i-1}}$ .

#### Proof of Theorem 1.3.4

We consider the mixture of methods  $\mathcal{B}^{\infty} = \{\bar{x}_i(\eta^{i-1}, t_i)\}_{i=1}^{\infty}$  and the method  $\tilde{\mathcal{B}} = \int_{\Delta^{\infty}} \mathcal{B}^{\infty} dF_{t^{\infty}}$ . From the definition of the distributions  $\Phi_{x|\eta^{i-1}}(\tilde{\mathcal{B}}, t)$  and the definition of  $\bar{x}_i(\eta^{i-1}, t_i)$  it can immediately be deduced (we shall not do this here) by induction over  $i$  that, for every  $f \in \mathcal{A}$  and every  $\omega^{\infty}, i$ , the distribution of the initial fragment of length  $i$  of the trajectory of  $\tilde{\mathcal{B}}$  on  $f$  with oracle noise  $\omega^{\infty}$  is precisely  $\Phi_{x^i|\omega^{\infty}}(\tilde{\mathcal{B}}, f)$ . Therefore the distribution of the trajectories of  $\tilde{\mathcal{B}}$  on  $f$  with oracle noise  $\omega^{\infty}$  is  $\Phi_{x^{\infty}|\omega^{\infty}}(\tilde{\mathcal{B}}, f)$ , and so  $\tilde{\mathcal{B}}$  is equivalent to  $\mathcal{B}$  in the sense required by Theorem 3.4.

Now let  $t^{\infty}(t)$  be a Borel function on  $\Delta$  with values in  $\Delta^{\infty}$ , the distribution of values of which (i.e. the distribution induced by the Lebesgue measure on  $\Delta$ ) is  $F_{t^{\infty}}$  (such a function exists by virtue of the lemma). It is clear that the mixture  $\tilde{\mathcal{B}}$  is equivalent to the mixture  $\mathcal{B}' = \int_0^1 \mathcal{B}' dt$  of methods  $\mathcal{B}' = \{\bar{x}_i(\eta^{i-1}, (t^{\infty}(t)))\}_{i=1}^{\infty}$ , and so  $\tilde{\mathcal{B}}$  is also equivalent to  $\mathcal{B}'$ . This proves Theorem 1.3.4.

#### Proof of the lemma.

Let  $X = \cup_{i=1}^{\infty} X_i^1$  be a partition of  $X$  into disjoint Borel sets of diameter  $\leq \frac{1}{2}$ , and let  $X_{i_1, \dots, i_k}^k = \cup_{i=k+1}^{\infty} X_{i_1, \dots, i_k, i}^{k+1}$  be a partition of  $X_{i_1, \dots, i_k}^k$  into disjoint Borel sets of diameter  $\leq \frac{1}{2^{k+1}}$ ,  $k = 1, 2, \dots$ . When  $X_{i_1, \dots, i_k}^k$  is not empty, let  $\xi_{i_1, \dots, i_k}^k$  be an arbitrary point of  $X_{i_1, \dots, i_k}^k$ . For an empty  $X_{i_1, \dots, i_k}^k$  the point  $\xi_{i_1, \dots, i_k}^k = \xi$ , where  $\xi$  is a fixed point of  $X$ . We write

$$t_{i_1, \dots, i_k}^k(y) = \Phi_{x|y}\{X_{i_1, \dots, i_k}^k\}.$$

Then  $t_{i_1, \dots, i_k}^k(y)$  is a non-negative Borel function of  $y$ .

We define the function

$$T_{i_1, \dots, i_k}^k(y), \quad i_j \geq 1, \quad j < k, i_k \geq 0,$$

in the following way. When  $k = 1$

$$T_0^1(y) = 0, \quad T_{i_1}^1(y) = \sum_{j=1}^{i_1} t_j^1(y);$$

suppose, further, that  $T_{i_1, \dots, i_k}^k(y)$  has already been defined. Then

$$T_{i_1, \dots, i_k, 0}^{k+1}(y) \equiv T_{i_1, \dots, i_{k-1}}^k(y),$$

$$T_{i_1, \dots, i_k, i}^{k+1}(y) = \sum_{j=1}^i t_{i_1, \dots, i_k, j}^{k+1}(y) + T_{i_1, \dots, i_{k-1}}^k(y).$$

It is clear that  $T_{i_1, \dots, i_k}^k(y)$  is a Borel function of  $y$ , and that  $T_{i_1, \dots, i_{k-1}, i}^{k+1}$  increases as  $i$  increases, and that

$$\lim_{i \rightarrow \infty} T_{i_1, \dots, i_{k-1}, i}^{k+1}(y) = T_{i_1, \dots, i_{k-1}}^{k+1}(y), \quad \lim_{i \rightarrow \infty} T_i^1(y) = 1.$$

All these relations follow from the obvious equality

$$\begin{aligned} T_{i_1, \dots, i_k, i}^{k+1}(y) &= \Phi_{x|y} \left\{ \left( \bigcup_{1 \leq j \leq i} X_{i_1, \dots, i_k, j}^{k+1} \right) \right. \\ &\quad \left. \cup \left( \bigcup_{1 \leq j \leq i_k-1} X_{i_1, \dots, i_{k-1}, j}^k \right) \cup \dots \cup \left( \bigcup_{1 \leq j \leq i_1-1} X_j^1 \right) \right\}. \end{aligned}$$

We define the functions

$$\varphi^k(s, y) = \begin{cases} \xi, & \text{if } s = 0, s = 1 \text{ or } s = T_{i_1, \dots, i_k}^k(y) \\ & \text{for some } i_1, \dots, i_k; \\ \xi_{i_1, \dots, i_k}^k, & \text{if } T_{i_1, \dots, i_{k-1}}^k(y) < s < T_{i_1, \dots, i_k}^k(y) \\ & \text{for some } i_1, \dots, i_k. \end{cases}^*$$

[\* Note that if such a set exists for a given  $s$ , then it is unique.]

Let us verify that  $\varphi^k(s, y)$  are Borel with respect to  $s, y$ . To do this, let  $s_1, s_2$  be real variables, and let

$$\theta_1(s_1, s_2, y) = \begin{cases} 0, & s_1 \neq s_2, \\ 1, & s_1 = s_2, \end{cases} \quad \text{and} \quad \theta_2(s_1, s_2, y) = \begin{cases} 0, & s_1 < s_2, \\ 1, & s_1 \geq s_2. \end{cases}$$

It is clear that  $\theta_1$  and  $\theta_2$  are Borel with respect to  $s, y$ . But then for a scalar Borel function  $T(y)$ , its *graph*, i.e. the set  $\{(s, y) | s = T(y)\}$  and its *epigraph*, i.e. the set  $\{(s, y) | s \geq T(y)\}$ , are both Borel, because the characteristic functions of these sets are  $\theta_1(T(y), s, y)$  and  $\theta_2(T(y), s, y)$ , and these functions are Borel with respect to  $s, y$  by virtue of the model argument in Section A.1.2.2. It is now clear that  $\varphi^k(s, y)$  is a Borel function.

We next verify that, for each fixed  $y = \bar{y}$ , when  $k \rightarrow \infty$  the function  $\varphi^k(s, y)$  converges (indeed, uniformly outside a certain countable set) to a function  $\varphi(s, y)$ . For, let  $L(\bar{y})$  be a countable set consisting of the ends of the interval  $\Delta$  and the values of all the functions  $T_{i_1, \dots, i_k}^k$  at the point  $\bar{y}$ , and let  $\Delta_{i_1, \dots, i_k}^k(\bar{y})$ ,  $i_1, \dots, i_k \geq 1$ , be the set  $\{s | T_{i_1, \dots, i_{k-1}, i_k-1}^k(\bar{y}) < s < T_{i_1, \dots, i_k}^k(\bar{y})\}$ . Clearly,

$$\Delta_{i_1, \dots, i_k}^k(\bar{y}) \supset \bigcup_{i \geq 1} \Delta_{i_1, \dots, i_k, i}^{k+1}(\bar{y});$$

and

$$\varphi^k(\Delta_{i_1, \dots, i_k}^k(\bar{y}), \bar{y}) \subset X_{i_1, \dots, i_k}^k.$$

Therefore on

$$\Delta \setminus L(\bar{y}) \subset \bigcup_{i_1, \dots, i_k \geq 1} \Delta_{i_1, \dots, i_k}^k(\bar{y})$$

(for every  $k \geq 1$ )  $\varphi^k(s, y)$  converges uniformly on  $s$ , and

$$\rho_x(\varphi^k(s, \bar{y}), \varphi(s, \bar{y})) \leq 2^{-k+1}$$

(because the diameter of  $X_{i_1, \dots, i_k}^k$  does not exceed  $2^{-k}$ ). On  $L(\bar{y})$  the sequence  $\{\varphi^k(s, \bar{y})\}_{k=1}^\infty$  is stabilised at every point, and so  $\varphi^k(s, y)$  certainly converges as indicated.

The function  $\varphi(s, y)$  is Borel (being the limit of Borel functions). We show that the distribution  $\Phi_{x|\bar{y}}$  of its values as a function of  $s \in \Delta$  is  $\Phi_{x|\bar{y}}$ . In view of the regularity of the measures  $\Phi_{x|y}$ ,  $\Phi_{x|\bar{y}}$ , it suffices to prove that if  $K \subset X$  is a compact set and if  $U$ , containing  $K$ , is an open set, then

$$\alpha \equiv \Phi_{x|\bar{y}}(K) \leq \Phi_{x|\bar{y}}(U)$$

(since both measures are probability measures, it follows from this that  $\Phi$  and  $\Phi$  coincide). We fix the  $K$ ,  $U$ , and  $\bar{y} \in Y$ , and let  $\delta > 0$  be such that the closed  $3\delta$ -neighbourhood of  $K$  lies in  $U$ . We choose  $m$  so that  $2^{-m} < \delta$ , and we consider all the sets  $X_{i_1, \dots, i_m}^m$  which intersect  $K$  (and which therefore lie in a  $\delta$ -neighbourhood of  $K$ ). The sum of the measures  $\Phi_{x|\bar{y}}$  of these sets is not less than  $\alpha$ . But each of these, let us say  $X_{i_1, \dots, i_m}^m$ , which has a non-zero measure, corresponds to a non-empty interval  $\Delta_{i_1, \dots, i_m}^m(\bar{y})$ . On  $\Delta_{i_1, \dots, i_m}^m(\bar{y})$  the function  $\varphi^m(s, \bar{y})$  takes values in  $X_{i_1, \dots, i_m}^m$ , and

$$\text{mes } \Delta_{i_1, \dots, i_m}^m(\bar{y}) = \Phi_{x|\bar{y}}(X_{i_1, \dots, i_m}^m).$$

Outside the countable set of points  $\Delta_{i_1, \dots, i_m}^m(\bar{y})$  we have

$$\rho_x(\varphi(s, \bar{y}), \varphi^m(s, \bar{y})) \leq 2^{-m+1} < 2\delta.$$

Thus, almost everywhere on  $\Delta_{i_1, \dots, i_m}^m(\bar{y})$  we have  $\varphi(s, \bar{y}) \in U$ . Therefore, putting

$\mathfrak{T} = \{i_1, \dots, i_m | X_{i_1, \dots, i_m}^m \cap K \neq \emptyset\}$ ,  
we obtain

$$\begin{aligned} \Phi_{x|\bar{y}}(U) &\geq \sum_{(i_1, \dots, i_m) \in \mathfrak{T}} \text{mes } \Delta_{i_1, \dots, i_m}^m(\bar{y}) \\ &= \sum_{(i_1, \dots, i_m) \in \mathfrak{T}} \Phi_{x|\bar{y}}(X_{i_1, \dots, i_m}^m) \geq \Phi_{x|\bar{y}}(K), \end{aligned}$$

as required, and the lemma is proved.

# Convex programming. linearly convergent methods for classes of general convex problems

In this chapter we describe the main subject of our investigation—convex programming problems. A classification of such problems is indicated. Methods are constructed for solving general (the widest possible) classes of problems of this kind; methods which have linear convergence (i.e. which converge at the rate of a geometrical progression). The rate of convergence of these methods is determined by the dimension of the space. The complexity of solving the corresponding classes of problems is realized by these methods asymptotically with respect to the accuracy (this is proved in Chapter 4). In the following chapter another family of methods will be constructed whose rate of convergence, although indeed no longer linear, does not depend explicitly on the dimension of the problem. Roughly speaking, methods of the second kind realize the complexity of solution of general classes of convex problems asymptotically with respect to the dimension. Combining these two groups of methods, in Chapter 4 we learn how to realize (more precisely, to ‘realize approximately’) the complexity of solution of general convex problems over a wide spectrum of the standard convex bodies  $G$  of arbitrary dimension and with arbitrary accuracy.

## 2.1 CONVEX SETS AND CONVEX PROBLEMS

We introduce the main classes, studied in this chapter, of convex problems and their corresponding informational mappings.

We start with some elementary facts about convex analysis, which we shall need. We shall try not only to introduce the material for reference purposes, but also to form a geometrical view of the objects in question which will be useful subsequently. Throughout,  $E$  will denote a real Banach space.

### 2.1.1

**DEFINITION.** Let  $G \subset E$ . The set  $G$  is said to be *convex* if it follows from  $x, y \in G$  and  $t \in [0, 1]$  that  $tx + (1-t)y \in G$ .

Geometrical interpretation: a set is convex if, whenever two points lie in it, the segment between them lies in it.

**Exercise 1** (on the definition). If  $G$  is convex and if  $x_1, \dots, x_s \in G$ , then every convex combination  $x = \sum_{i=1}^s a_i x_i$ ,  $a_i \geq 0$ ,  $\sum_{i=1}^s a_i = 1$  of the points  $x_i$  lies in  $G$ .

It is clear that the intersection of any family of convex sets in  $E$  is convex. The simplest example of a convex set is a subspace  $\{x \in E \mid \langle e | x - x_0 \rangle \leq 0\}$ , where  $e \in E^*$ ,  $e \neq 0$ , is a fixed element of  $E^*$ , and  $x$  is any point of  $E$ . This set is closed, as well as convex. By taking the intersection of a half-space with any family of sets, we again obtain a convex closed set. This construction enables us to obtain an arbitrary (distinct from  $E$ ), convex, closed subset of  $E$ .

### 2.1.2

**Theorem.** Let  $G \subset E$  be a convex closed set, and let  $x \in E \setminus G$ . Then there is a half-space which contains  $G$  but not  $x$ . In particular, if  $G \neq E$ , then  $G$  is the intersection of all the half-spaces which contain  $G$ .

The boundary of the half-space mentioned in the theorem is called the *hyperplane separating  $x$  and  $G$* .

The proof of this theorem (for the case where  $E$  is a Hilbert space) is given by the following useful exercises for the reader.

**Exercise 2.** Let  $E$  be a Hilbert space and  $G \subset E$  be a non-void, convex, closed set, and let  $\bar{x} \in E \setminus G$ . Let  $\{x_i\}$ ,  $x_i \in G$ , be a sequence minimizing  $\|\bar{x} - x\|$  over  $x \in G$ ;

$$\|\bar{x} - x_i\| \xrightarrow{i \rightarrow \infty} \inf_{x \in G} \|\bar{x} - x\| \equiv \rho(\bar{x}, G).$$

Prove that the sequence  $\{x_i\}$  converges to a point  $x^* \in G$ :  $\|\bar{x} - x^*\| = \rho(\bar{x}, G)$  and that the point  $x^*$  which satisfies the last relation (i.e. the point of  $G$  which is closest to  $\bar{x}$ ) is unique.

**Exercise 3.** With the conditions of Exercise 2, consider the half-space  $\{x \in E \mid \langle x - x^* | \bar{x} - x^* \rangle \leq 0\}$ . Prove that this half-space contains  $G$  but does not contain  $\bar{x}$ . Geometrically,  $G$  and  $\bar{x}$  lie on opposite sides of the hyperplane going through  $x^*$  and perpendicular to  $\bar{x} - x^*$ .

**Remark.** In the general case of a Banach space  $E$ , the Euclidean scheme of Exercises 2 and 3 does not suffice to prove Theorem 2.1.2. This theorem is equivalent to the fundamental Hahn–Banach theorem.

## 2.1.3

We now define the concept of a convex function.

**DEFINITION.** Let  $G \subset E$  be a non-void convex set, and let  $f: G \rightarrow \mathbb{R}$  be a real function on  $G$ . The function is said to be *convex on  $G$*  if, for all  $x, y \in G$ , and  $t \in [0, 1]$  the inequality

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \quad (1.1)$$

holds. An equivalent definition:  $f$  is convex on  $G$  if the restriction of  $f$  to any segment in  $G$  is a convex function on this segment.

**Exercise 4 (on the definition). Jensen's inequality.** Let  $G$  be convex, let  $f$  be convex on  $G$ , and let  $x = \sum_{i=1}^s \alpha_i x_i$ ,  $\alpha_i \geq 0$ ,  $\sum_{i=1}^s \alpha_i = 1$ , be a convex combination of points  $x_i \in G$ . Show that then

$$f(x) \leq \sum_{i=1}^s \alpha_i f(x_i).$$

By using the convexity of the function  $(-\ln t)$ ,  $t > 0$ , and Jensen's inequality, prove Cauchy's inequality

$$(t_1 \dots t_n)^{\frac{1}{n}} \leq \frac{t_1 + \dots + t_n}{n}, \quad t_i > 0.$$

**Exercise 5. Estimate of the modulus of continuity of a convex function by means of its oscillation.** Let  $f$  be defined and convex on a convex set  $G$ , let  $x, z \in G$ , and let  $y$  be a point on the segment  $[x, z]$ . Prove that then

$$|f(y) - f(x)| \leq \frac{\|x - y\|}{\|x - z\|} |f(z) - f(x)| \leq \frac{\|x - y\|}{\|x - z\|} V_G(f),$$

where

$$V_G(f) = \sup_G f - \inf_G f$$

is the variation of  $f$  on  $G$ . In particular, prove that if  $f$  is convex in a  $\rho$ -neighbourhood  $G_\rho$  of the set  $G$  and if  $V_{G_\rho}(f) < \infty$ , then  $f$  is a Lipschitz function on  $G$  with the Lipschitz constant  $(1/\rho)V_{G_\rho}(f)$ .

It is convenient to interpret the property of convexity of a function  $f$  in terms of its epigraph. We define the space  $E_+$  as  $E \times \mathbb{R}^1$ . The epigraph of a function  $f: G \rightarrow \mathbb{R}$  is a subset in  $E_+$  defined as  $\{(x, t) | x \in G, t \geq f(x)\}$ . A function  $f$  is convex if and only if its epigraph is a convex set in  $E_+$  (it is useful to think of the situation when  $E = \mathbb{R}^1$  or  $E = \mathbb{R}^2$ ).

A very simple example of a convex function on the whole of  $E$  is an *affine functional*  $f(x) = \langle x | e \rangle + c$ ,  $e \in E^*$ . For such a functional the inequality (1.1) becomes an equality. Affine functionals are the only possible continuous functions  $f$  such that both  $f$  and  $-f$  are convex on  $E$ .

Just as all convex, closed sets can be constructed from the simplest such sets (from half-spaces) by the operation of taking intersections, so all continuous

convex functions can be constructed from affine functions by the operation of taking the upper bound.

**Theorem.** Let  $f_\omega, \omega \in \Omega$ , be convex functions on a convex set  $G \subset E$ . Then the function  $\sup_\omega f_\omega(x)$  (provided that it is defined on  $G$ ) is also a convex function, and, if  $\alpha_\omega \geq 0$ , so is the function  $\sum_{\omega \in \Omega} \alpha_\omega f_\omega$  (in this formula only a finite number of the  $\alpha_\omega$  are non-zero).

The proof is left to the reader.

## 2.1.4

**Theorem.** Let  $G \subset E$  be a non-void, convex, closed set, and let  $f: G \rightarrow \mathbb{R}$  be a continuous, convex function on  $G$ . Further, let  $\text{int } G \neq \emptyset$  (or let  $E$  be finite-dimensional). Let  $A$  denote the set of all continuous, affine functions  $g$  such that  $g \leq f$  on  $G$ . Then

$$f(x) = \sup \{g(x) | g \in A\} \text{ for } x \in G. \quad (1.2)$$

This theorem is closely connected with Theorem 2.1.2. It is proved in the following series of exercises.

**Exercise 6.** (Reduction to the case  $\text{int } G \neq \emptyset$ .) Under the conditions of the theorem, let  $E$  be finite-dimensional, and let  $E^G$  be the minimal, affine subspace of  $E$  which contains  $G$ . Prove that the interior,  $\text{int}_0 G$ , relative to  $E^G$ , is not void. Would this still be true if  $E$  were not finite-dimensional?

**Exercise 7.** Under the conditions of the theorem, let  $x \in \text{int } G$ , and let  $x_+ = (x, t) \in E_+$ ,  $t < f(x)$ . Prove that the epigraph of  $f$  is a convex and closed set, and that the hyperplane separating it and  $x_+$  is the graph of a certain affine function in  $A$ . Hence deduce that (1.2) holds on  $\text{int } G$ .

**Exercise 8.** Prove that, under the conditions of the theorem, when  $\text{int } G \neq \emptyset$  the set  $\text{int } G$  is dense in  $G$ . Hence, and from the result of Exercise 7, deduce that the assertion of the theorem is true. (Hint. It is useful to remember that a convex function on an interval is semi-continuous above.)

## 2.1.5

We now define the fundamental concept of a support functional to a convex function.

**DEFINITION.** Let  $G \subset E$  be convex, and let  $f: G \rightarrow \mathbb{R}$  be a function. Any functional  $g \in E^*$  such that

$$f(x) - f(x_0) \geq \langle g | x - x_0 \rangle \text{ for } x \in G \quad (1.3)$$

is called a support functional to  $f$  on  $G$  at the point  $x_0$ .

*Geometrical interpretation:* for  $x \in G$  the graph of the affine function  $y = f(x_0) + \langle g | x - x_0 \rangle$  lies under the graph of the function  $f$  and touches the graph of  $f$  above the point  $x_0$ .

*Exercise 9 (on the definition).* Let  $x_0 \in \text{int } G$  and let  $f$  be differentiable at  $x_0$ . Prove that the support functional to  $f$  on  $G$  at  $x_0$  (if it exists) is then unique and coincides with the derivative of  $f$  at  $x_0$ .

*Exercise 10 (on the definition).* Let  $f$  be convex on a convex set  $G$  and let  $x_0 \in G$ . Let  $g$  be a support functional to  $f$  at  $x_0$  on some neighbourhood of  $x_0$  in  $G$ . Prove that  $g$  is also a support functional to  $f$  at  $x_0$  on the whole of  $G$ .

The set of support functionals to  $f$  on  $G$  at a point  $x$  is denoted by  $\partial_G f(x)$ . In the case of a Hilbert space  $E$  the elements of  $E^*$  are canonically identified with the elements of  $E$ , and so the support functionals can be regarded as vectors in  $E$ . In this situation they are called *sub-gradients*.

By (1.3) a support functional carries global information about the behaviour of  $f$  on  $G$ , information which, as we shall see, is extremely valuable for solving extremal problems. It should be emphasized that when  $f$  and  $G$  are convex, a support functional to  $f$  on  $G$  at the point  $x$  is determined by the behaviour of  $f$  in an arbitrarily small neighbourhood of  $x$  in  $G$ , or, as it is said, by the *germ* of  $f$  on  $G$  at the point  $x$  (the assertion of Exercise 10). So, in the convex case, knowledge of the local structure of  $f$  gives extremely valuable information about the *global* properties of  $f$ . This, it would seem, is what determines the objective ‘easiness’ of solving convex extremal problems as compared with multi-extremal problems.

Let us clarify whether there are sufficiently many support functionals to a convex function. This question is settled by the following theorem.

**Theorem.** Let  $G$  be convex, and let  $f: G \rightarrow \mathbb{R}$  be continuous at a point  $x_0 \in \text{int } G$  and be convex on  $G$ . Then the set  $\partial_G f(x)$  is non-empty.

*Proof.* By the assertion of Exercise 10 and the hypothesis that  $x_0 \in \text{int } G$ , we can replace  $G$  in the assertion by a ball  $V$  of sufficiently small radius with its centre at  $x_0$ . Since  $f$  is continuous at  $x_0$ , we can take  $f$  to be bounded in a  $\rho$ -neighbourhood of  $V$  (where  $\rho > 0$  is sufficiently small, just as the radius of  $V$  is). By the assertion of Exercise 5,  $f$  is continuous on  $V$ . Using Theorem 2.1.4, we can find a sequence of affine functionals  $g_i(x) = c_i + \langle g_i | x - x_0 \rangle$ ,  $g_i \in E^*$ , such that  $f(x) \geq g_i(x)$  on  $V$  and  $g_i(x_0) = c_i \xrightarrow{i \rightarrow \infty} f(x_0)$ . The linear functionals  $\langle g_i | x \rangle$  are bounded above in the ball  $V - x_0$  with centre at 0 by the same constant  $\sup_{V-x_0} f - \min_i c_i$ . Therefore their norms are uniformly bounded. If  $E$  is finite-dimensional, then  $\{g_i\}$  has a limit point  $g$  (in the general case one has to speak of the limit point in the sense of the weak topology in the dual space). It is clear that  $g$  is indeed a support functional to  $f$  on  $G$  at the point  $x_0$ .

*Exercise 11 (on the theorem).* Is the theorem still true if the hypothesis that  $f$  is convex is dropped? Or if the inclusion  $x_0 \in \text{int } G$  is dropped?

*Exercise 12 (on the theorem).* A converse of the theorem. Let  $G \subset E$  be convex, let  $\text{int } G \neq \emptyset$  and let  $f: G \rightarrow \mathbb{R}$  be a continuous function such that  $\partial_G f(x) \neq \emptyset$  for all  $x \in \text{int } G$ . Prove that  $f$  is convex on  $G$ .

*Exercise 13 (on the theorem).* Let  $G$  be convex,  $x_0 \in \text{int } G$ , and let  $f$  be convex on  $G$  and continuous at  $x_0$ . We define a function  $f'(x_0, h)$  of  $h \in E$  by the relation

$$f'(x_0, h) = \lim_{t \rightarrow +0} \frac{f(x_0 + th) - f(x_0)}{t}.$$

Prove that  $f'(x_0, h)$  is well defined, continuous, and convex, relative to  $h$ , and that  $f'(x_0 + \lambda h) = \lambda f'(x_0, h)$ ,  $\lambda \geq 0$ .

Prove that the set of support functionals to  $f'(x_0, h)$  as a function of  $h$  at the point  $h = 0$  coincides with  $\partial_G f(x_0)$ . Prove also that if  $g(h) = a + \langle g | h \rangle$  is an affine function such that  $g \in E^*$  and  $f'(x_0, h) \geq g(h)$ ,  $h \in E$ , then  $g$  is a support functional to  $f'(x_0, h)$  at the point  $h = 0$ . Hence and from Theorem 2.1.4 deduce that  $\partial_G f(x_0)$  is not empty (thus obtaining a new proof of Theorem 2.1.5).

*Remark.* A stronger result than the assertion of Theorem 2.1.5 can be deduced from the Hahn–Banach theorem; viz., under the conditions of Exercise 13, for every  $h \in E$  there is a  $g \in \partial_G f(x)$  such that  $\langle g | h \rangle = f'(x_0, h)$ . Thus  $\partial_G f(x)$  is adequately abundant: for every direction  $h$  there is a support functional to  $f$  at  $x_0$  with the same derivative as  $f$  in the direction  $h$ .

*Exercise 14 (on the theorem).* Let  $G$  be convex, let  $f$  be continuous and convex on  $G$ , and let  $\text{int } G \neq \emptyset$ . Prove that  $f$  satisfies a Lipschitz condition on  $G$  with a constant  $L$  if and only if, for every  $x_0 \in \text{int } G$  and  $g \in \partial_G f(x_0)$ , we have  $\|g\|_* \leq L$ .

*Exercise 15 (on the theorem).* Let  $G$  be convex, let  $f$  be a convex function satisfying a Lipschitz condition with constant  $L$  on  $G$ . Prove that  $\partial_G f(x) \neq \emptyset$  for all  $x \in G$ . Further, for every  $x \in G$ , a  $g \in \partial_G f(x_0)$  can be found such that  $\|g\|_* \leq L$ . (A reader who is unacquainted with such results in functional analysis as the Hahn–Banach theorem and the compactness of a ball in  $E^*$  with a suitable topology may restrict himself to the case of a Euclidean finite-dimensional space  $E$ .)

## 2.1.6

We now define the concept of strict convexity. For our purposes, the strongest definition is sufficient, and we need only consider the case of a Hilbert space.

**DEFINITION.** Let  $E$  be a Hilbert space,  $G \subset E$  a non-void, convex, closed set, and  $\alpha \geq 0$ . A function  $f: G \rightarrow \mathbb{R}$  is said to be  $\alpha$ -strictly convex on  $G$  if, for  $x, y \in G$ ,

$$2f\left(\frac{x+y}{2}\right) < -\frac{\alpha}{4}\|x-y\|^2 + f(x) + f(y)$$

and  $f$  is continuous on  $G$ .

*Remark.* 0-strictly convex functions on  $G$  are precisely continuous convex functions on  $G$ .

In Exercises 16–21,  $E$  is a Hilbert space, and  $G \subset E$  is convex, closed, and non-void.

**Exercise 16.** Let  $f_\omega$ ,  $\omega \in \Omega$ , be  $\alpha$ -strictly convex functions on  $G$ , and let  $f(x) = \sup_\omega f_\omega(x)$  be defined and continuous on  $G$ . Prove that  $f$  is  $\alpha$ -strictly convex on  $G$ .

**Exercise 17.** Let  $f, g$  be respectively  $\alpha$ -strictly and  $\beta$ -strictly convex on  $G$ . Prove that  $f+g$  is  $(\alpha+\beta)$ -strictly convex on  $G$ .

**Exercise 18.** Let  $f$  be  $\alpha$ -strictly convex on  $G$  and let  $0 \leq \beta \leq \alpha$ . Prove that the function  $\beta x^2/2$  is  $\beta$ -strictly convex on  $E$ , and  $f$  is  $(\beta x^2/2)$ -convex on  $G$ .

**Exercise 19.** Let  $f$  be  $\alpha$ -strictly convex on  $G$ , let  $x_0 \in G$ , and let  $g \in \partial_G f(x_0)$ . Prove that

$$f(y) \geq f(x_0) + \langle g | y - x_0 \rangle + \frac{\alpha}{2}(y - x_0)^2$$

when  $y \in G$ .

**Exercise 20.** Let  $f$  be  $\alpha$ -strictly convex on  $G$ ,  $\alpha > 0$ . Prove that  $f(x) \rightarrow +\infty$  when  $\|x\| \rightarrow \infty$ ,  $x \in G$ . Prove that there is a point  $x^*$  giving the minimum of  $f$  on  $G$  and that it is unique, and that the inequality

$$\|x - x^*\|^2 \leq \frac{4}{\alpha} (f(x) - f(x^*))$$

holds.

**Exercise 21.** Let  $f: G \rightarrow \mathbb{R}$  be continuous, and let  $\text{int } G \neq \emptyset$ . Let  $f$  have a second derivative  $\langle f''(x)h | h \rangle$  on  $\text{int } G$ . Prove that the following three statements are equivalent:

1.  $f$  is  $\alpha$ -strictly convex on  $G$ ;
2.  $\forall x, y \in \text{int } G: \langle f'(x) - f'(y) | x - y \rangle \geq \alpha(x - y)^2$ ;
3. the quadratic form in  $h$   $\langle f''(x)h | h \rangle - \alpha \langle h | h \rangle$  is non-negative definite for  $x \in \text{int } G$ .

The assertion of Exercise 20 shows the important rôle which the assumption of strict convexity plays in extremal problems: every sequence  $\{x_i\}$  minimizing  $f$  on  $G$  converges to the point  $x^*$  of the minimum of  $f$ , and  $\|x_i - x^*\|$  is effectively bounded by  $f(x_i) - f(x^*)$ .

### 2.1.7

Convex sets can be specified by means of convex functions. Let  $f$  be a continuous, convex function on  $E$ . Then, for any real  $a$ , the set  $\{x \in E | f(x) \leq a\}$  is convex and closed (check!). Conversely, if  $G \subset E$  is non-void, convex, and closed, then it can be expressed in the way mentioned, namely, as  $\{x \in E | \rho_{\parallel, \parallel}(x, G) \leq 0\}$ , where  $\rho_{\parallel, \parallel}(x, G) = \inf\{\|x - y\| | y \in G\}$  is the distance from  $x$  to  $G$ .

**Exercise 22.** Verify that, in the situation described above,  $g(x) \equiv \rho_{\parallel, \parallel}(x, G)$  is a continuous, convex function with Lipschitz constant 1.

**Exercise 23.** Show (at least in the case of a finite-dimensional  $E$ , although the assertion is always true) that  $\partial_G g(x)$  when  $x \notin G$  consists of functionals with norm 1. If  $\text{int } G \neq \emptyset$ , then also, when  $x \in G \setminus \text{int } G$ ,  $\partial_G g(x)$  contains functionals with norm 1.

Another way of specifying convex sets is to use the *Minkowski function*  $p(x)$ . Let  $G$  be convex and closed, and let  $x_0 \in \text{int } G$ . We define the function  $p(x)$  as  $\inf\{\lambda | x_0 + \lambda^{-1}(x - x_0) \in G\}$ . Geometrically,  $p(x)$  is the ratio of the segment from the point  $x$  to the point  $x_0$  to the segment from  $x_0$  to the intersection of the ray in the direction  $x - x_0$  with the boundary of  $G$ .

**Exercise 24.** Prove that, in the situation described above,  $p(x)$  is a continuous, convex function which is homogeneous to the first degree relative to the point  $x$  (i.e.  $p(\lambda(x - x_0) + x_0) = \lambda p(x)$ ,  $\lambda \geq 0$ ). Prove that  $G$  can be described as the set  $\{x | p(x) \leq 1\}$ .

### 2.1.8

We are now in a position to describe the main object of our investigation—a convex extremal problem.

**DEFINITION.** A problem

$$f_0(x) \rightarrow \min | x \in G \subset E, f_i(x) \leq 0, i = 1, \dots, m, \quad (1.4)$$

is said to be *convex* if  $G$  is a convex, closed, non-void subset of  $E$  and if the  $f_i$ ,  $i = 0, 1, \dots, m$ , are continuous, convex functions on  $G$ . Compatible convex problems are, as a rule, soluble.

**Theorem.** Let (1.4) be a convex problem, and compatible, let  $E$  be reflexive, and let the set

$$G_a = \{x \in G \mid f_0(x) \leq a, f_i(x) \leq 0, 1 \leq i \leq m\}$$

be non-void and bounded for some  $a = a_0$ . Then the problem (1.4) is soluble.

*Proof.* Let

$$f_{0*} = \inf \{f_0(x) \mid x \in G, f_i(x) \leq 0, i = 1, \dots, m\} \subset R \cup \{-\infty\}$$

be the optimal value of the objective functional in (1.4). We consider the sets  $G_a \neq \emptyset, f_{0*} < a \leq a_0$ . It is clear that they are convex and closed, and  $G_a \subset G_{a'}$  when  $a \leq a'$ . In addition the set  $G_{a_0}$  is bounded, by hypothesis. Since  $E$  is reflexive,  $G_{a_0}$  is compact in the weak topology of  $E$ , and  $G_a, f_{0*} < a \leq a_0$  is a centred family of its closed (in the weak topology) subsets. Therefore  $\cap_{a > f_{0*}} G_a \equiv G_* \neq \emptyset$ . Obviously,  $G_*$  is precisely the set of solutions of the problem (1.4).

*Exercise 25 (on the theorem).* Is the assertion still true if the boundedness condition on  $G_{a_0}$  is removed? Is it true if this condition is replaced, in the case of a Hilbert space  $E$ , by the requirement that the function  $\sum_{i=1}^m f_i$  be  $\alpha$ -strictly convex ( $\alpha > 0$ ) on  $G$ ? (In the latter case, the assertion of Exercise 19, Section 2.1.6, has to be used.)

In most applications  $E$  is reflexive (or even a Hilbert space) and either  $G$  is bounded or  $\sum_{i=0}^m f_i$  is  $\alpha$ -strictly convex on  $G$  with  $\alpha > 0$ , and so compatibility of a problem under consideration immediately implies its solubility.

## 2.2 CLASSES OF CONVEX EXTREMAL PROBLEMS

### 2.2.1

In accordance with the ideas described in Chapter 1, a class of extremal problems is specified by fixing three objects: a field of problems, an oracle, and a normalizing mapping. Actually, of these three objects only the first two are essential. The normalizing mapping is chosen relative to these two from considerations of mathematical naturalness and convenience. Thus, a classification of the fundamental objects of our investigation, i.e. of classes of convex problems, presupposes a classification both of sets of such problems (or, if preferred, of the problems themselves) and of the oracles used.

As regards the oracles, here we follow the customary classification of them into *oracles of zeroth order* (providing only the values of the functionals in the problem), *oracles of first order* (providing both the values of the functionals and the values of their first derivatives), and *oracles of higher order*. Additionally, we distinguish oracles of any of these types as being either *deterministic* or *stochastic*.

The customary view is that methods using oracles of order higher than the first cannot be widely used (at any rate, in solving problems of appreciable dimension), mainly because of difficulties associated with the construction of such oracles. On the theoretical side, study of methods of this kind is not particularly interesting, either. For these reasons we shall not investigate these methods at all thoroughly; a few remarks about them can be found in Section 8.1.1. *Methods of zeroth order* (using oracles of zeroth order) are undoubtedly of interest, because it is just these oracles which can be most easily constructed in practice. Some results about such methods are contained in Chapter 9; from them it follows that the possibilities of such methods are unfortunately rather limited.

From what we have said, the most interesting methods are those of the first order, and it is mainly with these that we shall be concerned. We point out that it is precisely these methods which occupy a central place in works on numerical optimization.

### 2.2.2

We now outline the classification of convex extremal problems which we shall adhere to. Any classification is based on the picking out and subsequent fixing of certain characteristics (the parameters) of a problem. Such characteristics are, in particular,

- (i) the *dimension* of the problem,  $n$ , i.e. the linear dimension of the space  $E$ ,  $n = 1, 2, \dots, \infty$ ;
- (ii) the *nature of the constraints* of the problem, according to which problems are classified into:
  - unconditional problems* ( $m = 0, G = E$ )
  - unconstrained problems* ( $m = 0, G$  arbitrary)
  - constrained problems* ( $m$  arbitrary,  $G$  arbitrary);
- (iii) the *smoothness properties* of the problem (i.e. the smoothness properties of the functions  $f_j$ ).

In most cases (and everywhere in this chapter) the basis of the classification will be smoothness properties of the problem. The other characteristics will determine the more precise classification of problems.

### 2.2.3

Let us carry out the classification of convex problems of the type (1.4) according to the degree of smoothness of the functions appearing in them. Let  $E, G \subset E$ , and  $m \geq 0$  be fixed,  $G$  being convex and closed. Then a problem (1.4) can be identified with a continuous  $(m+1)$ -dimensional vector-function  $f = (f_0, \dots, f_m)$  on  $G$ , whose components are each convex.

The minimal natural restriction (at least, for bounded  $G$ ) regarding the smoothness of  $f$  is the requirement of continuity (and this we now demand once and for all for all the problems considered) and the boundedness of  $f$  on  $G$ . The latter is equivalent to requiring that all the quantities

$$\mathbf{V}_0(f_i) \equiv \mathbf{V}_i(f) = \sup_{x \in G} f_i - \inf_{x \in G} f_i$$

shall be finite.

Accordingly we denote by  $C(G, E, m)$  the field of all convex problems of the form (1.4) with given  $G, E, m$  and finite  $\mathbf{V}_i(f), 0 \leq i \leq m$ . We equip this set with the normalizing mapping

$$r_i(f) = \begin{cases} \mathbf{V}_0(f), & i = 0, \\ \max \{0, \sup_G f_i\}, & i > 0. \end{cases}$$

The class of problems obtained by providing the field  $C(G, E, m)$  with an oracle  $\emptyset$  will be denoted by  $C^0(G, E, m)$  and called the *class of general convex problems*.

The norming which we have chosen for the error has a simple meaning. Suppose  $f \in C^0(G, E, m)$  is a compatible problem and we hand out a solution of it *a priori*—‘without looking’. Then the maximal possible absolute error of such a trivial method (laboriousness = 1) on  $f$  for the  $i$ th component is  $r_i(f)$ . Therefore the clause ‘a  $C^0(G, E, m)$ -method has accuracy  $v$  on  $f$ ’ means that this method is  $1/v$  times more accurate than the trivial method of search referred to (*cf.* the discussion in Section 1.2.3.4).

Additional restrictions on the moduli of continuity of the components  $f_i$  of a problem  $f$  can also be imposed. The most natural restriction is to require that the  $f_i$  be Lipschitz on  $G$ ; in other words, we demand that the quantities

$$L_{\|\cdot\|}(f_i) = \sup_{x \neq y, x, y \in G} \frac{|f_i(x) - f_i(y)|}{\|x - y\|}$$

be finite, where  $\|\cdot\|$  is the chosen norm in  $E$ . Correspondingly, we denote by  $C_{Lip} = C_{Lip}(G, E, \|\cdot\|, m)$  the field of all convex problems  $f$  of the form (1.4) with the given  $G, E, m$  and finite  $L_{\|\cdot\|}(f_i), 0 \leq i \leq m$ . (*Note:* When considering Lipschitz-convex problems we shall always suppose  $G$  to be bounded.) We provide this set of problems with the normalizing mapping

$$r_j(f) = \begin{cases} 2\rho_{\|\cdot\|}(G)L_{\|\cdot\|}(f_0), & j = 0, \\ \max \{0, \min_x f_j(x) + 2\rho_{\|\cdot\|}(G)L_{\|\cdot\|}(f_j)\}, & j > 0, \end{cases}$$

where  $\rho_{\|\cdot\|}(G)$  is the  $\|\cdot\|$ -radius of  $G$  (for compatible problems it is clear that  $r_j(f) = 2\rho_{\|\cdot\|}(G)L_{\|\cdot\|}(f_j)$  for all  $j$ ). The field of problems  $C_{Lip}(G, E, \|\cdot\|, m)$  provided with the oracle  $\emptyset$  and the indicated normalizing mapping will be denoted by  $C_{Lip}^0(G, E, \|\cdot\|, m)$  and called the *class of Lipschitz-convex problems*.

Clearly,  $r_j(f)$  are natural upper bounds for the quantities  $\sup_G f_0 - \inf_G f_0$  (for  $f = 0$ ) and  $[\max_x f_j(x)]_+, j > 0$ , i.e. of the normalizing factors of the problem  $f$ , regarded as an element of  $C(G, E, m)$ . So the interpretation of the meaning of the normalization is the same as before.

## 2.2.4

The restrictions on continuity mentioned relate only to the moduli of continuity of the  $f_i$ . It is possible to impose requirements such as continuous differentiability a specified number of times with restrictions on the modulus of continuity of the highest derivatives, and also combinations of these conditions with others such as strict convexity. The classes of smooth convex problems which thus arise are studied in Chapter 7. But in the next three chapters we focus our attention on classes of general (and Lipschitz) convex problems solved by means of an exact oracle.

We make a remark which is important for the whole of the sequel. The classification mentioned of problems according to their ‘smoothness’ is justified not by the fact that in practice we often encounter convex problems which are not Lipschitz, or Lipschitz-convex problems which are differentiable only a few times. Such a justification would be, so to say, the ‘pure water of demagogery’—rank deception by deliberate distortion of the facts. For the mere fact of a high degree of smoothness in a practical problem does not, as a rule, arouse any doubts. What must be borne in mind, however, is that smoothness does not, in itself, count for much; what is important is the values of the numerical parameters which characterize this smoothness (the values of the corresponding derivatives, and so on).

With the normalization of errors which is ‘natural’ for the given class of problems these parameters are taken into account explicitly (*cf.* the description given above of the normalizing factors for the classes of general and Lipschitz-convex problems). Moreover, sometimes these parameters directly affect the complexity or the class, or they may even have to be estimated *a priori*—without this it may be difficult, or even quite impossible, to solve the corresponding problems (see, for example, the situation in Chapter 7). Thus the numerical values of the parameters characterizing the smoothness of a problem do, in one way or another, affect the complexity of solving it. In this connection it is useful to have at our disposition methods which are sensitive only to the coarsest and most easily estimated parameters of the smoothness of a problem. Our classification of problems is aimed precisely at singling out the most natural and ‘not too delicate’ properties of the situation, and at constructing methods which are not too sensitive to the numerical characteristics of the other properties of the problem. Of course, the choice of what aspects of a situation are ‘vital’ is by no means unique, and so our subsequent considerations by no means exhaust the whole group of questions arising here.

We hope, however, that what follows will convince the reader of the reasonableness (though not, we repeat, of the completeness) of the chosen approach to the classification of convex extremal problems.

### 2.2.5

We now describe the oracles which we shall consider in studying general (or Lipschitz) convex problems. They will be deterministic oracles of the first order, providing the values and the support functionals to the components of the problem at the point under consideration. We shall not demand absolutely exact answers, but allow the oracle to make mistakes (but we shall demand that these mistakes shall not exceed the specified accuracy of the solution).

In order to demonstrate the naturalness of the formal definitions given below we start from the case of an exact oracle of the first order. In answer to a question about a problem  $f \in C(G, E, m)$  at a point  $x$  it must provide the values of  $f_i$  at the point  $x$  and also the support functionals  $\varphi_{i,x} \in \partial_G f_i(x)$  to them at  $x$  (on  $G$ ). If  $x$  is not an interior point of  $G$ , then  $\partial_G f_i(x)$  may be empty. Therefore, in considering classes of general complex problems, it is natural to restrict our attention to the case  $\text{int } G \neq \emptyset$  (in the finite-dimensional case this restriction is unimportant, because  $E$  can be replaced by the least affine subspace of  $E$  which contains  $G$ ), and we can suppose that the domain of questions of the oracle  $\mathcal{O}$  is  $\text{int } G$  and not the whole of  $G$ .

Let us make an aside. It is clear that the convention just made does not prevent us from using the whole set of definitions in Chapter 1, where the domain of questions to the oracle was always the whole domain  $G$ . For we may suppose that questions can be given on the boundary of  $G$  as well, but that the answer to such a question is a point fixed once and for all in the information space, this answer containing, in fact, no information whatsoever. We now return to the main stream of discourse.

Further, the oracle's answer about  $f_i$  at a point  $x$  can be regarded as an affine functional

$$\bar{g}_{i,x}(y) = f_i(x) + \langle \varphi_{i,x} | y - x \rangle, \quad \varphi_{i,x} \in \partial_G f_i(x). \quad (2.1)$$

Here  $\bar{g}_{i,x}(x) = f_i(x)$  and  $\bar{g}_{i,x}(y) \leq f_i(y)$  for all  $y \in G$ . It is now clear what is meant by 'the oracle makes a mistake by an amount  $\leq v_0$ '; this means that the oracle gives, not  $\bar{g}_{i,x}(x)$  but an affine functional  $g_{i,x}(x)$   $v_0$ -close to  $\bar{g}_{i,x}(x)$ . It turns out that it is sensible to measure 'closeness' by the maximal divergence of  $\bar{g}_{i,x}(y)$  and  $g_{i,x}(y)$  over  $y \in G$  (the closeness must, of course, be normalized by division by  $r_i(f)$ ). If the calculated divergence of  $\bar{g}_{i,x}$  and  $g_{i,x}$  obtained in this way is not greater than  $v_0$ , then changing from  $g_{i,x}$  to the functional  $\tilde{g}_{i,x}(y) = g_{i,x}(y) - v_0 r_i(f)$  will ensure that the inequality

$$f_i(y) \geq \tilde{g}_{i,x}(y) \quad (2.2)$$

holds if  $g_{i,x}(x)$  is close, as before, to  $f_i(x)$ . For technical reasons it is convenient to suppose from the start that the oracle's answers satisfy (2.2).

We pass on to the following definition, in which  $v_0 \geq 0$  is a parameter (the accuracy of the oracle).

**DEFINITION.** By an *oracle of type*  $(\text{int } G, v_0)$  for a field of problems  $C(G, E, m)$  we mean any deterministic oracle with  $\text{int } G$  as its domain of questions, and whose answers about a problem  $f$  at points  $x \in \text{int } G$  are sets of  $m+1$  affine functionals  $g_{i,x}(y)$ ,  $0 \leq i \leq m$ , which, for all  $f \in C(G, E, m)$  and for all  $x \in \text{int } G$ , have the following properties:

$$(1) \begin{cases} g_{i,x}(y) \leq f_i(y), & 0 \leq i \leq m, y \in G, \\ g_{i,x}(x) \geq f_i(x) - v_0 r_i(f), & 0 \leq i \leq m. \end{cases}$$

Similar considerations lead to the definition of an oracle for a field of problems  $C_{\text{Lip}}(G, E, \|\cdot\|, m)$ .

**DEFINITION.** By an *oracle of type*  $(G, v_0)$  for a field of problems  $C_{\text{Lip}} = C_{\text{Lip}}(G, E, \|\cdot\|, m)$  we mean any deterministic oracle with  $G$  as its domain of questions, and whose answers about a problem  $f$  at points  $x \in G$  are sets of affine functionals  $g_{i,x}(y)$  which, for all  $f \in C_{\text{Lip}} = C_{\text{Lip}}(G, E, \|\cdot\|, m)$  and for all  $x \in G$ , have the following properties:

$$(1) \begin{cases} g_{i,x}(y) \leq f_i(y), & 0 \leq i \leq m, y \in G, \\ g_{i,x}(x) \geq f_i(x) - 2v_0 \rho_{\|\cdot\|}(G) L_{\|\cdot\|}(f_i), & 0 \leq i \leq m; \end{cases}$$

$$(2) g_{i,x}(y) \text{ is Lipschitz with respect to } y \text{ with the constant } L_{\|\cdot\|}(f_i).$$

The second definition differs from the first in that, firstly, it is possible to ask questions over the whole of  $G$  and not just on  $\text{int } G$ . Further, we require that the norm of the support functional to  $f_i$  given by the oracle shall not exceed  $L_{\|\cdot\|}(f_i)$  (such a functional exists at each point of  $G$ , see Exercise 15 in Section 2.1).

The classes  $C^0(G, E, m)$  obtained by providing  $C(G, E, m)$  with an  $(\text{int } G, v_0)$ -oracle will be called *classes of type*  $C^0(G, E, m)$ . Similarly we define classes of type  $C_{\text{Lip}}^{v_0}(G, E, \|\cdot\|, m)$ .

In this chapter and the next we shall study deterministic methods of solving problems of these classes. It is important to notice that the methods proposed will not require *a priori* knowledge of the oracle  $\mathcal{O}$ . It is sufficient for them to know the type of the class in question (i.e. the list of the objects  $v_0, G, E, m$  for the first type, and  $v_0, G, E, \|\cdot\|, m$  for the second. The clause 'the class  $C$  is of the type  $C^0(G, E, m)$  resp.  $C_{\text{Lip}}^{v_0}(G, E, \|\cdot\|, m)$ ' will be written briefly as  $C \in C^0(G, E, m)$  resp.  $C \in C_{\text{Lip}}^{v_0}(G, E, \|\cdot\|, m)$ .

*Exercise 1.* Let  $G$  be a bounded, convex, closed set in  $E$ . Prove that every method of solving problems of a class of type  $C^0(G, E, m)$  naturally induces a

method of solving problems of the class of type  $C_{L_p}^0(G, E, \|\cdot\|, m)$  with the same (or better) characteristics of laboriousness and error on the class.

### 2.2.6

In conclusion we discuss the question of the appropriateness of referring a given extremal problem to one class or another of such problems. We point out that one and the same problem can be written in the form (1.4) by many equivalent methods. Suppose, for example, we have a problem written in the form (1.4) with the  $f_i$  continuous and convex on the whole of  $E$ . By introducing  $f_{m+1}(x) = \rho_{\parallel, i}(x, G)$  we can write down an equivalent problem but now without strict constraints:

$$f_0(x) \rightarrow \min |f_i(x)| \leq 0, \quad 1 \leq i \leq m+1. \quad (1.4.1)$$

In the new formulation  $G = E$ .

Other examples of equivalent formulations:

$$f_0(x) \rightarrow \min |x \in G, f^1(x)| \leq 0, \quad (f^1(x) = \max_{1 \leq j \leq m} f_j(x)), \quad (1.4.2)$$

$$f_0(x) \rightarrow \min |\tilde{f}(x)| \leq 0, \quad (\tilde{f}(x) = \max_{1 \leq j \leq m+1} f_j(x)), \quad (1.4.3)$$

$$u \rightarrow \min |f_0(x) - u| \leq 0, \quad f_j(x) \leq 0, \quad 1 \leq j \leq m, \quad (x, u) \in E_+, \quad x \in G. \quad (1.4.4)$$

We point out that these ways of writing the problem are equivalent only in the sense that the problems obtained have a common solution (more precisely, their solutions can be converted one into another). However, in relation to the classification mentioned, and indeed even as regards the possibilities of their numerical solution, these problems are by no means equivalent. Suppose, for instance, that the original problem was in the class of type  $C^0(G, E, m)$  with a bounded  $G$ , then the problems (1.4.2), (1.4.3), and (1.4.4) in general do not belong to a class of this type. Moreover, if the original problem were smooth and we had a method of solving it using this property, then the problems (1.4.2) and (1.4.3), say, might lose smoothness, and in such formulations the available method might not be applicable.

The circumstance we have just mentioned is neither a defect of the classification, nor a defect of numerical methods. The latter operate not with the original problems nor with classes of mutually equivalent formulations of them, but with completely definite formulations (statements) of these problems. The choice of such a formulation is the business of the mathematician. Similarly, the classification introduced for convex problems, which is intended for the analysis of numerical methods of solving them, classifies not the 'content of the problems' but the formal way in which they are written.

We stress that the most important stage in solving any problem of substance by exact methods is the stage of formalization, of theoretical study of the formal problem and its reduction to a problem which is solvable by numerical methods; this is a stage of essentially human, creative activity, and it does not yield (and seemingly will not yield for a long time to come) to mere formal study and automation. At this stage the specialist operates, studying to the full measure of his capabilities the specific nature of the actual problem, and pushing forward as far as possible along the road towards its analytical solution. But numerical methods, by the very meaning of their name, must be universal, adapted not to solving concrete problems, but to solving all problems of sufficiently wide classes. The purpose of the following study of the complexity of the traditional classes of extremal problems is, if you like, to work out an objective 'scale of values' of these classes. Guided by this scale of values, we shall be able to realize what we should strive for in the first, the 'human', stage of solving problems of substance.

Let us give an example. In a large number of problems of substance of the type (1.4), the strict constraints  $x \in G$  are either entirely absent ( $G = E$ ), or they are only 'weak'—the set  $G$  is unbounded. We saw earlier that, under weak hypotheses regarding  $f_i$ , it is always possible to remove strict constraints ((1.4.1) instead of (1.4)). It turns out, however, that the complexity of the class  $C(\{t \geq 0\}, R, 0)$ , say, is infinite, and so problems of this class cannot be 'solved effectively'. Therefore, if we are speaking of general convex problems, not only should the strict constraints not be removed, but, on the contrary, if they are not there originally, we should impose them, by adding boundedness of  $G$ . To do this, one must have an *a priori* estimate of the norm of the solution (i.e. one must increase, by means of some preliminary investigation, the store of information). We shall see later that not only the boundedness properties, but even the geometry of  $G$  as well, affect essentially the complexity of the corresponding class of convex problems.

## 2.3 THE CENTRES-OF-GRAVITY METHOD FOR SOLVING GENERAL CONVEX PROBLEMS

### 2.3.1

In this and the next section we describe a group of methods of solving general convex problems, and we estimate their laboriousness. All of them are methods of the first order. The idea of these methods is very simple. Suppose there is a convex problem  $f$  of the form (1.4), and a question is asked about it at a point  $x \in G$ . Suppose also that the answer to the question contains information about the values of the components of the vector  $f(x_0)$  and about a set of support functionals  $g_i$ ,  $i = 0, 1, \dots, m$ , to the functions  $f_i$  at the point  $x_0$ . Thus we

have obtained the affine functions

$$g_{i,x_0}^f \equiv f_i(x_0) + \langle g_i | x - x_0 \rangle$$

such that  $g_{i,x_0}^f(x) \leq f_i(x)$  on  $G$ , and  $g_{i,x_0}^f(x_0) = f_i(x_0)$ . Let  $v$  denote the required relative accuracy of the solution of the problem  $f$  (which belongs to one of the classes of convex problems described in the previous section), and let  $r_i(f)$  denote the normalizing factors of this class. It is possible that  $g_{i,x_0}^f(x_0) \leq vr_{i_0}(f)$  for some  $i_0 \geq i$ . We draw through the point  $x_0$  the level hyperplane of the function  $g_{i_0,x_0}^f(y)$ , and denote it by  $\Pi_{i_0}$ . The hyperplane  $\Pi_{i_0}$  divides  $G$  into two parts  $G_+$  and  $G_-$ , on one of which, let us say  $G_+$ , we have

$$g_{i_0,x_0}^f(x) \geq g_{i_0,x_0}^f(x_0).$$

Therefore  $G_+$  automatically contains no points  $x$  capable of being approximate solutions of  $f$  with an error  $v(x,f) \leq v$ . For, when  $x \in G$ , we have

$$f_i(x) \geq g_{i,x_0}^f(x) \geq g_{i,x_0}^f(x_0) > vr_{i_0}(f).$$

Thus, in the case considered, the answer to the question enables the part  $G_+$  of the domain  $G$  to be excluded from further consideration: points of  $G_+$  are automatically not solutions of  $f$  with the required accuracy.

Now let

$$f_i(x_0) \leq vr_i(f), \quad i = 1, \dots, m.$$

Then the point  $x_0$  itself satisfies (with the required accuracy) the constraints of the problem. We consider the level hyperplane  $\Pi_0$  passing through  $x_0$  of the affine function  $g_{0,x_0}^f(x)$ . As before we denote by  $G_+$  and  $G_-$  the parts into which  $\Pi_0$  divides  $G$ , and suppose

$$g_{0,x_0}^f(x) \geq g_{0,x_0}^f(x_0) \quad \text{in } G_+.$$

For the same reasons as before, the inequality  $f_0(x) \geq f_0(x_0)$  holds on  $G_+$ . Therefore if there is contained in  $G_+$  an approximate solution  $x'$  of the problem  $f$  with accuracy  $v(x',f) \leq v$ , then

$$f_0(x_0) \leq f_0(x') \leq f_* + vr_0(f)$$

(as always,  $f_*$  is the optimal value of the objective functional of the problem  $f$ ). We recall that the conditions

$$f_i(x_0) \leq vr_i(f), \quad 1 \leq i \leq m,$$

are satisfied, by hypothesis. Thus, if  $G_+$  does contain approximate solutions of  $f$  with the required accuracy, then  $x_0$  is already one such solution. But if  $G_+$  contains no such solutions, then *a fortiori* all of them lie in  $G_-$ . In all cases the domain  $G_+$  can be excluded from further consideration.

This argument shows that, having put the question about  $f$  at a point  $x$ , it is 'always possible' to pick out a part  $G_-$  of the domain  $G$  such that, if  $f$  has an

approximate solution with accuracy  $\leq v$  in  $G$ , then  $f$  also has such a solution in  $G_-$ . (We said 'always possible', but it would be more accurate to say 'always possible if the  $r_i(f)$  are known'. Later we shall have to adopt special measures to avoid the need for this information.) Here  $G_-$  is the part of  $G$  cut out by a certain hyperplane  $\pi$  passing through  $x_0$ . This hyperplane (and therefore the domain  $G_-$  as well) is determined by the local information about  $f$  at the point  $x_0$ .

Thus the answer to the question enables the solution to be localized to the domain  $G_-$ , part of the domain  $G$ . It is now possible to take  $G_-$  as the new domain of the functions appearing in the problem, to replace  $x_0$  by a suitable point of  $G_-$ , and to repeat the procedure already described, and so on. In passing from iteration to iteration, the localization of the solution becomes smaller and smaller. Intuitively it is clear that the localization of the solution in a 'sufficiently small' domain enables the result to be obtained with the required accuracy.

This simple geometrical idea just described forms the basis of the methods proposed below for solving general convex problems.

### 2.3.2 The method of centres of gravity (MCG)

We describe a method of solving problems of a class of the  $C^\infty(G, E, m)$  type. The idea of the method goes back to [18]. The following exposition is based on the authors' paper [30]. The method is applicable in the case where  $E$  is finite-dimensional, and  $G$  is bounded, convex, and closed. These conditions are now assumed to be satisfied. Instead of  $E$  we shall write  $R^n$ . It is further assumed that  $\text{int } G \neq \emptyset$ . This latter, under the hypotheses adopted, causes no loss of generality, since it is always possible to replace  $E$  by the least affine subspace of  $E$  which contains  $G$  (cf. Exercise 6 in Section 2.1).

Methods based on the geometrical idea described above really differ only in the rule for choosing the next point (at which a question about the problem will be posed) in the domain of localization of the solution, which is known at the present moment. In choosing this rule it must be ensured that the domain of localization will decrease from step to step 'as rapidly as possible'. A suitable measure of the 'size' of the domain has, of course, to be chosen. A useful measure, for the group of questions in which we are interested, turns out to be the Lebesgue volume of a domain.

Suppose we have  $\bar{G}$  as the current domain of localization of the solution and have chosen a point  $x \in \bar{G}$  as the point to be 'interrogated' at the current step. We do not know in advance precisely which of the hyperplanes passing through  $x$  will cut out the new domain of localization of the solution. With a badly chosen  $x$  (let us say,  $x \in \partial\bar{G}$ ) and a 'bad' plane of intersection (for example, a plane touching  $\partial\bar{G}$  when  $x \in \partial\bar{G}$ ), the new domain of localization may be only a little smaller than the existing one (in the extreme case

mentioned, the new domain would be the same as the old one). Intuitively it is clear that the closer  $x$  resembles the 'centre' of  $\tilde{G}$ , the greater will be the decrease in the domain of localization at this step even taking the 'most unfavourable' of the hyperplanes passing through  $x$  as the intersecting hyperplane.

A natural candidate for the rôle of 'centre' of a convex domain  $\tilde{G}$  is its centre of gravity (i.e. the centre of gravity of a mass uniformly distributed over the volume of  $\tilde{G}$ ), in other words, the point

$$x = \int_{\tilde{G}} y \, dy / \int_{\tilde{G}} dy.$$

A convex-programming method which realizes the geometrical idea introduced above and which uses as the next interrogated point the centre of gravity of the domain of localization of the solution obtained up to this time will be called the *method of centres of gravity* (MCG).

*Exercise.* Let  $\tilde{G}$  be a convex bounded body (the term 'body' means that  $\text{int } \tilde{G} \neq \emptyset$ ) in  $R^n$ , and let  $x$  be the centre of gravity of  $\tilde{G}$ . Prove that:

- (i) The definition of  $x$  does not depend on the choice of the Euclidean structure of  $R^n$  (in other words, the centre of gravity is an affine invariant of  $\tilde{G}$ : if  $A: R^n \rightarrow R^n$  is the inverse affine transformation, and if  $\tilde{x}$  is the centre of gravity of  $A\tilde{G}$ , then  $\tilde{x} = Ax$ ). Hence deduce that if  $\Delta$  is a simplex in  $R^n$  with vertices  $x_0, \dots, x_n$ , then the centre of gravity of  $\Delta$  is

$$x = (x_0 + \dots + x_n)/(n+1).$$

- (ii) If  $f$  is an affine functional on  $R^n$ , then

$$f(x) = \int_{\tilde{G}} f(y) \, dy / \int_{\tilde{G}} dy.$$

- (iii)  $x$  is an interior point of  $\tilde{G}$  (use (ii) and the assertion of Exercise 23 in Section 2.1).

The usefulness of the centre of gravity in the question interesting us is based on the following geometrical lemma (here and later  $|\cdot|_n$  denotes a Lebesgue  $n$ -dimensional volume).

### 2.3.2.1

**Lemma [21].** Let  $G$  be a convex, closed, bounded body in  $R^n$  with centre of gravity  $x$ , and let  $G_+$ ,  $G_-$  be the parts into which a hyperplane  $\Pi$  passing through  $x$  divides  $G$ . Then

$$|G_+|_n, |G_-|_n \leq \left(1 - \left(\frac{n}{n+1}\right)^n\right) |G|_n \leq \frac{e-1}{e} |G|_n.$$

*Geometrical commentary:* It is impossible to cut off by a plane passing through the centre of gravity of a convex body of volume 1 a part greater than that cut off from a simplex of equal volume by a plane passing through the centre of gravity parallel to one of the faces (in the latter case the part attached to the face is intended).

The lemma shows that the choice, as the interrogated point, of the centre of gravity of the constructed domain of localization of the solution really does guarantee a significant decrease of this domain—not less than  $(e-1)/e = a(\infty)$  times the volume, where

$$a(n) = 1 - \left(\frac{n}{n+1}\right)^n.$$

These considerations suffice for constructing the MCG. We consider a bounded, convex, closed body  $G \subset R^n$  and a class  $C$  of the  $C^{v_0}(G, R^n, m)$  type. Let  $v > v_0$  be the required accuracy of solution of problems of class  $C$ . We describe the MCG constructed to this accuracy. We shall take  $v < 1$ : every point of  $G$  is a solution of any problem  $f \in C$  with accuracy equal to 1, so that this assumption produces no loss of generality.

#### 2.3.2.2

The work of the MCG constructed to accuracy  $v$  on a problem  $f \in C$  consists in the construction of sequences  $\{x_i \in \text{int } G\}_{i=1}^{M_f+1}$  and  $\{G_i \subset G\}_{i=0}^{M_f}$ . Here  $M_f+1$  is the number of steps of the method on the problem  $f$  (the number is formed by the method itself), and  $G_i$  is a convex, closed, bounded body lying in  $G$  for all  $i < M_f$ . The set  $G_{M_f}$  is convex and closed and lies in  $G$ , but is not necessarily a body. For every  $i \leq M_f$  the point  $x_i$  is the centre of gravity of  $G_{i-1}$ , and  $G_0 = G$ .

As well as the above sequences the MCG constructs an auxiliary numerical sequence  $\{a_i\}_{i=0}^{M_f}$ . The number  $a_i$  is, roughly speaking, the recorded value of the objective functional of the problem being solved which is obtained after  $i$  steps, and  $a_0 = +\infty$ .

It will be proved below that the number  $M_f+1$  of steps of the MCG on any problem  $f \in C$  does not exceed the number  $\Phi_{n,\infty}(v-v_0)$ , where

$$\Phi_{n,\infty}(v) \equiv \left[ \frac{n \ln 1/v}{\ln 1/a(n)} \right] + 2 \leq \left[ \frac{n}{\ln e/(e-1)} \right] \ln 1/v + 2. \quad (3.1)$$

Here and elsewhere  $\lceil t \rceil$  is the least integer  $\geq t$ .

The work of the MCG constructed to accuracy  $v$ ,  $1 > v > v_0$ , on a problem  $f \in C$  is described by the following rules.

#### MCG 0. Initial setting

Put  $G_0 = G$ ,  $a_0 = +\infty$ . Go to MCG 1.

**MCG 1. i-th step**

At the  $i$ th step ( $1 \leq i \leq M_f$ ) there is a known convex, closed, bounded body  $G_{i-1} \subset G$ . On the  $i$ th step the following operations are carried out in succession:

**MCG 1.1.** The centre of gravity  $x_i$  of the domain  $G_{i-1}$  is constructed.

**MCG 1.2.** A question about the problem being solved is put to the oracle at the point  $x_i$ . (This is possible, because  $x_i \in \text{int } G_{i-1} \subset \text{int } G$ .)

Let  $g_i^j(y)$  be the affine functionals imparted by the oracle. Put  $r_i^j = \max\{0, \max_{y \in G} g_i^j(y)\}$ . If, for all  $j \geq i$ , the inequality  $(\max_{s \leq i} r_s^j)(v - v_0) \geq g_i^j(x_i)$  holds, put  $j(i) = 0$ ; otherwise put  $j(i)$  equal to some  $j \geq 1$  for which this inequality does not hold.

**MCG 1.3.** Put

$$a_i = \begin{cases} a_{i-1}, & j(i) \neq 0, \\ \min\{g_i^0(x_i), a_{i-1}\}, & j(i) = 0, \end{cases}$$

so that

$$a_i = \min\{g_s^0(x_s) \mid 1 \leq s \leq i, j(s) = 0\}.$$

**MCG 1.4.** Let

$$G_i = \{x \in G_{i-1} \mid g_i^j(x) < a_i, \quad g_i^j(x) \leq (v - v_0)(\max_{s \leq i} r_s^j), \quad 1 \leq j \leq m\}.$$

Take  $\tilde{G}_i$  to be the closure of the set  $G_i$ . If

$$|\tilde{G}_i|_n < (v - v_0)^n |G|_n, \quad (3.2)$$

go to MCG 2. Otherwise, complete the step (i.e. increase  $i$  by 1 and go to MCG 1).

**MCG 2. Rule for END and output of the result**

Suppose the switch to MCG 2 took place when  $i = M_f$ . The work of the method then stops, with output of the result

- \* if  $a_{M_f} = \infty$  (i.e. if  $j(s) > 0$  for all  $s \leq M_f$ ),
- $x_i$  if  $a_{M_f} < \infty$ ,  $i(i_0) = 0$  and  $g_{i_0}^0(x_{i_0}) = a_{M_f}$ .

**Commentary.** What is given out as the result is the best (according to the estimates  $g_i^0(x_i)$  of the numbers  $f_0(x_i)$ ) for all the points  $x_i$  considered which satisfy the constraints of the problem with the necessary accuracy (again according to the estimates  $g_i^j(x_i)$ ).

**2.3.2.3**

**Theorem.** When  $1 > v > v_0$  the method just described, constructed with accuracy  $v$ , ensures this accuracy on the class  $C$  with a laboriousness not exceeding  $\Phi_{n,\infty}(v - v_0)$ , where  $\Phi_{n,\infty}(v)$  is the function in (3.1).

**Proof.** (The reader is recommended to look into this well, since it is the first of a series of arguments of the same type.)

1°. We start with an estimate of the laboriousness of the method. From the rule for forming  $j(i)$  and  $a_i$  it follows that, for all  $i \leq M_f$ ,

$$g_i^{j(i)}(x_i) \begin{cases} \geq a_i, & j(i) = 0, \\ > (v - v_0) \max_{s \leq i} r_s^{j(i)}, & j(i) > 0, \end{cases}$$

i.e.  $\tilde{G}_i$  is a subset of part of  $G_{i-1}$  cut off from  $G_{i-1}$  by a hyperplane passing through the centre of gravity of  $G_{i-1}$ . By the lemma previously stated

$$|\tilde{G}_i|_n \leq a(n) |G_{i-1}|_n, \quad \text{i.e. } |\tilde{G}_i|_n \leq a^i(n) |G|_n.$$

Therefore (2.2) is satisfied not later than when

$$i = \left\lceil n \frac{\ln 1/(v - v_0)}{\ln 1/a(n)} \right\rceil + 1,$$

and this provides the required estimate of the laboriousness

$$M_f + 1 \leq \Phi_{n,\infty}(v - v_0).$$

2°. We now estimate the error of the method. We fix  $f \in C$ . Suppose first that  $f$  is incompatible. By the rule MCG 2 the result  $\bar{x}$  of applying the method to  $f$  is either the (correct) answer of incompatibility of the conditions of the problem or a point  $x_{i_0}$  such that, for all  $j \geq 1$

$$g_{i_0}^j(x_{i_0}) \leq (v - v_0) \max_{s \leq i_0} r_s^j \leq (v - v_0) \max\{0, \max_{x \in G} f_j(x)\} \quad (3.3)$$

(we have taken into account that, by virtue of the properties of the oracle,

$$r_s^j = \max_{x \in G} g_s^j(x) \leq \max_{x \in G} \{0, f_j(x)\}.$$

From (3.3) and the properties of the oracle

$$f_j(\bar{x}) = f_j(x_{i_0}) \leq (v - v_0) r_j(f) + v_0 r_j(f) \leq v r_j(f), \quad 1 \leq j \leq m. \quad (3.4)$$

Thus, for incompatible  $f$  we have  $v(x, f) \leq v$ , as required.

3°. Now let  $f$  be compatible, and let  $x^*$  be its solution. It is possible that the inequality

$$a_{M_f} \leq f_0(x^*) + (v - v_0) r_0(f) \quad (3.5)$$

holds. In that case the method ensures the required accuracy. For, it follows from (3.5) that  $a_{M_f} < \infty$ , and then it follows from MCG 2 that  $\bar{x} \neq *$ . But then  $\bar{x} = x_{i_0}$ , where  $j(i_0) = 0$  and  $a_{M_f} = g_{i_0}^0(x_{i_0})$ . As we have seen,  $j(i_0) = 0$  implies (3.4). It remains to check that  $f_0(x_{i_0}) \leq f_0(x^*) + vr_0(f)$ . But this follows at once from the following inequalities, which come from (2.5) and the definition of  $i_0$ :

$$f_0(x_{i_0}) \leq g_{i_0}^0(x_{i_0}) + v_0 r_0(f) \leq f_0(x^*) + vr_0(f).$$

4°. We now prove that (3.5) does in fact hold. We assume the contrary and derive a contradiction. Let

$$a_{M_f} > f_0(x^*) + (v - v_0)r_0(f). \quad (3.6)$$

It is possible that  $x^* \notin G_{M_f}$ . This would mean that, for some  $i_0$  and  $j_0 \in 0, \dots, m$ , we have

$$f_{j_0}(x^*) \geq g_{i_0}^{j_0}(x^*) \begin{cases} \geq a_{i_0}, & j_0 = 0, \\ > (v - v_0) \max_{s \leq i_0} r_s^{j_0}, & j_0 > 0. \end{cases} \quad (3.7)$$

(3.7) implies that  $j_0 = 0$ . For, when  $j_0 > 0$ , we have

$$g_{i_0}^{j_0}(x^*) > 0$$

by (3.7), and this is impossible, because  $f_{j_0}(x^*) \leq 0$ . Thus, in (3.7)  $j_0 = 0$ . But then (3.7) gives

$$f_* \geq g_{i_0}^0(x^*) \geq a_{i_0},$$

and this is impossible, because of (3.6). This means that  $x^* \in G_{M_f}$  and hence  $G_{M_f} \neq \emptyset$ .

We consider the body  $G$  obtained by a homothetic compression in the ratio  $1/(v - v_0)$  at the point  $x^*$ :

$$G = (v - v_0)(G - x^*) + x^*.$$

Since

$$|G_{M_f}|_n < (v - v_0)^m |G|_n$$

and  $G$  has the volume  $(v - v_0)^m |G|_n$ , the inclusion  $G_{M_f} \supset \text{int } G$  is impossible. So  $\text{int } G \not\subset G_{M_f}$ . This implies that there are  $v' < v_0 < v' < v$ , and a segment starting at  $x^*$  with its end  $z \in G$  such that the point  $y = (v' - v_0)(z - x^*) + x^*$  does not lie in  $G_{M_f}$ . By definition of  $G_{M_f}$  this implies that, for some  $i_0 \leq M_f$  and  $j_0 \in 0, \dots, m$ ,

$$g_{i_0}^{j_0}(y) \begin{cases} \geq a_{i_0}, & j_0 = 0 \\ > (v - v_0) \max_{s \leq i_0} r_s^{j_0}, & j_0 > 0. \end{cases} \quad (3.8)$$

At the same time

$$g_{i_0}^{j_0}(y) = (1 - (v' - v_0)) g_{i_0}^{j_0}(x^*) + (v' - v_0) g_{i_0}^{j_0}(z). \quad (3.9)$$

When  $j_0 > 0$  the right-hand side of (3.9) does not exceed

$$(v' - v_0) r_{i_0}^{j_0} \leq (v' - v_0) \max_{s \leq i_0} r_s^{j_0}.$$

Thus (3.8) and the assumption adopted that  $j_0 > 0$  would give

$$(v' - v_0) \max_{s \leq i_0} r_s^{j_0} > (v' - v_0) \max_{s \leq i_0} r_s^{j_0},$$

and this is impossible, because  $v' < v$ .

This means that  $j_0 = 0$ . But then (3.8) gives  $g_{i_0}^0(y) \geq a_{i_0}$ , and, by (3.9)

$$\begin{aligned} g_{i_0}^0(y) &\leq (1 - (v' - v_0)) f_0(x^*) + (v' - v_0) \max_{x \in G} f_0(x) \\ &\leq f_0(x^*) + (v' - v_0) r_0(f) \leq f_0(x^*) + (v - v_0) r_0(f), \end{aligned}$$

and since  $g_{i_0}^0(y) \geq a_{i_0}$ , we have

$$a_{M_f} \leq a_{i_0} \leq f_0(x^*) + (v - v_0) r_0(f),$$

which contradicts (3.6). The contradiction establishes the theorem.

## 2.4 SPECIAL VERSIONS OF THE MCG

This section is devoted to the exposition of versions of the MCG for special situations: for the case where the absolute error, not the relative error, is specified in advance, and for the case where the approximate solution must satisfy the constraints exactly (i.e. the error is measured in the scale  $v_*(x, f)$  and not in  $v(x, f)$ ).

### 2.4.1

We begin with the first of these two cases.

The MCG has been described on the hypothesis that the method must solve problems of some class  $C$  of the type  $C^o(G, R^n, m)$  and with the required relative error  $v$  specified *a priori*. In other words, the application to  $f$  of the MCG constructed on a relative accuracy  $v > v_0$  ensures our obtaining a result with an absolute error  $\leq vr_0(f)$ . If it is not the relative error required of the solution, but the absolute error  $\varepsilon_j$ , which is specified in advance, then, since the quantities  $r_j(f)$  are not known *a priori*, we cannot indicate that relative error, namely

$$v(f, \varepsilon) = \min_{0 \leq j \leq m} \varepsilon_j / r_j(f),$$

with which it is required to construct a method that will ensure the specified absolute error.

Thus the variant already described of the MCG is not suitable for ensuring specified absolute errors. It can, however, be modified so as to avoid this

difficulty. We shall describe the appropriate modification. It is obtained from the original version by making the following changes ( $\varepsilon_j > 0$  denotes the absolute error specified in advance).

1°. The rule for determining  $j(i)$  in MCG 1.2 is replaced by

$$j(i) = \begin{cases} 0, & \text{if } g_i^0(x_i) \leq \varepsilon_j, \quad 1 \leq j \leq m, \\ \text{any } j \geq 1 \text{ such that } g_i^0(x_i) > \varepsilon_j, & \text{if there is such a } j. \end{cases}$$

2°. The rule MCG 1.4 is replaced by the rule

MCG 1.4'. Put

$$r_i = \begin{cases} r_i^{(0)}, & j(i) > 0, \\ \max_{y \in G} g_i^0(y) - g_i^0(x_i), & j(i) = 0. \end{cases}$$

Let

$$v_i = \begin{cases} \frac{\varepsilon_{j(i)}}{r_i}, & \text{if } j(i) \neq 0, \\ \frac{\varepsilon_0}{\varepsilon_0 + r_i}, & \text{if } j(i) = 0. \end{cases}$$

We write

$$G_i = \begin{cases} \{x \in G_{i-1} \mid g_i^{(0)}(x) \leq \varepsilon_{j(i)}\} & \text{for } j(i) > 0, \\ \{x \in G_{i-1} \mid g_i^0(x) < g_i^0(x_i)\}, & j(i) = 0. \end{cases}$$

Further, let

$$v^i = \min_{0 \leq i \leq i} v_i, \quad G_i = \overline{(G_i)}.$$

( $\overline{(\cdot)}$  denotes closure.) Here we have taken  $v_0 = 1$ .

If  $|G_i|_n < (v^i)^n |G_0|_n$ , go to MCG 2; otherwise, increase  $i$  by 1 and go to MCG 1.

It will be proved that the method constructed in this way ensures the required absolute errors  $\varepsilon_j$  in the solution of problems  $f \in C$  'with accuracy up to the irremovable error of the oracle'. That is, for the result  $\bar{x}$  of applying the method to  $f \in C$  the inequality

$$\varepsilon(x, f) \leq \varepsilon + v_0 r(f) \tag{4.1}$$

holds, where  $\varepsilon = (\varepsilon_0, \dots, \varepsilon_m)$ , and  $r(f) = (r_0(f), \dots, r_m(f))$ . Thus, the errors of the solution with accuracy up to the 'irremovable errors of the oracle'  $v_0 r(f)$  are  $\varepsilon_j$ .

Suppose, let us say, it is known *a priori* that the required errors exceed the 'irremovable noise of the oracle'; for definiteness, let  $\varepsilon_j \geq 2v_0 r_j(f)$ . In this case

we can guarantee that the given constraints will be satisfied exactly to within the errors (it is sufficient to apply the method described, constructed with an error  $\varepsilon_j/2$ ). We lose this sort of possibility if the error of the oracle is comparable with the given specified accuracy of the solution, and even more so if it exceeds the latter. This effect does not discredit the method: such an oracle, for obvious objective reasons, is ill-adapted to (or is quite incapable of) solving problems to such high accuracy.

The laboriousness of the method described, as will be proved, admits the following estimate. Let  $v(f, \varepsilon) = \min_j \varepsilon_j / r_j(f)$  be the maximal relative error of the solution of the problem  $f$  that ensures the required relative errors, and let

$$\bar{v}(f, \varepsilon) = \frac{v(f, \varepsilon)}{1 + v(f, \varepsilon) + v_0}.$$

Then the bound for the laboriousness is  $\Phi_{n, \infty}(\bar{v}(f, \varepsilon))$ . We see that ensuring the specified absolute errors when the  $r_i(f)$  are not known *a priori* 'costs' only a little more than when the  $r_i(f)$  are known (the 'cost' in the first case is  $\Phi_{n, \infty}(\bar{v}(f, \varepsilon))$ , and in the second case  $\Phi_{n, \infty}(v(f, \varepsilon))$ ). Although  $\bar{v}(f, \varepsilon) < v(f, \varepsilon)$ , the ratio of these quantities is close to  $1 + v_0$  for small  $v(f, \varepsilon)$ ; similarly the ratio of the values of the laboriousness which correspond to these hypotheses regarding the *a priori* information about the problem is close to 1 for small  $v(f, \varepsilon)$ .

#### 2.4.1.1

We now prove that the method described does indeed satisfy the above assertions about the error and laboriousness. The proof, in the main, repeats the one already given. Let  $f \in C$  be the problem to be solved. We start with an estimate of the laboriousness. It is clear from the definition that, for all those  $i$  such that  $j(i) > 0$ , we have

$$v_i = \frac{\varepsilon_{j(i)}}{r_i} \geq \frac{\varepsilon_{j(i)}}{r_{j(i)}(f)} \geq v(f, \varepsilon) \geq \bar{v}(f, \varepsilon)$$

But if  $j(i) = 0$ , then

$$v_i = \frac{\varepsilon_0}{\varepsilon_0 + \max_{y \in G} g_i^0(y) - g_i^0(x_i)} \geq \frac{\varepsilon_0}{\varepsilon_0 + r_0(f)(1 + v_0)} \geq \bar{v}(f, \varepsilon)$$

(we have used the fact that

$$\max_{y \in G} g_i^0(y) - g_i^0(x_i) \leq \max_{y \in G} f_0(y) - f_0(x_i) + v_0 r_0(f) \leq (1 + v_0) r_0(f).$$

Thus,  $v_i \geq \bar{v}(f, \varepsilon)$ , and hence, as in Section 2.3.2.3, it follows that the laboriousness of the method does indeed not exceed  $\Phi_{n, \infty}(\bar{v}(f, \varepsilon))$ .

We now estimate the accuracy of the method. As before, the result  $\bar{x}$  of applying the method to  $f$  is either  $*$ , or it is a point of  $G$  such that the inequalities

$$f_j(\bar{x}) \leq \varepsilon_j + v_0 r_j(f), \quad 1 \leq j \leq m. \quad (4.2)$$

hold. Hence it follows that, for an incompatible  $f$ , the assertion about the accuracy of the method does indeed hold. Moreover,  $\bar{x} = *$  if and only if  $a_{M_f} = +\infty$ . If  $a_{M_f} < \infty$ , then  $\bar{x} = x_{i_0}$  for some  $i_0$ , and

$$g_{i_0}^0(\bar{x}) \leq a_{M_f}. \quad (4.3)$$

Hence it is clear that to prove (4.1) it suffices to prove the relation

$$a_{M_f} \leq f_* + \varepsilon_0. \quad (4.4)$$

Let us prove (4.4). For incompatible  $f$ , it is obviously true. Let  $f$  be compatible, and let  $x^*$  be its solution. By the definition of  $\tilde{G}_{M_f}$  for every  $x \in G \setminus \tilde{G}_{M_f}$ , there is an  $i = i(x) \leq M_f$  such that

$$g_i^{(0)}(x) \begin{cases} \geq g_i^{(0)}(x_i), & j(i) = 0, \\ > \varepsilon_{j(i)}, & j(i) > 0. \end{cases} \quad (4.5)$$

If  $x^* \notin \tilde{G}_{M_f}$ , then from (4.5) applied with  $x = x^*$  it follows that  $j(i) = 0$  and  $a_{M_f} \leq f_*$ , as required. Now let  $x \in \tilde{G}_{M_f}$ . Then, as above, there are  $v' < v^{M_f}$  and  $z \in G$  such that the point

$$y = x^* + v'(z - x^*)$$

is not in  $\tilde{G}_{M_f}$ . Then, by (4.5), for some  $i$

$$g_i^{(0)}(y) \geq \begin{cases} g_i^{(0)}(x_i), & j(i) = 0, \\ \varepsilon_{j(i)}, & j(i) \neq 0, \end{cases}$$

or

$$g_i^{(0)}(x^*) + v'(g_i^{(0)}(z) - g_i^{(0)}(x^*)) \geq \begin{cases} g_i^0(x_i), & j(i) = 0, \\ \varepsilon_{j(i)}, & j(i) \neq 0. \end{cases} \quad (4.6)$$

When  $j(i) > 0$ , (4.6) gives

$$\varepsilon_{j(i)} \leq v' r_i < v_i r_i = \varepsilon_{j(i)},$$

which is impossible. Thus  $j(i) = 0$  in (4.6), and (4.6) gives

$$(1 - v')g_i^0(x^*) + v'(g_i^0(z) - g_i^0(x_i)) \geq (1 - v')g_i^0(x_i),$$

or

$$g_i^0(x_i) \leq f_* + \frac{v'}{1 - v'} r_i \quad (4.7)$$

(we have used the fact that  $v_i \leq 1$  when  $j(i) = 0$ , and so  $v' < v_i \leq 1$ ). Since  $v' < v_i \leq 1$ , we have  $v'/(1 - v') \leq \varepsilon_0/r_i$ . Thus

$$g_i^0(x_i) \leq f_* + \varepsilon_0,$$

and  $a_{M_f} \leq f_* + \varepsilon_0$ . This completes the proof of (4.1), and, with it, validates the properties of the method under examination.

#### 2.4.1.2

*Remark.* We have described a version of the MCG suitable for solving general convex problems—problems of the  $C^0(G, R^n, m)$ -type classes, with a specified relative error in Section 2.3, and with a specified absolute error in Section 2.4. From the validation of the methods it is easily deduced that these versions are also suitable for solving Lipschitz-convex problems (classes of the  $C_{\text{Lip}}^0(G, R^n, \|\cdot\|, m)$  type) with the same errors (in the situation of this section the absolute errors in all cases, of course, are guaranteed with accuracy up to the irremovable errors of the oracle).

The assertion that methods which solve problems of the wider class (general convex problems) do not lose their characteristics on the narrower class of Lipschitz-convex problems might appear to be a tautology. But there is a rather nice point here. A class  $C$  of type  $C_{\text{Lip}}^0(G, E, \|\cdot\|, m)$  is not, generally speaking, a subclass of the class of type  $C^0(G, E, m)$  if  $v_0 > 0$ . Because, the error of an oracle for the 'Lipschitz' classes is measured on a coarser scale than for general convex problems, and so  $(G, v_0)$ —an oracle for the field of problems  $C_{\text{Lip}}(G, E, \|\cdot\|, m)$  regarded as a subset of the field of problems  $C(G, E, m)$ —cannot always be extended into an  $(\text{int } G, v_0)$ -oracle on the whole of  $C(G, E, m)$ .

#### 2.4.2 Strictly compatible problems

The method of centres of gravity does not, generally speaking, ensure that the constraints of the problem are satisfied exactly at the point found. This circumstance is not a defect of the method. As we shall see in due course (in (1.17) of Chapter 4), the  $*$ -complexity of a class  $C$  of type  $C^0(G, E, m)$  corresponding to any deterministic local oracle is infinite when  $m \geq 1$  and  $v < 1$  (i.e. no finite-step method can guarantee finding an admissible feasible point for every compatible problem  $f \in C$ ). If it is known in advance that a problem  $f$  is not only compatible but also that it has a 'sufficiently massive' set of admissible feasible points, then a suitable modification of the MCG will guarantee that the required  $*$ -error can be secured. As the 'measure of the massiveness' of the set of admissible feasible points of  $f$  it is natural to take the Lebesgue  $n$ -dimensional measure of this set ( $E = R^n$ ). In this connection we define the concept of a strictly compatible problem.

**DEFINITION.**  $f \in C(G, R^n, m)$  is said to be  $\alpha$ -strictly compatible if the set  $G_f$  of its feasible points has the bound

$$|G_f|_n \geq \alpha^n |G|_n.$$

**Exercise 1** (on the definition). Let  $f \in C(G, R^n, m)$  be a convex problem satisfying the Slater condition with a parameter  $\varepsilon > 0$ , i.e. for some  $x \in G$  we have  $f_j(x) \leq -\varepsilon r_j(f)$ ,  $1 \leq j \leq m$ . Prove that  $f$  is  $\alpha$ -strictly compatible, with  $\alpha = \varepsilon/(1+\varepsilon)$ .

We describe a version  $MCG_*$  of MCG which is suitable for solving strictly compatible problems with a specified  $*$ -error. Let  $C$  be a class of problems of type  $C^0(G, R^n, m)$ , and let  $C_\alpha$  be a subclass of  $C$  which contains precisely the  $\alpha$ -strictly compatible problems. This  $MCG_*$  itself determines when to stop work and put out the result. It is applicable to any problem from  $C(G, R^n, m)$ . If the problem is in  $C_\alpha$  for some  $\alpha > 0$ , then the method puts out the result for the problem not later than after

$$M(\alpha, v) = \Phi_{n, \infty}(\alpha v) \quad (4.8)$$

steps (it is assumed that the required accuracy  $v$ ,  $1 > v > 0$ , has been specified in advance), and the problem will have been solved with a  $*$ -error not greater than  $v$ . We stress that it is not required to be known in which of the  $C_\alpha$  classes the problem under solution lies. Further it is assumed that  $m \geq 1$ .

$MCG_*$  is in steps. At the  $i$ th step domains  $G_i$  and  $H_i$ ,  $H_i \supset G_i$ , and also a number  $a_i$  are constructed. Initially  $G_0 = H_0 = G$ , and  $a_0 = +\infty$ . The following operations take place in the  $i$ th step.

#### $MCG_*$ 1.

The centre of gravity  $x_i$  of the domain  $G_{i-1}$  is constructed. A question about the problem is put to the oracle at the point  $x_i$ . Let  $g_j^i(x)$  be the affine functionals communicated by the oracle. We put

$$j(i) = \begin{cases} 0, & \text{if } g_j^i(x_i) \leq 0, \quad 1 \leq j \leq m, \\ 1, & \text{otherwise} \end{cases}$$

Suppose, further, that

$$\begin{aligned} a_i &= \begin{cases} a_{i-1}, & j(i) = 1, \\ \min \{a_i, g_0^i(x_i)\}, & j(i) = 0, \end{cases} \\ H_i &= \{x \in H_{i-1} \mid g_j^i(x) \leq 0, \quad 1 \leq j \leq m\}, \\ G_i &= \{x \in G_{i-1} \mid g_j^i(x) \leq 0, \quad 1 \leq j \leq m, \quad g_i^0(x) < a_i\}, \\ \bar{G}_i &= (\overline{G_i}) \end{aligned}$$

#### $MCG_*$ 2.

The ratio  $|G_i|_n / |H_i|_n$  is calculated. If  $|G_i|_n \geq v^n |H_i|_n$ , then we go to the next step; otherwise the work of the method stops, and the result  $x_{i_0}$  is put out,

where  $i_0$  is such that  $j(i_0) = 0$  and

$$g_{i_0}^0(x_{i_0}) = \min \{g_s^0(x_s) \mid j(s) = 0, \quad s \leq M_f\} = a_{M_f}$$

(here  $M_f$  is the number of the step at which the method stops).

It is clear that  $MCG_*$  differs from the original MCG, apart from the rule for stopping work, only in that the rule for reducing the domain  $G_{i-1}$  (the rule for constructing  $G_i$ ) 'cuts off' from  $G_i$  all points at which  $g_j^i(x) > 0$  for some  $j$  ( $1 \leq j \leq m$ ), i.e. points at which the constraints are automatically not satisfied, while MCG 'cuts off' (when  $v_0 = 0$ ) only those points at which 'the constraints are not satisfied with the specified accuracy'.

#### 2.4.2.1

**Theorem.** Let  $v < 1$  and  $f \in C_\alpha$ . Then  $MCG_*$  constructed for a specified accuracy puts out the result  $\bar{x}$  not later than after  $M(\alpha, v)$  steps (cf. (4.8)), and  $v_*(\bar{x}, f) \leq v$ .

*Proof.* 1°. Let  $f$  be such that the method stops its work on  $f$  and gives out the result  $\bar{x}$ . We verify that then  $v_*(\bar{x}, f) \leq v$ . Let  $x^*$  be the solution of the problem  $f$ . We remark first that  $x^* \in H_{M_f}$ , where  $M_f + 1$  is the laboriousness of the method on  $f$ . For,  $H_0 = G$  and in  $H_i \setminus H_{i-1}$  we have  $\max_{1 \leq j \leq m} f_j(x) > 0$  by the definition of  $H_i$ . Therefore  $H_{M_f}$  contains the set of admissible feasible points of  $f$ , and so also  $x^*$ .

Further, if the method stops on the step  $M_f$ , then  $|G_{M_f}|_n < |H_{M_f}|_n$ . The latter is impossible if  $a_{M_f} = +\infty$ , and so  $a_{M_f} < \infty$ , and therefore  $a_{M_f} = \min \{f_0(x_s) \mid x_s \in G_f\}$  (by the rule for constructing  $a_i$  from  $a_{i-1}$  and the rule for choosing  $j(i)$ ). It is also clear that outside  $G_i$  in  $H_i$  either  $\max_{1 \leq j \leq m} f_j(x) > 0$ , or  $f_0(x) \geq a_i$ . Hence it follows that outside  $G_{M_f}$  in  $G$  either  $x \notin G_f$  or  $f_0(x) \geq a_{M_f} = f_0(\bar{x})$ . By construction  $f_j(\bar{x}) \leq 0$ , for  $1 \leq j \leq m$ , and it remains only to prove that  $f_0(x^*) > a_{M_f} + vr_0(f)$ . Suppose, on the contrary, that  $f_0(x^*) > a_{M_f} + vr_0(f)$ . We observe that then  $x^* \in G$ , for, as we have seen, we have  $f_0(x) \geq a_{M_f}$  everywhere on  $G_f \setminus G_{M_f}$ . Moreover, it is clear that  $G_{M_f}$  can be described in the form

$$G_{M_f} \supset \{x \in H_{M_f} \mid g_s^0(x) < a_s, \quad s \leq M_f\}.$$

Since  $|G_{M_f}|_n < v^n |H_{M_f}|_n$ , there are  $z \in H_{M_f}$  and  $y = vz + (1-v)x^* : y \notin G_{M_f}$ . Therefore, for some  $s$ , we have  $g_s^0(y) \geq a_s$  and *a fortiori*  $f_0(y) \geq a_s$ . But

$$f_0(y) - f_0(x^*) \leq v(f_0(z) - f_0(x^*)) \leq vr_0(f),$$

i.e.

$$a_{M_f} \leq a_s \leq f_0(y) \leq vr_0(f) + f_0(x^*).$$

This contradicts our assumption. Therefore  $v_*(\bar{x}, f) \leq v$ .

2°. We now estimate  $M_f$ , assuming that  $f \in C_\alpha$ . It was observed earlier that  $H_i \supset G_i$  for all  $i$ , and so  $|H_i|_n \geq \alpha^n |G_i|_n$  for all  $i$ . At the same time, as in the proof of Theorem 2.3.2.3,  $|G_i|_n \leq a'(n) |G|_n$ . Therefore

$$|G_i|_n / |H_i|_n \leq a'(n) / \alpha^n,$$

and this ratio becomes less than  $v^n$  as soon as  $i \geq M(\alpha, v) - 1$ . This implies that the method does come to an end on  $f$ , and not later than after  $M(\alpha, v)$  steps. The theorem has been proved.

$\text{MCG}_*$ , like MCG, can be adapted to finding an approximate solution of  $f \in C$  with a specified absolute  $*$ -error  $\varepsilon_0$  (and not, as above, for a specified relative  $*$ -error). To do this it is necessary, in addition to all the operations described above, to calculate in  $\text{MCG}_*$  1, at each step  $i$  with  $j(i) = 0$ , the number

$$v_i = \frac{\varepsilon_0}{\varepsilon_0 + \sup_{y \in G} g_i^0(y) - g_i^0(x_i)}$$

and also  $v^i = \min \{v_s \mid j(s) = 0, s \leq i\}$  (if  $j(s)$  were  $> 0$  for all  $s \leq i$ , then we would take  $v^i = 1$ ).

Further, the rule for constructing  $\tilde{G}_i$  in  $\text{MCG}_*$  1 is replaced by

$$\tilde{G}_i = \begin{cases} G_{i-1} \cap H_i, & j(i) = 1, \\ \{x \in G_{i-1} \cap H_i \mid g_i^0(x) < g_i^0(x_i)\}, & j(i) = 0. \end{cases}$$

Moreover, in  $\text{MCG}_*$  2 in the inequality  $|G_i|_n \geq v^n |H_i|_n$  determining the condition for continuing the process,  $v$  must be replaced by  $v^i$ .

The method modified in this way solves every strictly compatible problem in  $C$  with a  $*$ -error  $\leq \varepsilon$ , and its laboriousness on an  $\alpha$ -strictly compatible problem is bounded above by the number  $M(\alpha, \bar{v}(f, \varepsilon_0))$ , where

$$\bar{v}(f, \varepsilon_0) = \frac{\varepsilon_0}{\varepsilon_0 + r_0(f)} = \frac{v(f, \varepsilon)}{1 + v(f, \varepsilon)}$$

(here  $v(f, \varepsilon) = \varepsilon_0 / r_0(f)$  is the relative error which ensures the specified absolute error).

The proof that the modified MCG has the properties stated is left to the reader.

## 2.5 A REALIZABLE VERSION OF THE MCG

### 2.5.1

The method of centres of gravity, as a method of solving problems of the classes of type  $C^{v_0}(G, R, m)$ , cannot be essentially improved as regards laboriousness (in a certain precise sense, see Chapter 4). However, the

'computational cost' of every step of the method is extremely high: finding the centre of gravity, generally speaking, of an arbitrary convex body of some appreciable dimensionality is a more than complicated problem even with the most perfected computational aids. So, although the MCG can be used, it would be used only when the complication of obtaining information about the problem is extremely great (for example, if this information can be obtained only by high-costing 'live' experiments) so that economy in the number of the questions will compensate for the enormous computational complexity of a step. In the remaining cases (and today it is only with these that computational mathematics concerns itself) the extreme complexity of a step presents an insuperable obstacle to the use of MCG.

One would like to modify the method in such a way as to simplify a step to a reasonable level without inordinately increasing the laboriousness bound. Such a simplification is achieved below. We shall call the corresponding method, which was proposed in the authors' paper [31] the *modified method of centres of gravity* (MMCG).

The idea of the MMCG is easily arrived at by analysing the source of the computational difficulties in MCG—it is the 'arbitrariness of shape' of the successive domains  $G_i$  of localization of the solution. Let us assume that  $G$  itself is of the simplest form from the point of view of finding the centre of gravity of the body, say an ellipsoid (i.e. a ball in suitable co-ordinates). Then the first step of the MCG is easily effected, and  $G_1$  is a hemisphere (in the co-ordinates chosen). Finding the centre of gravity of a hemisphere is still quite simple. However,  $G_2$  can already have quite a complicated form, and so on; the computational difficulties rapidly grow as the iteration number of the step increases.

It is possible, however, to increase deliberately the size of the domain of localization which is being constructed, so as to ensure that it has a shape convenient for the computation. The simplest way of all is to deal only with ellipsoids. To achieve this, at the first step, after constructing  $G_1$ , we enclose this hemisphere in an ellipsoid  $\tilde{G}_1$  of the least possible volume, and regard  $\tilde{G}_1$  as the new domain of localization of the solution. Geometrically, the situation arising at the start of the second step is exactly the same as originally. Thus, by deliberately ignoring part of the information obtained at the step (increasing  $G_1$  to  $\tilde{G}_1$ ), we ensure the geometrical stability of the situation at all the steps and so make it possible to construct a simple recursive scheme of computation.

### 2.5.2

We preface the description of the MMCG by a simple geometrical lemma.

**Lemma.** Let  $V = \{x \in E^n \mid \|x\| \leq R\}$  be a ball of radius  $R$  in  $E^n$ , and let  $e \in E^n$  be a unit vector. The hemisphere  $V_e = \{x \in V \mid \langle x | e \rangle \leq 0\}$  can be enclosed in an

ellipsoid  $W$  of volume  $\beta^n(n)|V|_n$ , where

$$\beta(n) = \begin{cases} \frac{n}{\sqrt{n^2 - 1}} \left( \frac{n-1}{n+1} \right)^{1/2n}, & n > 1, \\ \frac{1}{2}, & n = 1. \end{cases} \quad (5.1)$$

We point out that

$$\beta(n) = 1 - c_n/n^2, \text{ where } c_n > 0 \text{ and } c_n \rightarrow \frac{1}{2} \text{ as } n \rightarrow \infty.$$

Let  $O_e$  denote the co-ordinate system (with the same volume element as the original Cartesian co-ordinate system  $O_0$  in  $R^n$ ) in which  $W$  is a ball with centre at 0. The radius of this ball will be  $\beta(n)R$ . The transformation of co-ordinates from  $O_e$  into  $O_0$  is given by the formula

$$x = A_{e,n}x' - \frac{R}{n+1}e. \quad (5.2)$$

Here

$$A_{e,n} = \left( \kappa^{n-1} - \frac{1}{\kappa} \right) P_e + \frac{1}{\kappa} I,$$

where  $P_e$  is the matrix of the ortho-projector on to the axis of the vector  $e$ , and  $I$  is the unit  $n \times n$  matrix, and

$$\kappa = \left( \frac{n+1}{n-1} \right)^{-\frac{1}{2n}}. \quad (5.3)$$

*Geometrically:* co-ordinates in  $O_e$  are obtained from the co-ordinates in  $O_0$  by shifting the origin to the point  $(-Re/(n+1))$ , changing the scale along the  $e$ -axis by the factor  $\kappa^{n-1}$ , and changing the scale in the orthogonal hyperplane by the factor  $1/\kappa$ .

*Exercise 1.* Prove the lemma. Show that the ellipsoid constructed in the lemma is the ellipsoid of least volume which contains  $V_e$ .

### 2.5.3

The modified method of centres of gravity (MMCG) is applicable to problems of classes of the  $C^{*0}(G, E, m)$  type in the case where  $E$  is finite-dimensional, and  $G$  is a bounded, convex, closed body in  $E = R^n$ . These conditions are now assumed to hold. Suppose also that we know a certain ellipsoid  $W_0$  containing  $G$ , and a number  $\beta$  such that

$$|W_0|_n \leq \beta^n |G|_n. \quad (5.4)$$

We let  $f_{m+1}(x)$  denote the function  $\rho_{1,1}(X, G)$ . Here  $\|\cdot\|$  is any norm on  $E$ . MMCG, like MCG, is applicable to any class  $C$  of the  $C^{*0}(G, R^n, m)$ -type. If

the required accuracy  $v$  of solution of problems of the class  $C$  satisfies the condition  $1 > v > v_0$ , then the method is capable of ensuring this accuracy.

The method, constructed for an accuracy  $v > v_0$  (we suppose  $v < 1$ ) consists of not more than

$$\Phi_{n,\infty}(v - v_0) = \left[ \frac{\ln(\beta/(v - v_0))}{\ln(1/(\beta(n)))} \right] + 2 \leq d_n n^2 \ln \frac{\beta}{v - v_0} + 2 \quad (5.5)$$

steps. Here  $d_n \rightarrow 2$  as  $n \rightarrow \infty$ . The number of steps  $M_f + 1$  of the method on a problem  $f \in C$  depends on  $f$  and on the required accuracy  $v$ .

We give an analytical description of MMCG accompanied by a geometrical commentary. MMCG consists in steps, at the  $i$ th one of which ( $1 \leq i \leq M_f$ ) we have a co-ordinate system  $O_{i-1}$  in  $R^n$  with origin at the point  $x_i$ , a ball  $W_{i-1}$  (in co-ordinates relative to  $O_{i-1}$ ) of radius  $r_{i-1}$  with centre at  $x_{i-1}$ , and a number  $a_{i-1}$ .  $O_0$  is the original co-ordinate system in which the ellipsoid  $W_0$  described in (5.4) is a ball with centre at 0; and the volume elements in all the co-ordinate systems  $O_{i-1}$  are the same.

From the analytical point of view, at the  $i$ th step we have a matrix  $B_{i-1}$  with determinant 1 (the matrix of the transformation from the co-ordinate system  $O_{i-1}$  into the system  $O_0$ ), and so  $B_0$  is the unit matrix  $I$ . We have also the point  $x_i$  (the origin of the co-ordinate system  $O_{i-1}$ ), and the number  $r_{i-1} > 0$  (the radius of  $W_{i-1}$  in the co-ordinate system  $O_{i-1}$ ). Points of  $R^n$  will be identified with the column-vectors of their co-ordinates in the system  $O_0$ .

Let  $v$ , with  $1 > v > v_0$ , be the required accuracy. The work of MMCG constructed to the accuracy  $v$  on problems of the class  $C$  is described by the following rules.

*MMCG 0. Initial setting.* Put  $a_0 = \infty$ ,  $B_0 = I$ ; choose  $O_0$ ,  $x_1$ ,  $r_0$  as described above. Go to MMCG 1.

### MMCG 1. The $i$ -th step.

At the  $i$ th step there are available: a matrix  $B_{i-1}$  with  $\det B_{i-1} = 1$ , numbers  $a_{i-1} \leq +\infty$ ,  $r_{i-1} \geq 0$ , and a co-ordinate system  $O_{i-1}$  with origin at the point  $x_i$ .  $B_{i-1}$  is the matrix of the transformation from the co-ordinate system  $O_{i-1}$  into the system  $O_0$ . At the  $i$ th step the following actions are taken in succession.

*MMCG 1.1.* If  $x_i \notin \text{int } G$ , then a support functional  $\eta_i^{m+1} \neq 0$  to the function  $\rho_{1,1}(x, G)$  at the point  $x_i$  is constructed (there is such a functional, because  $\text{int } G \neq \emptyset$ . Cf. Exercise 23, Section 2.1). We put

$$j(i) = m+1, g_i^{m+1}(y) = \rho_{1,1}(x_i, G) + \langle \eta_i^{m+1} | y - x_i \rangle,$$

and go to MMCG 1.3. But if  $x_i \in \text{int } G$ , then go to MMCG 1.2.

**MMCG 1.2.** A question about the problem  $f$  being solved is put to the oracle at the point  $x_i$ . Let  $g_j^i(y)$  be the affine functionals and  $\eta_j^i$  be their derivatives, communicated by the oracle. Put

$$\begin{aligned} r_j^i &= \max \{0, \max_{y \in G} g_j^i(y)\}, \quad 1 \leq j \leq m, \\ r_i^0 &= \max_{z \in G} g_i^0(z) - g_i^0(x_i). \end{aligned}$$

If  $g_j^i(x_i) \leq (v - v_0)r_j^i$  for all  $j \geq 1$ , put  $j(i) = 0$ ; otherwise, let  $j(i)$  be a  $j \geq 1$  such that  $g_j^i(x_i) > (v - v_0)r_j^i$ .

**MMCG 1.3.** Put  $\eta_i = \eta_i^{j(i)}$  ( $\eta_i$  is regarded as the column-vector of its co-ordinates relative to the basis in  $(R^n)$  corresponding to the  $O_0$ -system in  $R^n$ , i.e. to a Euclidean structure in which the co-ordinates  $O_{i-1}$  are Cartesian). Compute the column-vector  $\bar{p}_i = B_{i-1}^T \eta_i$  ( $T$  indicates the transpose).

If  $\bar{p}_i = 0$ , go to MMCG 1.4. If  $\bar{p}_i \neq 0$ , put

$$p_i = \bar{p}_i / \sqrt{\bar{p}_i^T \bar{p}_i}.$$

**Commentary.**  $x_i$  is the 'centre' of the next domain of localization of the solution; more precisely, the centre of the ellipsoid  $W_{i-1}$  embracing this domain ( $W_{i-1}$  is a ball with its centre at 0 in the co-ordinates  $O_{i-1}$ );  $p_i$  is the unit vector (in the  $O_{i-1}$  co-ordinate system) which is orthogonal (in the  $O_{i-1}$  Euclidean structure, i.e. the Euclidean structure in which the  $O_{i-1}$  co-ordinates are Cartesian) to the hyperplane which cuts off from  $W_{i-1}$  the next domain of localization.

**MMCG 1.4.** Put

$$a_i = \begin{cases} a_{i-1}, & j(i) > 0, \\ \min\{a_{i-1}, g_i^0(x_i)\}, & j(i) = 0. \end{cases}$$

If  $\bar{p}_i = 0$ , go to MMCG 2. Otherwise put

$$d_i = \begin{cases} g_i^0(x_i) - a_i, & j(i) = 0, \\ g_i^{j(i)}(x_i) - (v - v_0)r_j^i, & 1 \leq j(i) \leq m, \\ g_i^{m+1}(x_i), & j(i) = m+1. \end{cases}$$

Let

$$r_i = \beta(n) \sqrt{r_{i-1}^2 - \frac{d_i^2}{(\bar{p}_i^T \bar{p}_i)}}$$

(If  $d_i^2 / (\bar{p}_i^T \bar{p}_i) \geq r_{i-1}^2$ , go to MMCG 2.) Put

$$x_{i+1} = x_i - \left( \frac{d_i}{\sqrt{(\bar{p}_i^T \bar{p}_i)}} + \frac{r_i}{\beta(n)(n+1)} \right) B_{i-1} p_i,$$

$$B_i = B_{i-1} A_{p_i, n},$$

where  $A_{p_i, n}$  is the matrix defined in Lemma 2.5.2.

If  $r_i < ((v - v_0)/\beta)r_0$ , go to MMCG 2. Otherwise, finish the step (i.e. increase  $i$  by 1 and go to MMCG 1).

**Commentary.** As the next domain of localization the domain defined, in the  $O_{i-1}$  co-ordinates, as

$$U_i = \{x \in W_{i-1} | \bar{p}_i^T x_i \leq -d_i\}$$

can be taken (this follows at once from the definition of  $d_i$ ).  $U_i$  is contained in the  $O_{i-1}$ -ball  $W_i$  of radius  $\bar{r}_i = \sqrt{(r_{i-1}^2 - d_i^2) / (\bar{p}_i^T \bar{p}_i)}$  with its centre at the point  $-d_i p_i / \sqrt{(\bar{p}_i^T \bar{p}_i)}$  (everything relative to the  $O_{i-1}$  co-ordinates). In fact,  $U_i$  is contained in the hemisphere  $\bar{U}_i$  of this ball, which is cut off by the hyperplane passing through the centre and  $O_{i-1}$ -orthogonal to  $p_i$ .

By the lemma, the co-ordinates  $O_i$ , for which  $A_{p_i, n}$  is the matrix of the transformation from them into the system  $O_{i-1}$ , and whose origin (in the  $O_{i-1}$  system) is the point

$$v_i = -\frac{d_i p_i}{\sqrt{\bar{p}_i^T \bar{p}_i}} - \frac{\bar{r}_i p_i}{n+1},$$

has the property that the  $O_i$ -ball  $W_i$  of radius  $r_i$  and with centre at  $v_i$  contains  $\bar{U}_i$ .

The objects  $O_i$ ,  $r_i$ ,  $W_i$  are the objects which are constructed at the  $i$ th step.  $B_i = B_{i-1} A_{p_i, n}$  is the matrix for the transformation from the co-ordinates  $O_i$  into the  $O_0$  system, and  $x_i$  is the column of  $O_0$ -co-ordinates of the centre of  $W_i$ .

### MMCG 2.

Suppose the reversion to MMCG took place when  $i = M_f$ . The work of the method then ceases with the output of the result:

\* if  $j(s) > 0$  for all  $s \leq M_f$ ,

$x_{i_0}$ , where  $i_0$  is such that  $j(i_0) = 0$  and  $g_{i_0}^0(x_{i_0}) = a_{M_f}$ .

### 2.5.4

**Theorem.** Let  $C \in C^{\infty}(G, R^n, m)$  and  $1 > v > v_0$ . Then the method MMCG constructed for the relative accuracy  $v$  solves with this accuracy all problems in  $C$  with a laboriousness not exceeding

$$\Phi_{n, \infty}(v - v_0) = \left[ \frac{\ln(\beta(v - v_0))}{\ln(1/\beta(n))} \right] + 2. \quad (5.5)$$

**Remark.** Thus, as regards laboriousness, MMCG loses to MCG by a factor  $O(n)$  (asymptotically as  $v \rightarrow 0$ ). This is the cost of the 'conservatism' of the MMCG, of ensuring the stability of shape of the domain, constructed by the

method, of localization of the solution. We point out also that the bound for the laboriousness of the MMCG depends on  $\beta$  (which, clearly, may be supposed to be not greater than  $\alpha_{2,n}(G)$ , the minimal asphericity of  $G$  relative to all possible Euclidean normalizations of  $R^n$ ). It is known that  $\alpha_{2,n}(G) \leq n$ , and so, with a suitable choice of  $O_0, W_0$ , we can take  $\beta = n$ . It is true that such a choice may be difficult to achieve. Incidentally, the effect of  $\beta$  on the laboriousness is, after all, only logarithmic. Best of all would be to apply the MMCG when  $G$  is an ellipsoid (i.e. when it is possible to take  $W_0 = G$  and  $\beta = 1$ ).

We further remark that, *a posteriori*, the MMCG turned out to be a concrete variant of the general method of N.Z. Shor [28] of gradient descent with dilatation of the space. The geometrical ideas on which the MMCG is based lead to a completely determinate choice of the dilatation parameter and for regulation of the step. They enable convergence to be proved without any of the essential, additional hypotheses of the type found in [28] (this, it is true, is because MMCG is a concrete variant of N.Z. Shor's more general method).

*Proof of Theorem 2.5.4.* This is, in the main, a repetition of the proof of Theorem 2.3.2.3. We prove first a bound for the laboriousness. From the rule for forming  $r_i$  it follows that  $r_i \leq \beta(n)r_{i-1}$  for  $i \geq 1$ . Hence  $r_i \leq \beta^i(n)r_0$ . But when  $i < M_f$  we have  $r_i \geq ((v - v_0)/\beta)r_0$  in accordance with MMCG 1.4, and so

$$M_f \leq \left\lceil \frac{\ln(\beta/(v - v_0))}{\ln(1/\beta(n))} \right\rceil + 1,$$

as required.

We now prove the assertion about the accuracy of the method. As always, the result  $\bar{x}$  of applying the MMCG to  $f \in C$  is either \* (which happens if and only if  $a_{M_f} = +\infty$ ), or a point  $x_{i_0} \in G$  such that

$$f_j(x_{i_0}) \leq (v - v_0)r_{i_0}^j + v_0r_j(f) \leq vr_j(f), \quad 1 \leq j \leq m, \quad (5.6)$$

and

$$f_0(x_{i_0}) \leq a_{M_f} + v_0r_0(f). \quad (5.7)$$

Hence it is clear that, if  $f$  is incompatible, then  $v(\bar{x}, f) \leq v$ . Now let  $f$  be compatible, and let  $x^*$  be its solution. To prove the inequality  $v(\bar{x}, f) \leq v$ , we have to show that

$$a_{M_f} \leq (v - v_0)r_0(f) + f_* \quad (5.8)$$

Suppose, on the contrary, that (5.8) does not hold:

$$a_{M_f} > (v - v_0)r_0(f) + f_* \quad (5.9)$$

It is possible that  $g_i^{(i)}$  might be constant for some  $i \leq M_f$ . By definition of  $j(i)$ , this is impossible when  $j(i) = m + 1$ . If  $1 \leq j(i) \leq m$ , then, by definition of  $j(i)$ ,

we have  $g_i^{(i)}(x_i) > 0$  and  $f_{j(i)}(x) > 0$  when  $x \in G$ ; but this is impossible, because  $f$  is compatible, by hypothesis. Thus  $j(i) = 0$ , and if  $g_i^0(y)$  were constant, then we would have  $f_* \geq g_i^0(x_i) \geq a_i \geq a_{M_f}$ , which contradicts (5.9).

Thus, for all  $i$ ,  $1 \leq i \leq M_f$ , we may suppose that  $g_i^0(y)$  is not constant. Let  $G_0 = G$ , and

$$G_i = \begin{cases} \{x \in G_{i-1} \mid g_i^{(i)}(x) < g_i^{(i)}(x_i) - d_i\}, & j(i) \leq m, \\ \{x \in G_{i-1} \mid g_i^{(i)}(x) \leq g_i^{(i)}(x_i) - d_i\}, & j(i) = m+1. \end{cases}$$

From the rules MMCG 1.3–1.4 and the commentary on them it is clear that  $G_i \subset W_i$ . From these same rules it is obvious that, if the reversion to MMCG 2 took place because

$$\frac{d_{M_f}}{\bar{p}_{M_f}^T \bar{p}_{M_f}} \geq r_{M_f-1}^2,$$

then  $\text{int } G_{M_f}$  is empty. If this did not take place, then

$$|G_{M_f}|_n \leq |W_{M_f}|_n = \left(\frac{r_{M_f}}{r_0}\right)^n |W_0|_n < (v - v_0)^n |G|_n$$

(we have used the fact that

$$\frac{r_{M_f}}{r_0} < \frac{v - v_0}{\beta},$$

that the volume element in  $O_i$  is the same as that in  $O_0$ , and that  $|W_0|_n \leq (\beta)^n |G|_n$ ). Thus, in all cases,

$$|G_{M_f}|_n < (v - v_0)^n |G|_n. \quad (5.10)$$

Let  $|G_{M_f}|_n \leq (v' - v_0)^n$ ,  $v_0 \leq v' < v$ . It is possible that  $x^* \notin G_{M_f}$ . This would imply that, for some  $i$ ,

$$\begin{aligned} g_i^{(i)}(x^*) &> g_i^{(i)}(x_i) - d_i, & j(i) = m+1, \\ g_i^{(i)}(x^*) &\geq g_i^{(i)}(x_i) - d_i, & j(i) \leq m. \end{aligned} \quad (5.11)$$

This relation is impossible when  $j(i) = m+1$ , because  $x^* \in G$  (cf. the definition of  $g_i^{m+1}(y)$ ). So  $j(i) \leq m$ , and

$$f_{j(i)}(x^*) \geq g_i^{(i)}(x^*) \geq g_i^{(i)}(x_i) - d_i. \quad (5.12)$$

If  $1 \leq j(i) \leq m$ , then the right-hand side of (5.12) is not less than  $(v - v_0)r_i^{(i)} > 0$  (this comes from the definition of  $j(i)$ ), and so (5.12) is impossible. Therefore  $j(i) = 0$ , but then (5.11) gives  $f_* \geq a_i \geq a_{M_f}$ , which contradicts (5.9).

Thus  $x^* \in G_{M_f}$ . Since  $G_{M_f}$  is open in  $G$  and  $|G_{M_f}|_n \leq (v' - v_0)^n |G|_n$ , there is a  $z \in G$  and a  $y = (v' - v_0)z + (1 - (v' - v_0))x^*$  such that  $y \notin G_{M_f}$ . Then, for some  $i$ , (5.12) holds with  $x^*$  replaced by  $y$  (as before,  $j(i) \neq m+1$ ). If  $1 \leq j(i) \leq m$ , then

(5.12) gives

$$g_i^{j(i)}(y) \geq (v - v_0)r_i^{j(i)} > 0. \quad (5.13)$$

At the same time

$$g_i^{j(i)}(y) = (v' - v_0)g_i^{j(i)}(z) + (1 - (v' - v_0))g_i^{j(i)}(x^*) \leq (v' - v_0)r_i^{j(i)},$$

which contradicts (5.13). Thus  $j(i) = 0$ . But then  $g_i^0(y) \geq a_i$  from (5.12). At the same time

$$\begin{aligned} g_i^0(y) &= (v' - v_0)g_i^0(z) + (1 - (v' - v_0))g_i^0(x^*) \\ &\leq (v' - v_0)g_i^0(z) + (1 - (v' - v_0))f_0(x^*) \\ &\leq (v' - v_0)(\max_{z \in G} g_i^0(z) - f_0(x^*)) + f_* \leq (v' - v_0)r_0(f) + f_*. \end{aligned}$$

Therefore  $a_{M_f} \leq a_i \leq (v' - v_0)r_0(f) + f_*$ , which contradicts (5.9). The contradiction proves the theorem.

### 2.5.5

Like MCG, the MMCG can be adapted to solve problems  $f \in C \in C^{v_0}(G, R^n, m)$  with the absolute errors  $\varepsilon_j > 0$ ,  $0 \leq j \leq m$ , specified in advance. To do this, the method is modified as follows.

The rule for forming  $j(i)$  in MMCG 1.2 is replaced by

$$j(i) = \begin{cases} 0 & \text{if } g_i^j(x_i) \leq \varepsilon_j, \quad 1 \leq j \leq m \\ j \in 1, \dots, m & \text{such that } g_i^j(x_i) > \varepsilon_j \text{ if there is such a } j. \end{cases}$$

The rule MMCG 1.4 is replaced by

**MMCG 1.4'.** Put

$$\begin{aligned} a_i &= \begin{cases} a_{i-1}, & j(i) > 0, \\ \min \{a_{i-1}, g_i^0(x_i)\}, & j(i) = 0, \end{cases} \\ d_i &= \begin{cases} 0, & j(i) = 0, \\ g_i^{j(i)}(x_i) - \varepsilon_{j(i)}, & 0 < j(i) \leq m, \\ g_i^{m+1}(x_i), & j(i) = m+1. \end{cases} \end{aligned}$$

If  $\bar{p}_i = 0$ , go to MMCG2. Otherwise let

$$r_i = \beta(n) \sqrt{r_{i-1}^2 - \frac{d_i^2}{(\bar{p}_i^T \bar{p}_i)}},$$

$$x_{i+1} = x_i - \left( \frac{d_i}{\sqrt{\bar{p}_i^T \bar{p}_i}} + \frac{r_i}{\beta(n)(n+1)} \right) B_{i-1} p_i.$$

Define

$$\begin{aligned} r_i^0 &= \max_{z \in G} g_i^0(z) - g_i^0(x_i), \\ v_i &= \begin{cases} 1, & j(i) = m+1, \\ \frac{\varepsilon_{j(i)}}{r_i^{j(i)}}, & 1 \leq j(i) \leq m, \\ \frac{\varepsilon_0}{\varepsilon_0 + r_i^0}, & j(i) = 0. \end{cases} \end{aligned}$$

Let  $v^i = \min_{0 \leq s \leq i} v_s$  ( $v_0 = 1$ ). If  $r_i < (v^i r_0)/\beta$ , go to MMCG 2; otherwise increase  $i$  by 1 and go to MMCG 1.

We leave the reader to prove that the MMCG modified in this way has the following properties:

- (i) The result of applying it to any problem  $f \in C \in C^{v_0}(G, R^n, m)$ , regarded as a solution to  $f$ , has absolute errors which, up to the accuracy of the irremovable errors of the oracle, do not exceed the specified values  $\varepsilon_j$  (i.e. the condition (4.1) holds for the result  $\bar{x}$  of applying the method to  $f$ ).
- (ii) In addition, the laboriousness of the method described does not exceed  $\Phi_{n,\infty}(\bar{v}(f, \varepsilon))$ , where

$$\bar{v}(f, \varepsilon) = \frac{v(f, \varepsilon)}{1 + v_0 + v(f, \varepsilon)},$$

and  $v(f, \varepsilon) = \min_j \varepsilon_j / r_j(f)$  is the maximal relative error which ensures the specified absolute errors.

Both of the presented versions of the MMCG can, of course, be applied to problems in classes of the  $C_{\text{Lip}}^{v_0}(G, R^n, \|\cdot\|, m)$  type, just as they can to problems in classes of the  $C^{v_0}(G, R^n, m)$  type.

### 2.5.6

Let us say a few words about the practical realization of the method described. It requires the ability to calculate the values and derivatives of the functions  $f_{m+1}(x) = \rho_{1,1}(x, G)$ . In connection with the questions arising here the reader is referred to Section 3.3.3.4. In addition, the method requires the storing in memory, and the recurrent re-calculation of the  $n \times n$  matrix  $B_i$ . In accordance with Lemma 2.5.2 this re-calculation is carried out by means of the formula

$$B_{i+1} = \lambda_n B_i + \mu_n B_i P_{e_i},$$

where  $\lambda_n$  and  $\mu_n$  depend only on  $n$  quantities, expressions for which are given in (5.2)–(5.3), and  $P_{e_i}$  is the matrix of the ortho-projector on to the unit column-vector  $e_i = (e_i^1, \dots, e_i^n)$ . The elements of this matrix are  $p_{ki} = e_i^k e_i^l$ .

We point out that the product  $B_i P_{e_i}$  can be computed, not after  $O(n^3)$  arithmetical operations as is usually the case, but after only  $O(n^2)$  operations. For, it can be found in this way: compute the vector  $\bar{e}_i = B_i e_i$  and construct the matrix whose  $j$ th column is  $e_j / \bar{e}_i$ . This will be  $B_i P_{e_i}$ . It is clear that, in addition,  $O(n^2)$  arithmetical operations will be needed to obtain  $B_{i+1}$ . The expenditure in operations for organizing one step of the method (without counting the expenditure in operations of the oracle and in the comparisons needed to compute  $j(i)$ ) will also be of the order  $O(n^2)$ , and so the number of elementary operations for the organization of a method for an accuracy  $v$  will be  $O(n^4) \ln(\beta/v)$ .

### 2.5.7

We have seen that the methods with linear convergence which have been described in this chapter have, for a given accuracy, a laboriousness which is the greater, the greater the dimensionality of the problem under solution. Sometimes, however, there are grounds for hoping that the functionals of the problem, which, *a priori*, depend on a large number of variables, will actually be independent (or almost independent) of the great majority of these variables. More precisely, by a suitable (but not known in advance) transformation of the co-ordinate system it is possible to arrange things so that the components of the problem will depend only on a small proportion of all the variables. If we knew these co-ordinates in advance, then it would be possible from the very start to reduce the problem to one of smaller dimension and, in the case of the methods described in this chapter, to gain as regards laboriousness.

It turns out that, in practice, it is possible to obtain this effect without knowing in advance the 'essential' variables. That is, the methods MCG and MMCG can be given 'a form adaptive to the true dimensionality of the problem'. The corresponding construction is described in the authors' paper [22], and we refer the interested reader to this. Here it suffices to mention that the adaptive versions of these methods can be applied to any (Lipschitz-convex finite-dimensional) problem. The bound for the laboriousness of 'these versions on the classes  $C_{\text{Lip}}(G, E'', |\cdot|, m)$ , where  $|\cdot|$  is the Euclidean norm, is almost the same as for the non-adaptive methods. But then in solving problems with a small number of 'essential' variables the laboriousness of the adaptive versions turns out to be approximately the same as when the non-adaptive methods are used in the 'true' dimensionality of the given problem.

Losing nothing to the methods described in this chapter on the class of all convex problems of a given dimension, the adaptive versions thus gain markedly as regards laboriousness on the subclasses of problems of small 'essential' dimension, and it is this which accounts for their attractiveness.

# 3

## Methods of mirror descent

In this chapter a general construction is described which enables us to construct an extensive family of methods of convex optimization. A special feature of these methods is that their laboriousness does not depend explicitly on the dimension of the problem. Accordingly, it is sensible to use them for solving convex problems of high dimension. The idea of the construction is described in Section 3.1; Section 3.2 contains certain preliminary results. Methods of solving Lipschitz-convex problems are given in Section 3.3, and those for general convex problems are given in Section 3.4. The results of this chapter are based mainly on the authors' paper [23].

### 3.1 THE IDEA OF THE METHODS

#### 3.1.1

The methods of solution of general convex problems in the previous chapter cannot satisfy us completely in either the practical or the theoretical aspect. The practical drawbacks basically inherent in the MCG are obvious. As regards the theoretical aspect, we may point out the following. The MCG, as we shall see, cannot be improved essentially as regards laboriousness if we are considering the solution of general convex problems on *parallelopipeds*  $G$ . For arbitrary convex bodies  $G$ , essentially the method is incapable of improvement only in the asymptotics relative to the accuracy. Moreover, the 'accuracy starting at which the asymptotic behaviour is established' depends on the affine properties of  $G$ .

Suppose, for example,  $G$  is an  $n$ -dimensional ellipsoid. Then, as will be proved in Section 4.1, the complexity of the class of convex problems on  $G$  is of the form  $O(1/v^2)$ , with  $v \geq 1/\sqrt{n}$ , while the bound for the laboriousness of the MCG is  $O(n \ln(1/v))$ . Therefore, if it is required to solve convex problems on ellipsoids with a fixed accuracy  $v$ , and the dimension of the ellipsoid is sufficiently high ( $n \gg 1/v^2$ ), then there are no theoretical reasons for using the MCG. Thus, general convex problems on ellipsoids can be solved with a

laboriousness which does not depend at all on the dimension of the problem, whereas the laboriousness (and therefore the 'degree of non-optimality') of the MCG increases linearly with the dimension.

From the practical point of view the effect of a 'uniform (relative to the dimension) boundedness of the complexity' is very attractive, at any rate, as applied to economics problems. For, the requirements regarding the accuracy of solution of these problems are usually not too great, whereas their dimensionality may be very considerable. Therefore the construction of convex-programming methods which are 'not sensitive to the dimensionality' is quite a pressing question. It is precisely in this direction that the following analysis is carried out.

### 3.1.2

Clearly, it is impossible to put forward a method of solving general convex problems with a bound for the laboriousness independent of the dimension, unless definite hypotheses are made about the affine properties of the corresponding body  $G$ . For, we have already seen that the complexity of a class of general convex problems on a parallelepiped increases linearly with increase in the dimensionality of the latter. Thus, in constructing a method of convex optimization 'insensitive to dimensionality', we have somehow or other to distinguish the necessary affine properties of the domain  $G$  considered, and to understand which properties of ellipsoids ensure the uniform (relative to dimensionality) boundedness of the complexity of solutions of convex problems on ellipsoids. Anticipating a little, we shall give an answer to this question, which may possibly seem rather strange at first; but after the discussion given below, it will appear perfectly natural.

An ellipsoid can be regarded as a unit ball in a Euclidean space  $E$  with the Euclidean norm  $\|\cdot\|_2$ . We consider the space  $E^*$ ,  $\|\cdot\|_*$ , dual to  $E$  (it is, of course, canonically identified with  $E$ ,  $\|\cdot\|_2$ , but we shall see that, from the point of view of generalizations, it is important to go over to the dual space). On  $E^*$  there is a function  $V(\xi)$ , namely  $\frac{1}{2}\|\xi^2_*\|$  which has the following property.  $V(\xi)$  is uniformly differentiable (i.e. the function  $V(\xi)$  is uniformly continuous on bounded subsets of  $E^*$ ) and it increases at infinity faster than any linear function. In addition, its rate of growth and the modulus of continuity of the derivative do not depend on the dimension of  $E$ . There it is: everything is bound up with the existence of a function with these properties on  $E^*$ .

Let us make this assertion precise. Suppose there is a reflexive space  $E$  with a norm  $\|\cdot\|$ , whose dual admits the existence of a function  $V$  having the properties mentioned. We shall be able to associate with  $V$  a method of solving general convex problems on balls in  $E$ , a method with a bound for its laboriousness which depends (apart from on the accuracy) only on the modulus of continuity of  $V$  and on the rate of growth of  $V$  at infinity. We point

out that dimensionality does not appear explicitly in our assumptions. In fact, we impose restrictions only on the affine type of  $G$  (namely, in so far as this type determines the possibility of realizing  $G$  as a unit ball in some suitable norm).

The last remark is extremely important for understanding the idea underlying all that follows. For, we start with general convex problems—of a class determined in purely affine terms, irrespective of any norms, whereas all the subsequent exposition will be carried out in the language of norms. The methods proposed below will also be described in these terms. The reader should not be surprised by this lack of co-ordination. Normalization will, in a certain sense, be only an instrument. In its proper place it will be made clear that the possibility of using this instrument for solving general convex problems on a body  $G \subset R^n$  is connected with the affine properties of  $G$ .

### 3.1.3

Let us outline our plan of action. We start by considering classes of *Lipschitz-convex* problems on bounded, convex  $G$  in normed spaces  $E$ ,  $\|\cdot\|$ . Under definite hypotheses regarding  $E$ ,  $\|\cdot\|$  (they have, in fact, been formulated above) we learn how to solve the problems mentioned with an arbitrary accuracy, for a laboriousness which depends only on this accuracy, the Lipschitz constant of the problems being solved, the diameter of  $G$  in the norm considered, and certain characteristics of  $(E, \|\cdot\|)$  which do not include the dimension of  $E$  explicitly (the latter could even be infinite, for example). After this we learn how to solve *general* convex problems on bodies  $G$ . Further, in the bound for the laboriousness, the asphericity of  $G$  relative to the norm adopted in  $E$  will also appear. We remark that in this chapter the space  $E$  is not invariably assumed to be separable.

It would be possible to make the following objection to the programme we have just sketched out. Infinite-dimensional problems, from the computational point of view, are mere fictions. Numerical methods work actually only with finite-dimensional problems. But in the finite-dimensional case all norms are equivalent to one another and therefore to the Euclidean norm. It would appear to be sufficient to consider the Euclidean case, and that more general (and therefore more cumbersome and less transparent) considerations—these are all the works of the Devil!

However, this line of reasoning is deceptive. To explain this, let us consider first the case of Lipschitz problems. The fact itself that a convex function  $f$  is Lipschitz on a bounded, convex, closed body  $G \subset R^n$  does not depend on the choice of the norm in  $R^n$  which gives the metric; but  $\rho_{1,1}(G)$ , the radius of  $G$ , i.e. the radius of the minimal ball containing  $G$  and with its centre in  $G$ , and the Lipschitz constant  $L_{1,1}(f)$  of the function  $f$  depend essentially on the choice of the norm. The laboriousness of a method which is associated with a given norm and which is provided by the suggested approach and which ensures an

absolute error  $\varepsilon$  is of the form  $\Phi(\|\cdot\|; L_{1,1}(f) \rho_{1,1}(G)/\varepsilon)$ . When the norm  $\|\cdot\|$  changes, both arguments of  $\Phi$  change, and it is by no means necessary that the 'best' norm should be the Euclidean one (although for a fixed  $t$ , the function  $\Phi(\|\cdot\|, t)$  attains a minimum relative to  $\|\cdot\|$  actually for the Euclidean norm).

Let us give a rather effective example. Suppose we have to solve the problem

$$f(x) = \sum_{i=1}^n f_i(x_i) \rightarrow \min, \quad x_i \geq 0, \quad \sum_{i=1}^n x_i = 1,$$

where the  $f_i(x)$  are Lipschitz-convex functions (with Lipschitz constant 1) on the axis. A problem of this kind is, in a certain sense, very typical. To solve the problem we could apply a method associated with the Euclidean norm  $\|x\|_2 = \sqrt{\sum x_i^2}$  (the method would then be the usual gradient method). The Lipschitz constant of  $f$  in this norm will be  $\sqrt{n}$  (and, in general, it could not be less), and the diameter of the domain is equal to  $\sqrt{2}$ . In addition,  $\Phi(\|\cdot\|, t) = O(1/t^2)$ , and so the 'Euclidean' method with absolute accuracy  $\varepsilon$  will have a bound  $O(n/\varepsilon^2)$  for the laboriousness. A method of the same accuracy, associated with the norm  $\|x\| = \sum_{i=1}^n |x_i|$ , will have, under the same hypotheses, a bound  $O(\ln n/\varepsilon^2)$  for the laboriousness, winning appreciably, as we see. We shall meet a situation of this sort in Chapter 6, where it will become especially clear that we do not go beyond the Euclidean situation just for nothing.

In the case of general convex problems the situation is roughly the same. A method, associated with a norm  $\|\cdot\|$  of solving general convex problems on a body  $G$  has a bound for the laboriousness of the form  $\Phi(\|\cdot\|, \alpha_{1,1}(G)/v)$ , where  $v$  is the required relative accuracy, and  $\alpha_{1,1}(G)$  is the asphericity of  $G$  in the norm  $\|\cdot\|$ . The effect of the quantity  $\alpha_{1,1}(G)$  on the laboriousness may lead to the result that the best bound will correspond to a norm different from the Euclidean one. For example, if  $G$  is a 'hyper-octahedron'  $\{x \in R^n \mid \sum_{i=1}^n |x_i| \leq 1\}$ , then the best 'Euclidean' method of solving general convex problems on  $G$  with accuracy  $v$  will have a laboriousness of order  $O(n/v^2)$ , while a method corresponding to the norm  $\|\cdot\|_1$  will have a laboriousness only of order  $O(\ln n/v^2)$ . We believe that these arguments should convince the reader that the generality of the following considerations is not introduced merely for the sake of generality.

### 3.1.4

We now describe the idea of the proposed construction of methods. The simplest way to understanding the essence of the matter is to consider the problem of minimizing a convex Lipschitz function  $f$  on a Banach space  $(E, \|\cdot\|)$ . We suppose that on the space  $E^*$ ,  $\|\cdot\|_*$  dual to  $E$  there is defined a function  $V(\xi)$  which is uniformly differentiable and which increases at infinity

faster than any linear function. We further assume that  $f$  has a minimum at  $x^*$  and that  $E$  is reflexive. We intend to associate with  $V$  a method of minimizing  $f$ . To simplify the exposition of the idea of the construction, we describe a 'continuous analogue' of the method (the method itself is then obtained by a natural discretization).

Let  $\varphi \in E^*$ . Then  $V'(\varphi)$  is a linear functional on  $E^*$ , i.e. it is an element of  $E$  (because  $E$  is reflexive). Therefore we can consider a trajectory  $\varphi(t)$  of the differential equation

$$\frac{d\varphi}{dt} = -f'(V(\varphi(t))), \quad \varphi(0) = \varphi_0. \quad (1.1)$$

The mapping  $V'$  carries  $E^*$  into  $E$ , and  $f'$  carries  $E$  into  $E^*$ , and so the right-hand side is an element of  $E^*$ .

As well as the trajectory  $\varphi(t)$  we consider its 'shadow' or 'image'  $x(t) = V'(\varphi(t))$ . We introduce the function  $V_*(\varphi) = V(\varphi) - \langle \varphi | x^* \rangle$ , and we verify that it decreases along the trajectory  $\varphi(t)$ . For, we have

$$\begin{aligned} \frac{dV_*(\varphi(t))}{dt} &= \langle \varphi'(t) | V(\varphi(t)) - x^* \rangle = -\langle f'(V(\varphi(t))) | V(\varphi(t)) - x^* \rangle \\ &= \langle f'(x(t)) | x^* - x(t) \rangle \leq f(x^*) - f(x(t)) \leq 0. \end{aligned}$$

Further, we see that if  $\lim_{t \rightarrow \infty} f(x(t)) = f(x^*)$ , (i.e. if the trajectory  $x(t)$  badly approximates to  $x^*$  all the time along the functional), then

$$\frac{dV_*(\varphi(t))}{dt} \leq -\varepsilon < 0,$$

and so  $V_*(\varphi(t))$  decreases without bound. But this is impossible, because  $V(\varphi)$  increases at infinity faster than any linear function, and so  $V_*(\varphi)$  is bounded below on  $E^*$ . Thus  $\lim_{t \rightarrow \infty} f(x(t)) = f(x^*)$ , and the method, with a suitable rule for forming the result at the time  $t$ , converges along the functional.

It is now easy to 'discretize' this argument and to make it 'constructive', i.e. to indicate explicitly after what number of steps of the discretized method the accuracy  $v$  will be attained. It can be seen that this number of steps depends on the Lipschitz constant of  $f$ , the modulus of continuity of  $V$  in a suitable ball, and  $v$  (these constants together determine the step of the discretization, a step which does not lead to a divergence, 'dangerous' for our purposes, of the trajectories of the 'discretized' and 'continuous' methods), and also on  $\|x^*\|$  and the rate of growth of  $V$  at infinity. (These quantities determine the possible minimum value  $V_*$ , i.e. the time after which  $V_*(\varphi(t))$  would be able to decrease rapidly.)

We shall not carry out the corresponding analysis immediately; it will be expounded in a more general form later, and we shall also take the presence of constraints into account. In the construction of methods the simple ideas

which we have set out will, of course, be masked by a number of secondary details, and so the full descriptions might appear to be cumbersome. However, the authors do not have space to explain the nature of this or that detail. The reader will be able, if he wishes, to follow along the track from the elementary original premises to the resulting schemes.

Returning to the idea of the construction, we point out that the main motion actually takes place in the dual space. The motion in  $E$  is only a 'shadow' or 'image' of the main motion; it is this fact which has led us to choose the name, the *method of mirror descent*. The well-versed reader will at once notice that  $V(\varphi)$  (more precisely,  $V_*(\varphi)$ ) is simply the Lyapunov function of the method associated with  $V$ . The use of Lyapunov functions is the traditional method of proving convergence of iterative procedures. However, usually the procedure is already known, and for it a Lyapunov function is sought. We are working in the reverse order: we choose a Lyapunov function (it turns out that this can be almost any function on  $E^*$ ), and it is from this that we construct the method.

Let us now look at what happens to the proposed construction when  $E, \|\cdot\|$ , is a Hilbert space. In this case,  $E^*, \|\cdot\|_*$  is canonically identified with  $E$ , and we can take  $V(\varphi) = \frac{1}{2}\varphi^2$ , i.e.  $V'(\varphi) = \varphi$ . The process (1.1) becomes the process

$$\frac{dx(t)}{dt} = -f'(x(t)), \quad x(0) = x_0,$$

i.e. the usual gradient descent. The function  $V_*$  which decreases along the trajectory of descent is, up to a constant,  $\frac{1}{2}(x - x^*)^2$ . The convergence of the usual gradient method is proved precisely by following the behaviour of this function (i.e. the distance from  $x^*$ ) along the trajectories (we are thinking of the minimization of convex, but not necessarily smooth, functions).

If we try to extend directly this sort of scheme to prove the gradient method on a space  $E, \|\cdot\|$ , which is different from a Euclidean space, we get nowhere; on  $E$  there will be nothing like a decrease of the distance from  $x^*$ . Thus, from the general standpoint, it is a fortuitous circumstance that, in the case of a Hilbert space, the standard Lyapunov function of the gradient method is defined on  $E$  and not on  $E^*$  (because  $E$  is identified with  $E^*$ ) and that it has a simple geometrical meaning; this circumstance only obscures the true nature of the gradient method.

In conclusion we point out the following. The scheme outlined seems, at a first glance, to have a colossal freedom, namely, the choice of the function  $V$ . For example, in the Euclidean situation it would be possible, along with the usual gradient method, to consider also a method associated with, say, the function  $\frac{1}{2}x^2 + \cos^2|x|$ . The situation is thus fraught with a sort of chaos. However, nothing terrible happens. In those cases where we develop the general scheme into concrete methods, their right to exist will be based on their theoretical sub-optimality (which is proved by comparing the laboriousness of the methods with the lower bounds of the complexity of the corresponding

classes). It is true that the method associated, up to a constant, with the function  $\frac{1}{2}x^2 + \cos^2|x|$  is just as effective as the usual gradient method corresponding to  $\frac{1}{2}x^2$ . In situations of this kind it is sensible to prefer the more 'natural' methods, and this is just what we do.

We now end this rather extended 'introduction of the idea' and get down to business. We begin with a special study of the pair  $E, \|\cdot\|$ , which we are constructing.

## 3.2 REGULAR SPACES

In this section we shall make a particular study of a class of normed spaces that suit our purposes.

### 3.2.1

Let  $E$  be a real Banach space with a norm  $\|\cdot\|$ , and let  $E^*, \|\cdot\|_*$ , be the dual space.

**DEFINITION.** A space  $E, \|\cdot\|$ , is said to be *regular* if

- (1)  $E$  is reflexive;
- (2) there is a uniformly differentiable function  $V: E^* \rightarrow \mathbb{R}$  such that

$$V(\xi) \geq \theta(\|\xi\|_*), \quad \theta(t) \rightarrow \infty \text{ as } t \rightarrow \infty.$$

**Caution.** It should not be thought that the requirement (2) is always satisfied automatically. This is so, of course, in the finite-dimensional case (moreover, every finite-dimensional space satisfies (1), and so it is regular). In the infinite-dimensional case, (2) may not be satisfied even though (1) is satisfied.

For the following exposition it is convenient to normalize the functions  $V$  (which feature in the definition of regular spaces) in a particular way.

**DEFINITION.** Let  $E, \|\cdot\|$ , be a regular space. A function  $V: E^* \rightarrow \mathbb{R}$  is said to be a *regular function corresponding to  $E, \|\cdot\|$*  (or an  $(E, \|\cdot\|)$ -regular function) if

- (i)  $V(\varphi) \geq 0, V(0) = 0$ ;
- (ii)  $\|\varphi\|_* \leq \frac{1}{2} + V(\varphi)$ ;
- (iii)  $V(\varphi)$  has a derivative  $V'(\varphi)$  which is uniformly continuous and bounded over the whole of  $E^*$ .

It can be proved that if  $E, \|\cdot\|$ , is a regular space, then there are always regular functions corresponding to  $E, \|\cdot\|$ .

We give a method of constructing regular functions, a method which, as we shall soon see, is a universal one. Let  $\|\cdot\|$  be a norm on  $E$  such that the function  $\|\varphi\|_*$  is uniformly differentiable in a neighbourhood of the unit sphere in  $E^*$

(i.e. there is a function  $p(\xi): E^* \rightarrow E$  such that

$$\frac{1}{t} \| \varphi + t\xi \|_* - \| \varphi \|_* - t \langle \xi | p(\varphi) \rangle \underset{t \rightarrow 0}{\rightarrow} 0$$

uniformly with respect to  $\varphi$  and  $\xi$  with  $\| \varphi \|_* = \| \xi \|_* = 1$ ). In this case we may take as an  $(E, \|\cdot\|)$ -regular function

$$V_{E, \|\cdot\|}(\varphi) = \begin{cases} \frac{1}{2} \|\varphi\|_*^2, & \|\varphi\|_* \leq 1, \\ \|\varphi\|_* - \frac{1}{2}, & \|\varphi\|_* \geq 1. \end{cases} \quad (2.1)$$

*Exercise 1.* Prove that, under the above hypotheses, (2.1) does indeed define an  $(E, \|\cdot\|)$ -regular function.

We point out that the property that  $E, \|\cdot\|$ , is a regular space is, obviously, a property of the topology of  $E$  and not of the norm (i.e. it is preserved if  $\|\cdot\|$  is replaced by an equivalent norm  $\|\cdot\|'$ ). On the other hand, the property that  $V$  is an  $(E, \|\cdot\|)$ -regular function depends on the actual norm on  $E$  and not on the topology of  $E$ .

### 3.2.2

The purpose of this chapter is to describe the constructions which associate with  $(E, \|\cdot\|)$ -regular functions  $V$  methods of complex optimization (we shall call them *MD-methods*, from the name ‘methods of mirror descent’). The characteristics of the MD-methods associated with  $V$  depend on certain properties of  $V$ . These properties we are about to describe.

Let  $(E, \|\cdot\|)$  be a regular space, and let  $V(\varphi)$  be a corresponding regular function. We consider the mapping  $\varphi \mapsto V'(\varphi)$ , defined on  $E^*$ .  $V'(\varphi)$  is a continuous linear functional on  $E^*$ , i.e. an element of  $(E^*)^*$ . Since  $E$  is reflexive,  $(E^*)^*$  is canonically identified with  $E$ , and so we shall always suppose that  $V'(\varphi)$  takes values in  $E$ . By hypothesis, the function  $V'(\varphi): E^* \rightarrow E$  is uniformly continuous, and so we can speak of its modulus of continuity

$$\omega_V(t) = \sup \{ \|V'(\varphi) - V'(\psi)\| \mid \varphi, \psi \in E^*, \|\varphi - \psi\|_* \leq t \}.$$

We shall call the function  $\omega_V(t)$  the *modulus of smoothness* of  $V$ . It will play a fundamental rôle in estimating the laboriousness of the MD-method associated with  $V$ . It is clear from the definition that  $\omega_V(t)$  is a non-decreasing, non-negative function which is bounded on the half-line  $t \geq 0$ , is continuous, and tends to 0 as  $t \rightarrow +0$ .

The function inverse to  $\omega_V(t)$

$$\gamma_V(s) = \sup \{ t \mid \omega_V(t) \leq s \}$$

will play an essential part in what follows. It is clear that  $\gamma_V(s) > 0$  when  $s > 0$ , that  $\gamma_V(s)$  does not decrease as  $s$  increases, and that  $\omega_V(\gamma_V(s)) \leq s$ .

The laboriousness of the method associated with  $V$  is of order  $O(1/(v\gamma_V(t)))$ , i.e. it is the smaller, the greater  $\gamma_V(t)$  is, and therefore the smaller, the smaller  $\omega_V(t)$  is (the ‘smoother’  $V$  is). Accordingly we shall focus our attention on bounding  $\omega_V(t)$  below.

### 3.2.3

We now investigate the question of the stock of regular functions. It turns out that they can all be described in the standard terms of functional analysis. Namely, a Banach space  $E$  is regular if and only if the topology on  $E$  can be defined by means of a *uniformly-convex* norm  $\|\cdot\|$ . A norm  $\|\cdot\|$  is said to be *uniformly convex* if the *sagitta* of a not too small chord of the unit ball is itself not too small. (In the 2-dimensional case the *sagitta* is the versed sine. *Trans.*). More precisely, for a certain function  $\kappa(s)$  which is non-decreasing and positive when  $s > 0$ , the relation

$$\left\| \frac{x+y}{2} \right\| \leq 1 - \kappa(\|x-y\|)$$

must hold for all  $x, y$  with  $\|x\| = \|y\| = 1$ . The usual Hilbert norm is an example of a norm of this kind.

It is known (Shmulyan’s theorem, [6], Chapter 5) that a norm  $\|\cdot\|$  on a reflexive space  $E$  is uniformly convex if and only if the dual norm  $\|\cdot\|_*$  on  $E^*$  is uniformly differentiable in a neighbourhood of the unit sphere in  $E^*$  (cf. Section 3.2.1). Thus, as a regular function corresponding to a uniformly convex normed space  $E$  with a uniformly convex norm  $\|\cdot\|$  we can take  $V_{E, \|\cdot\|}(\cdot)$ . Conversely, it can be proved that, if  $E$  is regular, then  $E$  admits the introduction of a uniformly convex norm.

We now give some examples of regular spaces. First of all, a Hilbert space is regular (as already mentioned, its natural norm is uniformly convex; however, a regular function for a Hilbert space can easily be indicated directly). Consequently all finite-dimensional spaces are regular (they can be given the topology by means of the Euclidean norm). Another whole family of examples consists of the standard  $L_p(T, \mu)$  spaces.

The definition of these spaces is given in Section A.2.8. It is explained there that, when  $1 < p < \infty$  the spaces  $L_p(T, \mu)$  are reflexive. Also, the usual norm  $\|\cdot\|_p$  on an  $L_p(T, \mu)$  space is uniformly convex, and so the space is regular, for  $1 < p < \infty$ . As the corresponding regular functions we may take

$$V_{L_p, \|\cdot\|_p}(\varphi) = \begin{cases} \frac{1}{2} \|\varphi\|_q^2, & \|\varphi\|_q \leq 1, \\ \|\varphi\|_q - \frac{1}{2}, & \|\varphi\|_q \geq 1. \end{cases} \quad (q = p/(p-1)) \quad (2.2)$$

We shall denote these functions simply as  $V_p(\cdot)$ .

The modulus of smoothness  $\omega_{\nu_p}(t) \equiv \omega_p(t)$  and the inverse function  $\gamma_p(t) \equiv \gamma_{\nu_p}(t)$  are calculated in Section 3.5. It is shown there that the function  $\gamma_{\nu_p}(t)$  admits the following lower bound:

$$\gamma_p(t) \geq \begin{cases} c(p)t, & 1 < p \leq 2, \\ c(p)t^{p-1}, & p > 2. \end{cases} \quad (2.3)$$

Here  $c(p) > 0$  depends only on  $p$ . The form of this function will be indicated in Section 3.5.

When  $p = 1$  (and also when  $p = \infty$ ) the space  $L_p(T, \mu)$  is not reflexive (and so it is not regular) if its linear dimension is infinite. But if its dimension is finite and equal to  $n$ ,  $1 \leq n < \infty$ , then these spaces are, of course, regular. We shall be particularly concerned with the space  $L_1$ , i.e.  $L_1^{(n)}$ ,  $1 \leq n < \infty$  (see Section A.2.8).

The space dual to  $L_1^{(n)}$  is  $L_\infty^{(n)}$ . As an  $L_1^{(n)}$ -regular function we can take any one of the functions

$$V^*(\varphi) = \begin{cases} \frac{1}{2}\|\varphi\|_\tau^2, & \|\varphi\|_\tau \leq 1, \\ \|\varphi\|_\tau - \frac{1}{2}, & \|\varphi\|_\tau \geq 1. \end{cases} \quad 1 < \tau < \infty \quad (2.4)$$

It is sensible to choose  $\tau$  so as to maximize the smoothness of  $V^*$ . The appropriate analysis is carried out in Section 3.5. It leads to choosing  $\tau$  in the form  $c \ln n$  (we assume  $n > 1$ , and  $c$  is an absolute constant). For the corresponding function  $V \equiv V_{1,n}(\cdot)$  the following estimate holds:

$$\gamma_{V_{1,n}}(s) \geq c(1) \frac{s}{\ln n} \quad (2.5)$$

with an absolute constant  $c(1) > 0$ . Thus the smoothness characteristics of  $V_{1,n}$ , as would be expected, are spoiled as  $n$  increases, but only quite slowly.

### 3.2.4

In conclusion we formulate an elementary bound for the increment of the function  $V$ , a result which will be very useful later.

**Proposition.** Let  $E, \|\cdot\|$ , be a regular space, and let  $V$  be a regular function corresponding to  $E, \|\cdot\|$ . We introduce the function

$$V_*(\varphi) = V(\varphi) - \langle \varphi | x \rangle, x \in E.$$

Further let  $\varphi, \xi \in E^*$ . Then

$$V_*(\varphi + \xi) \leq V_*(\varphi) + \langle \xi | V_*(\varphi) \rangle + \|\xi\|_* \omega_V(\|\xi\|_*). \quad (2.6)$$

**Exercise 2.** Prove (2.6)

### 3.3 MD-METHODS ON CLASSES OF LIPSCHITZ-CONVEX PROBLEMS

In this section methods of mirror descent for solving Lipschitz-convex problems on regular spaces are constructed.

Let  $E, \|\cdot\|$ , be a regular space, let  $G$  be a bounded, convex, closed set in  $E$  of radius  $\rho_{1,1}(G)$ , and let  $m$  be a non-negative number. We shall study methods of solving classes of Lipschitz-convex problems, i.e. *classes of the type*  $C_{Lip}^*(G, E, \|\cdot\|, m)$  (see Section 2.2).

#### 3.3.1

We denote an  $(E, \|\cdot\|)$ -regular function by  $V(\cdot)$ . We shall describe the method of circular descent (denoted by  $MD_V$ ) associated with  $V$  of solving problems of a class  $C_{Lip} \in C_{Lip}^*(G, E, \|\cdot\|, m)$ . We shall describe the method on the assumption that  $\rho_{1,1}(G) = 1$  and that the 'centre' of  $G$  is 0. This is obviously always achievable by a change of variables, by a dilatation of, and shift in,  $G$ .

The laboriousness of  $MD_V$  with an accuracy of solution  $v > v_0$  is  $M_V(v - v_0)$ , where

$$M_V(v) = \left[ \frac{2}{v \gamma_V(v/4)} \right] + 2 \quad (3.1)$$

We introduce the following notation. Let  $\pi_G(x): E \rightarrow G$  be a function which carries a point  $x \in E$  into the point of  $G$  which is closest (in the metric defined by the norm in  $E$ ) to the point  $x$ . Further let  $\mu_G(x): E \rightarrow E^*$  associate with a point  $x \in E$  the support functional at  $x$  to the function  $\rho_{1,1}(\cdot, G)$ .

We pass on to the description of  $MD_V$  constructed for an error  $v > v_0$ . The case  $v \geq 1$  is trivial. With such a low demand on the accuracy of the solution, the problem can be solved in one step. Any point of  $G$  is a solution of any problem  $f \in C_{Lip}$  with error equal to 1.

So we can now suppose that the specified accuracy  $v$  satisfies  $1 > v > v_0$ . The work of a  $MD_V$  method constructed for an accuracy  $v$  with  $1 > v > v_0$  on a problem  $f \in C_{Lip}$  consists in constructing a sequence of points  $\varphi_i \in E^*$ ,  $0 \leq i \leq M_f$  (where  $M_f$  is the number of questions about  $f$  asked in the method), and the 'image'  $x_i \in E$ ,  $\bar{x}_i \in G$ , of this sequence. Auxiliary numerical sequences  $\{a_i\}$ ,  $\{b_i\}$ , with  $a_i \leq \infty$ ,  $b_i \geq 0$ , are also constructed. After  $M_f$  steps of the process ( $M_f$  is formed automatically) the result  $\bar{x}$  of the work of the method on  $f$  is constructed from the sequence  $\{\bar{x}_i\}$ .

#### 3.3.2

The rules defining  $MD_V$  are as follows.

**MD 0.** Initial setting.

Put  $\varphi_0 = 0$ ,  $i = 1$ ,  $a_0 = +\infty$ ,  $b_0 = 0$ , and go to MD 1.

**MD 1.**  $i$ -th step.

At the  $i$ th step a point  $\varphi_{i-1} \in E^*$ , numbers  $a_{i-1} \leq \infty$  and  $b_{i-1} \geq 0$  are available. In the  $i$ th step the following operations are carried out in turn.

**MD 1.1.** A point  $x_i = V'(\varphi_{i-1})$  is constructed, and also  $\bar{x}_i = \pi_G(x_i)$ . A question is put to the oracle at the point  $\bar{x}_i$  about the problem under solution. Let  $g_i^j(y)$  be the affine functionals, and  $\eta_i^j$  their derivatives with respect to  $y$ , communicated by the oracle.

**MD 1.2.** The numbers  $L_j(i) = \|\eta_i^j\|_*$  are calculated. The number  $j(i)$  is defined by:

$$j(i) = \begin{cases} 0, & \text{if } g_i^j(\bar{x}_i) \leq 2(v - v_0)L_j(i), 1 \leq j \leq m, \\ \text{any } j \geq 1 \text{ such that } g_i^j(\bar{x}_i) > 2(v - v_0)L_j(i), \\ & \text{if there is such a } j. \end{cases}$$

**MD 1.3.** The numbers  $a_i$  are defined:

$$a_i = \begin{cases} a_{i-1}, & j(i) > 0, \\ \min\{a_{i-1}, g_i^0(\bar{x}_i)\}, & j(i) = 0. \end{cases}$$

**MD 1.4** If  $\eta_i^{j(i)} = 0$  go to MD 2. Otherwise put  $\xi_i = \eta_i^{j(i)}/\|\eta_i^{j(i)}\|_*$  and

$$\delta_i = \frac{1}{\|\eta_i^{j(i)}\|_*} \begin{cases} g_i^0(\bar{x}_i) - a_i + 2(v - v_0)L_0(i), & j(i) = 0, \\ g_i^{j(i)}(\bar{x}_i), & j(i) > 0. \end{cases}$$

**MD 1.5.** Put  $\tilde{\zeta}_i = \xi_i + \mu_G(x_i)$ . If  $\tilde{\zeta}_i = 0$ , go to MD 2. Otherwise put  $\zeta_i = \tilde{\zeta}_i/\|\tilde{\zeta}_i\|_*$ ,  $\delta_i = \delta_i/\|\tilde{\zeta}_i\|_*$ .

**MD 1.6.** Define  $\rho_i = \gamma_V(\delta_i/2)$  and put  $\gamma_i = \delta_i\rho_i/2$ .

**MD 1.7.** Put  $\varphi_i = \varphi_{i-1} - \rho_i\zeta_i$ ,  $b_i = b_{i-1} + \gamma_i$ . If  $V(\varphi_i) - \|\varphi_i\|_* > -b_i$ , then go to MD 2. Otherwise, finish the step, i.e. increase  $i$  by 1 and go to MD 1.

**MD 2.** Rules for output of the result.

We recall that reversion to MD 2 takes place when  $i = M_f$ . Strictly speaking,  $M_f$ , i.e. the number of questions posed by the method about the problem  $f$ , is defined precisely as the iteration number of the step at which reversion to

MD 2 takes place. The work of the method stops at this point. The result is formulated as

$$\begin{cases} * & \text{if } j(i) > 0, \text{ for all } i \leq M_f, \\ \bar{x}_{i_0} & \text{if } j(i_0) = 0 \text{ and } g_{i_0}^0(\bar{x}_{i_0}) = a_{M_f} \text{ otherwise.} \end{cases}$$

By definition of  $a_i$ , such an  $i_0$  exists if  $j(i) = 0$  for at least one  $i \leq M_f$ .

*Remark.* The rule MD 1.2 does not, in general, determine  $j(i)$  uniquely. If, for a given  $i$ , there is a freedom of choice for  $j(i)$ , then it will be sensible to use it to ensure the maximum possible value of  $\gamma_i$ . In principle, such action will accelerate the descent.

In due course it will be shown that the method described, and the versions of it given below, guarantee the solution of every problem of the class  $C_{Lp}$  with a relative error  $\leq v$ . It will also be proved that its laboriousness does not exceed  $M_V(v - v_0)$ . Similar bounds for the laboriousness of other versions of MD methods will be formulated below in the descriptions of these versions, and these also will be proved in Section 3.3.5.

### 3.3.3

We now describe certain modifications of the method already constructed.

#### 3.3.3.1 A version with optimal choice of the shift $\rho_i$

The rule MD 1.6 can be replaced by the following rule:

**MD 1.6'.** Put  $r_i(t) = (\delta_i - \langle x_i | \zeta_i \rangle)t + V(\varphi_{i-1}) - V(\varphi_{i-1} - t\zeta_i)$ , and let  $t_i$  be the maximizing point of  $r_i(t)$  over all  $t \geq 0$  (if this maximum is not attained, then put, say, as in MD 1.6,  $t_i = \gamma_V(\delta_i/2)$ ).

Put  $\rho_i = t_i$  and  $\gamma_i = r_i(t_i)$ .

This modification, generally speaking, accelerates the process of finding the solution, but complicates the step. The laboriousness of this modified version does not exceed  $M_V(v - v_0)$ .

#### 3.3.3.2 The Hilbert-space version of the method

Now let  $E, \|\cdot\|$ , be a Hilbert space. In this case certain simplifications are possible, and these reduce  $M_D$  to a certain version of the standard gradient method. In the preceding description the following changes can be made:

- (i) put  $V(\varphi) = \frac{1}{2}\varphi^2$  (in the Hilbert-space case  $E$  and  $E^*$  are canonically identified, and we do not distinguish between them);

- (ii) in MD 1.7 replace the rule for forming  $\varphi_i$  by  $\varphi_i = \pi_G(\varphi_{i-1} - \rho_i \zeta_i)$  (then automatically we shall have  $\varphi_{i-1} = x_i = \bar{x}_i$ , and we can take  $\mu_G(x_i) = 0$ );
- (iii) in MD 1.6 replace the rule for choosing  $\rho_i, \gamma_i$  by  $\rho_i = \delta_i, \gamma_i = \delta_i^2/2$ .

In this situation MD 1.6' is identical with MD 1.6, and so in the Hilbert-space case we restrict consideration to a single version of the method, which we shall denote by  $\Gamma$ . The laboriousness of the method  $\Gamma$  constructed for an accuracy  $v, 1 > v > v_0$ , does not exceed  $M(v - v_0)$ , where

$$M(v) = \left\lceil \frac{1}{4v^2} \right\rceil + 2. \quad (3.2)$$

### 3.3.3.3 Adjustment for specified absolute errors

Up to now we have assumed that the relative error with which it was required to solve a problem in  $C_{Lip}$  was specified. Now suppose that it is not the relative error which is specified in advance, but the absolute errors  $\varepsilon_j > 0, j = 0, \dots, m$ , admissible for the approximate solution. It turns out that, with this formulation of the problem, the MD-methods are again capable of ensuring a solution with the requisite accuracy. More precisely, the versions of the MD-methods described above do require a certain modification for this purpose. First of all it is necessary to 'do a zeroth step'—to put a question about the problem under solution at the point 0. If it is found that

$$g_{j,0}(0) \leq \varepsilon_j, \quad 1 \leq j \leq m, \quad \text{and} \quad \min_{y \in G} g_{j,0}(y) \geq g_{j,0}(0) - \varepsilon_j,$$

then we put out 0 as the approximate solution of the problem. But if this is not so, but it turns out that for some  $j \geq 1$  we have  $\min_{y \in G} g_{j,0}(y) > 0$ , then we put out \* as the approximate solution. If neither of these possibilities occurs, then we apply, for solving  $f$ , any admissible (for the given  $E, \|\cdot\|$ ) one of the three versions of MD-methods of solving problems in  $C_{Lip} \subset C_{Lip}^{v_0}(G, E, \|\cdot\|, m)$ , modified in the following way.

Rule MD 1.2 should be replaced by MD 1.2':

$$j(i) = \begin{cases} 0 & \text{if } g_i(\bar{x}_i) \leq \varepsilon_j, \quad 1 \leq j \leq m, \\ \text{any } j \geq 1 \text{ such that } g_i(\bar{x}_i) > \varepsilon_j, & \text{if there is such a } j. \end{cases}$$

Also, in MD 1.4 the formula for  $\delta_i$  should be replaced by

$$\delta_i = \frac{1}{\|\eta_i^{(i)}\|} \begin{cases} g_i^0(\bar{x}_i) - a_i + \varepsilon_0, & j(i) = 0, \\ g_i^{(i)}(\bar{x}_i), & j(i) > 0, \end{cases}$$

The methods obtained in this way have the following property: their application to a problem  $f \in C_{Lip}$  provides a result  $\hat{x}$  which satisfies the conditions (4.1) of Chapter 2. Moreover, the laboriousness of the use of these

methods does not exceed respectively  $M_v(v(f, \varepsilon)), M(v(f, \varepsilon))$ , where

$$v(f, \varepsilon) = \min_{0 \leq j \leq m} \frac{\varepsilon_j}{r_j(f)}$$

is the maximum possible relative error which guarantees the required absolute error. Attention is drawn to the fact that, in the case where the 'zeroth step' is unsuccessful, i.e. when it does not lead to a result, then  $v(f, \varepsilon) \leq 1$ . Thus MD-methods are capable of ensuring the required absolute error with no less success than the methods described in the previous chapter.

### 3.3.3.4

Let us say a few words about the practical realization of the methods described. Of course, we can talk only of solving finite-dimensional problems. We restrict ourselves to the case when the spaces  $E, \|\cdot\|$ , are  $\ell_p^{(n)}$  with the norm  $\|\cdot\|_p, 1 \leq n < \infty$ . Algorithmic realization of MD-methods in this situation require the ability to solve three problems:

- (I) to find the derivative  $V'(\varphi)$  of a  $(\ell_p^{(n)}, \|\cdot\|_p)$ -regular function  $V$ ;
- (II) to construct the projection  $\pi_G(x)$  of a point  $x \in \ell_p^{(n)}$  on to a convex, closed, bounded set  $G \subset \ell_p^{(n)}$ ;
- (III) to construct the support functional  $\mu_G(x)$  to the function  $\rho_{\|\cdot\|_p}(\cdot, G)$ .

We emphasize that we are speaking only of problems connected with the organization of the methods themselves. We do not deal with questions of the algorithmic realization of the oracle, which have no direct connection, properly speaking, to the theory of optimization methods.

Let us analyse the problems (I)–(III). The first of them is very simple, at least in the case where the method in question is associated with a standard  $(\ell_p^{(n)}, \|\cdot\|_p)$ -regular function from Section 3.2.3. For, such functions admit explicit representation, from which it is easy to obtain analytic expressions for the co-ordinates of  $V'(\varphi)$ . The computations for these expressions require  $O(n)$  elementary operations (we include also in this number the computation of the standard functions—powers, exponentials, etc.).

Problem (II), generally speaking, can be far from simple. For, finding the point of  $G$  closest to  $x$ , in the metric of  $\|\cdot\|_p$ , is all the same as solving a particular problem of convex optimization which, generally speaking, is no less complicated than the original one. However, a special consideration comes to our aid here. If the domain  $G$  has a simple shape, then finding  $\pi_G(x)$  may be quite easy (examples will be given later), and to a large extent we are free to choose this domain.

Indeed, a domain of 'complicated shape' is generally defined by a certain number of inequalities; for convex problems they are convex. The term 'defined' here is to be understood not in the sense that there are such

inequalities (they always exist), but in the sense that the method of defining  $G$  is an algorithm which makes clear, for all  $x$ , whether the point  $x$  lies in  $G$  or not; an algorithm which is based on the explicit indication of a list of inequalities which points of  $G$  must satisfy. But in such a case there is nothing to prevent us from transferring the inequalities defining  $G$  to the set of constraints on the problem under solution. The problem can thereafter be regarded as being defined in a suitable domain  $G_1$  which contains  $G$  and which has a simple shape—simple, that is, from the point of view of calculating the projection  $\pi_{G_1}(x)$ .

Of course, with such an extension of  $G$  we increase the radius of the domain of definition of the problem, and this also increases the laboriousness of solving the problem with a given absolute accuracy. This is the price which has to be paid for wishing to deal with simple domains. Generally speaking, it is difficult to say whether, in the general case, this price is acceptable. The question of making a rational choice of  $G_1$  must be examined specially in each actual case. For MD-methods it is quite important, because their laboriousness depends on some power of the stipulated relative accuracy, and so increasing the size of  $G$  by  $k$  times leads to an increase in the laboriousness by  $k^s$  times, where  $s$  depends on  $p$ , and  $s \geq 2$ . On the other hand, as applied to methods having linear convergence (of the MMCG type, in which it is required to solve the projection problem for the Euclidean case), even a considerable ‘inflation’ of  $G$  does not lead to a substantial increase of the laboriousness.

As regards problem (III), it is closely connected with problem (II). For, it is clear that, if  $x \notin G$ , then  $\mu_G(x)$  is a support functional to the vector  $x - \pi_G(x) = \Delta_G(x)$  i.e. a functional  $\varphi \in E^*$  such that  $\|\varphi\|_* = 1$  and  $\langle \varphi | \Delta_G(x) \rangle = \|\Delta_G(x)\|$ . If we know how to solve problem (II), i.e. in particular, how to find the vector  $\Delta_G(x)$  for a given  $x$ , then, as a rule,  $\mu_G(x)$  can also be easily constructed.

Suppose, for example,  $1 < p < \infty$ . Then  $\|u\|_p$  is a differentiable function when  $u \neq 0$ , and so the support functional to a non-zero vector  $\Delta_G(x)$  (when  $x \notin G$ ,  $\Delta_G(x) \neq 0$ ) is simply the derivative with respect to  $u$  of the function  $\|u\|_p$  at the point  $u = \Delta_G(x)$ . This derivative can be calculated explicitly; its calculation from a given  $u$  requires  $O(n)$  elementary operations.

When  $p = 1$  matters are rather more complicated because the support functional to  $\Delta_G(x)$  may not be unique (at the same time, not every such functional is a support functional to  $\rho_{L_1}(\cdot, G)$  at  $x$ ), but in the examples given below this difficulty is overcome.

We now point out some very simple forms of the domains  $G$  for which the problems (II) and (III) can be solved explicitly. We suppose that  $E, \|\cdot\|$ , can be made to coincide with  $L_p^{(n)}$ ,  $1 \leq p < \infty$ .

1°.  $G$  is the  $L_p^{(n)}$ -ball  $\{\|x - x_0\|_p \leq r\}$ . We can suppose that  $r = 1$ ,  $x_0 = 0$ . In this case, when  $x \notin G$  we have  $\pi_G(x) = x/\|x\|_p$  (for  $x \in G$ , of course, we have

$\pi_G(x) = x$ ). Further, for  $x \in G$  we can take  $\mu_G(x) = 0$ . But if  $x \notin G$ , then

$$\mu_G(x) = \{(\mu_G(x))_i\}_{i=1}^n,$$

where

$$(\mu_G(x))_i = |\bar{x}_i|^{p-1} \operatorname{sgn}(\bar{x}_i), \quad \bar{x} = \frac{x}{\|x\|_p}.$$

(Here we have taken  $0^\circ = 0$ .)

2°.  $G$  is a ‘segment’ of an  $L_p^{(n)}$ -ball;  $G = \{\|x - x_0\|_p \leq r, x \geq x_0\}$ . We can suppose that  $x_0 = 0$ ,  $r = 1$ . For  $x \in L_p^{(n)}$  we put  $x_+ = (\max(0, x_1), \dots, \max(0, x_n))$  then

$$\pi_G(x) = \begin{cases} \frac{x_+}{\|x_+\|_p} & \text{for } \|x_+\|_p > 1, \\ x_+ & \text{for } \|x_+\|_p \leq 1 \end{cases}$$

and

$$(\mu_G(x))_i = |\bar{x}_i|^{p-1} \operatorname{sgn} \bar{x}_i, \quad \bar{x} = \frac{x - \bar{x}}{\|x - \bar{x}\|_p},$$

where

$$\bar{x} = \begin{cases} \frac{x_+}{\|x_+\|_p}, & \text{if } \|x_+\|_p > 1, \\ x_+ & \text{if } \|x_+\|_p \leq 1. \end{cases}$$

It can be seen that in the situations 1° and 2° the calculation of  $\pi_G(x)$  and  $\mu_G(x)$  requires  $O(n)$  elementary calculations. It is clear that in this case the realization of an MD-method requires  $O(n)$  elementary operations per step (here the operations carried out by the oracle and the operations of comparison needed for calculating  $j(i)$  have not been included. The number of the latter depends on the number of constraints in the problem being solved).

The list of ‘simple bodies’  $G$  could be extended, but we imagine that, in the light of all that has been said, the examples already given will suffice.

Of course, the remarks of this section apply to all versions of the MD-methods considered in this book (Chapters 3, 5, 6), and also to other methods which use the operation of projection.

### 3.3.4

Before proving the methods described, we consider the characteristics of their concrete versions corresponding to the case where  $E = L_p$ ,  $1 \leq p < \infty$ , and  $\|\cdot\|$  on  $E$  is  $\|\cdot\|_p$ . We have already looked at the Hilbert-space case (i.e.  $p = 2$ ) in Section 3.3.3.2. The corresponding bound for the laboriousness is

$$M(v - v_0) = \left[ \frac{1}{4(v - v_0)^2} \right] + 2.$$

In the general case, in accordance with (3.1), (2.3), (2.5), we obtain the following upper bounds for the laboriousness of MD-methods of solving Lipschitz-convex problems on the  $L_p$ -spaces with an accuracy  $v > v_0$ .

When  $1 < p < \infty$  (the method associated with  $V_p(\cdot)$ , see Section 3.2.3) the laboriousness of the method does not exceed

$$M_p(v - v_0) = \begin{cases} \left[ \frac{d(p)}{(v - v_0)^2} \right] + 2, & 1 < p \leq 2, \\ \left[ \frac{d(p)}{(v - v_0)^p} \right] + 2, & p > 2. \end{cases} \quad (3.3)$$

Here  $d(p)$  depends only on  $p$ . Moreover, for  $1 < p \leq 2$ , the estimate  $d(p) \leq d/(p-1)$  holds, where  $d$  is an absolute constant (see (2.3) and the estimate for  $c(p)$  is given in Section 3.5).

When  $p = 1$  and  $\dim L_1 = n$  with  $1 < n < \infty$  (i.e. when  $E = \mathbb{I}_1^{(n)}$ ), for the method associated with  $V_{1,n}(\cdot)$  (see Section 3.2.3), the laboriousness does not exceed

$$M_{1,n}(v - v_0) \leq \left[ \frac{d(1) \ln n}{(v - v_0)^2} \right] + 2. \quad (3.4)$$

MD-methods of solving Lipschitz-convex problems on  $L_p$ ,  $1 < p < \infty$ , associated with  $V(\cdot) = V_p(\cdot)$ , will be denoted by  $\text{MD}_p$ . Similarly, MD-methods of solving problems on  $\mathbb{I}_1^{(n)}$ , associated with  $V(\cdot) = V_{1,n}(\cdot)$  will be denoted by  $\text{MD}_{1,n}$  (this is the method  $\text{MD}_{1,n}$ -method promised in Section 3.1.3).

In Section 4.3 it will be proved that, if  $E = L_p$  and if  $n = \dim E$  is sufficiently large, and if  $G$  is a ball in  $E$ , then the laboriousness of  $\text{MD}_p$  (i.e.  $\text{MD}_{1,n}$  for  $p = 1$ ) cannot, in principle, be reduced more than a constant number of times (this constant depending only on  $p$ ). Thus, the MD-methods on classes of Lipschitz-convex problems are, in theory, sub-optimal, if the domain of the problems of the class is a body of large dimensionality in  $L_p$ . Precise formulations of this result are given in Sections 4.2 and 4.3.

**Exercise 1.** Let  $E, \|\cdot\|$ , be  $L_p, \|\cdot\|_p$  with  $1 \leq p \leq 2$  and let  $\dim E = n$  with  $1 < n < \infty$ . Prove that problems in the class  $C_{\text{Lip}} \in C_{\text{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$  can be solved with a bound for the laboriousness  $c \ln n / (v - v_0)^2$ , where  $c$  is an absolute constant. (For solving problems of the given class apply the  $\text{MD}_\nu$ -method associated with a function of the form  $V^\tau$  in (2.4). Choose the requisite  $\tau$  using the explicit form of the number  $c(p)$  from (2.3).)

### 3.3.5

We now pass on to the proof of the methods described.

**Theorem.** The following assertions hold.

- (i) Let  $E, \|\cdot\|$  be a regular space, and let  $V$  be a corresponding regular function. Both versions of the  $\text{MD}_\nu$ -methods constructed for an accuracy  $v$  with  $1 > \nu > v_0$  solve every problem in the class  $C_{\text{Lip}} \in C_{\text{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$  with a relative error  $\leq v$  in a number of steps  $\leq M_\nu(v - v_0)$ . A similar assertion (with  $M_\nu(\cdot)$  replaced by  $M(\cdot)$ ) is also true for the gradient method  $\Gamma$  in the case where  $E, \|\cdot\|$ , is a Hilbert space.
- (ii) In the case of the methods described in Section 3.3.3 and constructed for errors  $\varepsilon_j > 0$  ( $0 \leq j \leq m$ ), they solve every problem  $f$  in the class  $C_{\text{Lip}} \in C_0^{\nu_0}(G, E, \|\cdot\|, m)$  with absolute errors  $\leq \varepsilon_j + 2L_{1,1}(f_j)\rho_{1,1}(G)v_0$  and with laboriousness not exceeding  $M_\nu(v(f, \varepsilon))$  resp.  $M(v(f, \varepsilon))$ , where

$$v(f, \varepsilon) = \min_{0 \leq j \leq m} \frac{\varepsilon_j}{2L_{1,1}(f_j)\rho_{1,1}(G)}. \quad (3.5)$$

*Proof.* All these assertions are proved simultaneously. We outline the plan of the proof and at the same time explain the mechanism of the MD-methods. Suppose a compatible problem  $f \in C_{\text{Lip}} \in C_{\text{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$  is being solved and that  $x^*$  is its solution. First we make it clear that the process of solution does indeed break off at a certain moment  $M_f$  which has the requisite upper bound. This follows from the fact that, as can be proved, all the numbers  $y_i$  are greater than a certain positive number  $\bar{a}$  independent of  $i$ . Meanwhile the rule MD 1.7 breaks off the work of the method at a step  $i$  such that  $V(\varphi_i) - \|\varphi_i\|_* > -b_i \leq -i\bar{a}$ . The lower bound of the left hand side of this inequality over all the  $\varphi_i$  is finite and equal, let us say, to  $-T$ . Therefore the method automatically stops work before the moment  $]T/\bar{a}[$ . The explicit calculation of  $T$  and  $\bar{a}$  for the methods in question does indeed lead to the bounds stated above for the laboriousness.

We next explain how the assertion about the accuracy is justified. Let  $V_*(\varphi) = V(\varphi) - \langle \varphi | x^* \rangle$ . It is made clear that, if the method does not ensure the requisite accuracy, then, for all  $i \leq M_f$ , the inequalities  $\langle \zeta_i | x_i - x^* \rangle \geq \delta_i$  hold, i.e.  $\langle V_*(\varphi_{i-1}) | -\zeta_i \rangle \leq -\delta_i$ . Thus,  $(-\zeta_i)$  is the best direction of descent for  $V_*$ . The point  $\varphi_i$  is obtained from  $\varphi_{i-1}$  by a displacement in this direction. The rules for choosing the magnitude of the displacement and  $y_i$  are such that the decrease in  $V_*$  in going from  $\varphi_{i-1}$  to  $\varphi_i$  is not less than  $y_i$  (thus the meaning of  $y_i$  is that it is an *a priori* estimate of the decrease in  $V_*$  after one step). Therefore  $V_*(\varphi_i) \leq -b_i$ , and since  $V_*(\varphi) \geq V(\varphi) - \|\varphi\|_*$  (because  $\|x^*\| \leq 1$ ), so *a fortiori*  $V(\varphi_i) - \|\varphi_i\|_* \leq -b_i$ . But we know that for  $i = M_f$  this is not so (the rule for stopping). Thus the assumption that the method does not ensure the requisite accuracy leads to a contradiction.

We now carry out a strict argumentation for the assertions of the theorem.

1°. We prove that the laboriousness of the methods admits the required bounds. To do this, we satisfy ourselves that for all the methods considered the inequality  $\delta_i \geq a > 0$ ,  $i < M_f$ , holds. This is indeed true for  $i = M_f$  provided

that  $\eta_i^{(0)} \neq 0$  and  $\zeta_i \neq 0$ . Here  $a$  is a parameter depending on the situation (i.e. on which of the assertions (i)–(ii) is being proved and for precisely which version). In fact, for both versions of the MD $\nu$ -method we can take

$$a = v - v_0. \quad (3.6)$$

For, from the rule for choosing  $j(i)$  and the rules for determining  $\delta_i$  it follows that  $\delta_i \geq 2(v - v_0)$ . Moreover  $\|\zeta_i\|_* \leq 2$  and so  $\delta_i \geq \delta_i/2 \geq v - v_0$ .

For the gradient method  $\Gamma$  we can take

$$a = 2(v - v_0) \quad (3.7)$$

(for the same reasons as above. It has only to be taken into account that in this case  $\delta_i = \delta_i$  because  $\mu_G(x_i) = 0$ ).

In the situation of assertion (ii) we can take  $a = v(f, s)$  if the modification of MD $\nu$  is being applied (for in this case

$$\delta_i \geq \frac{\varepsilon_{j(i)}}{\|\eta_i^{(0)}\|_*} = \frac{2\varepsilon_{j(i)}}{2\|\eta_i^{(0)}\|_* \rho_{1,1}(G)} \geq \frac{2\varepsilon_{j(i)}}{2L_{j(i)}(f)\rho_{1,1}(G)} \geq 2v(f, s),$$

and  $\delta_i \geq \delta_i/2$ ; here and later  $L_j(f) = L_{1,1}(f_j)$ . If the method  $\Gamma$  is being used, then  $a = 2v(f, s)$  (for the same reasons as above).

2°. We shall now assume that the work of the method under examination has not stopped after the  $(N+1)$ th step,  $N \geq 1$ . This means that all the time in the first  $N$  steps

$$V(\varphi_i) - \|\varphi_i\|_* \leq -b_i, \quad i \leq N. \quad (3.8)$$

We shall deduce from this that  $N$  cannot be too large.

(1) For the versions of the MD $\nu$ -methods which correspond to the situations (i) and (ii) we have, using the rule MD1.6,

$$\gamma_i = \frac{\rho_i \delta_i}{2} \geq \frac{a \gamma_\nu(a/2)}{2}$$

because of the rule for choosing  $\rho_i$  ( $\rho_i = \gamma_\nu(\delta_i/2)$ ) and from the inequality  $\delta_i \geq a$ ,  $i < M_f$ , proved above. Therefore

$$V(\varphi_N) - \|\varphi_N\|_* \leq -\frac{1}{2}N a \gamma_\nu\left(\frac{a}{2}\right).$$

But the left-hand side of this inequality is not less than  $-\frac{1}{2}$  (because of Section 3.2.1(ii)). Therefore

$$N \leq \frac{1}{a \gamma_\nu(a/2)},$$

i.e.

$$M_f \leq \frac{1}{a \gamma_\nu(a/2)} + 1$$

and the laboriousness of the methods being investigated does not exceed

$$M_f + 1 \leq \frac{1}{a \gamma_\nu(a/2)} + 2,$$

which, by virtue of the definition of  $a$ , gives the required bound for the laboriousness.

(2) For the MD $\nu$ -methods, in the situation (i) or (ii), which use the rule MD1.6' we have, in view of the assumption in 2.4, for  $i \leq N$ ,

$$\begin{aligned} r_i(t) &= (\delta_i - \langle x_i | \zeta_i \rangle)t + V(\varphi_{i-1}) - V(\varphi_i - t\zeta_i) \\ &\geq (\delta_i - \langle x_i | \zeta_i \rangle)t + \langle V'(\varphi_{i-1} | \zeta_i) \rangle t - \omega_\nu(t)t \geq \delta_i t - \omega_\nu(t)t. \end{aligned}$$

Hence it is clear that

$$\max_{t \geq 0} r_i(t) \geq r_i\left(\gamma_\nu\left(\frac{\delta_i}{2}\right)\right) \geq \frac{\delta_i}{2} \gamma_\nu\left(\frac{\delta_i}{2}\right),$$

i.e. that in this case too

$$\gamma_i \geq \frac{a}{2} \gamma_\nu\left(\frac{a}{2}\right),$$

and this leads to the required bound for  $M_f$ .

(3) Finally, for the gradient method the required bound is obtained in exactly the same way as in (2) above.

3°. Thus the bounds for the laboriousness have been proved. We pass on to the proof of the assertions about the error. We consider first the situation (ii) and suppose that the result was obtained on the 'zeroth step' (see Section 3.3.3.3). This result is 0 if

$$g_{j,0}^f(0) \leq \varepsilon_j, \quad 1 \leq j \leq m, \quad \text{and} \quad \min_{y \in G} g_{j,0}^f(y) \geq g_{j,0}^f(0) - \varepsilon_0;$$

in this case

$$g_{j,0}^f(0) \leq \varepsilon_j, \quad 1 \leq j \leq m, \quad \text{and} \quad \min_G f_0(y) \geq g_{j,0}^f(0) - \varepsilon_0.$$

Therefore in this case

$$f_0(0) \leq f_* + \varepsilon_0 + r_0(f)v_0, \quad f_j(0) \leq \varepsilon_j + r_j(f)v_0,$$

as promised. Moreover, this result is \* if

$$\exists j \geq 1: \min_{y \in G} g_{j,0}^f(y) > 0,$$

and in this case  $f$  is incompatible. Thus if the result is obtained at the zeroth step, then it satisfies the requisite demand regarding the accuracy.

We now assume that this is not the case, and that in the situations (i) and (ii) the methods being considered operate ‘non-trivially’, according to the rules MD 0–MD 2.

4°. Suppose the problem  $f$  which is being solved is incompatible. By the rule MD 2 the output result  $\bar{x}$  is either the (correct) answer that  $f$  is incompatible or it is a point of the form  $\bar{x}_i$  such that, for  $1 \leq j \leq m$ ,

$$g_{i_0}^j(\bar{x}_{i_0}) \leq \begin{cases} 2(v - v_0)L_j(i) & \text{in the situation (i),} \\ \varepsilon_j & \text{in the situation (ii).} \end{cases}$$

Since  $L_j(i) \leq L_j(f) = \frac{1}{2}r_j(f)$  (we recall that  $\rho_{1,1}(G) = 1$ ), it follows from this and from the properties of the oracle that, when  $\bar{x} \neq *$  and for all  $j \geq 1$ ,

$$f_j(\bar{x}) \leq \begin{cases} vr_j(f) & \text{in the situation (i),} \\ \varepsilon_j + v_0 r_j(f) & \text{in the situation (ii).} \end{cases} \quad (3.9)$$

For an incompatible problem  $f$  these inequalities mean that the demands concerning the accuracy of the result are indeed satisfied.

5°. Now suppose that  $f$  is compatible, and let  $x^*$  be its solution. We shall prove that if, for some  $i \leq M_f$ , we had

$$\eta_i^{j(0)} = 0 \quad (\text{this is possible only when } i = M_f) \quad (3.10)$$

or

$$\langle \eta_i^{j(0)} | \bar{x}_i - x^* \rangle < \delta_i \|\eta_i^{j(0)}\|_* \quad \text{and} \quad \|\eta_i^{j(0)}\|_* \neq 0, \quad (3.11)$$

then the method will have the required accuracy.

For, in the second case we have, from the definition of  $\delta_i$ ,

$$\begin{aligned} g_i^{j(0)}(\bar{x}_i) - g_i^{j(0)}(x^*) &= \langle \eta_i^{j(0)} | \bar{x}_i - x^* \rangle \\ &< \begin{cases} g_i^0(\bar{x}_i) - a_i + 2(v - v_0)L_0(i), & j(i) = 0, \text{ situation (i),} \\ g_i^0(\bar{x}_i) - a_i + \varepsilon_0, & j(i) = 0, \text{ situation (ii),} \\ g_i^{j(0)}(\bar{x}_i), & j(i) > 0. \end{cases} \end{aligned}$$

Suppose that  $j(i) \neq 0$ . Then

$$g_i^{j(0)}(\bar{x}_i) - g_i^{j(0)}(x^*) \geq g_i^{j(0)}(\bar{x}_i)$$

and the last inequality gives

$$g_i^{j(0)}(\bar{x}_i) < g_i^{j(0)}(\bar{x}_i),$$

which is impossible. Thus  $j(i) = 0$ . In that case,

$$g_i^0(\bar{x}_i) - f_0(x^*) < \begin{cases} g_i^0(\bar{x}_i) - a_i + 2(v - v_0)L_0(i), & \text{situation (i),} \\ g_i^0(\bar{x}_i) - a_i + \varepsilon_0, & \text{situation (ii).} \end{cases}$$

Thus in the case (3.11) we have, for the given  $i$ ,

$$a_i \leq \begin{cases} f_0(x^*) + (v - v_0)r_0(f), & \text{situation (i),} \\ f_0(x^*) + \varepsilon_0, & \text{situation (ii).} \end{cases} \quad (3.12)$$

It is easy to see that these same assertions are also true when  $\eta_i^{j(0)} = 0$ , i.e. in the case (3.10). We deduce from them the required bound for the error of the result. Under the conditions (3.12) the result  $\bar{x}$  is a point  $\bar{x}_{i_0}$ , where  $i_0 \leq M_f$ , such that  $j(i_0) = 0$  and  $g_{i_0}^0(\bar{x}_{i_0}) = a_{M_f} \leq a_i$ . From the relation  $j(i_0) = 0$  we have already derived in 4° the relation (3.9). Moreover,

$$f_0(\bar{x}_{i_0}) \leq g_{i_0}^0(\bar{x}_{i_0}) + r_0(f)v_0 \leq a_{M_f} + r_0(f)v_0,$$

which, in view of (3.12), gives

$$f_0(\bar{x}_{i_0}) \leq \begin{cases} f_* + vr_0(f), & \text{situation (i),} \\ f_* + \varepsilon_0 + v_0 r_0(f), & \text{situation (ii).} \end{cases} \quad (3.13)$$

The relations (3.9) and (3.13) prove the required assertions about the accuracy of the result.

6°. To complete the proof it suffices to convince ourselves that for a certain  $i \leq M_f$  either (3.10) or (3.11) is indeed satisfied. Suppose, on the contrary, that, for all  $i \leq M_f$ ,

$$\eta_i^{j(0)} \neq 0 \quad \text{and} \quad \langle \eta_i^{j(0)} | \bar{x}_i - x^* \rangle \geq \delta_i \|\eta_i^{j(0)}\|_*; \quad (3.14)$$

we shall lead (3.14) to a contradiction. Indeed, it follows from (3.14) that

$$\langle \xi_i | \bar{x}_i - x^* \rangle \geq \delta_i.$$

We shall prove that then

$$\xi_i + \mu_G(x_i) = \zeta_i \neq 0 \quad \text{and} \quad \langle \zeta_i | x_i - x^* \rangle \geq \delta_i. \quad (3.15)$$

In the Hilbert-space case this is obvious because, in such a space,  $x_i = \bar{x}_i$  and  $\mu_G(x_i) = 0$ , i.e.  $\zeta_i = \xi_i$ .

In the general case we have

$$\begin{aligned} \langle \zeta_i | x_i - x^* \rangle &= \langle \xi_i | x_i - x^* \rangle + \langle \mu_G(x_i) | x_i - x^* \rangle \\ &= \langle \xi_i | \bar{x}_i - x^* \rangle + \langle \xi_i | x_i - \bar{x}_i \rangle + \langle \mu_G(x_i) | x_i - x^* \rangle \\ &\geq \langle \xi_i | \bar{x}_i - x^* \rangle - \|x_i - \bar{x}_i\| + \rho_{1,1}(x_i, G) = \langle \xi_i | \bar{x}_i - x^* \rangle \geq \delta_i \end{aligned}$$

(we have used the fact that, by definition of  $\mu_G$ , we have

$$\langle \mu_G(x_i) | x_i - u \rangle \geq \rho_{1,1}(x_i, G) \quad \text{for all } u \in G.$$

Thus (3.15) holds. We saw in 1° that  $\delta_i \neq 0$  if it is generally well-defined (i.e. if  $\eta_i^{j(0)} \neq 0$ ). So  $\zeta_i \neq 0$  and  $\langle \zeta_i | x_i - x^* \rangle \geq \delta_i$ , or, what comes to the same thing,

$$\langle V'_{\bullet}(\varphi_{i-1}) | \zeta_i \rangle \geq \delta_i. \quad (3.16)$$

From (3.16) we shall deduce that for all  $i \leq M_f$

$$-V_*(\varphi_i) + V_*(\varphi_{i-1}) \geq \gamma_i, \quad (3.17)$$

i.e. since  $V_*(0) = 0$ ,

$$V_*(\varphi_i) \leq -b_i, \quad i \leq M_f. \quad (3.18)$$

Since  $V_*(\varphi_i) \geq V(\varphi_i) - \|\varphi_i\|_*$ , so (3.18) implies that

$$V(\varphi_i) - \|\varphi_i\|_* \leq -b_i. \quad (3.19)$$

The last inequality is impossible when  $i = M_f$  (we already know that the work of the method cannot stop because  $\eta_{M_f}^{(M_f)} = 0$  or  $\zeta_{M_f} = 0$  and so at the instant  $M_f$  an inequality contradicting (3.19) must hold).

Thus, to obtain the required contradiction, it suffices to derive (3.17) from (3.16). This is done as follows.

(1) Suppose we are dealing with MD $_\nu$ -methods in the situation (i) or (ii) and the rule MD 1.6 is used. The function

$$\tilde{r}_i(t) = -V_*(\varphi_{i-1} - t\zeta_i) + V_*(\varphi_{i-1})$$

is equal to 0 when  $t = 0$  and it has a derivative with respect to  $t$  which is equal to  $\langle \zeta_i | x_i - x^* \rangle$ , i.e. is  $\geq \delta_i$ . It differs by a linear addend from  $V(\varphi_{i-1}) - V(\varphi_{i-1} - t\zeta_i) \equiv \tilde{r}_i(t)$ . Therefore, by definition of  $\omega_\nu(t)$ , we have, when  $0 \leq t$ ,

$$\frac{d}{dt} \tilde{r}_i(t) \geq \delta_i - \omega_\nu(t).$$

from the definition of  $\rho_i$  we obtain that, when  $0 \leq t \leq \rho_i$ ,

$$\frac{d}{dt} \tilde{r}_i(t) \geq \delta_i/2.$$

From this (3.17) follows immediately, as required.

(2) The analysis of the situation with MD $_\nu$ -methods in the cases (i), (ii) when the rule MD 1.6' is used differs in no way from the preceding. For, the function  $r_i(t)$ , by its definition, differs from  $\tilde{r}_i(t)$  by a linear addend, and  $\tilde{r}_i(t) \geq r_i(t)$  when  $t \geq 0$ . So, in this case,

$$V_*(\varphi_{i-1}) - V_*(\varphi_i) = \tilde{r}_i(\rho_i) \geq r_i(\rho_i) = \gamma_i,$$

as required.

(3) The Euclidean case (i.e. the gradient method) is analysed in exactly the same way as in (2)).

Thus we have derived (3.17) from (3.16) for all the situations in which we are interested. The proof of the theorem is completed.

### 3.4 MD-METHODS ON CLASSES OF GENERAL CONVEX PROBLEMS

We shall show that MD-methods can also be applied to general convex problems provided only that  $G$  is assumed to be convex, closed, bounded, and to have an interior (i.e.  $G$  is to be a body).

Let  $E, \|\cdot\|$ , be a regular space, and let  $G \subset E$  be a non-void, convex, bounded body in  $E$  with asphericity  $\alpha_{E,\|\cdot\|}(G) = \alpha$ . We shall describe methods of solving problems of classes of the  $C^{v_0}(G, E, m)$  type. We shall give several versions of the methods, as we did in the previous section. We shall assume that  $G$  is contained in a ball of radius 1 with its centre at 0 and that it contains a ball of radius  $1/(2\alpha)$  with the same centre. This can obviously be achieved by a similarity transformation of  $G$ , and so these presuppositions in no way limit the generality of the constructions.

#### 3.4.1

Let  $v > v_0$  be the required relative accuracy of solution of problems in  $C \in C^{v_0}(G, E, m)$ ; we shall consider later the case when the absolute errors of the solution are specified instead of the relative error. We shall see that to solve problems of the class  $C$  with an accuracy  $v > v_0$  is the same, as regards laboriousness, as to solve Lipschitz-convex problems (problems of a class  $\in C_{Lip}^{v_0}(G, E, \|\cdot\|, m)$ ) with an accuracy  $\bar{v}(v) = v_0 + (v - v_0)/(4\alpha)$ .

As always, every point  $x \in G$  is a solution of any problem  $f \in C$  with an error  $\leq 1$ . Nevertheless, for reasons which will become clear later, it will be convenient in Section 3.4.2.3 to consider the range of variation of  $v$  to be from  $v_0$  to  $2\alpha$ , and to give a description of a method for all values of  $v$  in this range.

#### 3.4.2

We are about to describe MD-methods of solving general convex problems. We start with the most general case, where all we assume as regards  $E, \|\cdot\|$ , is regularity.

Let  $\pi_G(x): E \rightarrow G$  and  $\mu_G(x): E \rightarrow E^*$  be defined as mappings which respectively associate with a point  $x \in E$  the point  $x \in G$  closest to  $x$  ('closest' in the norm of  $E$ ), and the support functional to  $\rho_{E,\|\cdot\|}(\cdot, G)$  at the point  $x$ . Let  $V(\varphi)$  be an  $(E, \|\cdot\|)$ -regular function corresponding to  $E, \|\cdot\|$ . We shall describe an MD-method of solving problems in  $C$  associated with  $V$  (we shall denote the method by  $\overline{MD}_\nu$ ).

The work of  $\overline{MD}_\nu$  constructed for an accuracy  $v$  with  $2\alpha > v > v_0$  on a problem  $f \in C$  consists in the construction of a sequence  $\varphi_i \in E^*$  ( $0 \leq i \leq M_f$ ) ( $M_f$  is the number of questions of the method on  $f$ ) and its 'image'  $x_i \in E$ ,  $\bar{x}_i \in \text{int } G$ . In the course of the work two auxiliary numerical sequences  $\{a_i\}$  and

$\{b_i\}$ , with  $a_i \leq +\infty$  and  $b_i > 0$ , are also constructed. After  $M_f$  steps ( $M_f$  is formed by the method itself), the result  $\bar{x}$  is formed from the sequence  $\{\bar{x}_i\}$ . The laboriousness of the method on  $f$  is thus  $M_f + 1$ . It will be proved that this quantity does not exceed  $M_{V,G}(v - v_0)$ , where

$$M_{V,G}(v) = M_V\left(\frac{v}{4\alpha}\right) = \left\lceil \frac{8\alpha}{v\gamma_V(v/16\alpha)} \right\rceil + 2 \quad (4.1)$$

The work of the method  $\overline{MD}_f$  of a problem  $f$  is defined by the following rules.

#### $\overline{MD}_f$ 0. Initial setting.

Put  $\varphi_0 = 0$ ,  $i = 1$ ,  $a_0 = +\infty$ ,  $b_0 = 0$ , and let  $\theta$  be defined by the relation  $1/\theta - 1 = v - v_0$ . Then  $0 < \theta < 1$ . Go to  $\overline{MD}_f$  1.

#### $\overline{MD}_f$ 1. $i$ -th step.

At the  $i$ th step there are available: a point  $\varphi_{i-1}$ , and numbers  $a_{i-1} \leq +\infty$  and  $b_{i-1} \geq 0$ . In the  $i$ th step the following operations are carried out in succession.

$\overline{MD}_f$  1.1. Points  $x_i = V'(\varphi_{i-1})$  and  $\bar{x}_i = \theta \pi_G(x_i)$  are constructed. At the point  $\bar{x}_i$  a question is put to the oracle about the problem being solved (this is possible because  $\theta < 1$  and  $\theta \pi_G(x) \in \text{int } G$ ). Let  $g_i^j(y)$  be the affine functionals communicated by the oracle, and let  $\eta_i^j$  be their derivatives with respect to  $y$ .

#### $\overline{MD}_f$ 1.2. The numbers

$$r_j(i) = \begin{cases} \sup_{y \in G} g_i^0(y) - g_i^0(\bar{x}_i), & j = 0, \\ [\sup_{y \in G} g_i^j(y)]_+ - [g_i^j(\bar{x}_i)]_+, & j \geq 1. \end{cases}$$

are calculated. A number  $j(i)$  is defined by

$$j(i) = \begin{cases} 0, & \text{if } g_i^j(\bar{x}_i) \leq (v - v_0)r_j(i), \quad 1 \leq j \leq m, \\ j \geq 1 & \text{such that } g_i^j(\bar{x}_i) > (v - v_0)r_j(i), \text{ if there is such a } j. \end{cases}$$

#### $\overline{MD}_f$ 1.3. $a_i$ is defined by

$$a_i = \begin{cases} a_{i-1}, & j(i) > 0, \\ \min\{a_{i-1}, g_i^0(\bar{x}_i)\}, & j(i) = 0. \end{cases}$$

$\overline{MD}_f$  1.4. If  $\eta_i^{j(i)} = 0$ , go to  $\overline{MD}_f$  2. Otherwise put

$$\begin{aligned} \xi_i &= \frac{\eta_i^{j(i)}}{\|\eta_i^{j(i)}\|_*}, \\ \delta_i &\equiv -(v - v_0) \langle \xi_i | \bar{x}_i \rangle + \delta_i \\ &= \frac{1}{\|\eta_i^{j(i)}\|_*} \begin{cases} g_i^0(\bar{x}_i) - a_i + (v - v_0)r_0(i), & j(i) = 0, \\ g_i^{j(i)}(\bar{x}_i), & j(i) > 0. \end{cases} \end{aligned}$$

$\overline{MD}_f$  1.5. Put  $\zeta_i = \xi_i + \mu_G(x_i)$ . If  $\zeta_i = 0$ , then go to  $\overline{MD}_f$  2. Otherwise, put

$$\zeta_i = \frac{\zeta_i}{\|\zeta_i\|_*}, \quad \delta_i = \frac{\delta_i}{\|\zeta_i\|_*}.$$

#### $\overline{MD}_f$ 1.6. Put

$$\rho_i = \gamma_V\left(\frac{\delta_i}{2}\right), \quad \gamma_i = \frac{\delta_i \rho_i}{2}.$$

(Instead of the rule  $\overline{MD}_f$  1.6 the following rule  $\overline{MD}_f$  1.6' can be used:

$\overline{MD}_f$  1.6'. Put  $r_i(t) = (\delta_i - \langle x_i | \zeta_i \rangle)t + V(\varphi_{i-1}) - V(\varphi_{i-1} - t\zeta_i)$ . Maximize  $r_i(t)$  over all  $t \geq 0$ . If this maximum is attained, then put it equal to  $\gamma_i$ , and the optimal value of  $t$  is  $\rho_i$ . But if the maximum is not attained, then put

$$\rho_i = \gamma_V\left(\frac{\delta_i}{2}\right), \quad \gamma_i = \frac{\delta_i \rho_i}{2}.$$

$\overline{MD}_f$  1.7. Put  $\varphi_i = \varphi_{i-1} - \rho_i \zeta_i$  and  $b_i = b_{i-1} + \gamma_i$ . If  $V(\varphi_i) - \|\varphi_i\|_* > -b_i$ , then go to  $\overline{MD}_f$  2. Otherwise, finish the step (i.e. increase  $i$  by 1 and go to  $\overline{MD}_f$  1).

#### $\overline{MD}_f$ 2. Rule for output of the result.

Suppose that reversion to  $\overline{MD}_f$  2 took place when  $i = M_f$ . The work of the method then stops. The result is formed as

- \* if  $j(i) > 0$ ,  $1 \leq i \leq M_f$ ,  
 $\bar{x}_{i_0}$  if  $j(i_0) = 0$  and  $g_{i_0}^0(\bar{x}_{i_0}) = a_{M_f}$  otherwise.

(By the definition of  $a_i$  (see  $\overline{MD}_f$  1.3) there is such an  $i_0$  if  $j(i) = 0$  for at least one  $i \leq M_f$ .)

In the last case, also put

$$\bar{r}_0(f) = \sup_{f(i) = 0} r_0(i),$$

and in the first case put  $\bar{r}_0(f) = 0$ . Put

$$\bar{r}_j(f) = \max_{1 \leq i \leq M_f} r_j(i), \quad j \geq 1$$

(The quantities  $\bar{r}_j(f)$  are required later in Section 3.4.3.2.)

*Remark.* In the choice of  $j(i)$  according to rule  $\overline{MD}$  1.2 there is, generally speaking, a certain freedom. It is expedient to use this so as to maximize  $y_i$ .

*Exercise.* Compare the descriptions of  $MD_v$  and  $\overline{MD}_v$ . In what way do they differ?

### 3.4.3

As in the previous section, we postpone for a time the proof of the method in order to give a number of versions of the method which correspond to various assumptions about the situation. The properties of these versions which are stated below will be proved in Section 3.4.6.

#### 3.4.3.1

In the Hilbert-space case (where we do not distinguish between  $E$  and  $E^*$ ) some simplifications can be made. These simplifications consist in the following:

- (a) as  $V(\varphi)$  we take  $\frac{1}{2}\varphi^2$ ;
- (b) in  $\overline{MD}$  1.7 the rule for forming  $\varphi_i$  is replaced by  $\varphi_i = \pi_G(\varphi_{i-1} - \rho_i \zeta_i)$  (here we shall have automatically  $x_i = \varphi_{i-1}$  and we can suppose that  $\mu_G(x_i) = 0$ ).

In  $MD$  1.6 the rules for choosing  $\rho_i$  and  $y_i$  are replaced by  $\rho_i = \delta_i$ ,  $y_i = \delta_i^2/2$  ( $\overline{MD}$  1.6' then becomes equivalent to  $\overline{MD}$  1.6). The method so obtained is a version of the ordinary gradient method, and we shall denote this version by  $\Gamma$ . The method  $\Gamma$  constructed for an accuracy  $v$ , with  $2x > v > v_0$ , solves all problems in  $C \in C^{v_0}(G, E, m)$  with an accuracy  $v$  and with a laboriousness not exceeding  $M^G(v - v_0)$ , where

$$M^G(v) = \left\lceil \frac{4x^2}{v^2} \right\rceil + 2. \quad (4.2)$$

#### 3.4.3.2

Up to now we have been concerned with ensuring a specified relative error. We now consider the case where it is not the relative, but the absolute, error which is specified in advance, i.e. the vector  $\epsilon$  of the absolute errors  $\epsilon_j > 0$  admissible for the result of the work of the method is specified. As in the case of Lipschitz-convex problems, each of the three methods described above (the two  $MD_v$  versions and the method  $\Gamma$ ) can, after suitable modification, be applied to solve problems of the class  $C \in C^{v_0}(G, E, m)$  with specified absolute errors (the phrase 'each of the three methods' presupposes, of course, that  $E, \|\cdot\|$ , satisfy the conditions for applicability of the method considered).

We shall show how this is done for the given method—let us call it  $\mathcal{B}$ . We fix on some continuous function  $N(v)$  which gives an upper bound for the laboriousness of the method  $\mathcal{B}$ , on the class  $C$ , constructed for an accuracy  $v_0 + v$ . We put  $v^0 = 2x$ , and let  $i_0$  be the first natural number  $i$  such that  $2^{i_0} \geq N(v^0)$ . We let  $v^i$ ,  $i = 1, 2, \dots$ , denote the least  $v$  such that  $N(v) = 2^{i_0+i}$ . Then  $v^0 > v^1 > v^2 > \dots$

Let  $\mathcal{B}(v^0)$  be a 2-step method of solving problems in  $C$  which is defined in the following way. A question about the problem  $f$  under solution is put to the oracle at the point 0. If, for some  $j \geq 1$  we then obtain  $g_{f,0}^j(y) > 0$  for all  $y \in G$ , then the problem is declared to be incompatible. In the opposite case the point 0 is given out as its approximate solution. Let

$$\tilde{r}_j^{(0)}(f) = \begin{cases} [\sup_{y \in G} g_{f,0}^j(y)]_+, & j \geq 1, \\ \sup_{y \in G} g_{f,0}^j(y) - g_{f,0}^j(0), & j = 0. \end{cases}$$

Also let  $\mathcal{B}(v^i)$  denote the method  $\mathcal{B}$  constructed for an accuracy  $v^i$ ,  $1 \leq i$ .

We now consider a method  $\tilde{\mathcal{B}}$  defined as follows. Let  $\epsilon = (\epsilon_0, \dots, \epsilon_m)$  be the admissible absolute errors of solutions of problems in  $C \in C^{v_0}(G, E, m)$ . The work of the method  $\tilde{\mathcal{B}}$  consists in the successive application to the problem  $f \in C$  under solution of procedures  $\mathcal{B}(v^0), \mathcal{B}(v^1), \dots$ . In the application of the  $s$ th procedure the numbers

$$\delta_j(s, f) = v^s \tilde{r}_j^{(s)}(f), \quad 0 < j \leq M \quad (4.3)$$

are fixed. (The numbers  $\tilde{r}_j^{(s)}(f)$ ,  $s > 0$ , are obtained as a result of applying  $\mathcal{B}(v^s)$  to  $f$  in accordance with the rule  $\overline{MD}$  2). The work of the method  $\tilde{\mathcal{B}}$  on the problem  $f$  ends with the application of the first procedure  $\mathcal{B}(v^s)$  for which the result  $\delta_j(s, f) \leq \epsilon_j$ ,  $0 \leq j \leq m$ , is obtained (there is such a  $j$  because, from the properties of the oracle and the definition of  $\tilde{r}_j^{(s)}(f)$ , it is clear that

$$\tilde{r}_j^{(s)}(f) \leq r_j(f), \quad j \geq 1, \quad \tilde{r}_0^{(s)}(f) \leq (1 + v_0)r_0(f),$$

and  $v^s \rightarrow_{s \rightarrow \infty} 0$ ). It turns out that the method just described ensures a solution of any problem  $f \in C$  with absolute errors  $\leq \bar{\epsilon}_j = \epsilon_j + r_j(f)v_0$ . In other words, the absolute error of the method, as always in similar situations, coincides with the required absolute error up to 'the irremovable error of the oracle' (cf. Sections 3.3.3 and 2.4.1). The laboriousness of the method described does not exceed  $8N(v, (f, \epsilon))/(1 + v_0)$ , where

$$v(f, \epsilon) = \min_{0 \leq j \leq m} \epsilon_j / r_j(f)$$

is the maximum possible relative error which ensures the specified absolute errors (notice that it is quite possible for  $v(f, \epsilon)$  to be  $\geq 1$ ).

We are now in a position to explain why it was required to define  $\overline{\text{MD}}$ -methods even for accuracies  $v > 1$ . The point is that methods of this kind enter into the method  $\mathcal{B}$  (for some finite value of the index  $i$  it is quite possible for  $v^i$  to be  $\geq 1$ ). If  $v(f, \epsilon) \geq 1$  too, then the work of  $\mathcal{B}$  may, perhaps, stop on one of the procedures with  $v^i \geq 1$ . In that case we would be able to put out at once as the solution any point of  $G$ , but we are not actually able to do this because we do not know how to find  $v(f, \epsilon)$  from the known  $\epsilon$ . Therefore we are obliged to take certain actions in all cases.

Of course, we could start the method  $\mathcal{B}$  with a procedure of the type  $\mathcal{B}(v)$  with  $v < 1$ , but then we would waste a large number of steps, possibly, in constructing the required solution. Such an effect, of course, could take place only when  $v(f, \epsilon) \geq 1$ .

### 3.4.4

We write down upper bounds for the laboriousness of MD-methods of solving general convex problems on the standard  $L_p$ -spaces. These bounds are derived directly from (4.1) and (3.3)–(3.4).

$1 < p < \infty$ . Method associated with  $V(\cdot) = V_p(\cdot)$  (we shall call it  $\overline{\text{MD}}_p$ ); the laboriousness of solving problems in  $C \in C^{\circ 0}(G, E, m)$  on  $G$  with a relative error  $v > v_0$  does not exceed

$$M_{p,G}(v - v_0) \leq \begin{cases} \left[ \frac{d(p)\alpha_{\parallel \cdot \parallel}^2(G)}{(v - v_0)^2} \right] + 2, & 1 < p \leq 2, \\ \left[ \frac{d(p)\alpha_{\parallel \cdot \parallel}^p(G)}{(v - v_0)^p} \right] + 2, & p > 2 \end{cases} \quad (4.4)$$

(in the case  $p = 2$  when applying the method  $\tilde{\Gamma}$  the bound is given in (4.2)).

$p = 1$  and  $\dim E = n$ ,  $1 < n < \infty$ . Method associated with  $V_{1,n}(\cdot)$ . (we shall call it  $\overline{\text{MD}}_{1,n}$ ). A bound for the laboriousness is determined by the inequality

$$M_{1,n,G}(v - v_0) \leq \left[ \frac{d(1)\alpha_{\parallel \cdot \parallel}^2(G) \ln n}{(v - v_0)^2} \right] + 2. \quad (4.5)$$

In (4.4) and (4.5) the quantity  $d(p) < \infty$  depends only on  $p$ . We point out that  $\overline{\text{MD}}_{1,n}$  is the method promised in Section 3.1.3.

### 3.4.5

Let us discuss the scheme of application of the MD-methods. We have presented above a construction which associates with a certain norm on  $E$  (more accurately, with a certain function corresponding to this norm) a

method of solving general convex problems in the classes  $C^{\circ 0}(G, E, m)$  on convex bodies  $G \subset E$ . We now remark that this is not a very natural situation: classes of the  $C^{\circ 0}(G, E, m)$  type are defined only in terms of the topology of  $E$ , irrespective of the choice of the actual norm which gives this topology. In other words, the class  $C \in C^{\circ 0}(G, E, m)$  is connected not with the actual norm  $\|\cdot\|$  on the given linear space  $E$ , but with a whole class of mutually equivalent norms. Therefore, the possibilities of methods of solving problems in the class  $C$  do not have to depend on the (in general, arbitrary) choice of a norm in the given equivalence class of norms. They must be determined by the linear-topological properties of  $G$  (in the finite-dimensional case, by the affine properties of  $G$ ), and not by the metrical properties of this body. But at the same time MD-methods are connected with completely concrete metrics. There is thus a certain disparity of ideas between the methods we are recommending and the 'spirit' of the problems we are considering.

In actual fact there is no real disparity here at all: in a *proper* application of MD-methods to the solution of general convex problems the possibilities of the methods are determined by the linear-topological properties of  $G$ . To explain what we mean by a 'proper' application of MD-methods, let us consider the only practically important (and formally very simple) case where  $E$  is finite-dimensional, i.e.  $E$  is the space  $R^n$ ,  $1 < n < \infty$ . In this case all norms on  $E$  are equivalent, and so any MD-method associated with any norm on  $E$  can be used for solving problems in  $C$ . We fix on any norm  $\|\cdot\|$  on  $R^n$  and a regular function, corresponding to it, say  $V$ , on the dual space (which is identified as a linear space with  $R^n$ ). There is, of course, more than one choice for  $V$ . But we know that, to reduce the laboriousness of the method, we should choose a regular function  $V$  which is as smooth as possible (i.e. which has as large a  $\gamma_V(\cdot)$  as possible).

Together with the particular norm  $\|\cdot\|$  we can consider all possible norms  $\|x\|^A = \|Ax\|$ , where  $A$  is an invertible  $n \times n$  matrix. To each of them there corresponds a regular function  $V^A(\varphi) = V((A^*)^{-1}\varphi)$ . Any of the  $\overline{\text{MD}}_{V,A}$ -methods can be used to solve problems in  $C$ ; their bounds for the laboriousness will be

$$N_V \left( \frac{v - v_0}{4\alpha_{\parallel \cdot \parallel}^2(G)} \right),$$

where  $N_V(v)$  depends only on  $V$  and  $v$ , and  $\alpha_{\parallel \cdot \parallel}^2(G)$  is the asphericity of  $G$  relative to  $\|\cdot\|^A$  (or, what comes to the same thing, the asphericity of  $AG$  relative to  $\|\cdot\|$ ). A 'proper' application of  $\overline{\text{MD}}_{V,A}$ -methods consists in applying that one of the  $\text{MD}_{V,A}$ -methods for which the bound for the laboriousness is minimal, i.e. corresponding to that  $A$  for which the asphericity  $\alpha_{\parallel \cdot \parallel}^2(AG)$  is minimal (in other words,  $AG$  is 'minimally aspheric' relative to  $\|\cdot\|$ ). The corresponding optimal value of  $\alpha_{\parallel \cdot \parallel}^2(AG)$  is a purely affine characteristic of  $G$ . We can thus say that the possibilities of the method of mirror descent associated with  $V$  in

actual fact are determined by the affine properties of  $G$ . Similar considerations apply to other versions of the  $\overline{MD}$ -methods.

Let us consider, for example, what we are actually given by the  $\overline{MD}_p$  methods—the methods of mirror descent associated with  $\|\cdot\|_p$  on  $R^n$  (corresponding to the identification of  $R^n$  with  $I_p^{(n)}$ ). We define for all  $p$ ,  $1 \leq p < \infty$ , the affine characteristic  $\alpha_p(G)$  of a convex, bounded, closed body  $G \subset R^n$  as

$$\inf \{\alpha_{\|\cdot\|_p}(AG) \mid A \text{ is an invertible } n \times n \text{ matrix}\}.$$

Let  $A_G^p$  be the corresponding optimal value of  $A$  (it is easy to see that it exists). A 'proper' application of  $\overline{MD}_p$  to a class  $C \in C^{\circ 0}(G, R^n, m)$  requires our finding  $A_G^p$  and applying the method associated with the  $I_p$ -norm  $\|A_G^p x\|_p$ . With this approach the bounds for the laboriousness of  $\overline{MD}_p$ -methods on the class  $C$  will be

$$\begin{aligned} \frac{d(1) \ln n \alpha_1^2(G)}{(v - v_0)^2}, & \quad \text{if } p = 1, \\ \frac{d(p) \alpha_p^2(G)}{(v - v_0)^2}, & \quad \text{if } 1 < p \leq 2, \\ \frac{d(p) \alpha_p^p(G)}{(v - v_0)^p}, & \quad \text{if } 2 < p < \infty. \end{aligned}$$

These bounds, as assumed, depend on the affine properties of  $G$ .

The next step in the optimization of the method must consist in choosing the optimal  $p$  (minimizing the bound). Strictly speaking, we have no need to restrict ourselves to the scale of  $I_p$ -norms and the methods corresponding to them; other norms on  $R^n$  could be brought in. Of course, the solution of the problem stated of making a 'proper' choice of an MD-method of solving a given class of general convex problem is, in the general case, an 'expensive' procedure, even if we consider only the gamut of  $I_p$ -methods. How to calculate  $\alpha_p(G)$  and find  $A_G^p$  is completely obscure. However, for the frequently encountered bodies of 'simple form' these problems may turn out to be simple, and what is more, no one is compelling us to solve them exactly. However, it must be remembered that to neglect such an 'optimization of the norm with which the method is connected' will lead to a loss as regards the laboriousness of the method compared with an 'optimal' MD-method.

Similar remarks may be made in connection with MD-methods of solving classes of Lipschitz-convex problems. These classes are, of course, defined in terms of a concrete norm  $\|\cdot\|$  on  $E$ . But the store of problems in a class of type  $C_{Lip}^{\circ 0}(G, E, \|\cdot\|, m)$  depends by definition only on  $G$  and the topology of  $E$ . Thus, if  $\|\cdot\|'$  is equivalent to  $\|\cdot\|$ , i.e. if

$$0 < \underline{a} = \inf_{x \neq 0} \frac{\|x\|}{\|x\|'} \leq \sup_{x \neq 0} \frac{\|x\|}{\|x\|'} = \bar{a} < \infty,$$

then the store of problems is the same in any two classes of the type

$$C_{Lip}^{\circ 0}(G, E, \|\cdot\|, m) \text{ and } C_{Lip}^{\circ 0}(G, E, \|\cdot\|', m).$$

The classes can differ only in the values of  $v_0$  (i.e. in the accuracy of the oracle) and in the definition of the normalizing factors. Thanks to this, methods of solving problems of one of these classes can be applied (observing the necessary precautions) to problems of the other class as well. We shall show how this is done, supposing for simplicity that it is required to solve problems of a class of type  $C_{Lip}^0(G, E, \|\cdot\|, m)$  (i.e. that we have an exact oracle available), and assuming that  $G$  is a body.

We consider a concrete class  $C$  of the type  $C_{Lip}^0(G, E, \|\cdot\|, m)$ , and let the oracle  $\mathcal{O}$  correspond to it. The class  $C$  regarded as a set of problems equipped with the oracle is at the same time a certain class  $C'$  of the type  $C^0(G, E, \|\cdot\|', m)$  (also regarded as a set of problems equipped with an oracle). The classes  $C$  and  $C'$  differ from one another only in the normalizing mapping. [Strictly speaking, for this it would be necessary to require of the oracle  $\mathcal{O}$  that the set  $\{\psi(x, f) \mid x \in G\}$  be contained in the closure of the set  $\{\psi(x, f) \mid x \in \text{int } G\}$  (otherwise  $\psi$  might not satisfy condition (2) in the definition of a  $(G, 0)$ -oracle relative to  $\|\cdot\|'$ ; see Section 2.2.5. However, this restriction on  $\mathcal{O}$  is of a purely academic nature, and of course, it is satisfied in all reasonable situations.]

Let  $L_{\|\cdot\|}(f_j), L_{\|\cdot\|'}(f_j)$  be the Lipschitz constants of  $f_j$  relative to  $\|\cdot\|$  and  $\|\cdot\|'$  resp. It is clear that

$$L_{\|\cdot\|}(f_j) \leq \bar{a} L_{\|\cdot\|'}(f_j) \text{ and } \rho_{\|\cdot\|}(G) \leq \underline{a}^{-1} \rho_{\|\cdot\|'}(G).$$

Let

$$\kappa = \frac{\rho_{\|\cdot\|}(G)}{\bar{a} \rho_{\|\cdot\|'}(G)}.$$

Then, for all  $f_j$ ,

$$\kappa \leq \frac{L_{\|\cdot\|}(f_j) \rho_{\|\cdot\|}(G)}{L_{\|\cdot\|'}(f_j) \rho_{\|\cdot\|'}(G)} \text{ and } \kappa \geq \frac{\underline{a}}{\bar{a}} \equiv a^{-1} (\|\cdot\|, \|\cdot\|').$$

It is clear that an approximate solution of the problem  $f \in C$ , regarded as a problem in  $C'$  with a relative error (in the sense of the second class)  $\bar{v} = \kappa v$ ,  $v > 0$ , is, at the same time, a solution of  $f$  with a relative error  $v$  in the sense of the first class. Thus, a method of solving problems of the class  $C'$  with a relative error  $\kappa v$  induces a method for problems of  $C$  with the same laboriousness and with the relative error  $v$ . In particular, if the method mentioned is an  $\overline{MD}$ -method associated with  $\|\cdot\|'$  and having a bound  $M_{\|\cdot\|'}(\bar{v})$  for the laboriousness, where  $\bar{v}$  is the accuracy of solution of problems in  $C'$ , then it induces a method of solving problems in  $C$  with the bound for laboriousness

$$M_{\|\cdot\|'}(\kappa v) \leq M_{\|\cdot\|'}(a^{-1} (\|\cdot\|, \|\cdot\|') v)$$

Applied to MD-methods, a construction of this kind is possible for all  $v_0$ , and not just for  $v_0 = 0$ . The bound for the laboriousness of a  $C_{\text{Lip}}^{v_0}(G, E, \|\cdot\|, m)$ -method induced by an MD-method of the class  $C_{\text{Lip}}^{v_0}(G, E, \|\cdot\|, m)$  with laboriousness  $M_{\|\cdot\|}(v - \kappa v_0)$  will be  $M_{\|\cdot\|}(\kappa(v - v_0))$ .

Summarizing, we can say that, although classes of Lipschitz-convex problems of the type  $C_{\text{Lip}}^{v_0}(G, E, \|\cdot\|, m)$ , in contrast to classes of general convex problems, are defined in terms of a concrete norm on  $E$ , nevertheless there is freedom in the choice of an MD-method for solving them. This freedom comes from the possibility of applying MD-methods associated with any norms which are equivalent to the original norm. In part, this freedom will be obvious from the very start (a regular function  $V'$  associated with  $\|\cdot\|'$  can be changed by obvious similarity transformations into a regular function  $V$  corresponding to  $\|\cdot\|$ ; and the bound for the laboriousness of  $\text{MD}_V$  is approximately equal to  $M_V((v - v_0)/a(\|\cdot\|, \|\cdot\|'))$ ). However, the freedom presented by the approach mentioned above is rather wider; the method there described changes the bound  $M_V(v - v_0)$  for the laboriousness into  $M_V(\kappa(v - v_0))$ , and  $\kappa$  is  $\geq 1/(a(\|\cdot\|, \|\cdot\|'))$  (it can also happen that  $\rho_{\|\cdot\|}(G)/\rho_{\|\cdot\|'}(G) \gg a$ ). However, if  $G$  is a  $\|\cdot\|$ -ball, then both approaches coincide, for then  $\rho_{\|\cdot\|}(G)/\rho_{\|\cdot\|'}(G) = a$  and  $\kappa = 1/(a(\|\cdot\|, \|\cdot\|'))$ .

### 3.4.6

We now turn to the proof of the results already announced for the methods described.

**Theorem.** The following assertions are true.

- (i) Let  $E, \|\cdot\|$  be a regular space, and let  $V$  be a corresponding regular function. Both versions of the method  $\overline{\text{MD}}_V$  constructed for a relative accuracy  $v$ , where  $2\alpha > v > v_0$ , ensure this accuracy in the solution of any problem in the class  $C \in C^v(G, E, m)$  with a laboriousness  $\leq M_{V,G}(v - v_0)$ . If  $E, \|\cdot\|$  is a Hilbert space, then similar results hold for  $\Gamma$  (with  $M_{V,G}$  replaced by  $M^C$ ).
- (ii) Each of the methods in Section 3.4.3.2 (which are admissible for the given space  $E, \|\cdot\|$ ), constructed for an absolute error vector  $\varepsilon = (\varepsilon_0, \varepsilon_1, \dots, \varepsilon_m) > 0$ , solves every problem  $f \in C \in C^{v_0}(G, E, m)$  with an absolute error  $\leq \bar{\varepsilon}_j = \varepsilon_j + v_0 r_j(f)$  and with laboriousness  $\leq 8\bar{N}(v(f, \varepsilon)/(1 + v_0))$ . Here  $v(f, \varepsilon) = \min_{0 \leq j \leq m} \varepsilon_j/r_j(f)$ , and  $\bar{N}(v)$  is the bound for laboriousness corresponding to a method (from Section 3.4.3.2) constructed for a relative error  $v$ .

*Proof.* I. We start by proving (i). The proof is on the same lines as for the case of Lipschitz convex problems (*cf.* proof of Theorem 3.3.5).

1°. The bounds for the laboriousness of the methods are derived from the inequalities, 'obvious' in view of  $\overline{\text{MD}}$  1.4,

$$\delta_i \geq a \quad \text{if } \eta_i^{j(i)} \neq 0, \quad (4.6)$$

where  $a = (v - v_0)/4\alpha$  for the  $\text{MD}_V$ -method, and  $a = (v - v_0)/(2\alpha)$  for the method  $\Gamma$ .

[The authors kindly explain 'obvious' by the following note—*Trans.*: For, by  $\overline{\text{MD}}$  1.4, when  $j(i) = 0$  we have, from the definition of  $r_0(i)$ , that

$$\begin{aligned} \delta_i &= g_i^0(\tilde{x}_i - a_i)(\|\eta_i^{j(i)}\|_*)^{-1} + (v - v_0) \left\{ \sup_{x \in G} \langle \xi_i | x - \tilde{x}_i \rangle + \langle \xi_i | \tilde{x}_i \rangle \right\} \\ &\geq (v - v_0) \sup_{x \in G} \langle \xi_i | x \rangle. \end{aligned}$$

But if  $j(i) > 0$ , then we have, from the definition of  $j(i)$  and  $r_{j(i)}(i)$ ,

$$r_{j(i)}(i) = \sup_{x \in G} \langle \eta_i^{j(i)} | x - \tilde{x}_i \rangle,$$

and so, by  $\overline{\text{MD}}$  1.4, we obtain

$$\begin{aligned} \delta_i &= (v - v_0) \langle \xi_i | \tilde{x}_i \rangle + (\|\eta_i^{j(i)}\|_*)^{-1} g_i^{j(i)}(\tilde{x}_i) \\ &\geq (v - v_0) \langle \xi_i | \tilde{x}_i \rangle + (\|\eta_i^{j(i)}\|_*)^{-1} \sup_{x \in G} \langle \eta_i^{j(i)} | x - \tilde{x}_i \rangle \\ &= (v - v_0) \sup \{ \langle \xi_i | x \rangle | x \in G \}. \end{aligned}$$

Thus we always have

$$\delta_i \geq (v - v_0) \sup_{x \in G} \langle \xi_i | x \rangle.$$

Hence  $\delta_i \geq (v - v_0)/(2\alpha)$ , because  $\|\xi_i\|_* = 1$  and  $G$  contains a ball of radius  $1/(2\alpha)$  with its centre at 0.]

We recall that the laboriousness bounds for the methods  $\text{MD}_V$  and  $\Gamma$  were derived from inequalities just like (4.6); the difference from the Lipschitz case is that  $a$  is now  $4\alpha$  times smaller than before. The derivation of the bounds from (4.6) repeats the corresponding argument in Section 3.3 word for word.

2°. As regards the error bounds, in Section 3.3 they were all derived from a single conditional assertion

(A): if on a given compatible problem  $f$  the error of the method considered exceeds the value  $v$ , then, for all  $i \leq M_f$ ,

$$\eta_i^{j(i)} \neq 0 \quad \text{and} \quad \langle \xi_i | \tilde{x}_i - x^* \rangle \geq \delta_i, \quad \tilde{x}_i = \pi_G(x_i), \quad (4.7)$$

where  $x^*$  is the solution of  $f$ .

The relation (4.7) was later shown to lead to a contradiction, thus proving that the premise (A) cannot hold. In the case now being considered, the conditional assertion in (A) is again true. This is what we are about to prove.

But we shall not show that (4.7) leads to a contradiction, because the argument is word for word the same as in Section 3.3.

So we shall prove (A), or, rather, we shall in fact prove the assertion

- (A'): Let  $f \in C^{\circ\circ}(G, E, m)$  be a compatible problem which does not satisfy the following condition (B):  
(B): the result  $\bar{x}$  of applying the method considered to the problem  $f$  is different from  $*$ , and for it

$$g_{j,\bar{x}}^f(\bar{x}) \leq (v - v_0)\bar{r}_j(f) + f_{j*},$$

where

$$f_{j*} = \begin{cases} 0, & \text{if } j \geq 1, \\ f_*, & \text{if } j = 0. \end{cases}$$

Then for all trajectories of the method on  $f$  the relations (4.7) hold.

The assertion (A') proves the bounds formulated in (i) for the relative error of the methods. For, it has already been pointed out that the relations (4.7) cannot hold for all  $i \leq M_f$ . Suppose  $f$  is compatible. In view of (A'),  $f$  cannot fail to satisfy condition (B), i.e. for all  $j$  with  $0 \leq j \leq m$ ,

$$g_{j,\bar{x}}^f(\bar{x}) \leq (v - v_0)\bar{r}_j(f) + f_{j*}.$$

As pointed out in Section 3.4.3.2, when  $j \geq 1$  we have  $\bar{r}_j(f) \leq r_j(f)$ , i.e.  $f_j(\bar{x}) \leq vr_j(f)$  for  $j \geq 1$ . Further, recalling the definition of  $\bar{r}_0(f)$ , we obtain from (B) when  $j = 0$  that, for a certain  $i_1$  with  $j(i_1) = 0$ ,

$$\begin{aligned} g_{0,\bar{x}}^f(\bar{x}) &\leq (v - v_0) \left( \sup_{y \in G} g_{0,\bar{x}_{i_1}}^f(y) - g_{0,\bar{x}_{i_1}}^f(\bar{x}_{i_1}) \right) + f_* \\ &\leq (v - v_0) (f_0 - g_{0,\bar{x}}^f(\bar{x})) + f_*, \end{aligned}$$

(here we have used the fact that, by  $\overline{MD}$  2,  $g_{0,\bar{x}}^f(\bar{x}) \leq g_{0,\bar{x}_{i_1}}^f(\bar{x}_{i_1})$ ).

Hence

$$(1 + (v - v_0))(g_{0,\bar{x}}^f(\bar{x}) - f_*) \leq (v - v_0) (\sup_G f_0 - f_*) \leq (v - v_0)r_0(f).$$

Thus

$$g_{0,\bar{x}}^f(\bar{x}) \leq f_* + (v - v_0)r_0(f) \text{ and } f_0(\bar{x}) \leq f_* + vr_0(f).$$

Summing the bounds obtained for  $f_j(\bar{x})$ ,  $j \geq 1$ , and  $f_0(\bar{x})$ , we find that  $v(\bar{x}, f) \leq v$ , as required.

But if  $f$  is incompatible, then, from  $\overline{MD}$  2 and the relations  $r_j(i) \leq r_j(f)$  for  $j \geq 1$ , it follows at once that  $v(\bar{x}, f) \leq v$ .

So (i) has been proved.

3°. So now we shall prove (A'). Let  $f$  be an object satisfying the premise (A'), and let  $x^*$  be the solution of  $f$ . We shall satisfy ourselves that, for all  $i \leq M_f$ , we

have

$$\langle \eta_i^{(i)} | \bar{x}_i - x^* \rangle \geq \kappa_i, \quad (4.8)$$

where

$$\kappa_i = \begin{cases} g_i^{(0)}(\bar{x}_i) - a_i + (v - v_0)r_0(i), & \text{for } j(i) = 0, \\ g_i^{(i)}(\bar{x}_i), & \text{for } j(i) \neq 0. \end{cases}$$

For, if we had

$$g_i^{(i)}(\bar{x}_i) < g_i^{(i)}(x^*) + \kappa_i, \quad (4.9)$$

then would  $j(\bar{i}) = 0$ . Because, when  $j(\bar{i}) > 0$ , the second term on the right-hand side of (4.9) would be equal to  $g_i^{(i)}(x^*)$ , and the first term would be non-positive; but then (4.9) would be impossible.

Thus  $j(\bar{i}) = 0$ . In that case (4.9) gives

$$g_i^0(\bar{x}_i) < f_0(x^*) + g_i^0(\bar{x}_i) - a_i + (v - v_0)r_0(\bar{i}),$$

i.e.

$$a_{M_f} \leq f_* + (v - v_0)\bar{r}_0(f).$$

Hence and from  $\overline{MD}$  2 it immediately follows that  $f$  satisfies (B), contrary to the definition of  $f$ . It is easy to see that the hypothesis  $\eta_i^{(i)} = 0$  for some  $i \leq M_f$  leads to a similar contradiction. Thus we may suppose that (4.8) is satisfied and that

$$\eta_i^{(i)} \neq 0 \quad \text{for } i \leq M_f. \quad (4.10)$$

4°. From (4.8) and (4.10) we deduce the required assertion (4.7). In these circumstances we have, for every  $i \leq M_f$ ,

$$\langle \xi_i | \bar{x}_i - x^* \rangle \geq \kappa_i (\|\eta_i^{(i)}\|_\infty)^{-1} = \delta_i.$$

Hence

$$\begin{aligned} \langle \xi_i | \bar{x}_i - x^* \rangle &= \langle \xi_i | \bar{x}_i - x^* \rangle + (\theta^{-1} - 1) \langle \xi_i | \bar{x}_i \rangle \\ &\geq \delta_i + (v - v_0) \langle \xi_i | \bar{x}_i \rangle \equiv \delta_i, \end{aligned}$$

as required (we have used the fact that  $\theta^{-1} - 1 = v - v_0$  by the definition of  $\theta$ ). Thus, we have proved (A'), and with it, the part (i) of the theorem.

II. It remains to prove (ii). With the notation of Section 3.4.3.2 we have for the result  $\hat{x}_i$  of a procedure  $\mathcal{B}(v^i)$  applied to a problem  $f \in C^{\circ\circ}(G, E, m)$  that: if  $f$  is compatible, then  $\hat{x}_i \neq *$ , and when  $\hat{x}_i \neq *$  we have

$$g_{0,\hat{x}_i}^f(\hat{x}_i) \leq f_* + v^i \bar{r}_0^i(f), \quad g_{j,\hat{x}_i}^f(\hat{x}_i) \leq v^i \bar{r}_j^i(f), \quad j \geq 1$$

(when  $i > 0$ , this follows from (i) for compatible  $f$ , and from  $\overline{MD}$  2 for incompatible  $f$ ; we leave the proof for the case  $i = 0$  to the reader).

From the rule in Section 3.4.3.2 for stopping, the work of the method ends at the moment (let its index be  $i_f$ ) of the end of the first procedure for which  $v^i \bar{r}_j^i(f) \leq \varepsilon_j$  ( $0 \leq j \leq m$ ). Therefore the result of applying  $\mathcal{B}$  to  $f$  satisfies the

requirements to the necessary accuracy. We estimate the laboriousness of  $\bar{N}$  on  $f$ . It is possible that  $i_f = 0$ . Then the laboriousness of  $\bar{N}$  on  $f$  is equal to 2, i.e. it is automatically less than  $8\bar{N}(\nu(f, \epsilon)/(1 + \nu_0))$ . Now let  $i_f > 0$ . Then

$$2^{i_f+i_0-1} = \bar{N}(\nu^{i_f-1}) < \bar{N}\left(\frac{\nu(f, \epsilon)}{1 + \nu_0}\right),$$

since  $\nu^{i_f-1} > \nu(f, \epsilon)/(1 + \nu_0)$ . At the same time, the laboriousness of  $\bar{N}$  on  $f$  is equal to

$$2 + \sum_{s=1}^{i_f} 2^{i_0+s} \leq 2^{i_0+i_f+2} = 8 \cdot 2^{i_f-1+i_0} \leq 8\bar{N}\left(\frac{\nu(f, \epsilon)}{1 + \nu_0}\right),$$

as required. The theorem has now been proved.

### 3.5 SOME ADDITIONAL PROOFS

In the following proofs we use without special citation the ideas and notation of the relevant sections of the main text.

We prove the estimates (2.6) of Section 3.2 and calculate the corresponding constants  $c(p)$ .

1°. Let  $E = L_p(T, \mu)$  and let  $1 < p < \infty$ . Then  $E^* = L_q(T, \mu)$  with  $q = p/(p-1)$ . Let  $r(\varphi)$  denote the function  $\|\varphi\|_q$  on  $E^*$ . Then  $r(\varphi)$  is a convex, homogeneous (of degree 1) function on  $E^*$ . It is easily seen that the support functional  $r'(\varphi) \in E$  to  $r(\varphi)$  is uniquely determined (when  $\varphi \neq 0$ ) and that  $r'(\lambda\varphi) = r'(\varphi)$  when  $\lambda > 0$ . If  $r(\varphi) = 1$ , however, then

$$[r'(\varphi)](t) = |\varphi(t)|^{q-1} \operatorname{sgn} \varphi(t). \quad (5.1)$$

We shall prove the following relation: if  $\varphi, \psi \in L_q$  and if  $\|\varphi\|_q = \|\psi\|_q = 1$ , then

$$\|r'(\varphi) - r'(\psi)\|_p \leq \begin{cases} \frac{2^{(2-p)/p}}{p-1} \|\varphi - \psi\|_q, & \text{if } q > 2, \\ 2\|\varphi - \psi\|_q^{1/(p-1)}, & \text{if } q \leq 2. \end{cases} \quad (5.2)$$

For, let  $\|\varphi\|_q = \|\psi\|_q = 1$ . We define

$$T_+ = \{t | \operatorname{sgn} \varphi(t) = \operatorname{sgn} \psi(t)\}, T_- = T \setminus T_+.$$

Writing

$$\Delta(t) = |[r'(\varphi)](t) - [r'(\psi)](t)|,$$

we obtain from (5.1) that when  $t \in T_+$

$$\Delta(t) = |\varphi(t)|^{q-1} - |\psi(t)|^{q-1}|.$$

Suppose first that  $q \leq 2$ . Then the function  $s^{q-1}$  is concave for  $s \geq 0$ , and so,

### Some additional proofs

when  $q \leq 2$ ,

$$|s_1|^{q-1} - |s_2|^{q-1} \leq |s_1 - s_2|^{q-1}.$$

Thus, when  $q \leq 2$ , we have, for  $t \in T_+$ ,

$$\Delta(t) \leq |\varphi(t) - \psi(t)|^{q-1}.$$

Further, it is clear that when  $t \in T_-$  the inequality

$$\Delta(t) \leq 2|\varphi(t) - \psi(t)|^{q-1}.$$

holds. Thus, always when  $q \leq 2$ ,

$$\Delta(t) \leq 2|\varphi(t) - \psi(t)|^{q-1}.$$

Hence

$$\int \Delta^p(t) d\mu(t) \leq 2^p \int |\varphi(t) - \psi(t)|^{(q-1)p} d\mu(t) = 2^p \|\varphi - \psi\|_q^q.$$

Therefore

$$\|\Delta(t)\|_p \leq 2 \|\varphi - \psi\|_q^{q/p},$$

as required in (5.2) when  $q \leq 2$ .

Now let  $q > 2$ . The function  $s^{q-1}$  is then convex for  $s \geq 0$ , and so, when  $s_1 \geq s_2 \geq 0$

$$s_1^{q-1} - s_2^{q-1} \leq \left[ \frac{d}{ds} \Big|_{s=s_1} (s^{q-1}) \right] (s_1 - s_2) = (q-1)s_1^{q-2}(s_1 - s_2).$$

Therefore, when  $t \in T_+$  we have, putting  $h(t) = \max \{|\varphi(t)|, |\psi(t)|\}$ ,

$$\Delta(t) \leq (q-1)h^{q-2}(t)|\varphi(t) - \psi(t)|. \quad (5.3)$$

It is clear that (5.3) is also true for  $t \in T_-$ . Therefore

$$\int \Delta^p(t) d\mu(t) = (q-1)^p \int h^{(q-2)p}(t)|\varphi(t) - \psi(t)|^p d\mu(t).$$

Putting

$$\sigma \equiv \frac{q}{p} > 1 \quad \text{and} \quad \tau \equiv \frac{\sigma}{\sigma-1} = \frac{q}{(q-2)p}$$

and applying Hölder's inequality, we obtain

$$\begin{aligned} \int \Delta^p(t) d\mu(t) &\leq (q-1)^p \left[ \int h^{(q-2)p\tau}(t) d\mu(t) \right]^{\frac{1}{\tau}} \left[ \int |\varphi(t) - \psi(t)|^{p\sigma} d\mu(t) \right]^{\frac{1}{\sigma}} \\ &= (q-1)^p \left[ \int h^q(t) d\mu(t) \right]^{\frac{1}{\tau}} \left[ \int |\varphi(t) - \psi(t)|^q d\mu(t) \right]^{\frac{1}{\sigma}} \\ &\leq (q-1)^p 2^{\frac{1}{\tau}} \|\varphi - \psi\|_q^{\frac{q}{\sigma}} = (q-1)^p \cdot 2^{\frac{(q-2)p}{q}} \|\varphi - \psi\|_q^p, \end{aligned}$$

i.e.

$$\|\Delta(t)\|_p \leq (q-1) \cdot 2^{\frac{q-2}{q}} \|\varphi - \psi\|_q = \frac{2^{\frac{2-p}{p}}}{p-1} \|\varphi - \psi\|_q,$$

which completes the proof of (5.2).

2°. We now estimate the modulus of smoothness  $\omega_p$  of the function  $V_p(\varphi)$ .

Let  $\varphi, \psi \in L_q$  and let  $\|\varphi - \psi\|_q = t$ . We estimate  $\|V'_p(\varphi) - V'_p(\psi)\|_p$ . We may suppose that  $\|\psi\|_q \geq \|\varphi\|_q$ . Let

$$\bar{\varphi} = \varphi / \|\varphi\|_q, \quad \bar{\psi} = \psi / \|\psi\|_q$$

(with  $\bar{\varphi} = 0$  when  $\varphi = 0$ ,  $\bar{\psi} = 0$  when  $\psi = 0$ ). Further suppose that  $\|\varphi\|_q > 0$ . Then

$$\begin{aligned} \|\bar{\varphi} - \bar{\psi}\|_q &= \left\| \frac{\|\psi\|_q \varphi - \|\varphi\|_q \psi}{\|\varphi\|_q \|\psi\|_q} \right\|_q \\ &\leq \frac{\|\varphi - \psi\|_q \|\varphi\|_q + \|\varphi\|_q \|\psi - \varphi\|_q}{\|\varphi\|_q \|\psi\|_q} = \frac{2\|\varphi - \psi\|_q}{\|\psi\|_q}. \end{aligned}$$

This inequality is clearly also true for  $\varphi = 0$  (here we take  $0/0 = 0$ ). So

$$\|\bar{\psi} - \bar{\varphi}\|_q \leq \frac{2\|\psi - \varphi\|_q}{\|\psi\|_q}. \quad (5.4)$$

Suppose first that  $\|\varphi\|_q \geq 1$ . Then, from (5.4),

$$\|\bar{\varphi} - \bar{\psi}\|_q \leq 2\|\varphi - \psi\|_q.$$

At the same time,  $V'_p(\varphi) = r'(\bar{\varphi})$ ,  $V'_p(\psi) = r'(\bar{\psi})$ , and *a fortiori*, in view of (5.2), the inequality

$$\|V'_p(\varphi) - V'_p(\psi)\|_p \leq 2 \begin{cases} \frac{2 \cdot 2^{\frac{(2-p)}{p}}}{p-1} \|\varphi - \psi\|_q, & 1 < p \leq 2, \\ 2 \cdot 2^{\frac{1}{p-1}} \|\varphi - \psi\|_q^{\frac{1}{p-1}}, & p > 2. \end{cases} \quad (5.5)$$

holds.

Now let  $\|\psi\|_q \geq 1$ ,  $\|\varphi\|_q \leq 1$ . Then, as before,

$$\|\bar{\varphi} - \bar{\psi}\|_q \leq 2\|\varphi - \psi\|_q.$$

At the same time,

$$V'_p(\psi) = r'(\bar{\psi}), \quad V'_p(\varphi) = \|\varphi\|_q r'(\bar{\varphi})$$

and

$$\begin{aligned} \|V'_p(\psi) - V'_p(\varphi)\|_p &\leq (1 - \|\varphi\|_q) + \|r'(\bar{\varphi}) - r'(\bar{\psi})\|_p \\ &\leq \|\varphi - \psi\|_q + \|r'(\bar{\varphi}) - r'(\bar{\psi})\|_p. \end{aligned}$$

Moreover, it is clear that  $\|V'_p(\varphi) - V'_p(\psi)\|_q \leq 2$  for all  $\varphi$  and  $\psi$ . Hence and from (5.2) the inequality (5.5) immediately follows.

The case  $\|\psi\|_q \leq 1$  remains to be examined. In this case

$$V'_p(\varphi) - V'_p(\psi) = (\|\varphi\|_p - \|\psi\|_q) r'(\bar{\varphi}) + \|\psi\|_q (r'(\bar{\varphi}) - r'(\bar{\psi}))$$

(here  $r'(0) = 0$ ). Thus, by (5.2) and (5.4)

$$\|V'_p(\varphi) - V'_p(\psi)\|_p$$

$$\leq \|\varphi - \psi\|_q + \|\psi\|_q \begin{cases} \frac{2 \cdot 2^{\frac{(2-p)}{p}}}{p-1} \frac{\|\varphi - \psi\|_q}{\|\psi\|_q}, & p \leq 2, \\ 2 \cdot 2^{\frac{1}{p-1}} \frac{\|\varphi - \psi\|_q^{\frac{1}{p-1}}}{\|\psi\|_q^{\frac{1}{p-1}}}, & p > 2, \end{cases}$$

and hence, as above (5.5) follows. Thus

$$\omega_p(t) \leq \begin{cases} \frac{2^{\frac{2+p}{p}}}{p-1} t, & 1 < p \leq 2, \\ 2^{\frac{2p-1}{p}} t^{\frac{1}{p-1}}, & p > 2. \end{cases} \quad (5.6)$$

Hence it follows that

$$\gamma_p(s) \geq \begin{cases} (p-1) 2^{-\frac{2+p}{p}} s, & 1 < p \leq 2, \\ 2^{\frac{(1-2p)(p-1)}{p}} s^{p-1}, & p > 2, \end{cases}$$

and (2.6) is proved. Moreover, we can take

$$c(p) = \begin{cases} (p-1) 2^{-\frac{2+p}{p}} \geq \frac{1}{6}(p-1), & 1 < p \leq 2, \\ 2^{\frac{(1-2p)(p-1)}{p}} \geq 4^{-(p-1)}, & p > 2. \end{cases} \quad (5.7)$$

3°. We now consider the case  $p = 1$ . Let  $E = l_1^{(n)}$  with  $1 < n < \infty$ , let  $\tau \geq 2$ , and  $\bar{p} = \tau/(\tau-1)$ . For  $\varphi, \psi \in l_\infty^{(n)}$ , we have, by (5.6),

$$\|(V')'(\varphi) - (V')'(\psi)\|_{\bar{p}} \leq \frac{8}{\bar{p}-1} \|\varphi - \psi\|_1.$$

Further, it is clear that

$$\|\varphi - \psi\|_1 \leq n^{1/\tau} \|\varphi - \psi\|_\infty$$

and

$$\|(V')'(\varphi) - (V')'(\psi)\|_1 \leq n^{1/\tau} \|(V')'(\varphi) - (V')'(\psi)\|_{\bar{p}}.$$

Thus

$$\|(V')'(\varphi) - (V')'(\psi)\|_1 \leq n^{2/\tau} \frac{8}{\bar{p}-1} \|\varphi - \psi\|_\infty. \quad (5.8)$$

We now choose  $\tau$  so as to minimize  $n^{2/\tau}/(\bar{p}-1)$  under the condition  $\tau \geq 2$ . Unconditional minimization (without considering the condition  $\tau \geq 2$ ) gives

$$\tau = \ln n + \sqrt{\ln^2 n - 2 \ln n}. \quad (5.9)$$

To simplify the choice of  $\tau$ , we take an approximate value of the root,  $\tau = 2 \ln n$ . With this choice of  $\tau$ , the quantity  $n^{2/\tau} 8/(\bar{p}-1)$  will be equal to

$$8e(\tau-1) = 8e(2 \ln n - 1) \leq 16e \ln n.$$

Thus, for the choice  $\tau = 2 \ln n$  (we suppose  $n > 2$ ) we obtain, in (2.11),  $c(1) = 1/(16e)$ . When  $n = 2$  we can take  $\tau = 2$ , and then (2.11) will hold with  $c(1)$  as indicated. Thus, for the space  $\mathbb{I}_1^{(n)}$  the regular function should be chosen from the condition

$$\tau = \max \{2, 2 \ln n\}; \text{ and then } c(1) = 1/(16e). \quad (5.10)$$

# 4

## Complexity of classes of general convex problems (exact first-order oracle)

In this chapter lower bounds are constructed for the complexity of classes of general convex problems in Section 4.1 and of classes of Lipschitz-convex problems in Section 4.2. These bounds are compared with the upper bounds of the complexity which are provided by the results of the two previous chapters about the laboriousness of the methods of optimization described therein. Such comparisons enable us to characterize the behaviour of the complexity functions of the classes considered, and to trace out the domains of theoretical optimality (more precisely, of sub-optimality) of methods of solving the general convex and Lipschitz-convex problems in Chapters 2 and 3 (Section 3.3). In Sections 4.4–4.6 we give the proofs of the bounds formulated in Sections 4.1 and 4.2. The results in this chapter are based mainly on the authors' paper [30].

### 4.1 COMPLEXITY OF CLASSES OF GENERAL CONVEX PROBLEMS

In this section we formulate and prove two-sided estimates for the stochastic and deterministic complexities of classes of general convex problems.

#### 4.1.1

First we introduce certain geometrical characteristics of convex bodies in Banach spaces. These characteristics enable these bodies to be compared with  $L_p$ -balls, and in the final count make it possible to formulate the required bounds for the complexity of classes of convex problems. Let  $E$  be a real Banach space, and let  $n$ ,  $1 \leq n \leq \infty$ , be its dimension; let  $G$  be a bounded, convex, closed body in  $E$ , and let  $\|\cdot\|$  be a norm on  $E$ .

We start by defining the  $k$ -dimensional  $p$ -asphericity  $\alpha_{p,k}(\|\cdot\|)$  of a norm  $\|\cdot\|$

on  $E$ . Here  $k$  is a natural number not exceeding  $\dim E$ , and  $p$  is a parameter,  $1 \leq p \leq \infty$ . The quantity  $\alpha_{p,k}(\|\cdot\|)$  is defined as follows. We consider a  $k$ -dimensional subspace  $E'$  of  $E$ . On  $E'$  we can specify a  $\|\cdot\|_p$ -norm in many different ways (each of these ways corresponds to fixing a basis in  $E'$ ). Suppose such a norm has been fixed on. In relation to that norm the unit ball (relative to the norm  $\|\cdot\|$ ) of the space  $E'$  has a certain asphericity  $\alpha$ . The lower bound of these asphericities over all possible choice of the  $\|\cdot\|_p$ -norm on  $E'$  will be called the  $p$ -asphericity of  $E'$  and will be denoted by  $\alpha_p(E')$ . The lower bound of the quantities  $\alpha_p(E')$  over all the  $k$ -dimensional subspaces of  $E$  is called the  $k$ -dimensional  $p$ -asphericity  $\alpha_{p,k}(\|\cdot\|)$  of  $E$ .

The quantity  $\alpha_{p,k}(\|\cdot\|)$  is a characteristic of the space  $E$ ,  $\|\cdot\|$ . We now introduce an analogous characteristic of the convex subsets of  $E$ . We put

$$\alpha_{p,1\|\cdot\|,k}(G) = \alpha_{p,k}(\|\cdot\|)\alpha_{1\|\cdot\|}(G), \quad (1.1)$$

$$\alpha_{p,k}(G) = \inf \{ \alpha_{p,1\|\cdot\|,k}(G) \mid \|\cdot\|' \text{ is an admissible norm on } E \}. \quad (1.2)$$

We also put

$$\alpha(G) = \inf \{ \alpha_{1\|\cdot\|}(G) \mid \|\cdot\|' \text{ is an admissible norm on } E \}.$$

We point out that  $\alpha_{p,1\|\cdot\|,k}(G)$  is a metric characteristic of  $G$  (moreover, it does not change under homothetic transformations of  $G$ ), whereas  $\alpha_{p,k}(G)$  is an 'affine' characteristic of  $G$  (it does not change under continuous, affine transformations of  $G$ ). Some useful properties of these characteristics are formulated in the following exercises.

*Exercises.* Let  $E$ ,  $\|\cdot\|$ ,  $G$  be as above, and let  $p, p' \in [1, \infty]$ .

1. Prove that

$$\left. \begin{aligned} \alpha_{p,1\|\cdot\|,k}(G) &\leq k^{\frac{1}{p}-\frac{1}{p'}} \alpha_{p',1\|\cdot\|,k}(G), \\ \alpha_{p,k}(G) &\leq k^{\frac{1}{p}-\frac{1}{p'}} \alpha_{p',k}(G). \end{aligned} \right\} \quad (1.3)$$

2. Let  $\dim E = n < \infty$ . Prove that among the simplexes contained in  $G$  there is one of maximum volume. The vertices of this simplex lie on  $\partial G$ . Moreover, the hyperplanes passing through a vertex of the simplex parallel to the side opposite the vertex are support hyperplanes of  $G$ .

3. Deduce from Exercise 2 that, when  $\dim E = n < \infty$ ,

$$\alpha_{\infty,n}(G) \leq 2n, \quad (1.4)$$

4. Deduce from Exercises 1–3 that always

$$\left. \begin{aligned} (1) \quad \alpha_{\infty,1\|\cdot\|,k}(G) &\leq 2k\alpha_{1\|\cdot\|}(G), \\ (2) \quad \alpha_{p,1\|\cdot\|,k}(G) &\leq 2k^{1+\frac{1}{p}}\alpha_{1\|\cdot\|}(G), \quad 1 < p < \infty; \end{aligned} \right\} \quad (1.5)$$

if  $\dim E = n < \infty$ , then

$$(3) \quad \alpha_{p,\infty}(G) \leq n^{1+\frac{1}{p}}, \quad 1 \leq p \leq \infty.$$

In fact, when  $p = 2$ , more precise bounds exist [14]: for all  $k \leq \dim E$ ,

$$\alpha_{2,1\|\cdot\|,k}(G) \leq \sqrt{k}\alpha_{1\|\cdot\|}(G); \quad (1.6)$$

if  $\dim E = n < \infty$ , then

$$\alpha_{2,n}(G) \leq n. \quad (1.7)$$

Moreover, Dvoretzky [8] has proved that, for every integer  $k$  and number  $\delta > 0$ , there is an  $n = n(k, \delta)$  such that, if  $\dim E > n(k, \delta)$ , then

$$\alpha_{2,1\|\cdot\|,k}(G) \leq (1 + \delta)\alpha_{1\|\cdot\|}(G). \quad (1.8)$$

In particular, if  $\dim E = \infty$ , then

$$\alpha_{2,1\|\cdot\|,k}(G) = \alpha_{1\|\cdot\|}(G). \quad (1.9)$$

#### 4.1.2

We formulate the first of the main results about the complexity of classes of convex problems, but first we introduce some notation.

Let  $\Phi_{n,p}(v)$ ,  $1 \leq p < \infty$ , denote the upper bound for the laboriousness of the  $MD_p$ -method, constructed for an accuracy  $v$ , of solving Lipschitz-convex problems of a class of the type

$$C_{\text{Lip}}^0(G, L_p(T, \mu), \|\cdot\|_p, m),$$

corresponding to the case  $\dim L_p(T, \mu) = n$ ,  $n = 1, 2, \dots, \infty$ . Actually, if  $p \neq 1$ , the bound  $\Phi_{n,p}(v)$  does not depend on  $n$ .

Further, let  $\Phi_{n,\infty}(v)$  be the upper bound for the laboriousness of the MCG method, constructed for an accuracy  $v < 1$ , of solving problems of a class of the type  $C^0(\cdot, R^n, m)$ ,  $n = 1, 2, \dots, \infty$ . For  $n = \infty$ , the bound  $\Phi_{n,\infty}(v)$  is defined to be  $\infty$ .

Corresponding to the results of Sections 2.3 and 3.3, we have:

$$\lceil p \rceil = 1,$$

$$\Phi_{n,1}(v) = \begin{cases} \frac{D(1) \ln(n+1)}{v^2}, & 1 \leq n < \infty, \\ +\infty, & n = \infty; \end{cases}$$

$$1 < p \leq 2,$$

$$\Phi_{n,p}(v) = \frac{D(p)}{v^2};$$

$$2 < p < \infty,$$

$$\Phi_{n,p}(v) = \frac{D(p)}{v^p};$$

$$p = \infty,$$

$$\Phi_{n,\infty}(v) = D(\infty)n \ln \frac{2}{v}. \quad \square \quad (1.10)$$

Here the number  $D(p) < \infty$  depends only on  $p$ ,  $1 \leq p \leq \infty$ .

We also introduce a set of functions  $\varphi_{n,p}(v)$  defined for  $1 > v > 0$ , for all  $n = 1, 2, \dots, \infty$  and for all  $p, 1 \leq p \leq \infty$ :

$$\sqrt{p} = 1, 1 \leq n < \infty,$$

$$\varphi_{n,1}(v) = \begin{cases} 1, & v \geq \frac{1}{4}, \\ d(1) \frac{\ln(n+1)}{v^2}, & \frac{1}{4} > v \geq n^{-1/8}, \\ \frac{d(1)}{v^2}, \min\{\frac{1}{4}, n^{-1/8}\} > v \geq n^{-1/2}, \\ d(1)n, & v < \min\{n^{-1/2}, \frac{1}{4}\}; \end{cases}$$

$$p = 1, n = \infty,$$

$$\varphi_{n,1}(v) = \begin{cases} 1, & v \geq \frac{1}{4}, \\ +\infty, & v < \frac{1}{4}; \end{cases}$$

$$1 < p < \infty,$$

$$\varphi_{n,p}(v) = d(\infty) \min(n, \varphi_p(v)),$$

where

$$\varphi_p(v) = \begin{cases} 1, & v \geq \frac{1}{4}, \\ d(p)v^{-\max(2, p)}, & v < \frac{1}{4}; \end{cases}$$

$$p = \infty, 1 \leq n \leq \infty,$$

$$\varphi_{n,\infty}(v) = \begin{cases} 1, & v \geq \frac{1}{4}, \\ d(\infty)n \ln\left(\frac{1}{v}\right), & v < \frac{1}{4}; \end{cases}$$

$$p = \infty, n = \infty,$$

$$\varphi_{n,\infty}(v) = \begin{cases} 1, & v \geq \frac{1}{4}, \\ \infty, & v < \frac{1}{4}. \end{cases} \quad (1.11)$$

In (1.11) the numbers  $d(p)$  are positive and depend only on  $p, 1 \leq p \leq \infty$ . The set of these numbers for all  $p$  determines completely the family of functions  $\varphi_{n,p}(v)$ . For  $2 < p < \infty$ ,  $d(p) \geq 4^{-p}$ .

Let  $C$  be any class of problems of the type  $C^0(G, E, m)$ . Let  $N(v)$ , resp.  $N_*(v)$ , denote the deterministic complexity function resp.  $*$ -complexity function (see Section 1.4.4) of this class. Further, let  $\mathcal{O}$  be any local, deterministic oracle for the field of problems  $C(G, E, m)$ , and let  $N^\circ(v)$  resp.  $\tilde{N}^\circ(v)$  be the deterministic complexity resp. stochastic complexity of the class of problems  $C^\circ$  obtained from  $C$  by replacing the original oracle by the oracle  $\mathcal{O}$ .

The following theorem holds.

**Theorem on the bound for the complexity of classes of general convex problems.** There is a function  $d(p) > 0$ ,  $1 \leq p \leq \infty$ , such that for the

functions  $\varphi_{n,p}(v)$  corresponding to it (as in (1.11) the following assertions hold: the functions  $N(v)$ ,  $N_*(v)$ ,  $N^\circ(v)$ ,  $\tilde{N}^\circ(v)$  have the following bounds.

(Note: if  $\mathcal{O}$  is local but not necessarily deterministic, then the lower bounds for  $N^\circ(\cdot)$  in part A of the theorem are still valid, but we shall not prove this.)

#### A. Lower bounds.

A1. For every natural number  $k \leq n$  and for every  $p, 1 \leq p \leq \infty$ , we have

$$N^\circ(v) \geq \varphi_{k,p}(\alpha_{p,k}(G)v) \quad (1.12)$$

and

$$N^\circ(v) \geq \frac{\varphi_{k,p}(8\alpha_{p,k}(G)v)}{1 + [\ln] \varphi_{k,p}(8\alpha_{p,k}(G)v) [+]} \quad (1.13)$$

A2. Let  $E = L_p(T, \mu)$  and  $1 < p < \infty$ . There is a  $k_p(t) < \infty$  for  $t > 0$  such that, when  $n \geq k_p(\alpha_{1-p}(G)v)$ , we have

$$\tilde{N}^\circ(v) \geq \varphi_{n,p}(28\alpha_{1-p}(G)v). \quad (1.14)$$

A3. There is a  $\bar{k}_2(t) < \infty$  for  $t > 0$  such that, when  $n > \bar{k}_2(\alpha_{1-p}(G)v)$ , we have

$$\tilde{N}^\circ(v) \geq \varphi_{\infty,2}(\alpha_{1-p}(G)v). \quad (1.15)$$

A4. If  $E$  is not reflexive, then

$$N^\circ(v) = +\infty \text{ for } v \leq v(G), \text{ where } v(G) > 0. \quad (1.16)$$

A5. If  $m > 0$ , then

$$N_*(v) = +\infty \text{ for } v < 1. \quad (1.17)$$

#### B. Upper bounds.

B1. When  $n < \infty$

$$N(v) \leq \min \left\{ \Phi_{n,\infty}(v); \inf_{1 \leq p < \infty} \Phi_{n,p} \left( \frac{v}{4\alpha_{p,n}(G)} \right) \right\}. \quad (1.18)$$

B2. If  $E = L_p(T, \mu)$ ,  $1 < p < \infty$ :

$$N(v) \leq \Phi_{\infty,p} \left( \frac{v}{4\alpha_{1-p}(G)} \right). \quad (1.19)$$

The proofs of the group of assertions A of the theorem are proved in Sections 4.4–4.6. The assertions B we already know. For, the bound (1.18) is connected with the possibility of applying to problems in  $C$  the MCG method from Section 2.3 (this leads to the inequality  $N(v) \leq \Phi_{n,\infty}(v)$ ) and methods from the family of  $\overline{MD}_p$ -methods,  $1 \leq p < \infty$  (see Section 3.4.5); in that section the number  $\alpha_{p,n}(G)$  in (1.18) was denoted simply by  $\alpha_p(G)$ .

Table 1

$G$	$N(v) \leq$	$N(v) \geq$
—	$\Phi_\infty(v) \equiv c_\infty n \left\lceil \ln \left( \frac{1}{v} \right) \right\rceil$	$\lambda_\infty \Phi_\infty(v)$ for $v \leq n^{-2}$
$\alpha_{p,n}(G) \leq \alpha$ , $1 < p < \infty$ , $s = \max(2, p)$ . $\bar{\alpha} = \min\{2n, \alpha n^{1/p}\}$	$\Phi_\infty(v) \equiv c_\infty n \left\lceil \ln(1/n) \right\rceil$	$\lambda_\infty \Phi_\infty(v)$ for $\bar{\alpha}v < \frac{1}{4}$ and $v \leq \bar{\alpha}^{-2}$
	$\alpha^{2s} \Phi_p(\alpha v) \equiv c_p \frac{\alpha^s}{v^s}$	$\lambda_p \Phi_p(\alpha v)$ for $\alpha v \geq n^{-1/s}$
$\alpha_{1,n}(G) \leq \alpha$	$\Phi_\infty(v) \equiv c_\infty n \left\lceil \ln(1/v) \right\rceil$	$\lambda_\infty \Phi_\infty(v)$ for $v \leq n^{-2}$
	$\alpha^4 \Phi_1^{(n)}(\alpha v) \equiv c_1 \ln n \cdot \frac{\alpha^2}{v^2}$	$\lambda_1 \Phi_1^{(n)}(\alpha v)$ for $\frac{1}{4} > \alpha v \geq n^{-1/8}$
$\alpha_{\infty,n}(G) \leq \alpha$	$\Phi_\infty(v) \equiv c_\infty n \left\lceil \ln(1/v) \right\rceil$	$\frac{\lambda_1}{(\alpha v)^2}$ for $n^{-1/8} > \alpha v \geq n^{-1/2}$

$\check{N}^\phi(v) \geq$	Asymptotic behaviour of $N(v)$	
	as $v \rightarrow 0$	as $n \rightarrow \infty$
$\lambda_\infty \frac{\Phi_\infty(v)}{\ln \Phi_\infty(v)}$ for $v \leq n^{-2}$	$\frac{n \ln(1/v)}{\mu_\infty} \leq N(v)$ $\leq \mu_\infty n \ln(1/v)$ for $v \leq n^{-2}$	—
$\lambda_\infty \frac{\Phi_\infty(v)}{\ln \Phi_\infty(v)}$ for $\bar{\alpha}v \leq \frac{1}{32}$ and $v \leq \bar{\alpha}^{-2}$	$\frac{n \ln(1/v)}{\mu_\infty} \leq N(v)$ $\leq \mu_\infty n \ln(1/v)$	$\lambda_p \Phi_p(\alpha v) \leq N(v)$ $\leq \Phi_p(\alpha v) \alpha^{2s}$
$\lambda_p \frac{\Phi_p(\alpha v)}{1 + [\ln \Phi_p(\alpha v)]_+}$ for $\alpha v \geq n^{-1/s}$	for $\bar{\alpha}v < \frac{1}{4}$ and $v \leq \bar{\alpha}^{-2}$	for $\alpha v \geq n^{-1/s}$
$\lambda_p \Phi_p(\alpha v)$ for $n \geq k_p(\alpha v)$		
$\lambda_\infty \frac{\Phi_\infty(v)}{\ln \Phi_\infty(v)}$ for $v \leq n^{-2}$		
$\lambda_1 \frac{\Phi_1^{(n)}(\alpha v)}{1 + [\ln \Phi_1^{(n)}(\alpha v)]_+}$ for $\frac{1}{32} > \alpha v > n^{-1/8}$	$\frac{n \ln(1/v)}{\mu_\infty} \leq N(v)$ $\leq \mu_\infty n \ln(1/v)$	$\lambda_1 \Phi_1^{(n)}(\alpha v) \leq N(v)$ $\leq \alpha^4 \Phi_1^{(n)}(\alpha v)$
$\frac{\lambda_1}{(\alpha v)^2 \ln(1/\alpha v)}$ for $n^{-1/8} > \alpha v \geq n^{-1/2}$	for $v \leq n^{-2}$	for $\alpha v \geq n^{-1/8}$ and $\frac{1}{4} > \alpha v$
$\frac{\lambda_1}{(\alpha v)^2}$ for $n > \bar{k}_2(\alpha v)$		
$\lambda_\infty \frac{\Phi_\infty(v)}{\ln \Phi_\infty(v)}$ for $\alpha v < \frac{1}{32}$ and $v \leq \alpha^{-2}$	$\frac{n \ln(1/v)}{\mu_\infty} \leq N(v)$ $\leq \mu_\infty n \ln(1/v)$ for $\alpha v < \frac{1}{4}$ and $v \leq \alpha^{-2}$	$\frac{n \ln(1/v)}{\mu_\infty} \leq N(v)$ $\leq \mu_\infty n \ln(1/v)$ for $\alpha v > \frac{1}{4}$ and $v \leq \alpha^{-2}$

The assertions about the laboriousness of  $\overline{MD}_p$ -methods lead to the bounds

$$N(v) \leq \Phi_{n,p} \left( \frac{v}{4\alpha_{p,n}(G)} \right), \quad 1 \leq p < \infty.$$

Finally, (1.19) follows from the possibility of applying  $\overline{MD}_p$ , see Section 3.4.4.

#### 4.1.3

Theorem 4.1.2 contains quite a lot of information about the complexity of classes of convex problems, but the information is, of necessity, given in an inconvenient and highly condensed form. We have given the bounds (1.12)–(1.19) in a more convenient tabular form. Table I for the bounds applies to the only important case from the practical point of view, where  $E = R^n$  is finite-dimensional. The infinite-dimensional considerations, which are of a rather academic character, are given in Section 4.1.5.

In Table I the first column gives the geometrical characteristics of  $G$ , and the other columns give the bounds for the deterministic complexity  $N(v)$  and the stochastic complexity  $\tilde{N}(v)$  of a class  $C$  of the type  $C^0(G, R^n, m)$ . Only a lower bound is given for  $\tilde{N}(v)$ . The upper bounds for  $\tilde{N}(v)$  are the same as for  $N(v)$ , because, clearly,  $\tilde{N}(v) \leq N(v)$ . The lower bounds for  $n(v)$  and  $\tilde{N}(v)$  are also valid for all the complexities  $N^\circ(v)$ ,  $\tilde{N}^\circ(v)$ . In the table  $c(p)$ ,  $\lambda(p)$  and  $\mu(p)$  denote certain functions of  $p$  which are positive and finite for all  $p$ . The table itself needs no special justification; it is an immediate consequence of the bounds (1.12)–(1.19). Finally, it is assumed that  $0 < v < 1$  everywhere in the table.

#### 4.1.4

We comment on the results given, by drawing some obvious consequences from the data in Table I.

The complexity functions of a class  $C$  of the type  $C^0(G, R^n, m)$  depend on the accuracy  $v$  and on the parameters, i.e. on the list of objects which characterize the class  $C$ , i.e. primarily on  $G$ , but also on the oracle  $\emptyset$  corresponding to the class  $C$ , and on  $m$ .

We do not know how to calculate the complexity exactly (this obviously is a hopeless problem for arbitrary general situations); we can only estimate it. Our estimates coincide with the complexity in the best case with an accuracy up to an absolute multiplicative constant. It should be remarked that to strive for more here is not worthwhile; it must be remembered that the informational complexity itself is a rather rough characteristic of the intuitively understood complexity of a computational method.

Corresponding to what has been said, the natural questions about the complexity which we should answer (and which we can answer in a wide range of situations) are questions about the nature of the dependence of the complexity on the parameters which define it, i.e. about the asymptotic behaviour of the complexity with respect to these parameters. Let us see what can be said in this connection.

#### 4.1.4.1

We consider first the question of the asymptotic behaviour of the complexity as  $v \rightarrow 0$  (it is precisely to this sort of analysis that study of the effectiveness of numerical methods is traditionally restricted). It is well-known that, for small  $v$ ,  $N(v) \geq O(\ln(1/v))$  (this estimate can be obtained even from an analysis of the one-dimensional case). The results of Theorem 4.1.2 make this fact more precise and give exhaustive information about the asymptotic behaviour with respect to  $v$  of the complexity: namely, for all sufficiently small  $v$  (viz. for  $v < v(G) > 0$ ) we have that  $N(v) \sim n \ln(1/v)$  up to an absolute multiplicative constant. (The notation  $a \sim b$  means that for certain absolute positive constants  $c$  and  $d$  we have  $d \leq b/a \leq c$ .)

The moment when the asymptotic behaviour is established (i.e. the value of  $v(G)$ ) depends on the affine properties of the body  $G$ . For parallelepipeds  $G$ , the asymptotic behaviour is established 'immediately':  $v(G) = \frac{1}{4}$ . In the general case,  $v(G)$  depends on the 'acubicity' of  $G$ —on the value of  $\alpha_{\infty,n}(G) \equiv \alpha$ . Namely,  $v(G) = \min\{1/(4\alpha), 1/\alpha^2\}$ . Moreover, for all bodies  $G$  of a given dimension, there is a common moment of establishment of the asymptotic behaviour  $\tilde{v}(n) = n^{-2}$  (this happens because of the inequality  $\alpha_{\infty,n}(G) \leq 2n$ , see (1.4)).

The results regarding the stochastic complexity  $\tilde{N}(v)$  are not so precise. The asymptotic behaviour as  $v \rightarrow +0$  of the upper bound differs but little from that of the lower bound, it is true; when  $v \rightarrow +0$ , the estimates

$$\frac{n \ln(1/v)}{\ln(n \ln(1/v))} \lesssim \tilde{N}(v) \lesssim n \ln \frac{1}{v}$$

hold. (Here  $a \lesssim b$  means that  $b/a$  is not less than an absolute positive constant.) In particular, asymptotically as  $v \rightarrow +0$ ,

$$\tilde{N}(v) \gtrsim \frac{N(v)}{\ln N(v)},$$

i.e. a randomization of search at high accuracies, if it enables the effectiveness of methods to be increased at all, does so only 'logarithmically'. We remark further that, independently of the shape of the convex body  $G \subset R^n$ , the moments of establishment of all the asymptotic formulae mentioned are automatically bounded below by a quantity of the order  $n^{-2}$ .

## 4.1.4.2

We now consider how the complexity depends on the other parameters. As regards the actual choice of the  $(\text{int } G, 0)$ -oracle and the number of constraints  $m$ , the bounds in the theorem do not react to these at all (this is quite natural, if one thinks about it). So we just have to explain the rôle of  $G$ . Strictly speaking,  $G$  is an extremely inconvenient parameter (because non-numerical), and so it is very difficult to trace out the part it plays from its general aspect. The sensible thing to do is to pick out certain numerical characteristics of  $G$  (they must be affinely invariant because of the affine nature of the classes considered), and to see how the complexity depends on these characteristics.

The most important (and the simplest) of such characteristics is the dimension  $n$  of the body  $G$ . Both theoretically and practically it is very useful to know how the complexity behaves asymptotically as  $n \rightarrow \infty$ . Unfortunately, it turns out that a simple answer cannot be given to this question (because the complexity depends on  $G$  as a 'whole', and not just on its dimension).

For, if  $G$  is a parallelepiped of dimension  $n$  (i.e.  $\alpha_{\infty,n}(G) = 1$ ), and if  $v < \frac{1}{4}$ , then, as  $n \rightarrow \infty$ ,

$$N(v) \sim n \ln \frac{1}{v}, \quad \tilde{N}(v) \gtrsim \frac{n}{\ln n} \ln \frac{1}{v},$$

i.e. the complexity increases with increase in dimension linearly, or almost linearly. On the other hand, if  $G$  is an  $l_p^{(n)}$ -ball, i.e.  $\alpha_{p,n}(G) = 1$ , and if  $1 < p < \infty$ , and  $v < 1/4$ , then, as  $n \rightarrow \infty$ ,

$$N(v) \underset{p}{\sim} \tilde{N}(v) \underset{p}{\sim} v^{-\max(2,p)}$$

(here  $a \sim_x b$  means that  $b/a$  is bounded above and below by a quantity which depends only on  $x$ ;  $\gtrsim_x$  is to be understood similarly.) Thus in the second case the complexity is bounded uniformly with respect to  $n$  (by a quantity which depends only on  $v$  and  $p$ ). But if  $G$  is an  $l_1^{(n)}$ -ball and if  $v < \frac{1}{4}$ , then, as  $n \rightarrow \infty$ ,

$$N(v) \sim \tilde{N}(v) \sim \frac{\ln n}{v^2};$$

here we observe a growth in complexity with growth of  $n$  which is not linear but logarithmic. We point out that, in the last two cases, the value of  $n$  from which the indicated asymptotic behaviour of the complexity operates depends on  $v$  (and  $p$ ). Moreover, for the stochastic complexity the asymptotic behaviour is established, generally speaking, later than for the deterministic complexity.

In view of what has been said, it is impossible without additional hypotheses about  $G$  to answer the question of how the complexity changes with increase of the dimension. It is still necessary to fix in the requisite way the 'affine type' of  $G$ . Let us consider this problem, supposing that  $G$  is of the type of an  $l_p^{(n)}$ -ball (more precisely, we suppose that  $\alpha_{p,n}(G) \leq \alpha$  and we examine the behaviour of the complexity as  $G$  changes, taking  $\alpha$  as a parameter). The lower and upper

bounds which are obtained in this case for the complexity depend on  $G$  only through the dimension of  $G$ . When  $\alpha v < 1/32$  and  $n \rightarrow \infty$  we have

$$\begin{aligned} \frac{n \ln (1/v)}{\ln(n \ln(1/v))} &\lesssim \tilde{N}(v) \leq N(v) \underset{p}{\sim} n \ln \frac{1}{v}, \quad p = \infty, \\ N(v) \underset{\alpha,p}{\sim} \tilde{N}(v) \underset{\alpha,p}{\sim} &\begin{cases} 1/v^{\max(2,p)}, & 1 < p < \infty, \\ \ln n/v^2, & p = 1, \end{cases} \end{aligned} \quad (1.20)$$

i.e. the asymptotic behaviour of the complexities as  $n \rightarrow \infty$  is the same as for  $l_p^{(n)}$ -balls (of course, the divergence between the upper and lower bounds of the complexity now also depends on  $\alpha$ ; it is the greater, the greater  $\alpha$  is).

## 4.1.4.3

It is clear from the above considerations that the complexity depends rather strongly on a factor which is so ill-suited for examination, i.e. on the 'affine type of  $G$ ', and therefore in the general case it is difficult to rely on the possibility of obtaining constructive and exhaustive results. However, for a certain rather wide spectrum of concrete bodies  $G$ , namely, for  $l_p^{(n)}$ -balls, it is possible, using the bounds in the table, to obtain an 'almost complete' idea of the complexity. We give the corresponding results.

*The case  $p = \infty$ .* Here the picture is the most complete: for all  $v$  ( $v < 1/32$ ) the relations

$$N(v) \sim n \ln \frac{1}{v} \quad \text{and} \quad \tilde{N}(v) \gtrsim \frac{N(v)}{\ln N(v)}$$

hold.

*The case  $1 < p < \infty$ .* Here (as in the case  $p = 1$ ) the results are rather less precise. They consist of the following (we suppose  $v < 1$ ):

(i) For high accuracies, viz. when  $v \leq n^{-2/p}$  the following relation holds:

$$N(v) \underset{p}{\sim} n \ln \frac{1}{v}.$$

(ii) For low accuracies, viz. when  $v \geq n^{-1/\max(p,2)}$ , the relation

$$N(v) \underset{p}{\sim} \frac{1}{v^{\max(2,p)}}$$

holds.

(iii) For intermediate accuracies, viz. when  $n^{-1/\max(2,p)} > v > n^{-2/p}$ , the two-sided estimates

$$n \underset{p}{\lesssim} N(v) \underset{p}{\lesssim} n \ln \frac{1}{v}$$

holds.

Thus, with increase of the accuracy, the complexity at first increases like  $1/v^{\max(2, p)}$  and it ‘does not feel’ the value of  $n$ . This growth continues until the complexity reaches the level of  $n$ . Here the dimension first begins ‘to be felt’. After a certain band of indeterminacy of values of  $v$  (an indeterminacy in our knowledge, of course) in which the complexity is included between the levels  $\sim_p n$  and  $\sim_p n \ln n$ , the complexity begins to vary like  $n \ln 1/v$  and it ‘forgets’ about the geometry of  $G$  (i.e., strictly, it forgets about  $p$ ).

We remark that, as regards the stochastic complexity, for all  $v < 1$  the estimate

$$\frac{N(v)}{\ln N(v) + 1} \stackrel{p}{\lesssim} \tilde{N}(v) \leq N(v)$$

holds, i.e. all that we have said about  $N(v)$  is true, with certain corrections, for  $\tilde{N}(v)$  also. We further notice that we have an estimate for  $N(v)$  which is accurate up to a factor  $\sim_p \ln N(v)$ .

*The case  $p = 1$ .* Here the situation is qualitatively the same as for  $p > 1$ . Namely ( $v < 1/32$ ):

(i) for high accuracies ( $v \leq n^{-2}$ ) we have

$$N(v) \sim n \ln \frac{1}{v};$$

(ii) for very low accuracies ( $v \geq n^{-1/2}$ )

$$N(v) \sim \frac{\ln(n+1)}{v^2};$$

(iii) for intermediate accuracies ( $n^{-1/8} > v > n^{-2}$ ) matters stand thus: at first (when  $v > n^{-1/2}$ ) the estimates

$$\frac{1}{v^2} \lesssim N(v) \lesssim \frac{\ln(n+1)}{v^2}$$

hold, and then (when  $n^{-2} \leq v < n^{-1/2}$ ) the following estimates hold:

$$n \lesssim N(v) \lesssim n \ln n.$$

Thus, the difference from the case  $1 < p < \infty$  is that the dimension ‘is felt’ from the very start, at low accuracies, but only weakly. ‘The band of indeterminacy’ is relatively wider than for  $p > 1$ , but in it (and then also everywhere) we know  $N(v)$  with accuracy up to a factor  $\sim \ln N(v)$ . Finally, for all  $v < 1/32$ , the estimate (1.21) holds.

#### 4.1.5

Let us examine the infinite-dimensional situation. Here the complexity of classes of general convex problems can be either finite for all  $v > 0$ , or infinite

for all sufficiently small  $v > 0$ . In every case, if  $G$  is a convex, bounded, closed body in  $E$  and if  $\dim E = \infty$ , then by A3 of Theorem 4.1.2:

$$N(v), \tilde{N}(v) \gtrsim \left( \frac{1}{\alpha(G)v} \right)^2,$$

i.e. the complexity on any infinite-dimensional space admits the same lower bound as in the Hilbert-space case. This lower bound *a fortiori* points to a regular—regular, that is, with regard to the nature of the dependence on  $v$ —asymptotic behaviour of the complexity as  $v \rightarrow +0$ , if  $E$  is regular and if  $V$  is a regular function corresponding to  $E$  such that  $\omega_V(t) \leq O(t)$ , as occurs in the Hilbert-space case.

Further, if  $E$  is regular, then  $N(v)$  is finite for all  $v > 0$ . Finally, in the case where  $E = L_p(T, \mu)$ ,  $1 < p < \infty$ , and  $\dim E = \infty$ , we can obtain more complete results. Namely, by (1.12), (1.14), and (1.19), we have

$$N(v) \underset{\alpha_{I \cdot I_p(G), p}}{\sim} \tilde{N}(v) \underset{\alpha_{\| \cdot \|_p(G), p}}{\sim} \frac{1}{v^{\max(2, p)}}.$$

## 4.2 COMPLEXITY OF CLASSES OF LIPSCHITZ-CONVEX PROBLEMS

In this section we carry out the same examination regarding classes of Lipschitz-convex problems as we did in the previous section for classes of general convex problems. The results in the main are close to those of Section 4.1. As before,  $E$  is a real Banach space,  $G$  is a bounded, convex, closed body in  $E$ , and  $\| \cdot \|$  is a norm on  $E$ . Further, let  $0 < v < 1$ .

### 4.2.1

We begin by formulating the main results about the complexity of the classes of problems in question.

Let  $C_{\text{Lip}}$  be any class of problems of the type  $C_{\text{Lip}}^0(G, E, \| \cdot \|, m)$ . Let  $N(v)$  resp.  $N_*(v)$  denote the deterministic complexity function resp.  $*$ -complexity function (see Section 1.4.4) of this class. Further, let  $\mathcal{O}$  be an arbitrary, local, deterministic oracle for the field of problems  $C_{\text{Lip}}(G, E, \| \cdot \|, m)$ , and let  $N^\circ(v)$ ,  $\tilde{N}^\circ(v)$  be the deterministic resp. stochastic complexity of the class of problems  $C^\circ$  obtained from  $C_{\text{Lip}}$  by replacing the original oracle by the oracle  $\mathcal{O}$ .

**Theorem on a bound for the complexity of classes of Lipschitz-convex problems.** There is a function  $d(p) > 0$ ,  $1 \leq p \leq \infty$ , such that the following assertions hold for the functions  $\varphi_{n,p}(v)$  associated by (1.11) with  $d(p)$ . The functions  $N(v)$ ,  $N_*(v)$ ,  $N^\circ(v)$ ,  $\tilde{N}^\circ(v)$  have the following bounds.

**A. Lower bounds.**

A.1. For every natural number  $k \leq n$  and every  $p$  with  $1 \leq p \leq \infty$ , we have

$$N^e(v) \geq \varphi_{k,p}(\alpha_{p,1\cdot 1\cdot k}(G)v) \quad (2.1)$$

and

$$\tilde{N}^e(v) \geq \frac{\varphi_{k,p}(8\alpha_{p,1\cdot 1\cdot k}(G)v)}{1 + [\ln] \varphi_{k,p}(8\alpha_{p,1\cdot 1\cdot k}(G)v)[ ]_+}. \quad (2.2)$$

A.2. Let  $E = L_p(T, \mu)$ ,  $\|\cdot\| = \|\cdot\|_p$ ,  $1 < p < \infty$ . There is a  $k_p(t) < \infty$ ,  $t > 0$ , such that, when  $n \geq k_p(\alpha_{\|\cdot\|_p}(G)v)$ , we have

$$\tilde{N}^e(v) \geq \varphi_{n,p}(28\alpha_{1\cdot 1\cdot p}(G)v). \quad (2.3)$$

A.3. There is a  $\bar{k}_2(t) < \infty$  for  $t > 0$  such that, for  $n > \bar{k}_2(\alpha_{1\cdot 1}(G)v)$ , we have

$$\tilde{N}^e(v) \geq \varphi_{\infty,2}(\alpha_{1\cdot 1}(G)v). \quad (2.4)$$

A.4. If  $E$  is not reflexive, then

$$N^e(v) = +\infty \text{ when } v < v(G), \text{ where } v(G) > 0. \quad (2.5)$$

A.5. If  $m > 0$ , then

$$N_*(v) = +\infty \text{ when } v < 1 \quad (2.6)$$

**B. Upper bounds.**

B1. When  $n < \infty$

$$N(v) \leq \left\{ \Phi_{n,\infty}(v), \inf_{1 \leq p < \infty} \Phi_{n,p}\left(\frac{v}{\alpha_{p,n}(\|\cdot\|)}\right) \right\}. \quad (2.7)$$

B2. When  $E = L_p(T, \mu)$ ,  $1 < p < \infty$ ,

$$N(v) \leq \Phi_{\infty,p}(v). \quad (2.8)$$

The upper bounds are valid whether or not  $G$  is a body.

(Note: If  $\mathcal{C}$  is local but not necessarily deterministic, then the lower bounds for  $N^e(\cdot)$  in part A of the theorem are still valid, but we shall not prove this.)

We point out the differences between the results of this theorem and those in Theorem 4.1.2. In the lower bounds the only difference is that all the affine invariants of  $G$ —the numbers  $\alpha_{p,k}(G)$  in Section 4.1.2—are replaced in Section 4.2.1 by the metric invariants  $\alpha_{p,1\cdot 1\cdot k}(G)$ . In the upper bounds, the characteristics which depend on  $G$ ,  $4\alpha_{p,n}(G)$  and  $4\alpha_{1\cdot 1}(G)$ , are replaced respectively by the characteristic of the norm  $\|\cdot\|$ , i.e.  $\alpha_{p,n}(\|\cdot\|)$ , and by 1. Both the lower and the upper bounds have now become less. Of course, this is as it should be, because the class of problems  $C_{Lip}$  is narrower than  $C$ . The proof of

the upper bounds  $B$  is carried out in the same way as in Section 4.1.2 except that the results about the laboriousness of the MCG and MD<sub>p</sub> methods must be quoted not for the class of general convex problems, but for the class of Lipschitz convex problems, viz. those in Sections 2.4.1.2. and 3.3. The lower bounds A are proved in Sections 4.4–4.6.

**4.2.2**

As in Section 4.1, we present an explicit summary in tabular form of the error bounds given in the theorem. In Table 2 it is assumed that  $n < \infty$ . Compared with Table 1, the new table has an extra column showing the hypothesis regarding the structure of the norm  $\|\cdot\|$  on  $E$  which is considered, that is, the norm in terms of which the class in which we are interested is defined. The rest of the table and the method of justifying it are the same as in Section 4.1.

**4.2.3**

The commentary on Table 2 is, in the main, the same as that given in Section 4.1. We shall run through these comments briefly, dwelling mainly on the special features for the Lipschitz case as compared with the general case.

**4.2.3.1 Asymptotic behaviour as  $v \rightarrow +0$ .**

As in Section 4.1, when  $v < v_{1\cdot 1}(G) > 0$  we have  $N(v) \sim N \ln 1/v$  and  $\tilde{N}(v) \gtrsim N(v)/\ln N(v)$ . The special feature of the Lipschitz case is that  $v_{1\cdot 1}(G)$  depends on both the affine properties of  $G$  and on the metrical properties of  $G$  relative to the norm  $\|\cdot\|$ . In particular,  $v_{1\cdot 1}(G)$  admits no positive minorant depending only on  $n$ , as the quantity  $v(G)$  did in Section 4.1.4.1. In fact,  $v_{1\cdot 1}(G)$  is the less, generally speaking, the bigger  $\alpha_{1\cdot 1}(G)$  is (we can take  $v_{1\cdot 1}(G) = \min\{1/32; (\alpha_{1\cdot 1}(G)n)^{-2}\}$ ). This, too, is natural. Suppose, for example, that  $G$  is a  $\delta$ -neighbourhood (in the metric defined by  $\|\cdot\|$ ) of a segment of unit length. For small  $\delta$  a solution of problems of a class of the type  $C_{Lip}^0(G, E, \|\cdot\|, m)$  with an error  $v \gg \delta$  can be reduced to solving one-dimensional problems. Moreover, the complexity will behave, for such  $v$ , like  $\ln 1/v$ . The smaller  $\delta$  is, the wider will be the range of values of  $v$  in which this ‘pathological’ bound for the complexity of problems of high dimension will operate.

It is clear why such an effect does not arise when general convex problems are considered. The functionals of such problems can vary arbitrarily rapidly (the only restriction is their variation on  $G$ ). For this reason there are no grounds in the previous example for reducing the problems to one-dimensional problems, however small  $\delta$  might be. The same thing can be said

Table 2

$\ \cdot\ $	$G$	$N(v) \leq$	$N(v) \geq$
—	—	$\Phi_\infty(v) \equiv c_\infty n \ln(1/v)$	—
$\alpha_{p,n}(\ \cdot\ ) \leq \alpha$ , $1 < p < \infty$ , $s = \max(2, p)$ , $\bar{\alpha} = \min\{2n, \alpha n^{1/p}\}$	$\alpha_{\ \cdot\ }(G) \leq \beta$	$\Phi_\infty(v) \equiv c_\infty n \ln(1/v)$ for $\beta\bar{\alpha}v < \frac{1}{32}$ and $v \leq (\beta\bar{\alpha})^{-2}$	$\lambda_\infty \Phi_\infty(v)$ for $\beta\bar{\alpha}v < \frac{1}{4}$ and $v \leq (\beta\bar{\alpha})^{-2}$
$\alpha_{1,n}(\ \cdot\ ) \leq \alpha$	$\alpha_{\ \cdot\ }(G) \leq \beta$	$\alpha^s \Phi_p(v) \equiv c_p \alpha^s / v^s$	$\lambda_p \Phi_p(v) (\beta\alpha)^{-s}$ for $\alpha\beta v \geq n^{-1/s}$
$\alpha_{\infty,n}(\ \cdot\ ) \leq \alpha$	$\alpha_{\ \cdot\ }(G) \leq \beta$	$\Phi_\infty(v) \equiv c_\infty n \ln(1/v)$ for $v \leq (n\beta)^{-2}$	$\lambda_\infty(v)$ for $v \leq (n\beta)^{-2}$
		$\alpha^2 \Phi_1^{(n)}(v) \equiv c_1 \ln n \alpha^2 / v^2$	$\lambda_1 \Phi_1^{(n)}(\alpha\beta v)$ for $\alpha\beta v < \frac{1}{4}$ and $\alpha\beta v \geq n^{-1/2}$
			$\lambda_1 / (\alpha\beta v)^2$ for $\alpha\beta v < \frac{1}{4}$ and $\alpha\beta v \geq n^{-1/2}$
		$\Phi_\infty(v) \equiv c_\infty n \ln(1/v)$	$\lambda_\infty \Phi_\infty(v)$ for $\alpha\beta v < \frac{1}{4}$ and $v \leq (\beta\alpha)^{-2}$

in another way: the class of general convex problems is defined in affine terms, and therefore it is meaningless to study how its complexity is affected by the ‘narrowness’ of  $G$  in some directions—this ‘narrowness’ is not an affine invariant of  $G$ . In the affine sense, a body  $G$  cannot be ‘too narrow’ (say,  $\alpha_{\infty,n}(G) \leq 2n$ ), which is what determines the existence of ‘absolutely high accuracies’ for classes of general problems in a given dimension—accuracies, that is, starting from which the complexity of these classes behave like  $n \ln 1/v$ .

$\tilde{N}^\epsilon(v) \geq$	Asymptotic behaviour of $N(v)$	
	as $v \rightarrow 0$	as $n \rightarrow \infty$
—	—	—
$\frac{\Phi_\infty(v)}{\ln \Phi_\infty(v)}$ for $\beta\bar{\alpha}v < \frac{1}{32}$ and $v \leq (\beta\bar{\alpha})^{-2}$	$(n/\mu_\infty) \ln(1/v)$ $\leq N(v)$ $\leq \mu_\infty n \ln(1/v)$	$\lambda_p(\beta\alpha)^{-s} \Phi_p(v)$ $\leq N(v)$ $\leq \alpha^s \Phi_p(v)$
$\lambda_p(\beta\alpha)^{-s} \frac{\Phi_p(v)}{[\ln \Phi_p(\beta\alpha v)]_+ + 1}$ for $\beta\bar{\alpha}v \geq n^{-1/s}$ and $v \leq (\beta\alpha)^{-2}$		for $\alpha\beta v \geq n^{-1/s}$
$\lambda^s(\beta\alpha)^{-s} \Phi_p(v)$ for $n \geq k_p(\beta\alpha v)$		
$\lambda_\infty \Phi_\infty(v) / \ln \Phi_\infty(v)$ for $v \leq (n\beta)^{-2}$	$(n/\mu_\infty) \ln(1/v)$ $\leq N(v)$ $\leq \mu_\infty n \ln 1/v$	$\lambda_1 \ln n / (\alpha\beta v)^2$ $\leq N(v)$ $\leq c_1 \ln n \cdot \alpha^2 / v^2$
$\lambda_1 \Phi_1^{(n)}(\alpha\beta v)$ for $\alpha\beta v < \frac{1}{32}$ and $\alpha\beta v \geq n^{-1/2}$	for $v \leq (n\beta)^{-2}$	for $\alpha\beta v < \frac{1}{4}$ and $\alpha\beta v \geq n^{-1/2}$
$\lambda_1 / (\alpha\beta v)^2 \ln 1/\alpha\beta v$ for $\alpha\beta v < \frac{1}{4}$ and $\alpha\beta v \geq n^{-1/2}$		
$\lambda_1 / (\alpha\beta v)^2$ for $n > \bar{k}_2(\alpha\beta v)$		
$\lambda_\infty \Phi_\infty(v) / \ln \Phi_\infty(v)$ for $\alpha\beta v < \frac{1}{32}$ and $v \leq (\beta\alpha v)^{-2}$	$(n/\mu_\infty) \ln(1/v)$ $\leq N(v)$ $\leq \mu_\infty n \ln(1/v)$ for $\alpha\beta v < \frac{1}{4}$ and $v \leq (\beta\alpha)^{-2}$	$(n/\mu_\infty) \ln(1/v)$ $\leq N(v)$ $\leq \mu_\infty n \ln(1/v)$ for $\alpha\beta v < \frac{1}{4}$ and $v \leq (\beta\alpha)^{-2}$

## 4.2.3.2

As regards the dependence of the complexity on the dimension of  $G$ , the difference from the results in Section 4.1.4.2 is that now it is necessary to fix not  $\alpha_{p,n}(G)$  but  $\alpha_{p,\|\cdot\|,n}(G)$ , i.e. to examine the behaviour of  $N(v)$  and  $\tilde{N}(v)$  as functions of  $n$  under the condition  $\alpha_{p,\|\cdot\|,n}(G) \leq \alpha$ . Under such a condition the relation (1.20) is preserved in the Lipschitz case as well.

## 4.2.3.3

As regards the bounds for the complexity on  $\mathbb{I}_p^{(n)}$ -balls, if we assume that  $G$  is an  $\mathbb{I}_p^{(n)}$ -ball and that  $\|\cdot\|$  is precisely  $\|\cdot\|_p$ , then everything that was said in Section 4.1.4.3 can be repeated word for word.

## 4.2.4

Regarding the infinite-dimensional situation, everything that was said in Section 4.1.5 remains true for the Lipschitz case, provided that, instead of  $\alpha(G)$  which featured there we must put  $\alpha_{1,p}(G)$ .

### 4.3 RECOMMENDATIONS ABOUT THE USE OF METHODS OF SOLVING GENERAL CONVEX OR LIPSCHITZ-CONVEX PROBLEMS

## 4.3.1

In this section the facts about complexity, contained in the two previous sections, are transformed into recommendations about the use of MCG and  $\overline{MD}_p$ . The basis for these recommendations is the coincidence (with an accuracy up to details of little theoretical importance) of the bounds for the laboriousness of these methods with the bounds for the complexity of the classes of problems under consideration.

The facts about the use of  $\overline{MD}_p$  are given under the supposition that the norm  $\|\cdot\|$ , with which the method is associated is optimal for the  $G$  under consideration in the sense of Section 3.4.5, i.e. that for this norm  $\alpha_{1,p}(G) = \alpha_{p,n}(G)$ . The recommendations about the use of the methods are given in tabular form. The first column fixes the hypotheses about the properties of  $G$ . The second column indicates the recommended method, and the third the conditions under which it is recommended. The fourth column gives the upper bound  $M(v)$  for the laboriousness of the recommended method, and the fifth the potential boundaries for reducing the laboriousness in the class of deterministic and stochastic methods (i.e. the upper bounds of the ratios  $M(v)/N(v)$  and  $M(v)/\tilde{N}(v)$ , where  $N(v)$  and  $\tilde{N}(v)$  are the respective complexity functions of the class considered). We point out that the recommendation for the use of MCG has a purely theoretical character, since it is not clear how to realize this method for cases of arbitrary and considerable dimensionality. However, use of MMCG instead of MCG leads to obvious changes in the table.

## 4.3.2

We start with the class of general convex problems. We consider a class of problems  $C$  of the type  $C^0(G, R^n, m)$ . Of course, theoretically we should also

consider the infinite-dimensional case, but here it is sufficient to refer the reader to Section 4.1.5, where essentially we have already formulated recommendations about the use of MD-methods and assertions (for  $E = L_p(T, \mu)$ ) about their sub-optimality.

The recommendations about the use of the methods of solving general convex problems are set out in Table 3.

## 4.3.3

We make some comments about the data in Table 3. First of all, if the specified accuracy is sufficiently high, ( $v \leq v(G)$ ,  $v(G) \geq n^{-2}$ ), then theoretically MCG is sub-optimal. Its laboriousness ( $\sim n \ln(1/v)$ ) in the class of deterministic methods cannot, in principle, be reduced more than  $\sim 1$  times, and in the class of stochastic methods by not more than  $\sim \ln(n \ln(1/v))$  times.

But if the specified accuracy  $v$  is not too high, then we are unable to give a simple answer to the question of a theoretically sub-optimal method of solving general convex problems. We know only that, if  $G$  is of the type of an  $\mathbb{I}_p^{(n)}$ -ball ( $\alpha_{p,n}(G) \equiv \alpha$  not too great), then, in a known range of variation of  $v$  (this range is  $\alpha v \geq n^{-1/s}$ ,  $s = \max(2, p)$ , if  $1 < p < \infty$ ), it is reasonable to use  $\overline{MD}_p$ .

[Note: In this discussion we are restricting our attention to the case  $1 < p < \infty$ . The case  $p = \infty$  is simple; here all the accuracies are high:  $v(G) = \min(1/32, \alpha^{-2})$ . The case  $p = 1$  is constructed qualitatively the same as the case  $1 < p < \infty$ .]

Moreover, the laboriousness of this method ( $\sim_p \alpha^s/v$ ) cannot be reduced more than by  $\sim_p \alpha^{2s}$  times in the class of deterministic methods, and by  $\sim_p \alpha^{2s} \ln \alpha^s/v^s$  times in the class of stochastic methods. In the range of intermediate accuracies (for  $1 < p < \infty$  this range is  $\{v | \alpha v < n^{-1/s}, v > v(G)\}$ ) we have recommended use of MCG, although the theoretical justification for this is weaker than for high accuracies; the laboriousness of MCG here is not reduced more than by  $\sim_p \ln(n \ln(1/v))$  times in the class of deterministic methods and by  $\sim_p \ln^2(n \ln(1/v))$  times in the class of stochastic methods.

We remark that, with large  $\alpha_{p,n}(G)$  the grounds for what has been said in favour of  $\overline{MD}_p$  are not very trustworthy, because the boundaries for the potentially possible improvement of the method depend in this case on the value of  $\alpha_{p,n}(G)$ . This is quite natural—domains  $G$  which are badly approximated by  $\mathbb{I}_p^{(n)}$ -balls for all  $p$  are ‘not picked up’ by the scale of  $L_p$ -methods and bounds. But in our own justification we observe that it is clear, from the results given, how strongly the complexity of the class considered depends on the affine properties of  $G$ . Moreover, it is difficult to take into account both constructively and sufficiently completely such a nebulous concept as the ‘affine properties of  $G$ ’.

We further remark that the recommendations given regarding the use of  $\overline{MD}_p$ , even on  $\mathbb{I}_p^{(n)}$ -balls, are ‘proven with accuracy up to a factor  $\lambda(p)$ ’ which,

## Complexity of classes of general convex problems

Table 3

Conditions on $G$	Recommended method	Conditions under which the method is recommended	Upper bound for the laboriousness of the method $M(v)$	Potential boundary for reduction of the laboriousness in the class of deterministic methods, $\leq$	Potential boundary for reduction of the laboriousness in the class of stochastic methods, $\leq$
$\alpha_{p,n}(G) \leq \alpha$ , $1 < p < \infty$ , $s = \max\{2, p\}$ , $\bar{\alpha} = \min\{2n, \alpha n^{1/p}\}$	$MCG$	$v \leq n^{-2}$	$3n] \ln(1/v)[$	$\lambda_\infty$ for $v \leq n^{-2}$	$\lambda_\infty \ln M(v)$ for $v \leq n^{-2}$
	$\overline{MD}_p$	$\alpha v \geq n^{-1/s}$	$c_p^s \frac{\alpha^s}{v^p}$	$\lambda_p \alpha^{2s}$ for $\alpha v \geq n^{-1/s}$	$\lambda_p \alpha^{2s} [\ln M(v)]$ for $\alpha v \geq n^{-1/s}$
	$MCG$	$\alpha v < n^{-1/s}$	$3n] \ln(1/v)[$	$\lambda_\infty$ for $\bar{\alpha}v < \frac{1}{32}$ and $v \leq \bar{\alpha}^{-2}$	$\lambda_\infty \ln M(v)$ for $\bar{\alpha}v < \frac{1}{32}$ and $v < \bar{\alpha}^{-2}$
$\alpha_{1,n}(G) \leq \alpha$	$\overline{MD}_{1,n}$	$\alpha v \geq n^{-1/2}$	$\frac{c_1 \alpha^2 \ln(n+1)}{v^2}$	$\lambda_p \ln M(v)$ for $\alpha v < n^{-1/s}$ $\lambda_1 \alpha^4$ for $\frac{1}{4} > \alpha v \geq n^{-1/2}$	$\lambda_p \ln^2 M(v)$ for $\alpha v < n^{-1/s}$ $\lambda_1 \alpha^4 \ln^2 M(v)$ for $\alpha v \geq n^{-1/2}$
	$MCG$	$\alpha v < n^{-1/2}$	$3n] \ln(1/v)[$	$\lambda_\infty$ for $v \leq n^{-2}$	$\lambda_\infty \alpha^4 \ln M(v)$ for $n > k_2(\alpha v)$
$\alpha_{\infty,\alpha}(G) \leq \alpha$	$MCG$		$3n] \ln(1/v)[$	$\lambda_1 \ln M(v)$ for $v < n^{-1/2}$	$\lambda_1 \ln^2 M(v)$ for $\alpha v < n^{-2}$ $\lambda_\infty \ln M(v)$ for $v \leq \alpha^{-2}$ and $\alpha v < \frac{1}{32}$

## Methods of solving general convex or Lipschitz-convex problems

generally speaking, can be large. For theoretical analysis this is possibly sufficient, but in practice a more detailed analysis is required (this remark also applies to the following discussions). Let us give an example. Suppose it is required to solve general convex problems on  $I_p^{(n)}$ -balls. It is assumed that the accuracy of solution  $v$  is not too high, that  $n$  is quite large, and that  $p$  is close to 1. In this case the bound for the laboriousness of  $\overline{MD}_p \sim 1/(p-1)v^2$ . But in fact, by the results of Section 3.3.1, when  $1 \leq p \leq 2$  these problems can be solved with a laboriousness  $\sim \ln n/v^2$ . Clearly, if  $1/(p-1) \gg \ln n$ , then the second method is indeed preferable to the first.

## 4.3.4

We make one more remark about  $\overline{MD}_p$ -methods,  $1 < p < \infty$ . Their sub-optimality on bodies  $G \subset I_p^{(n)}$  of asphericity  $\alpha$  was established (for large  $n$ ) with accuracy up to a factor of the form  $f(p, \alpha)$ . In other words, they are unimprovable only as regards the order of dependence on  $v$ , if  $n$  is sufficiently large. But in fact the laboriousness of  $\overline{MD}_p$  strongly depends on  $\alpha$ . It is natural to ask whether the ‘sensitiveness of the method to  $\alpha$ ’ can be reduced. It turns out, generally speaking, that it is impossible to do this. For, suppose  $G$  is a convex, closed, bounded body in  $R^n$  with an asphericity  $\leq \alpha$ , and that  $C$  is a class of the type  $C^0(G, R^n, m)$  corresponding to a local, deterministic oracle. Let  $N(v)$  denote the deterministic complexity function of this class, and let  $N_p(v, \alpha)$  be the upper bound of the functions  $N(v)$  over all possible classes  $C$  of the type described (i.e. classes  $C$  corresponding to all possible  $n$ , with  $G \subset R^n$ , and  $\alpha_{p,n}(G) \leq \alpha$ ). From the properties of  $\overline{MD}_p$  it follows that

$$N_p(v, \alpha) \lesssim \left(\frac{\alpha}{v}\right)^{\max(2, p)}.$$

It turns out that a lower bound for  $N_p(v, \alpha)$  of the type

$$N_p(v, \alpha) \gtrsim \left(\frac{\alpha}{v}\right)^{\max(2, p)}, \quad v < \frac{1}{\alpha}. \quad (3.1)$$

is also valid.

Thus a reduction in the laboriousness of  $\overline{MD}_p$  immediately on all bodies with a  $p$ -asphericity  $\leq \alpha$  is possible by not more than  $\sim_p 1$  times. Of course, on some bodies of this kind a greater reduction in the laboriousness of  $\overline{MD}_p$  is possible.

So, in a certain sense,  $\overline{MD}_p$ -methods achieve the maximum possible in the  $L_p$ -scale of characterization of convex bodies.

## Exercise 1. Prove (3.1).

< Suppose first that  $p \geq 2$ . For  $\alpha = 1$  (3.1) follows directly from (1.12), and so we can restrict our attention to the case  $\alpha > 1$ . Let  $\bar{\alpha} = \alpha/8$ . We find  $p'$  from the equation  $(1/4v)^{p'} = (\bar{\alpha}/v)^p$ . It is clear that this equation has a root  $p' > p$ . Let

$n = \lceil (\bar{\alpha}/v)^p \rceil$ , and let  $G$  be an  $\mathbb{I}_p^{(n)}$ -ball. We verify that  $\alpha_{p,n}(G) \leq \alpha$ . For, by (1.3), we have

$$\begin{aligned}\alpha_{p,n}(G) &\leq n^{\frac{1}{p}-\frac{1}{p'}} \leq 2^{\frac{1}{p}-\frac{1}{p'}} \left(\frac{1}{4v}\right)^{\frac{p'-p}{p}} = 2(4^{p'} \cdot \bar{\alpha}^p)^{\frac{1}{p}} \cdot 4^{-\frac{p'-p}{p}} \\ &= 2 \cdot 4 \cdot 4^{-\frac{p'}{p}} \cdot 4^{\frac{p'}{p}} \bar{\alpha} = 8\bar{\alpha} = \alpha.\end{aligned}$$

On the other hand, in view of A1 in Theorem 4.1.2 and (1.11), the complexity of solution of problems of a class  $C$  of type  $C^0(G, R^n, m)$  is not less than

$$d(\infty) \min\left(n, \frac{4^{-p'}}{v^{p'}}\right) = d(\infty) \min\left\{\left(\frac{1}{4v}\right)^{p'}, \left(\frac{1}{4v}\right)^{p'}\right\} = d(\infty) \frac{\bar{\alpha}^p}{v^p},$$

which is what is required in (3.1).

Now let  $1 < p < 2$ . Again it suffices to examine the case  $\alpha \geq 16$ . Let  $G_0^{(n)}$  be the unit ball in the space  $\mathbb{I}_p^{(n)}$ . As was shown in (1.8), for a given  $k$  there is a natural number  $n(k)$  such that  $G_0^{(n)}$  admits a  $k$ -dimensional section  $G_0^{n,k}$  (by a plane  $E^{n,k}$  going through the centre of  $G_0^{(n)}$ ) the asphericity of which relative to a suitable Euclidean norm  $|\cdot|$  on  $E^{n,k}$  is not greater than 2. We can suppose that the inequality  $2|x| \geq \|x\|_p \geq |x|$  holds on  $E^{n,k}$ .

Now let  $\bar{\alpha} = \alpha/16$  and let  $p'$  be chosen so that  $(1/(4v))^{p'} = (\bar{\alpha}/v)^2$ . Then  $p' > 2$ . We put  $k = \lceil (\bar{\alpha}/v)^2 \rceil$  and introduce a norm  $\|\cdot\|_{p'}$  on  $E^{n,k}$  ( $n > n(k)$ ) such that the unit ball in  $E^{n,k}$  relative to this norm—let us call this unit ball  $G_1^{n,k}$ —contains the  $|\cdot|$ -ball of radius 1 and with centre at 0, and is contained in the  $|\cdot|$ -ball of radius  $k^{1/2-1/p'}$  with the same centre. Clearly,  $G_1^{n,k} \supset G_0^{n,k}$ . Let  $G^{(n)}$  denote the convex hull of  $G_0^{(n)}$  and  $G_1^{n,k}$ . Clearly,  $G^{(n)}$  is a centrally symmetric, convex body with a  $\|\cdot\|_{p'}$ -asphericity less than or equal to

$$2k^{\frac{1}{p}-\frac{1}{p'}} \leq 2\left(\frac{\bar{\alpha}^2}{v^2}\right)^{\frac{1}{p}-\frac{1}{p'}} \leq 4\left(\frac{1}{4v}\right)^{p'\left(\frac{1}{2}-\frac{1}{p'}\right)} = 4 \cdot 4^{\frac{p'}{2}} \bar{\alpha} \cdot 4^{-\frac{p'}{2}+1} = 16\bar{\alpha} = \alpha.$$

So  $\alpha_{p',n}(G^{(n)}) \leq \alpha$ .

On the other hand,  $G^{(n)}$  is the unit ball in  $R^n$  relative to the norm  $\|\cdot\|$  for which  $\alpha_{p',k}(\|\cdot\|) = 1$  (because the section of  $G^{(n)}$  by the plane  $E^{n,k}$  is precisely  $G_1^{n,k}$ ). Therefore  $\alpha_{p',k}(G^{(n)}) = 1$  and, in view of A1 in Theorem 4.1.2, the complexity  $N(v)$  of a class of type  $C^0(G^{(n)}, R^n, m)$  admits the estimate

$$\begin{aligned}N_p(\alpha, \delta) &\geq N(v) \\ &\geq d(\infty) \min\left\{k, \frac{1}{(4v)^p}\right\} = d(\infty) \min\left\{\left(\frac{\bar{\alpha}}{v}\right)^2, \left(\frac{1}{4v}\right)^p\right\} \geq d(\infty) \left(\frac{\bar{\alpha}}{v}\right)^2,\end{aligned}$$

which is what is required in (3.1).  $\rangle$

**Exercise 2.** Let  $G$  be an  $\mathbb{I}_p^{(n)}$ -ball,  $1 \leq p < \infty$ . Is there a function  $\varphi(n)$ , with  $\varphi(n)/n \rightarrow 0$  as  $n \rightarrow \infty$ , such that the complexity  $N(v)$  of classes of the

type  $C^0(G, R^n, m)$  satisfies the estimate  $N(v) \leq f(v)\varphi(n)$  (uniformly with respect to  $p$ )?

The same question for the range  $1 \leq p \leq 2$ .

### 4.3.5

We now bring in Table 4 of recommendations about the use of methods for solving Lipschitz-convex problems (of classes of the type  $C_{Lip}^0(G, R^n, \|\cdot\|, m)$ ). As regards the infinite-dimensional situation, see Section 4.2.5. In the table it is assumed that the application of  $MD_p$ -methods to solving problems of the classes considered is carried out as indicated in Section 3.4.5.

Any comments on this table would be, from the qualitative point of view, close to those made earlier for classes of general convex problems. We leave it to the reader to discover and explain the differences in the results arising from the restriction of the class of problems compared with classes of general convex problems.

### 4.3.6

Having acquainted himself with our recommendations regarding the use of methods for solving general complex and Lipschitz-convex problems, the reader will have a right to be curious why we say nothing about the conditions for using the well-known traditional methods of convex programming, except for the gradient method. The point here is that there are by no means many of the standard methods which are suitable for solving arbitrary non-smooth convex problems; as a rule, convergence of the traditional methods—at any rate, the ‘constructive’ convergence, with estimates of the rate—is established on the assumption that the components in the problem are smooth. So far as we know, apart from the gradient method studied in detail earlier, there is only one other method which can be regarded as ‘standard’, and that is Kelly’s method. As we shall soon satisfy ourselves, this method is badly inadmissible from the point of view of its ‘laboriousness-error’ characterization.

We describe the simplest variant of this method which can be applied to solving problems without constraints—let us say, to problems of the class  $C_{Lip}^0(G, R^n, 0)$ . The method works as follows. Suppose a problem is being solved of minimizing a convex Lipschitz function  $f$  in a convex domain  $G \subset R^n$ , and suppose we have already put questions about  $f$  at the points  $x_1, \dots, x_k$  and have obtained the values of  $f(x_i)$  and of the support functionals  $f'(x_i)$ ,  $i = 1, 2, \dots, k$ , to the function  $f$  at these points. We consider the function

$$f^*(x) = \max_{1 \leq i \leq k} \{f(x_i) + \langle f'(x_i)|x - x_i\rangle\};$$

Table 4

Conditions on $\ \cdot\ _G$	Recommended method	Conditions under which the method is recommended	Upper bound for the labouriousness of the method $M(v)$	Potential boundary for reduction of the labouriousness in the class of deterministic methods, $\leq$	Potential boundary for reduction of the labouriousness in the class of stochastic methods, $\leq$
$\alpha_{\infty}(\ \cdot\ ) \leq \alpha$ , $\alpha_{\parallel}(\ G\ ) \leq \beta$ , $1 < \rho < \infty$ , $s = \max\{2, \rho\}$ , $\bar{\alpha} = \min\{2n, \alpha n^{1/\rho}\}$	MD <sub>v</sub>	$v\alpha\beta \geq n^{-1/s}$	$c_p \left(\frac{\alpha}{v}\right)^s$	$\lambda_n \alpha^2 \beta^s$ for $\alpha\beta v \geq n^{-1/s}$	$\lambda_p \alpha^{2s} \beta^s$ for $\alpha\beta v \geq n^{-1/s}$
	MCG	$v\alpha\beta < n^{-1/s}$	$c_{\infty} n] \ln(1/v)[$	$\lambda_p \ln(\beta n)$ for $\alpha\beta v < n^{-1/s}$	$\lambda_p \ln(\beta n) \ln M(v)$ for $\alpha\beta v < n^{-1/s}$
	MD <sub>1,n</sub>	$v\alpha\beta \geq n^{-1/2}$	$c_1 \alpha^2 \frac{\ln(n+1)}{v^2}$	$\lambda_1 \alpha^4 \beta^2$ for $\frac{1}{4} > \alpha\beta v > n^{-1/8}$ for $\alpha\beta v \geq n^{-1/2}$	$\lambda_1 \alpha^4 \beta^2 \ln(n+1)$ for $\alpha\beta v \geq n^{-1/8}$ for $\frac{1}{2} > \alpha\beta v \geq n^{-1/8}$
	MCG	$v\alpha\beta < n^{-1/2}$	$c_{\infty} n] \ln(1/v)[$	$\lambda_1 \ln(\beta n)$ for $\alpha\beta v \leq n^{-1/2}$	$\lambda_1 \ln(\beta n) \ln M(v)$ for $\alpha\beta v < n^{-1/2}$
		$\alpha_{\infty}(\ \cdot\ ) \leq \alpha$ , $\alpha_{\parallel}(\ G\ ) \leq \beta$			
		$\alpha_{\infty}(\ \cdot\ ) \leq \alpha$ , $\alpha_{\parallel}(\ G\ ) \leq \beta$			

clearly, it satisfies the condition  $f^*(x) \leq f(x)$  and it coincides with  $f$  at the 'interrogated' points. In the Kelly method the point where  $f^k$  is minimum on  $G$  is chosen as the next point  $x_{k+1}$ . Strictly speaking, this description does not yet define the method; we have to indicate how  $x_{k+1}$  is to be chosen when there is more than one point at which  $f^k$  has a minimum.

It is convenient in what follows to suppose that if the set of minima of  $f$  intersects the boundary of  $G$ , then  $x_{k+1}$  is chosen on the boundary. It turns out that then the Kelly method has a laboriousness which increases exponentially with the dimension.

*Exercise 3.* Let  $n \geq 3$  and let  $C$  be the class of problems of the form

$$f(x) \rightarrow \min |x \in V_n$$

on the unit ball  $V_n$  of the space  $E^n$ , the problems being generated by convex Lipschitz functions with Lipschitz constant 1 on  $V_n$ . Prove that, if  $M(\epsilon)$  is the number of steps in the 'truncated' Kelly method which solves all problems in  $C$  with an absolute error  $\epsilon < 10^{-2}$ , then

$$M(\epsilon) \geq c(n)(1/\epsilon)^{(n-2)/2}, \text{ where } c(n) > 0.$$

< Let  $x^1, \dots, x^n$  be Cartesian co-ordinates in  $E^n$ . Consider the problem

$$f_e(x) = \max\{|x^1|, g_e(x)\},$$

where

$$g_e(x) = [-1 + 2\epsilon + \sqrt{((x^1)^2 + \dots + (x^n)^2)}]_+.$$

Let  $e_1, \dots, e_n$  be the basic unit vectors corresponding to the chosen co-ordinates, and let the method begin its work at the point  $e_1$ . Let  $S$  denote the set

$$\left\{ x \in V_n | x^1 = 0, \sum_{i=2}^n (x^i)^2 = 1 \right\}.$$

It is easy to see that the work of the Kelly method on  $f_e$  takes place in the following way. The first point interrogated is  $e_1$ , the second is  $(-e_1)$ ; the method then starts to 'sound' the points of  $S$ . 'Knowing' the value of  $f_e$  at a point  $x \in S$ , the method excludes from further consideration the part of  $S$  defined by the relation  $\{y \in S | \langle y|x \rangle \geq 1 - 2\epsilon\}$ ; the next point at which the method 'interrogates'  $f_e$  is some one of the points of  $S$  not yet excluded. This processes continues, at least while there are still points of  $S$  not yet excluded. During this 'sounding' process, the error of the successive approximations regarded as solutions of the problem  $f_e$  will clearly not be less than  $2\epsilon$ .

Thus the Kelly method 'truncated with accuracy  $\epsilon$ ' is obliged to carry out the whole 'sounding' stage. On the other hand, elementary considerations relating to an estimate of the total area of the excluded parts of  $S$  show that the number of questions in the 'sounding' stage admits the required lower bound. >

It is interesting to notice that the method of centres of gravity is constructed 'almost' in the same way as the Kelly method; the only difference is that, as the next point, it is not the point where  $f^k$  has a minimum which is chosen, but the centre of gravity of the region  $\{x \in G | f^k(x) \leq a_k\}$ , where  $a_k$  is the 'record' value of  $f$  (i.e. the lowest value of those found up to the present step). The choice of  $x_{k+1}$  in the Kelly method would seem to be more natural than that in MCG. It is surprising that this more natural rule should lead to an incomparably worse result.

#### 4.3.7

In conclusion we present a number of exercises which contain the answers to certain natural questions about the complexity of classes of general convex and Lipschitz convex problems.

*Exercise 4. Complexity of a class of strictly compatible convex problems.* Let  $G$  be a convex, closed, bounded body in  $R^n$ . Prove that the  $*$ -complexity  $N_\theta^*(v)$  of a class  $C_\theta(G, R^n, m)$  relative to any local deterministic oracle satisfies the estimate

$$\Phi_{n,\infty}(\theta v) \geq N_\theta^*(v) \geq \max\{\varphi_{n,\infty}(\alpha_{\infty,n}(G)v), \varphi_{n,\infty}(\alpha_{\infty,n}(G)\theta)\}.$$

Hence deduce that  $MCG_*$  is a sub-optimal method of solving problems in all classes  $C_\theta$  in the range  $\theta \leq \alpha_{\infty,n}^{-2}(G)$ ,  $v \leq \alpha_{\infty,n}^{-2}(G)$ .

*Exercise 5. The complexity of solution of convex problems 'with respect to an argument' is infinite when  $n > 1$ .* Consider a class  $A$  of functions of two variables on the square

$$K^2 = \{x | |x^1| \leq 1, |x^2| \leq 1\},$$

formed by functions of the form

$$f_\varphi(x^1, x^2) = \max\{\varphi(x^1), ax^2 + b\},$$

where  $|a|, |b| \leq 1$ , and  $\varphi$  is a Lipschitz function on the axis with Lipschitz constant 1. Prove that, for any local deterministic oracle  $\mathcal{O}$  for the class  $A$  and any  $M$ -step method  $\mathcal{B}^M$  of minimizing functions in  $A$  using  $\mathcal{O}$ , with any natural number  $M$ , there is a function  $f \in A$  which has the following property:  $f$  attains its minimum on  $K^2$  at a unique point  $x_f$ , and if  $x^*$  is the result of applying  $\mathcal{B}^M$  to  $f$ , then  $|x^* - x_f| \geq 1$ . Prove a similar result for all bodies  $G$  of dimension not less than 2.

< Consider a subclass  $\bar{A} \subset A$  formed by functions which are independent of  $x^2$ . There is (see Section 4.5.1 below) a  $\varphi \in \bar{A}$  such that all the trajectories of  $\mathcal{B}^M$  lie in the set  $\varphi > 0$ , whereas  $\min_{|x^1| \leq 1} \varphi(x')$  is attained at a unique point  $x^1 = c_\varphi$  and is negative, say  $= -\varepsilon$ , where  $\varepsilon > 0$ . Let  $x^*$  be the result of

applying  $\mathcal{B}^M$  to  $\varphi$  and let  $(x^*)^2$  be  $\geq 0$  (the case  $(x^*)^2 < 0$  is treated similarly). We put  $a = \varepsilon/2$ ,  $b = -\varepsilon/2$ , and consider the function

$$f(x^1, x^2) = \max\{\varphi(x^1), ax^2 + b\}.$$

It is clear that  $f$  attains its minimum on  $K^2$  at a unique point  $(c_\varphi, -1) = x_f$ , and  $f$  coincides with  $\varphi$  in the domain  $\varphi > 0$ . Hence  $x^*$  is the result of applying  $\mathcal{B}^M$  to  $f$ , and so  $|x_f - x^*| \geq 1$ .

*Exercise 6. The complexity of solving Lipschitz problems in unbounded domains is infinite.* Let  $A$  be the class of functions  $f$  on the half-line  $t \geq 0$  which are Lipschitz with Lipschitz constant 1, are convex, and such that  $f(0) = 0$  and  $\lim_{t \rightarrow \infty} f(t) = -1$ . Prove that, for any local, deterministic oracle  $\mathcal{O}$  for the class  $A$  and for any deterministic method  $\mathcal{B}^M$  using the oracle  $\mathcal{O}$  for minimizing functions in  $A$ , there is an  $f \in A$  such that  $f(x^*) > -\frac{1}{2}$ , where  $x^*$  is the result of applying  $\mathcal{B}^M$  to  $f$ . Extend the result to the case of arbitrary unbounded, convex bodies  $G$  (distinct from the whole space).

< We prove by induction over  $p$  that for every  $\mathcal{B}^p$  there is a function  $f_p$  in  $A$  such that, on all trajectories  $\{x_i\}$  of the work of  $\mathcal{B}^p$  on  $f_p$ ,

$$f(x_i) > -\frac{1}{2} + 1/(10p)$$

(the result of applying  $\mathcal{B}^p$  to  $f_p$  is regarded as the  $p$ th point of the trajectory).

The case  $p = 1$  is obvious. The inductive step  $p \rightarrow p+1$ : let the method  $\mathcal{B}^{p+1}$  be given, and let  $f_p$  be that function in  $A$  for which the trajectory of the first  $p$  points  $x_i$  of the work of  $\mathcal{B}^p$  lies in the set

$$\{f_p(x) > -\frac{1}{2} + 1/(10p)\}.$$

Let  $x_{p+1}$  be the  $(p+1)$ th point of the work of  $\mathcal{B}^{p+1}$  on  $f_p$ . It is possible that  $x_{p+1} \leq \max_{1 \leq i \leq p} x_i$ . In that case, put  $f_{p+1} = f_p$ . But if  $x_{p+1} > \max_{1 \leq i \leq p} x_i$ , then there is a Lipschitz function  $\varphi$  with Lipschitz constant 1 which decreases with increasing  $x$  and is such that

$$\varphi(x_{p+1}) > -\frac{1}{2} + 1/(10(p+1)),$$

whereas

$$\varphi(x_i) < -\frac{1}{2} + 1/(10p), \quad 1 \leq i \leq p.$$

Put  $f_{p+1} = \max\{f_p, \varphi\}$ . Then  $f_{p+1}$  will be the required function. >

#### 4.4. PROOF OF LOWER BOUNDS FOR THE COMPLEXITY. I

##### 4.4.1

In this section we begin the proof of the lower bounds for the complexity of classes of general convex or Lipschitz-convex problems, which were formulated in Theorems 4.1.2 and 4.2.1. Let us make a few general remarks in this

connection. First of all, there is a certain connection between the bounds A1–A5 for the Lipschitz case (Theorem 4.2.1) and the corresponding results of Theorem 4.1.2. For, let  $\mathcal{O}$  be an arbitrary deterministic local oracle for the field of problems  $C(G, E, m)$ . Whichever the norm  $\|\cdot\|$  on  $E$  may be (of those which can specify the topology of  $E$ ), the class of problems which is obtained by providing  $C(G, E, m)$  with the standard normalizing factors and the oracle  $\mathcal{O}$  contains the class of problems which is obtained by the similar operation on the field of problems  $C_{\text{Lip}}(G, E, \|\cdot\|, m)$ . Moreover, the normalizing factors of the second class on problems of that class majorize the normalizing factors for the first class. It is therefore clear that a lower bound, which is valid for all the oracles mentioned, of the complexity of the second class is also valid for the first class. Therefore, however the norm  $\|\cdot\|$  on  $E$  compatible with the topology of  $E$  may be chosen, the estimates A1–A5 in Theorem 4.2.1 corresponding to that norm are also true for the functions  $N^o(\cdot)$ ,  $\bar{N}^o(\cdot)$ ,  $N_*(\cdot)$  of Theorem 4.1.2. On the other hand, the upper bounds of these estimates over all possible norms  $\|\cdot\|$  on  $E$  coincide exactly with the right-hand sides of the estimates A1–A5 in Theorem 4.1.2 (it should be remembered that  $\varphi_{n,p}(t)$  is right-continuous for  $t > 0$ ). Thus, it is sufficient to consider the Lipschitz case. Further, it is clear that to solve a problem with constraints ( $m > 0$ ) is in no way simpler than to solve a problem without constraints ( $m = 0$ ), i.e. for the proof of A1–A4 (starting from this point, we shall always have in view the estimates of Theorem 4.2.1) we can suppose that  $m = 0$  and this is what we shall now do.

And one last remark. The proof of the estimates obtained is somewhat tedious. In order not to encumber the exposition with secondary details, we shall set out the purely technical stages rather summarily, but at the same time we shall explain the ideas of the construction fully.

#### 4.4.2

We now pass on from general remarks to the proof of the lower bounds for the complexity. We introduce a characteristic  $k_{E,\|\cdot\|}(\varepsilon)$  of a Banach space  $E$  with a norm  $\|\cdot\|$ , which will be useful in the following exposition. We fix  $E$  and  $\|\cdot\|$ . Let  $k$  be a natural number. We consider a set  $\xi = (\xi_1, \dots, \xi_k)$  of  $k$  functionals belonging to  $E^*$ , with norms  $\leq 1$ , and form  $2^k$  functions on  $E$  of the form

$$f_{\omega,\xi}(x) = \max_{1 \leq i \leq m} \{\omega_i \langle \xi_i | x \rangle\}, \quad \omega_i = \pm 1, \quad \omega = (\omega_1, \dots, \omega_k). \quad (4.1)$$

Let

$$g(\xi, \omega) = \inf_{x: \|x\| \leq 1} f_{\omega,\xi}(x), \quad (4.2)$$

$$\bar{g}(\xi) = \sup_{\omega} g(\xi, \omega), \quad (4.3)$$

$$h_{E,\|\cdot\|}(k) = - \inf_{\xi = (\xi_1, \dots, \xi_k), \|\xi_i\|_* \leq 1} \bar{g}(\xi).$$

We note further that, by virtue of von Neumann's lemma (see Section 6.1.1.),

$$\begin{aligned} g(\xi, \omega) &= \inf_{\|x\| \leq 1} \max_{1 \leq i \leq k} \{\langle \omega_i \xi_i | x \rangle\} = \inf_{\|x\| = 1} \sup_{\lambda = (\lambda_1, \dots, \lambda_k)} \{\langle \sum \lambda_i \omega_i \xi_i | x \rangle\} \\ &= \sup_{\lambda \geq 0, \sum \lambda_i = 1} \inf_{\|x\| = 1} \{\langle \sum \lambda_i \omega_i \xi_i | x \rangle\} = \sup_{\lambda \geq 0, \sum \lambda_i = 1} (-\|\sum \lambda_i \omega_i \xi_i\|_*), \end{aligned}$$

and so

$$\bar{g}(\xi) = - \inf_{\sum |\lambda_i| = 1} \|\sum \lambda_i \xi_i\|_*. \quad (4.4)$$

The function  $\bar{g}(\xi)$  thus has a simple geometrical interpretation: its value on the set  $\xi$  is the distance in  $E^*$  from 0 to the boundary of the 'hyper-octahedron' with vertices at  $\pm \xi_i$ .

It is clear from what has been said that  $h_{E,\|\cdot\|}(k)$  is a non-increasing function of  $k$ . Moreover,  $h_{E,\|\cdot\|}(k) = 0$  if  $k > \dim E$  (why?), whereas  $h_{E,\|\cdot\|}(1) = 1$ . From the last fact, for every  $\varepsilon < \frac{1}{4}$  there is a  $k$  such that  $h_{E,\|\cdot\|}(k) > 4\varepsilon$ .

We put

$$k_{E,\|\cdot\|}(\varepsilon) = \sup \{k | h_{E,\|\cdot\|}(k) > 4\varepsilon\}. \quad (4.5)$$

But if  $\varepsilon \geq \frac{1}{4}$ , then we put  $k_{E,\|\cdot\|}(\varepsilon) = 1$ .

The function  $k_{E,\|\cdot\|}(\varepsilon)$  is a non-increasing function of  $\varepsilon$ , defined for  $\varepsilon > 0$ , and taking non-negative integer values and possibly the value  $+\infty$ . Its rôle in the estimation of the complexity of convex programming relates to the main proposition of this section, which we deal with next.

#### 4.4.3

Let  $W$  denote the unit ball in the space  $E, \|\cdot\|$ , and let  $A(E)$  be the set of functions expressible in the form

$$f(x) = \max_{1 \leq i \leq N} \theta_i(\langle \xi_i | x \rangle), \quad (4.6)$$

where  $\xi_i \in E^*$ ,  $\|\xi_i\|_* \leq 1$ , and  $\theta_i(t)$  is a convex Lipschitz function with Lipschitz constant 1 on the axis. The following propositions hold.

**Proposition 1.** Let  $\mathcal{O}$  be an arbitrary deterministic, local oracle for the field of problems  $A(E)$ , let  $\varepsilon > 0$ , and  $M$  be a natural number, and let  $\mathcal{M}^M$  be an  $M$ -step deterministic method, using the oracle  $\mathcal{O}$ , of minimizing functions in  $A(E)$ . Then, if

$$M < k_{E,\|\cdot\|}(\varepsilon), \quad (4.7)$$

then there is in the class  $A(E)$  a function  $f$  such that

$$\inf_W f(x) < -4\varepsilon \quad \text{and} \quad f(\bar{x}) > 0, \quad (4.8)$$

where  $\bar{x}$  is the result of applying  $\mathcal{M}^M$  to  $f$ .

**Proposition 2.** Under the same hypotheses, let  $\mathcal{B}^M$  be any stochastic method of minimizing functions in  $A(E)$ , using the oracle  $\mathcal{O}$ , and such that  $I(\mathcal{B}^M, f) \leq M$  for all  $f \in A(E)$ . Then, if

$$64M < \frac{k_{E,1,1}(8\epsilon)}{1 + \log_2 k_{E,1,1}(8\epsilon)} \equiv \bar{k}(\epsilon), \quad (4.9)$$

there is a function  $f$  in the class  $A(E)$  such that

$$\inf_{x \in W} f(x) < 0 \quad \text{and} \quad \mathbf{M} \max(f(\bar{x}_f), 0) \equiv \tilde{\epsilon}(\mathcal{B}^M, f) > 4\epsilon, \quad (4.10)$$

where  $\bar{x}_f$  is the (random) result of applying  $\mathcal{B}^M$  to  $f$ , and  $\mathbf{M}$  is the mean over the distribution of this result induced by the distribution of trajectories of  $\mathcal{B}^M$  on  $f$ . When  $\bar{x} = *$  or  $\emptyset$ , take  $f(\bar{x}_f) \equiv 1$ .

*Proof.* 1. We can suppose  $\epsilon < \frac{1}{4}$ , for otherwise (4.7) would be impossible. Let  $k$  be a natural number such that  $k \leq k_{E,1,1}(\epsilon)$ . This implies that  $h_\epsilon(k) > 4\epsilon$ . Choose  $\delta > 0$  such that  $h_\epsilon(k) > 4\epsilon + \delta$ . By the definition of  $h_\epsilon$  the last inequality implies there is a set  $\xi_1, \dots, \xi_k \in E^*$  with  $\|\xi_i\|_* \leq 1$  such that, for all  $\omega = (\omega_1, \dots, \omega_k)$ ,  $\omega_i = \pm 1$ , we have

$$\min_{\|x\|_* \leq 1} \max_{1 \leq i \leq k} \{ \langle \omega_i \xi_i | x \rangle \} < -4\epsilon - \delta. \quad (4.11)$$

We choose  $k$  numbers  $\delta_1, \dots, \delta_k$  such that  $\delta > \delta_1 > \dots > \delta_k > 0$ . Let  $\mathcal{B}^M$  be a deterministic method, using the oracle  $\mathcal{O}$  of minimizing functions in  $A(E)$ , with  $M < k$ . It suffices to prove that there is in the class  $A(E)$  a function which has the property (4.8) in relation to  $\mathcal{B}^M$ .

We construct this function by means of a  $k$ -step construction. Clearly,  $\mathcal{B}^M$  can be regarded as a  $k$ -step method, the result of applying which to any function in  $A(E)$  is the last ( $k$ th) point at which the method puts a question about the problem to the oracle. Accordingly we shall now denote the method by  $\mathcal{B}^k$ . Let  $I_0 = \{1, \dots, k\}$ .

*First step.* Let  $x_1$  be the first point at which the method puts a question. Choose the number greatest in absolute value among the set of numbers  $\langle \xi_i | x_1 \rangle$ ,  $i \in I_0$ ; let the corresponding  $i$  be  $i(1)$ . We put

$$\omega_1 = \begin{cases} +1, & \langle \xi_{i(1)} | x_1 \rangle \geq 0, \\ -1, & \langle \xi_{i(1)} | x_1 \rangle < 0, \end{cases}$$

$I_1 = I_0 \setminus \{i(1)\}$ , and let  $f_1(x) = \omega_1 \langle \xi_{i(1)} | x \rangle + \delta_1$ . The first step is finished.

*(s+1)-th step. Notation:* after  $s$  steps of the construction let the items known be:

- (a) the mutually distinct numbers  $i(j)$ ,  $1 \leq j \leq s$ , from the set  $I_0$ , and the set  $I_s = I_0 \setminus \{i(1), \dots, i(s)\}$ ;

- (b) the numbers  $\omega_j = \pm 1$ ,  $1 \leq j \leq s$ ;
- (c) the functions  $f_j(x) = \max_{1 \leq i \leq j} \{ \omega_i \langle \xi_{i(j)} | x \rangle + \delta_i \}$ ;
- (d) the points  $x_1, x_2, \dots, x_{s+1}$ , where  $x_{j+1}$  is the  $(j+1)$ th point of the work of  $\mathcal{B}^k$  on  $f_j$ .

In the  $(s+1)$ th step we find the number greatest in absolute value among the numbers  $\langle \xi_i | x_{s+1} \rangle$ ,  $i \in I_s$ ; let the corresponding index be  $i(s+1)$ . We put

$$\omega_{s+1} = \begin{cases} +1, & \langle \xi_{i(s+1)} | x_{s+1} \rangle \geq 0, \\ -1, & \langle \xi_{i(s+1)} | x_{s+1} \rangle < 0, \end{cases} \quad I_{s+1} = I_s \setminus \{i(s+1)\},$$

and

$$f_{s+1}(x) = \max_{1 \leq i \leq s+1} \{ \omega_i \langle \xi_i | x \rangle + \delta_i \}.$$

Let  $x_{s+2}$  be the  $(s+2)$ th step of the work of  $\mathcal{B}^k$  on  $f_{s+1}$  ( $x_{s+2}$  is defined only for  $s+1 < k$ ). The  $(s+1)$ th step of the construction is finished.

After  $k$  steps of this construction the functions  $f_1, \dots, f_k$ , and the numbers  $\omega_i = \pm 1$ ,  $1 \leq i \leq k$ , will have been obtained. Moreover,  $f_i \in A(E)$ , and

- (1)  $f_k(x) = \max_{1 \leq i \leq k} \{ \omega_i \langle \xi_{i(k)} | x \rangle + \delta_i \}$ ,  $f_1 \leq f_2 \leq \dots \leq f_k$ ;
- (2)  $x_{s+1}$  is the  $(s+1)$ th step of the work of  $\mathcal{B}^k$  on  $f_s$ ,  $1 \leq s \leq k-1$ ;  $x_1$  is the first point of the work of  $\mathcal{B}^k$ ;
- (3)  $\langle \omega_i \xi_{i(l)} | x_l \rangle \geq \max_{1 \leq i \leq k} \langle \omega_i \xi_{i(l)} | x_l \rangle$ ,  $1 \leq l \leq k$ , and  $\langle \omega_i \xi_{i(l)} | x_l \rangle \geq 0$  (this comes from the rules for forming  $i(l)$  and  $\omega_i$ ).

Let us satisfy ourselves that  $x_k$  is the result of applying  $\mathcal{B}^k$  to  $f_k$  and that  $f_k$  is the function required by the proposition. We check the first statement. By the lemma Section 1.5.3) on indistinguishability it suffices to show that  $f_s = f_k$  in a neighbourhood of  $x_s$ ,  $l \geq s$ . We prove that  $f_s \equiv f_k$  in a neighbourhood of  $x_s$ ,  $1 \leq s \leq k$ . Since  $f_s \leq f_{s+1} \leq \dots \leq f_k$ , it will follow from this that, for  $s, s' \geq t$ , we shall have  $f_s \equiv f_{s'}$  in a neighbourhood of  $x_t$ , as we require.

We have

$$f_s(x) \geq \langle \omega_s \xi_{i(s)} | x \rangle + \delta_s,$$

$$f_k(x) = \max \{ f_s(x), \max_{s < t \leq k} \{ \langle \omega_t \xi_{i(t)} | x \rangle + \delta_t \} \} \equiv \max \{ f_s(x), f^{*}(x) \}.$$

The first relation implies that

$$f_s(x_s) \geq |\langle \xi_{i(s)} | x_s \rangle| + \delta_s,$$

while for all  $t > s$

$$\omega_t \langle \xi_{i(t)} | x_s \rangle \leq |\langle \xi_{i(t)} | x_s \rangle|$$

by (3). Since  $\delta_s \leq \delta_t$ , still holds for  $s < t$ , we have  $f_s(x_s) > f^{*}(x_s)$ . Therefore, in a neighbourhood of  $x_s$ ,  $f_k \equiv \max \{ f^{*}, f_s \}$  coincides with  $f_s$ , as required.

We next verify that  $f_k$  is the required function. It is clear that  $f_k \in A(E)$ . Further,  $x_k$  is the result of applying  $\mathcal{B}^k$  to  $f_k$ , and

$$f_k(x_k) \geq \omega_k \langle \xi_{i(k)} | x_k \rangle + \delta_k > 0.$$

On the other hand,

$$\begin{aligned} \inf_{\|x\| \leq 1} f_k(x) &= \inf_{\|x\| \leq 1} \max_{1 \leq i \leq k} \{ \langle \omega_i \xi_{i(i)} | x \rangle + \delta_i \} \\ &\leq \inf_{\|x\| \leq 1} \max_{1 \leq i \leq k} \{ \langle \omega_i \xi_{i(i)} | x \rangle \} + \delta < -4\varepsilon - \delta + \delta = -4\varepsilon \end{aligned}$$

(we have used the fact that  $i(1), \dots, i(k)$  is a permutation of the integers  $1, \dots, k$ , and the relation (4.11)). The proposition has been proved.

## 2. We now prove proposition 2.

1°. Let  $\mathcal{B}^M, \varepsilon$  be the objects in the formulation of this part. (4.9) cannot hold for  $\varepsilon \geq 1/32$ , and so we shall suppose that  $\varepsilon < 1/32$ . Let an integer  $k > 0$  and functionals  $\xi_1, \dots, \xi_k \in E^*$  with  $\|\xi_i\|_* \leq 1$ , and also a number  $\delta > 0$ , all be chosen so that

$$\frac{k}{64 \log_2(2k)} > M$$

and

$$\min_{x \in W} \max_{1 \leq i \leq k} \langle \omega_i \xi_i | x \rangle < -32\varepsilon - \delta.$$

It follows from the last relation that, when

$$\Delta = \{\bar{\delta} | \bar{\delta} = (\delta_1, \dots, \delta_k), 0 < \delta_i < \delta\}$$

and

$$\Omega = \{\bar{\omega} = (\omega_1, \dots, \omega_k) | \omega_i = \pm 1\}$$

we have

$$\min_{x \in W} f_{\bar{\omega}, \bar{\delta}}(x) = \min_{x \in W} \{ \max_{1 \leq i \leq k} (\omega_i \langle \xi_i | x \rangle + 32\varepsilon + \delta_i) \} < 0. \quad (4.12)$$

Let  $p(\bar{\omega})$  denote the equi-probable distribution on  $\Omega$ .

2°. Let  $X^R$  be an arbitrary set of  $R$  distinct points of the interval  $(0, \delta)$ , and let

$$\Delta_R = \{\bar{\delta} \in \Delta | \delta_i \in X^R, 1 \leq i \leq k\}.$$

Consider the set of problems  $A^R = \{f_{\bar{\omega}, \bar{\delta}} | \bar{\omega} \in \Omega, \bar{\delta} \in \Delta_R\}$ ; we provide this set with an equi-probable distribution  $\sigma(\bar{\omega}, \bar{\delta}) = p(\bar{\omega})q_R(\bar{\delta})$ . Let  $\mathcal{O}$  be the oracle in the formulation of the proposition, and let  $\bar{\mathcal{O}}$  be an oracle whose answer about the function  $f_{\bar{\omega}, \bar{\delta}}$  at a point  $x$  consists in indicating the numbers  $f_{\bar{\omega}, \bar{\delta}}(x)$  and the support functional  $f'_{\bar{\omega}, \bar{\delta}}(x)$ , of the form  $\pm \xi_1, \dots, \pm \xi_k$ , to the function  $f_{\bar{\omega}, \bar{\delta}}(x)$  at the point  $x$  (there always is such a support functional).

Let

$$\tilde{\varepsilon} = \max_{\bar{\omega}, \bar{\delta}} \{ \tilde{\varepsilon}(\mathcal{B}^M, f_{\bar{\omega}, \bar{\delta}}) | \bar{\omega} \in \Omega, \bar{\delta} \in \Delta \}.$$

Then

$$\left. \begin{aligned} \int_{A^R} l(\mathcal{B}^M, f_{\bar{\omega}, \bar{\delta}}) dp(\bar{\omega}) dq_R(\bar{\delta}) &\leq \frac{k}{64 \log_2(2k)}, \\ \int_{A^R} \tilde{\varepsilon}(\mathcal{B}^M, f_{\bar{\omega}, \bar{\delta}}) dp(\bar{\omega}) dq_R(\bar{\delta}) &\leq \tilde{\varepsilon}. \end{aligned} \right\} \quad (4.13)$$

As in the proposition of Section 1.5.2, there is a deterministic, regular method  $\mathcal{B}'^R$  such that

$$\left. \begin{aligned} \int_{A^R} l(\mathcal{B}'^R, f_{\bar{\omega}, \bar{\delta}}) dp(\bar{\omega}) dq_R(\bar{\delta}) &\leq \frac{k}{32 \log_2(2k)}, \\ \int_{A^R} \tilde{\varepsilon}(\mathcal{B}'^R, f_{\bar{\omega}, \bar{\delta}}) dp(\bar{\omega}) dq_R(\bar{\delta}) &\leq 2\tilde{\varepsilon}. \end{aligned} \right\} \quad (4.14)$$

3°. We construct from the method  $\mathcal{B}'^R$  working with the oracle  $\mathcal{O}$  a deterministic method  $\mathcal{B}_R$  using the oracle  $\bar{\mathcal{O}}$ , in the following way. The first point looked for with  $\mathcal{B}_R$  is the same as for  $\mathcal{B}'^R$ . We describe the construction for the  $(s+1)$ th point sought. Let  $x_1, \dots, x_s$  be the initial fragment of the trajectory of  $\mathcal{B}_R$  on  $f \in A^R$ , and let  $h_1(y), \dots, h_s(y)$  be the affine functionals communicated by the oracle  $\bar{\mathcal{O}}$  at the points  $x_1, \dots, x_s$  respectively, i.e. the functionals  $f(x_j) + \langle f'(x_j) | y - x_j \rangle$ , where  $f'(x)$  is the support functional communicated by the oracle  $\mathcal{O}$ . We form the function  $f^s(y) = \max_{1 \leq i \leq s} h_i(y)$ , and we introduce as the  $(s+1)$ th point of the trajectory of  $\mathcal{B}_R$  on  $f$  the  $(s+1)$ th point of the trajectory of  $\mathcal{B}'^R$  on  $f^s$  (this latter point is determined uniquely by  $f^s$ , i.e. by the information obtained by  $\mathcal{B}_R$  from  $\bar{\mathcal{O}}$  in the first  $s$  steps of the work of  $\mathcal{B}_R$ ). If here  $x_{s+1} = *$  or  $\emptyset$  is obtained, then, starting from the index  $j = s+2$ , we put  $x_j = \emptyset$ .

4°. We shall prove that, if  $R$  is sufficiently large, then the initial fragments of the trajectories of  $\mathcal{B}_R$  and  $\mathcal{B}'^R$  (of length  $k$ ) coincide with each other on the 'majority' of the problems in  $A^R$ . That is, if  $a(R)$  is the probability, calculated from the distribution  $\sigma(\cdot)$ , of the event

$$Q(R) \equiv \{f_{\bar{\omega}, \bar{\delta}} | \text{the initial fragments of length } k \text{ of the trajectories of } \mathcal{B}_R \text{ and } \mathcal{B}'^R \text{ on } f_{\bar{\omega}, \bar{\delta}} \text{ are distinct from each other}\},$$

then  $a(R) \rightarrow 0$  as  $R \rightarrow \infty$ .

Postponing for a time the proof of this fact, we deduce from it the required assertion. For, let

$$A_1 = \left\{ (\bar{\omega}, \bar{\delta}) \in \Omega \times \Delta_R | l(\mathcal{B}'^R, f_{\bar{\omega}, \bar{\delta}}) \leq \frac{k}{4 \log_2(2k)} \right\}.$$

By (4.14) we then have  $\sigma(A_1) \geq 7/8$ . Now let  $A_2(R) = A_1 \setminus Q(R)$ . Since  $\sigma(Q(R)) = a(R) \rightarrow 0$  as  $R \rightarrow \infty$ , there is an  $\bar{R}$  such that  $\sigma(A_2(\bar{R})) > 6/7$ . It is

clear from the construction that the result of applying  $\mathcal{B}_R$  to  $f_{\bar{\omega}, \bar{\delta}} \in A_2(R)$  is obtained not later than at the step with iteration number  $\bar{s} = \lceil (K/4) \log_2(2k) \rceil$ , and that

$$\int_{A_2(R)} \tilde{\varepsilon}(\mathcal{B}_R, f_{\bar{\omega}, \bar{\delta}}) d\sigma(\bar{\omega}, \bar{\delta}) = \int_{A_2(R)} \tilde{\varepsilon}(\mathcal{B}^{t_k}, f_{\bar{\omega}, \bar{\delta}}) d\sigma(\bar{\omega}, \bar{\delta}) \leq 2\bar{\varepsilon}.$$

By the last inequality and the relation  $\sigma(A_2(\bar{R})) \geq 6/7$  we have, for some  $\delta_0 \in \Delta_R$ , on putting  $\Omega_{\delta_0} = \{\bar{\omega} | f_{\bar{\omega}, \delta_0} \in A_2(\bar{R})\}$ , that

$$\left. \begin{aligned} \int_{\Omega_{\delta_0}} \tilde{\varepsilon}(\mathcal{B}_R, f_{\bar{\omega}, \delta_0}) dp(\bar{\omega}) &\leq 3\bar{\varepsilon}, \\ \int_{\Omega_{\delta_0}} dp(\bar{\omega}) &\geq \frac{1}{2}. \end{aligned} \right\} \quad (4.15)$$

5°. We now consider the work of  $\mathcal{B}_R$  on the set of problems  $A_{\delta_0} = \{f_{\bar{\omega}, \delta_0} | \bar{\omega} \in \Omega\}$ . On this set it is clear that

$$f_{\bar{\omega}, \delta_0}(x) = \langle f'_{\bar{\omega}, \delta_0}(x) | x \rangle + \delta_{i(f'_{\bar{\omega}, \delta_0}(x)), 0} + 32\varepsilon,$$

where  $i(f'_{\bar{\omega}, \delta_0}(x)) = j$  if  $f'_{\bar{\omega}, \delta_0}(x) = \pm \xi_j$ . Thus on  $A_{\delta_0}$  the whole answer of the oracle  $\bar{\mathcal{O}}$  to a question at the point  $x$  can be recovered from one of its components—from the support functional  $f'(x)$  and, of course, from the point  $x$ . Thus in examining the work of  $\mathcal{B}_R$  on  $A_{\delta_0}$  we can suppose that this method uses an oracle  $\bar{\mathcal{O}}$  which communicates, not the pair  $f(x), f'(x)$ , but just  $f'(x)$ . Correspondingly, the answer of  $\bar{\mathcal{O}}$  to any question is a point of a fixed set of  $2k$  elements. Hence it follows, since  $\mathcal{B}_R$  is deterministic, that there are not more than  $(2k)^s$  different initial fragments of length  $s$  of the trajectories of  $\mathcal{B}_R$  on problems  $f \in A_{\delta_0}$ . Recalling the definition of  $\Omega_{\delta_0}$ , we hence deduce that there cannot be more than

$$(2k)^{k/(4 \log_2(2k))} = 2^{k/4}$$

different results which can be given by  $\mathcal{B}_R$  on all possible problems  $f_{\bar{\omega}, \delta_0}$ ,  $\bar{\omega} \in \Omega_{\delta_0}$ .

On the other hand, it is clear that the sets  $\{x \in W | f_{\bar{\omega}, \delta_0}(x) < 32\varepsilon\}$  for different  $\bar{\omega}$  are disjoint, i.e. for not more than  $2^{k/4}$  problems  $f$  in  $\Omega_{\delta_0}$  can the inequality  $\tilde{\varepsilon}(\bar{x}_j, f) < 32\varepsilon$  hold, where  $\bar{x}_j$  is the result of applying  $\mathcal{B}_R$  to  $f$ . Hence and from the second relation in (4.15) it follows that

$$\int_{\Omega_{\delta_0}} \tilde{\varepsilon}(\mathcal{B}_R, f_{\bar{\omega}, \delta_0}) dp(\bar{\omega}) \geq 32\varepsilon \left( \frac{1}{2} - \frac{2^{k/4}}{2^k} \right),$$

i.e. by the first relation in (4.15),

$$\bar{\varepsilon} \geq 10\varepsilon \left( \frac{1}{2} - 2^{-3k/4} \right).$$

Hence it immediately follows that, when  $k > 5$ , we have

$\bar{\varepsilon} > 4\varepsilon$ . This, by virtue of (4.12) and the definition of  $\bar{\varepsilon}$ , proves (4.10). But if  $k \leq 5$ , (4.9) is obviously impossible, and so (4.10) has been proved.

6°. To complete the proof we have still to prove that  $a(R) \rightarrow 0$  as  $R \rightarrow \infty$ . To do this, we introduce the concept of a  $t$ -regular  $f \in A^R$  ( $t = 0, 1, \dots$ ). By definition, every  $f$  is 0-regular, and  $f$  is  $(t+1)$ -regular if and only if it is  $t$ -regular and  $f^{(t+1)} \equiv f$  in a neighbourhood of  $x_{t+1}$ , where  $x_{t+1}$  is the  $(t+1)$ th point of the work of  $\mathcal{B}_R$  on  $f$  (by definition, all  $f \in A^R$  coincide with one another in neighbourhoods of  $*$  and  $\emptyset$ ). In other words,  $f$  is  $t$ -regular if, for all  $s$  with  $1 \leq s \leq t$ , we have  $f = f^s$  in a neighbourhood of  $x_s$ . By the lemma on indistinguishability (Section 1.5.3), and the definition of  $\mathcal{B}_R$ , it follows at once that, if  $f$  is  $t$ -regular, then the initial fragments of length  $t+1$  of the trajectories of  $\mathcal{B}^R$  and  $\mathcal{B}_R$  on  $f$  coincide with one another.

Every answer of the oracle  $\bar{\mathcal{O}}$  to a question put by  $\mathcal{B}_R$  is an affine functional of the form

$$\eta(y) = \omega \langle \xi_j | y \rangle + \delta, \quad \omega = \pm 1, j \in 1, \dots, k, \delta \in \mathbb{R}.$$

We identify  $\eta$  with the triple  $(\omega, j, \delta)$ .

Let  $\eta^s = (\eta_1, \dots, \eta_s)$  denote the set of answers of the oracle  $\bar{\mathcal{O}}$  to the questions put by  $\mathcal{B}_R$  in the first  $s$  steps of the work of  $\mathcal{B}_R$ , let  $H^s$  be the set of all such sets, and let

$$\mathcal{F}_R(\eta^s) = \{f \in A^R | \text{the information about } f \text{ in the first } s \text{ steps of the work of } \mathcal{B}_R \text{ on } f \text{ is } \eta^s\}.$$

Also, for the answer  $\eta = (\omega, j, \delta)$  of the oracle  $\bar{\mathcal{O}}$  to a question about  $f \in A^R$  let  $i(\eta) = j$ , and let  $i(\eta^s)$  be the subset

$$\{j \in (1, 2, \dots, k) | j = i(\eta_l) \text{ for some } l \leq s\}.$$

We shall calculate the number  $L_R(t)$  of functions  $f \in A^R$  which are not  $t$ -regular. We recall that we have to prove that

$$a(R) \equiv L_R(k)/2^k R^k \xrightarrow[R \rightarrow \infty]{} 0.$$

Let  $N_R(t)$  be the number of  $(t-1)$ -regular functions which are not  $t$ -regular, and let  $N_R(\eta^{t-1}, t)$  be the number of such functions in  $\mathcal{F}_R(\eta^{t-1})$ . Then, clearly,  $L_R(0) = 0$ , and

$$L_R(t+1) = L_R(t) + \sum_{\eta^t \in H^t} N_R(\eta^t, t+1). \quad (4.16)$$

Further, if  $f$  is a  $(t+1)$ -regular function in  $\mathcal{F}_R(\eta^t)$ , then this means that, for certain values  $i_1 \neq i_2$ , we have

$$\omega_{i_1} \langle \xi_{i_1} | x_{t+1} \rangle + \delta_{i_1} = \omega_{i_2} \langle \xi_{i_2} | x_{t+1} \rangle + \delta_{i_2}, \quad (4.17)$$

and moreover, at least one of the numbers  $i_1, i_2$  does not lie in  $i(\eta^t)$ . Fixing  $\eta^t$

uniquely determines  $\delta_j, \omega_j, j \in i(\eta')$ . Therefore it is easy to figure out that, for a fixed  $\eta'$ , (4.17) can hold for not more than  $\varphi(k)R^{k-1|i(\eta')|-1}$  functions  $f \in \mathcal{F}_R(\eta')$ . Here  $|i(\eta')|$  is the number of elements in  $i(\eta')$ , and  $\varphi(k) < \infty$  does not depend on  $R$ . Noting that the number of distinct  $\eta'$  with a given  $i(\eta')$  and a non-void  $\mathcal{F}_R(\eta')$  does not exceed  $\varphi(k)R^{|i(\eta')|}$ , we obtain that

$$L_R(t+1) \leq L_R(t) + \varphi^2(k)R^{k-1}N(k),$$

where  $N(k)$  is the number of distinct subsets of the set  $1, 2, \dots, k$ . Thus

$$L_R(t) \leq t\varphi^2(k)N(k)R^{k-1},$$

and this gives the required relation  $a(R) \rightarrow 0$  as  $R \rightarrow \infty$ . The proposition is proved.

#### 4.4.4

As a corollary we obtain bounds for  $N^\varepsilon(v)$  and  $\bar{N}^\varepsilon(v)$ —the deterministic and stochastic complexities of the class of problems of problems which is obtained from a class  $C_{Lip}$  of the type  $C_{Lip}^0(G, E, \|\cdot\|, m)$  by replacing the oracle of the class by an arbitrary deterministic local oracle  $\emptyset$ .

**Corollary.** With the notation of Section 4.4.2 the bounds

$$N^\varepsilon(v) \geq k_{E, \|\cdot\|}(a_{1,1}(G)v) \quad (4.18)$$

and

$$\bar{N}^\varepsilon(v) \geq \frac{1}{64} \frac{k_{E, \|\cdot\|}(8a_{1,1}(G)v)}{\log_2(2k_{E, \|\cdot\|}(8a_{1,1}(G)v))} \quad (4.19)$$

hold.

*Proof.* Without loss of generality we can suppose that  $G$  contains the unit ball  $W$  in the space  $E, \|\cdot\|$ , and is contained in a ball of radius  $a_{1,1}(G)$ ; strictly speaking, we should say, of radius  $\alpha$  with  $\alpha > a_{1,1}(G)$ . But since  $\alpha$  can be chosen arbitrarily close to  $a_{1,1}(G)$ , and since the right-hand sides of (4.18)–(4.19) are right-continuous in  $v$ , the result of the following argument will not be affected. As mentioned in Section 4.4.1, we can suppose that  $m = 0$ .

We now consider problems from the field  $A(E)$ . Let  $v > 0$  and let  $\varepsilon = a_{1,1}(G)v$ . Write  $N^\varepsilon(v) = N$ . Suppose that (4.18) is false. This means that there is a deterministic method  $\mathcal{B}^M$ , with a laboriousness on a class  $C_{Lip}$  less than  $k_{E, \|\cdot\|}(\varepsilon)$ , which solves all problems in  $C_{Lip}$  with an error  $\leq v$ .

We now apply Proposition 1 of Section 4.4.3 to the objects  $\emptyset, \mathcal{B}^M$ , and  $\varepsilon$ . By this proposition, for a certain  $f \in A(E)$ ,

$$\inf_w f(x) < 0 \quad \text{and} \quad f(\bar{x}) > 4\varepsilon = 4a_{1,1}(G)v,$$

where  $\bar{x}$  is the result of the work of  $\mathcal{B}^M$  on  $f$ . Since  $W \subset G$  by hypothesis, we

have

$$v(\mathcal{B}^M, f) > 4\varepsilon/2L_{\|\cdot\|}(f)\rho_{1,1}(G).$$

But  $\rho_{1,1}(G) \leq 2a_{1,1}(G)$  and  $L_{1,1}(f) \leq 1$  (the latter because of the properties of  $A(E)$ ), i.e.  $v(\mathcal{B}^M, f) > v$ , contrary to the definition of  $\mathcal{B}^M$ . This contradiction proves (4.18); (4.19) is proved similarly (in its proof Proposition 2 of Section 4.4.3 is used instead of Proposition 1 on which (4.18) is based). The corollary is proved.

#### 4.4.5

We are now in a position to prove the estimate A1 (for  $1 \leq p < \infty$ ). To do this it suffices to establish the following fact.

##### 4.4.5.1

The characteristic  $k_{E, \|\cdot\|}(t)$  has the following property.

Let  $E'$  be a subspace of  $E$ , let  $\|\cdot\|'$  be a norm on  $E'$  corresponding to which the  $\|\cdot\|$ -norm on  $E'$  has asphericity  $\beta$ . Then

$$k_{E, \|\cdot\|}(t) \geq k_{E', \|\cdot\|'}(\beta t). \quad (4.20)$$

##### 4.4.5.2

If  $1 \leq p < \infty$ , then, for a suitable choice of the value of  $d(p) > 0$ , the following estimates hold for the corresponding functions  $\varphi_{n,p}$  (defined from (1.11)):

$$k_{1^{(n)}, \|\cdot\|_p}(t) \geq \varphi_{n,p}(t) \quad (1 \leq n < \infty). \quad (4.21)$$

Let us explain how, for  $1 \leq p < \infty$ , the validity of the bounds (2.1) and (2.2) follows from Sections 4.4.5.1 and 4.4.5.2 and (4.18)–(4.19). Firstly, from the definition of  $a_{p, \|\cdot\|, k}(G)$  and the right-continuity with respect to  $v$  of the right-hand sides of (4.18)–(4.19), we obtain from Section 4.4.5 and (4.18)–(4.19) that

$$N^\varepsilon(v) \geq k_{1^{(n)}, \|\cdot\|_p}(a_{p, \|\cdot\|, k}(G)v)$$

and

$$\bar{N}^\varepsilon(v) \geq \frac{k_{1^{(n)}, \|\cdot\|_p}(8a_{p, \|\cdot\|, k}(G)v)}{64 \log_2(2k_{1^{(n)}, \|\cdot\|_p}(8a_{p, \|\cdot\|, k}(G), v))}$$

for all natural numbers  $k \leq \dim E$ . Secondly, substitution in the last estimates of the estimates of Section 4.4.5.2 leads to (2.1) and (2.2). So it remains to prove Sections 4.4.5.1 and 4.4.5.2.

*Proof of Section 4.4.5.1.* It is clear that the unit ball  $V'_*$  in the space  $(E')^*$  relative to the norm  $\|\cdot\|'_*$  has asphericity  $\beta$  relative to the  $\|\cdot\|_*$ -unit ball  $V_*$  in the same space. Without loss of generality we can assume that  $V_* \subset V'_* \subset \beta V_*$ .

Now let  $k_{E,1,1}(\beta\varepsilon) = m$ . If  $m = 1$ , then (4.20) holds; but if  $m > 1$ , i.e. if  $\beta\varepsilon < \frac{1}{4}$ , then from the definition of  $k_{E,1,1}$  it follows that there are  $m$  functionals  $\xi_1, \dots, \xi_m \in V'_*$  such that, for all  $\lambda$  with  $\sum_0^m |\lambda_i| = 1$ , we have  $\|\sum \lambda_i \xi_i\|_*' > 4\beta\varepsilon$ . Since  $V'_* \subset \beta V_*$ , we have  $\|\xi_i\|_* \leq \beta$ . Put  $\zeta_i = \xi_i/\beta$ . Then  $\zeta_i$  is a functional whose norm is  $\leq 1$  on  $(E', \|\cdot\|)$ . Therefore  $\zeta_i$  can be extended from  $E'$  into a functional  $\xi_i$  on  $(E, \|\cdot\|)$  having a norm  $\leq 1$ . Clearly, since  $V_* \subset V'_*$ , we have

$$\|\sum \lambda_i \xi_i\|_* \geq \|\sum \lambda_i \zeta_i\|_*' > \frac{4\beta\varepsilon}{\beta} = 4\varepsilon$$

if  $\sum |\lambda_i| = 1$ , and this gives  $k_{E,1,1}(\varepsilon) \geq m$ . The assertion (4.20) is proved.

#### Proof of Section 4.4.5.2.

1° Let  $p = \infty$ . Then  $k_{l_p^{(n)}}(t) = n$  for  $t < \frac{1}{4}$ ; this follows immediately from a consideration of the functionals  $\langle \xi_i | x \rangle = x^i, 1 \leq i \leq n$ .

2° Let  $2 < p < \infty$ . Then

$$k_{l_p^{(n)}}(t) \geq \frac{1}{2} \min \left\{ n, \frac{4^{-p}}{t^p} \right\}.$$

To see this, it suffices to verify that

$$k_{l_p^{(n)}}(t) \geq \frac{4^{-p}}{t^p} \cdot \frac{1}{2} \text{ when } 2 < t^{-p} \cdot 4^{-p} \leq n.$$

So let  $l$  be the greatest integer less than  $t^{-p} \cdot 4^{-p}$ . Then  $l > \frac{1}{2}t^{-p} \cdot 4^{-p}$  and  $l < n$ . Consider the  $l$  co-ordinate functionals  $\langle \xi_i | x \rangle = x^i, 1 \leq i \leq l$  on  $l_p^{(n)}$ . Let  $\sum_1^l |\lambda_i| = 1$ . Put  $\xi = \sum_1^l \lambda_i \xi_i$ . Then, for  $q = p/(p-1) < 2$ , we have

$$\|\xi\|_q = \left( \sum_1^l |\lambda_i|^q \right)^{1/q} \geq \left( l \left( \frac{1}{l} \right)^q \right)^{1/q} = l^{-(q-1)/q} = l^{-1/p} > 4t,$$

i.e.

$$k_{l_p^{(n)}, 1, l_p}(t) \geq l > \frac{1}{2}(t^{-p} 4^{-p}),$$

as we wished to verify.

3° Let  $1 \leq p \leq 2$ . Then

$$k_{l_p^{(n)}}(t) \geq \frac{1}{4} \min \left\{ n, \frac{1}{(4t)^2} \right\}.$$

To see this, it suffices to verify that

$$k_{l_p^{(n)}}(t) \geq \frac{1}{4} \left( \frac{1}{4t} \right)^2 \text{ when } 4 < (1/4t)^2 \leq n.$$

Let  $s$  be the greatest power of 2 such that  $2^s < (1/4t)^2$ . Then  $2^s \geq \frac{1}{2}(1/4t)^2$ .

Consider the subspace  $R^{(s)} = l_q^{(n)}$ , where  $q = p/(p-1)$ , consisting of vectors, all of whose co-ordinates, except on the first  $2^s$ , are equal to 0. We observe that in this space there is an orthogonal basis consisting of vectors whose first  $2^s$  co-ordinates are equal to  $\pm 1$ . For, we can take  $n = 2^s$ . For  $s = 0$ , there is such a basis. Then, if  $\xi_1, \dots, \xi_m$  is such a basis for a given  $s$ ,  $\{\xi_1, \xi_1\}, \{\xi_1, -\xi_1\}, \dots, \{\xi_{2^s}, \xi_{2^s}\}, \{\xi_{2^s}, -\xi_{2^s}\}$  is the necessary basis for  $s+1$ , and so the assertion is true, by induction.

Let  $\xi_1, \dots, \xi_m, m = 2^s$ , be the corresponding basis for  $R^{(s)}$ , and let  $\xi_i = m^{-1/q} \xi_i, 1 \leq i \leq m$ , so that  $\|\xi_i\|_q = 1$ . Also let  $\sum_1^m |\lambda_i| = 1$  and  $\xi = \sum_1^m \lambda_i \xi_i$ . Then

$$\|\xi\|_2^2 = \sum_1^m \lambda_i^2 \|\xi_i\|_2^2 = m^{1-2/q} \sum_1^m \lambda_i^2 \geq m^{-2/q}.$$

Writing  $\theta_j$  for the moduli of the co-ordinates of  $\xi$ , we obtain

$$\sum_1^m \theta_j^q = \sum_1^m (\theta_j^2)^{q/2} \geq m \left( \frac{\sum \theta_j^2}{m} \right)^{q/2} \geq m^{1+q(-1-2/q)/2} = m^{-q/2}$$

(we have used the fact that  $q \geq 2$ ). Hence

$$\|\xi\|_q \geq m^{-1/2} = 2^{-s/2} > 4t,$$

and so

$$k_{l_p^{(n)}}(t) \geq m \geq \frac{1}{4} \left( \frac{1}{4t} \right)^2.$$

4° It remains to consider the case  $p = 1$ . By the results in 3° it suffices to verify that

$$k_{l_1^{(n)}}(t) \geq a \frac{\ln(n+1)}{t^2} \text{ when } \frac{1}{4} > t \geq n^{-1/8},$$

where  $a$  is an absolute constant. We put  $m_a = \lceil a \ln n / t^2 \rceil$ , and consider  $m_a$  vectors  $\xi_1, \dots, \xi_{m_a}$  from  $l_\infty^{(n)}$  whose co-ordinates are chosen randomly and independently of one another from the numbers  $\pm 1$  with probability  $\frac{1}{2}$ . Simple probability estimates (which we omit) show that there is a small number  $a > 0$ , a small  $t_0 > 0$ , and a large  $n_0 < \infty$  such that, for all  $n > n_0$  and  $t < t_0$ ,  $t > n^{-1/8}$ , the probability that a randomly chosen set  $\xi_1, \dots, \xi_{m_a}$  should satisfy the condition

$$\left\{ \left\| \sum_1^{m_a} \lambda_i \xi_i \right\| > 4t \text{ when } \sum_1^{m_a} |\lambda_i| = 1 \right\}$$

is positive. This gives the estimate

$$k_{l_1^{(n)}}(t) \geq a \ln n / t^2 \text{ when } n > n_0 \text{ and } n^{-1/8} \leq t \leq t_0.$$

Further, for  $t \geq t_0$  and all  $n > 2$ , we have  $k_{l_1^{(n)}}(t) \geq \lceil \ln n \rceil$ . To see this, consider a

set of  $\lceil \ln n \rceil$  vectors of  $\ell_\infty^{(n)}$  the first  $2^{\lceil \ln n \rceil}$  co-ordinates of which are equal to  $\pm 1$ ; the sets of co-ordinates with different subscripts, when the subscripts change from 1 to  $2^{\lceil \ln n \rceil}$ , run through all possible  $2^{\lceil \ln n \rceil}$  sets of the  $\lceil \ln n \rceil$  numbers  $\pm 1$ . Moreover, for  $n \leq n_0$  and  $t < \frac{1}{4}$ , we have

$$k_{\ell_1^{(n)}}(t) \geq \frac{1}{4} \min \left\{ n, \frac{1}{(4t)^2} \right\}$$

by the results of 3°. Taken together, these facts show that

$$k_{\ell_1^{(n)}}(t) \geq \tilde{a} \frac{\ln(n+1)}{t^2} \text{ when } \frac{1}{4} > t \geq n^{-1/8}, \tilde{a} > 0.$$

The assertion in Section 4.4.5.2 has now been proved.

## 4.5 PROOF OF LOWER BOUNDS FOR THE COMPLEXITY. II

Our efforts so far have led to the proof of the bounds (2.1)–(2.2) for all values of  $p$  satisfying the inequality  $1 \leq p < \infty$ . Our objective in this section is to prove (2.1) and (2.2) for  $p = \infty$ . On the way A5 will also be proved.

### 4.5.1

Let  $E, \|\cdot\|$  be a Banach space, let  $n$  be a natural number  $\leq \dim E$ , and let  $\xi^n = (\xi_1, \dots, \xi_n) \in E^*$  be  $n$  linearly independent functionals from  $E^*$ . Let  $K(\xi^n)$  denote the set  $\{x \in E \mid |\langle \xi_i | x \rangle| \leq 1, 1 \leq i \leq n\}$ , and let  $A(\xi^n)$  denote a family of functions of the form  $f(x) = \max_{1 \leq i \leq n} \theta_i(\langle \xi_i | x \rangle)$ , where the  $\theta_i(t)$  are Lipschitz functions (with Lipschitz constants = 1) on the axis. We formulate a proposition whose rôle will be similar to that of those in Section 4.4.3.

**Proposition.** There is an absolute constant  $d(\infty) > 0$  such that the following assertion holds. Let

$$\varphi_{n,\infty}(t) = \begin{cases} 1, & t \geq \frac{1}{4}, \\ d(\infty)n \ln \frac{1}{t}, & t < \frac{1}{4}. \end{cases}$$

Further, let  $\mathcal{O}$  be an arbitrary deterministic local oracle for the field of problems  $A(\xi^n)$ . Also let  $\mathcal{B}^M$  and  $\mathcal{B}^M$  be the corresponding deterministic and stochastic methods of minimizing functions from  $A(\xi^n)$  using the oracle  $\mathcal{O}$ , and let

$$I(\mathcal{B}^M, f) \leq M \text{ and } I(\mathcal{B}^M, f) \leq M \text{ for all } f \in A(\xi^n).$$

Let  $\varepsilon > 0$ . Then:

I. for the method  $\mathcal{B}^M$ : if

$$[M] < \varphi_{n,\infty}(\varepsilon), \varepsilon > 0, \quad (5.1)$$

### Proof of lower bounds for the complexity. II

then, for some  $f \in A(\xi^n)$ , we have

$$\inf_{x \in K(\xi^n)} f(x) < -4\varepsilon \text{ and } f(\bar{x}) > 0, \quad (5.2)$$

where  $\bar{x}$  is the result of applying  $\mathcal{B}^M$  to  $f$ .

II. For the method  $\mathcal{B}^M$ : if

$$[M] < \frac{\varphi_{n,\infty}(8\varepsilon)}{1 + \ln [\varphi_{n,\infty}(8\varepsilon)]}, \quad (5.3)$$

then, for some  $f \in A(\xi^n)$ , we have

$$\inf_{x \in K(\xi^n)} f(x) < 0 \text{ and } \mathbf{M}\max\{f(\bar{x}_f), 0\} \equiv \tilde{\epsilon}(\mathcal{B}^M, f) > 4\varepsilon, \quad (5.4)$$

where  $\bar{x}_f$  is the (random) result of applying  $\mathcal{B}^M$  to  $f$ , and  $\mathbf{M}$  is the mean over its distribution induced by the distribution of the trajectories of  $\mathcal{B}^M$  on  $f$ . For  $\bar{x}_f = *$  or  $\emptyset$ , we take  $f(\bar{x}_f) = 1$ .

### 4.5.2

Postponing for a time the proof of this proposition, we derive from it the bounds (2.1)–(2.2) and the result A5.

Let  $G \subset E$  and let  $k$  be a natural number, and let  $\alpha_{1,\parallel}(G) = \alpha$  and  $\alpha_{\infty,k}(\|\cdot\|) = \beta$ . As in Section 4.4.4, we can suppose that  $G$  is contained in a ball of radius  $\alpha$  and contains the unit ball with centre at 0. We can further suppose that on a certain subspace  $E' \subset E$  an  $l_\infty$ -norm  $\|\cdot\|_\infty$  is introduced such that  $(1/\beta)\|x\|_\infty \leq \|x\| \leq \|x\|_\infty$ . We consider  $k$  functionals  $\xi_1, \dots, \xi_k$  from  $(E')^*$  which are co-ordinate functionals in the  $l_\infty^{(k)}$ -structure introduced on  $E'$ . Then  $\|\xi_i\|_{\infty,*} = 1$ , and so the  $\xi_i$ , regarded as functionals on  $E, \|\cdot\|$ , have a norm  $\leq \beta$ . Therefore the  $\xi_i$  can be extended from  $E'$  on to  $E$  as functionals  $\xi_i$  with norms  $\leq \beta$ . Since the restrictions of the  $\xi_i$  to  $E'$  are linearly independent, so the  $\xi_i$  themselves,  $i = 1, \dots, k$ , are linearly independent. Now let  $N^\circ(v)$  and  $\tilde{N}^\circ(v)$  be the deterministic and stochastic complexities of the class of problems which is obtained by re-equipping a class  $C_{\text{Lip}}$  of the  $C_{\text{Lip}}^0(G, E, \|\cdot\|, m)$  type with a deterministic local oracle  $\mathcal{O}$ . We shall prove that

$$N^\circ(v) \geq \varphi_{k,\infty}(\alpha\beta v) \quad (5.5)$$

and

$$\tilde{N}^\circ(v) \geq \frac{\varphi_{k,\infty}(8\alpha\beta v)}{1 + \ln [\varphi_{k,\infty}(8\alpha\beta v)]}. \quad (5.6)$$

We can, of course, suppose that  $m = 0$ . Let  $\mathcal{B}^M$  be a deterministic method of solving problems of the class considered, with accuracy  $v$  and laboriousness  $M$ . We assume that  $[M] < \varphi_{k,\infty}(\alpha\beta v)$  and derive a contradiction; this will then establish (5.5). Suppose, then, that  $[M] < \varphi_{k,\infty}(\alpha\beta v)$ . We apply to  $\mathcal{O}$ ,

$\xi^n = (\xi_1, \dots, \xi_n)$ , and  $\mathcal{B}^M$  the first proposition in Section 4.5.1. According to the latter, there is an  $f \in A(\xi^n) \subset C_{\text{Lip}}(G, E, \|\cdot\|, M)$  for which (5.2) holds. We observe that, for  $f \in A(\xi^n)$ , we have

$$\inf_{K(\xi^n)} f(x) = \inf_{K(\xi^n) \cap E'} f(x) = \inf_{x \in E', \|x\|_\infty \leq 1} f(x) \geq \inf_{x \in E', \|x\| \leq 1} f(x) \geq \inf_G f(x),$$

i.e.  $\inf_G \bar{f}(x)$  is also  $< 0$ . But then it follows from (5.2) that  $\varepsilon(\mathcal{B}^M, \bar{f}) > 4\alpha\beta\nu$ . On the other hand, it follows from the definition of  $A(\xi^n)$  that  $\bar{f}$  is Lipschitz with a constant  $\leq \max_i \|\xi_i\|_* \leq \beta$ . Since  $G$  lies in a  $\|\cdot\|$ -ball of radius  $\alpha$ , we have  $\rho_{\|\cdot\|}(G) \leq 2\alpha$ , i.e.

$$v(\mathcal{B}^M, \bar{f}) \geq \frac{\varepsilon(\mathcal{B}^M, \bar{f})}{2 \cdot 2\alpha \cdot \beta} > v,$$

which contradicts the definition of  $\mathcal{B}^M$ . This contradiction proves (5.5).

The relation (5.6) is proved in exactly the same way, except that the second proposition in Section 4.5.1 is used instead of the first. We see, incidentally, that whatever the deterministic method  $\mathcal{B}^M$  with finite laboriousness on the class considered may be, it is not able to obtain, for every  $f$  from a class  $C_{\text{Lip}}$  with  $\inf_G f < 0$ , a result  $\bar{x}$  such that  $f(\bar{x}) \leq 0$ . Hence it is clear that, when  $m > 0$ , no deterministic method of finite laboriousness will be able to find an admissible feasible point for every compatible problem of the class. This proves A5.

We have thus completed the programme mentioned for this section, except that we have not yet proved the proposition in Section 4.5.1. To this proof we now turn.

### 4.5.3

We start with the proof of assertion I of the proposition. It suffices to prove it under the hypothesis that  $M$  is a natural number, and then it suffices to prove that the assertion I is true for

$$\varphi_{n, \infty}(\varepsilon) = \begin{cases} 1, & \varepsilon \geq \frac{1}{4}, \\ \max \left[ n, \frac{n}{3} \ln \frac{1}{80\varepsilon} \right], & \varepsilon < \frac{1}{4}. \end{cases} \quad (5.7)$$

I°. Assertion I is trivially true if  $\varepsilon \geq \frac{1}{4}$ . Suppose, then, that  $\varepsilon < \frac{1}{4}$ . We verify that (5.2) is valid when  $M < n$ . Indeed, for all  $\omega_i = \pm 1$ , we have

$$\inf_{K(\xi^n)} \max_{1 \leq i \leq n} \{ \langle \omega_i \xi_i | x \rangle \} = -1$$

because of the linear independence of the  $\xi_i$  and by the definition of  $K(\xi^n)$ . By means of the construction in the proof of Proposition 4.4.3.1 applied to the set  $\xi_1, \dots, \xi_n$ , we satisfy ourselves that, when  $M < n$ , the required function can

already be found among the functions of the form

$$\max_{1 \leq i \leq n} \{ \langle \omega_i \xi_i | x \rangle + \delta_i \}, \quad \omega_i = \pm 1.$$

2°. It is now sufficient to prove that (5.2) is true when

$$M < \frac{n}{3} \ln \frac{1}{80\varepsilon},$$

if the right-hand side of this last inequality is greater than  $n$ . We shall prove that it is even true when

$$M < \frac{n}{3 \ln 2} \ln \frac{1}{80\varepsilon}. \quad (5.8)$$

3°. We consider a certain family of functions on the axis. Let  $0 < \delta < 10^{-4}$ . We put

$$\left. \begin{aligned} \gamma &= \gamma(\delta) = \frac{1}{4} - \delta, \\ c &= c(\delta) = \frac{3}{56} - \frac{3}{4}\delta, \\ q &= q(\delta) = (\frac{1}{2} - 7\delta)/(1 - 4\delta). \end{aligned} \right\} \quad (5.9)$$

Suppose, further, that  $\tau$  is such that  $0 < \tau < \delta$ . We put

$$\theta_r(t) = \begin{cases} -c + q|t + \frac{1}{4}|, & -\infty < t \leq -\delta, \\ -c + q|t + \frac{1}{4}| + \tau[t + \delta]_+, & t \geq -\delta. \end{cases}$$

Here  $[t]_+ = \max\{0, t\}$ . The graph of  $\theta_r(t)$  is sketched in Figure 4.1. It can be immediately verified that:

- (i)  $\theta_r(t)$  is a Lipschitz-convex function with Lipschitz constant 1 and having corners at the points  $-\frac{1}{4}, -\delta$ ;

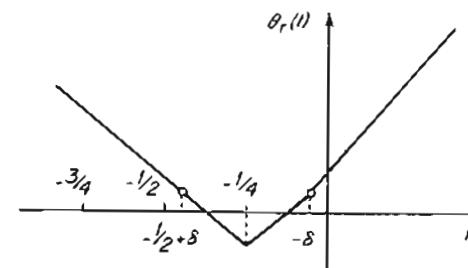


Figure 4.1

(ii) the following inequalities hold:

$$\left. \begin{array}{l} \theta_r(t) < |t| - \frac{3}{7}, \quad |t| \geq 1, \\ \theta_r(t) \geq |t| - \frac{3}{7}, \quad |t + \frac{1}{4}| \leq \gamma; \end{array} \right\} \quad (5.10)$$

(iii)  $\min_r \theta_r(t) = \theta_r(-\frac{1}{2}) = -c(\delta)$ . At the points  $(-\frac{1}{2} + \delta)$  and  $-\delta$  we have  $\theta_r(t) = \frac{4}{3}c(\delta)$ ;

(iv) if  $0 \leq t, r' < r$ , then  $\theta_{r'}(t) < \theta_r(t)$  and  $\theta_{r'}(t) > \theta_r(-t)$ . Moreover,  $\theta_r(t)$  increases when  $t \geq 0$ .

4°. We now formulate and prove a lemma, from which the proposition formulated easily follows.

**Lemma.** Let  $\mathcal{O}$  be an arbitrary deterministic local oracle for the class  $A(\xi^n)$ , let  $0 < \delta < 10^{-4}$ , and let  $q, \gamma, c$  be defined by (5.9). Let  $M$  be a natural number, and let  $\mathcal{B}^M$  be a deterministic method, using the oracle  $\mathcal{O}$ , for minimizing functions from  $A(\xi^n)$ . Then, for every integer  $p$  such that  $0 \leq pn \leq M$ , a set of objects can be found, consisting of:

- (i) a set  $K_p$  of the form  $\{x \mid |\langle \xi_i | x \rangle - c_p^i| \leq \gamma^p\}$ ;
- (ii) functions  $f_p \in A(\xi^n)$  such that

$$f_p(x) = \left\{ q^p \max_{1 \leq i \leq n} |\langle \xi_i | x \rangle - c_p^i| - \frac{3}{7} \left( \frac{7c}{3} \right)^p \right\} \begin{cases} \geq 0 & \text{outside } K_p, \\ = 0 & \text{on } K_p, \end{cases} \quad (J_p)$$

(iii) points  $x_1, \dots, x_{pn}$  which are the first  $pn$  points of the realization of  $\mathcal{B}^M$  on  $f_p$  and which do not lie in  $K_p$ .

**Proof.** This is by induction over  $p$ . For  $p = 0$  the assertions of the lemma are trivially true: we can take  $K_p = K(\xi^n)$  (i.e.  $c_0^i \equiv 0$ ) and

$$f_0(x) = \max_{1 \leq i \leq n} |\langle \xi_i | x \rangle| - \frac{3}{7}.$$

Now let the assertion of the lemma be true for  $p = l$  with  $(l+1)n \leq M$ , and let  $K_l, f_l, x_1, \dots, x_n$  be the corresponding set of objects. We shall construct the set corresponding to  $p = l+1$ . To abbreviate the notation we associate with a point  $x \in E^n$  of its 'co-ordinates'  $x^l = \langle \xi_i | x \rangle - c_p^i$  (these are not real co-ordinates; the correspondence  $x \rightarrow (x^1, \dots, x^n)$  is not one-to-one). The required set of objects will be constructed as the result of  $n$  steps, in the  $j$ th of which there will be available certain functions  $f^1, \dots, f^j$ , pairwise distinct numbers  $i(1), \dots, i(j) \in I_0 = \{1, \dots, n\}$ , a set  $I_j = I_0 \setminus \{i(1), \dots, i(j)\}$ , and also the numbers  $\omega_j = \pm 1$  and points  $y_1, \dots, y_j \in E$ .

We put  $f^0 = f_l$  and let  $\tau_i$  be chosen so that  $\delta > \tau_1 > \dots > \tau_n > 0$ . Suppose that  $s$  steps of the construction have already been done, and that we have the objects  $f^s, I_s$ . At the  $(s+1)$ th step the following operations are carried out.

- (1) a point  $y_{s+1}$  is determined as the  $(nl+s+1)$ th point of the work of  $\mathcal{B}^M$  on  $f^s$ ;

- (2) From the co-ordinates  $y_{s+1}^i$  of the point  $y_{s+1}$  which have indices  $i \in I_s$ , the co-ordinate greatest in absolute value is chosen; let its index be  $i(s+1)$ , and let

$$\omega_{s+1} = \begin{cases} -1, & y_{s+1}^{i(s+1)} < 0, \\ +1, & y_{s+1}^{i(s+1)} \geq 0. \end{cases}$$

We put  $I_{s+1} = I_s \setminus \{i(s+1)\}$ .

- (3) The function

$$f^{s+1}(x) = \max \{f^s(x), (yq)^s \theta_{I_{s+1}}(\omega_{s+1} x^{i(s+1)}/\gamma^l)\} \quad (5.11)$$

is defined. The step is finished. We note that, by (5.11),

$$f_l = f^0 \leq f^1 \leq \dots \leq f^n. \quad (5.12)$$

After  $n$  steps the function  $f^n \equiv f_{l+1}$ , the points  $y_1, \dots, y_n$ , the numbers  $\omega_1, \dots, \omega_n$ , and the permutation  $i(1), \dots, i(n)$  will all have been obtained. We put

$$K_{l+1} = \left\{ x \mid \left| \omega_j x^{i(j)} + \frac{\gamma^l}{4} \right| \leq \gamma^{l+1}, \quad 1 \leq j \leq n \right\}, \quad (5.13)$$

i.e.  $c_{l+1}^i = -\omega_{j^{-1}(i)} \gamma^l / 4 + c_p^i$ , where  $j^{-1}(i)$  is the permutation inverse to  $i(j)$ , and  $x_{ln+1} = y_l$ ,  $1 \leq i \leq n$ .

A direct verification, which we omit, will show that  $K_{l+1}$ ;  $f_{l+1}$ ;  $x_1, \dots, x_{(l+1)n}$  are the required objects. The lemma is proved.

It is now easy to prove assertion I in Proposition 5.5.1. For, let  $M$  be a natural number, and let  $\mathcal{B}^M$  be a method of minimizing functions from  $A(\xi^n)$  using the oracle  $\mathcal{O}$ . We find the least  $M' > M$  which is a multiple of  $n$ . Without loss of generality we can suppose that  $\mathcal{B}$  is an  $M'$ -step method, the result of applying which to any problem from  $A(\xi^n)$  is the last (the  $M'$ th) interrogated point. Applying the lemma proved above to the method  $\mathcal{B}'^M = \mathcal{B}^M$  with  $p = M'/n$  and any  $\delta$ , we find a function  $f_p \in A(\xi^n)$  and a 'cube'  $K_p$  such that the condition  $(J_p)$  holds, and the whole trajectory of  $\mathcal{B}'^M$  on  $f_p$  lies outside  $K_p$ . By  $(J_p)$ ,  $f > 0$  outside  $K_p$ , because on the boundary of this set

$$f = (yq)^p - \frac{3}{7} \left( \frac{7c}{3} \right)^p = \left( \frac{7c}{3} \right)^p - \frac{3}{7} \left( \frac{7c}{3} \right)^p > 0.$$

Moreover, at the 'centre'  $\bar{x}$  of the set  $K_p$ —at any point at which  $\langle \xi_i | \bar{x} \rangle = c_p^i$ ,  $1 \leq i \leq n$ , we have, again by  $(J_p)$ , that  $f_p(\bar{x}) = -\frac{3}{7}(7c/3)^p$ . Since  $M' \leq M+n$ , so

$$f_p(\bar{x}) \leq -\frac{3}{7} \left( \frac{7c}{3} \right)^{M/n+1}.$$

Now let  $\varepsilon > 0$  and  $M$  be given so that

$$\frac{3}{28\varepsilon} > 8^{\frac{M}{n}+1}. \quad (5.14)$$

Then, and with a sufficiently small  $\delta > 0$ , we shall have

$$\frac{3}{7} \left( \frac{7c(\delta)}{3} \right)^{\frac{M}{n}+1} < 4\epsilon.$$

For such a  $\delta$ , the preceding argument proves that there is an  $f \in A(\xi^n)$  such that

$$\inf_{K(\xi^n)} f(x) \leq -\frac{3}{7} \left( \frac{7c(\delta)}{3} \right)^{\frac{M}{n}+1} < -4\epsilon$$

and  $f > 0$  along the trajectory of the work of  $\mathcal{B}^M$  on  $f$  (with the addition to the trajectory of the result of this work). It remains to notice that (5.14) is equivalent to the inequality

$$\frac{M}{n} < \frac{1}{3 \ln 2} \ln \frac{3}{224\epsilon},$$

which is automatically satisfied when

$$M < \frac{n}{3 \ln 2} \ln \frac{1}{80\epsilon},$$

which is what we require. Assertion I of Proposition 4.5.1 is now proved.

#### 4.5.4

We shall now prove assertion II.

1° We define a function  $\theta(t)$  by the relation

$$\theta(t) = \begin{cases} |t| - \frac{1}{2}, & |t| \geq \frac{1}{2}, \\ \frac{1}{8}t - \frac{1}{16}, & -\frac{1}{8} \leq t \leq \frac{1}{2}, \\ \frac{1}{16} \left| t + \frac{1}{4} \right| - \frac{11}{128}, & -\frac{3}{8} \leq t \leq -\frac{1}{8}, \\ -\frac{5}{8} \left( t + \frac{1}{2} \right), & -\frac{1}{2} \leq t \leq -\frac{3}{8}. \end{cases}$$

Then  $\theta(t)$  is Lipschitz convex with constant 1 on the axis.

2° Given an interval  $\Delta$  of the real line, let  $\Delta^{+1}$  and  $\Delta^{-1}$  be the intervals, of length  $|\Delta|/4$ , which have their midpoints at the midpoints of the right, resp. left halves of  $\Delta$ . Let  $N$  be a fixed natural number, and let

$$T = \{\omega = (\omega_1, \dots, \omega_N) : \omega_i = \pm 1\}.$$

By an embedding of  $\omega$  we shall mean a sequence of intervals  $\Delta_1(\omega), \dots, \Delta_N(\omega)$

defined in the following recursive way:

$$\Delta_0(\omega) \equiv [-\frac{1}{2}, \frac{1}{2}], \quad \Delta_{i+1}(\omega) = \Delta^{\omega_{i+1}}(\Delta_i(\omega)).$$

Further, let  $t_i(\omega)$  be the midpoint of  $\Delta_i(\omega)$ .

Let  $T^n$  denote the set  $\{\bar{\omega} = (\omega^1, \dots, \omega^n) : \omega^i \in T\}$ . Further, let  $\delta > 0$  be a fixed number, and  $R$  be a natural number. Let  $X_R$  denote any  $R$ -element subset of the set  $(-\delta, 0)$ , and let  $S_R$  denote the set of  $n \times (N+1)$ -dimensional vectors whose co-ordinates lie in  $X_R$ . Vectors from  $S_R$  will be denoted by  $\tau = \{\tau_j^i, 0 \leq j \leq N, 1 \leq i \leq n\}$ . Let  $x^i$  be the function defined as  $\langle \xi_i | x \rangle, 1 \leq i \leq n$ .

With every ordered pair  $(\bar{\omega}, \tau) \in T^n \times S_R$  we associate the functions  $f_{\bar{\omega}, \tau}^j(x)$ ,  $0 \leq j \leq N$ , defined as

$$f_{\bar{\omega}, \tau}^0(x) = \max_{1 \leq i \leq n} \{ |x^i| - \frac{1}{2} + \tau_0^i \} \equiv \max_{1 \leq i \leq n} \{ \theta_{i+1}^0(x^i) + \tau_0^i \} \quad (5.15)$$

and

$$f_{\bar{\omega}, \tau}^j(x) = \max_{1 \leq i \leq n} \{ \theta_{i+1}^j(x^i - t_{j-1}(\omega^i)) + \tau_j^i \}, j \geq 1. \quad (5.16)$$

Here, for  $j \geq 1$

$$\begin{aligned} \theta_{i+1}^j(t) &\equiv \theta_{i+1}^j(-t) \\ &\equiv \frac{1}{8^{2(j-1)} \alpha_0 \dots \alpha_{j-1}} \theta(-4^{j-1} t) + \sum_{s=0}^{j-2} \frac{8^{-2s}}{\alpha_0 \dots \alpha_s} \theta\left(-\frac{1}{8}\right). \end{aligned} \quad (5.17)$$

In this expression  $\alpha_0 = 1 < \alpha_1 < \dots$  are fixed numbers, and  $\sum_0^\infty a_i \equiv 0$ ,  $\prod_1^\infty b_i = 1$ .

3° We establish certain properties of the functions  $f_{\bar{\omega}, \tau}^j$ . Let  $\tilde{f}_{\bar{\omega}}^j$  be defined by the same formulae as for  $f_{\bar{\omega}, \tau}^j$ , but with  $\tau = 0$ . Then the functions  $\tilde{f}_{\bar{\omega}}^j$  are such that, for  $0 \leq j \leq N-1$ ,

$$\tilde{f}_{\bar{\omega}}^{j+1} - \tilde{f}_{\bar{\omega}}^j \begin{cases} \geq 0 & \text{on } K_j(\bar{\omega}), \\ \leq 0 & \text{outside } K_j(\bar{\omega}), \end{cases} \quad (5.18)$$

where

$$K_j(\bar{\omega}) = \{x \in E | x^i \in \Delta_j(\omega^i), 1 \leq i \leq n\}.$$

In (5.18) the inequality outside  $K_j(\bar{\omega})$  is strict when  $j > 0$ . Further, let

$$\bar{f}_{\bar{\omega}}(x) = \max_{0 \leq j \leq N} \tilde{f}_{\bar{\omega}}^j(x), a = \min_{x \in K(\xi^n)} \tilde{f}_{\bar{\omega}}(x)$$

(it is easy to see that the right-hand side does not depend on  $\bar{\omega}$ ), and

$$\tilde{f}_{\bar{\omega}}(x) = \bar{f}_{\bar{\omega}}(x) - \frac{1}{2}a, \quad \tilde{f}_{\bar{\omega}}^j(x) = \bar{f}_{\bar{\omega}}^j(x) - \frac{1}{2}a.$$

Then

$$\tilde{f}_{\bar{\omega}}(x) = \tilde{f}_{\bar{\omega}}^{j+1}(x) \text{ in } K_j(\bar{\omega}) \setminus K_{j+1}(\bar{\omega}), \quad 0 \leq j \leq N-1 \quad (5.19)$$

$$\left. \begin{aligned} \min_{x \in K(\xi^n)} \tilde{f}_{\bar{\omega}}(x) &< 0, \\ \tilde{f}_{\bar{\omega}}(x) &> \frac{1}{4 \cdot 64^N \alpha_0 \dots \alpha_{N-1}} \text{ for } x \notin K_N(\bar{\omega}). \end{aligned} \right\} \quad (5.20)$$

The inequalities (5.18)–(5.20) are proved by direct verification, starting from the definition of  $\mathcal{O}(t)$ . We leave this work to the reader.

4° We write  $\mathcal{F}_R = \{(\bar{\omega}, t) | \bar{\omega} \in T^k, t \in S_R\}$ . We provide  $\mathcal{F}_R$  with an equi-probability distribution  $p(\bar{\omega}, t) = r(\bar{\omega})q_R(t)$ . For each  $(\bar{\omega}, t) \in \mathcal{F}_R$  let

$$f_{\bar{\omega}, t}(x) = \max_{0 \leq j \leq N} f_{\bar{\omega}, t}^j(x) - \frac{a}{2} \equiv \max_{0 \leq j \leq N} \tilde{f}_{\bar{\omega}, t}^j(x).$$

It is easy to see that  $f_{\bar{\omega}, t}(x) \in A(\xi^n)$ .

Now let

$$\mathcal{B}^M = \int_0^1 \mathcal{B}' dt$$

be the method in the formulation of Proposition 4.5.1. Let  $\tilde{\varepsilon}$  denote the quantity

$$\sup_{(\bar{\omega}, t) \in \cup_R \mathcal{F}_R} \tilde{\varepsilon}(\mathcal{B}^M, f_{\bar{\omega}, t}).$$

Then, for every  $R$ ,

$$\int_{\mathcal{F}_R} l(\mathcal{B}^M, f_{\bar{\omega}, t}) d\mathcal{B}^M d\bar{\omega} dq_R(t) \leq M$$

and

$$\int_{\mathcal{F}_R} \tilde{\varepsilon}(\mathcal{B}^M, f_{\bar{\omega}, t}) d\mathcal{B}^M d\bar{\omega} dq_R(t) \leq \tilde{\varepsilon}. \quad (5.21)$$

Proceeding as in the proof of Proposition 1.5.2 we deduce from (5.21) the existence of a deterministic regular method  $\mathcal{B}'^R$  for which

$$\int_{\mathcal{F}_R} l(\mathcal{B}'^R, f_{\bar{\omega}, t}) d\mathcal{B}'^R d\bar{\omega} dq_R(t) \leq 2M$$

and

$$\int_{\mathcal{F}_R} \tilde{\varepsilon}(\mathcal{B}'^R, f_{\bar{\omega}, t}) d\mathcal{B}'^R d\bar{\omega} dq_R(t) \leq 2\tilde{\varepsilon}. \quad (5.22)$$

5° Now let  $\bar{\mathcal{O}}$  be an oracle for the field of problems  $\{f_{\bar{\omega}, t} | (\bar{\omega}, t) \in \cup_R \mathcal{F}_R\} \equiv \mathfrak{U}$ , the oracle being organized in the following way. The answer of the oracle  $\bar{\mathcal{O}}$  to a question about a problem  $f \in \mathfrak{U}$  at a point  $x \in E$  is a set of four quantities

$$\bar{j} = \bar{j}(\bar{\omega}, t, x), \bar{i} = \bar{i}(\bar{\omega}, t, x), \tau_{\bar{j}}^{\bar{i}}, \omega_{\bar{j}}^{\bar{i}},$$

where  $\bar{j} = \bar{j}(\bar{\omega}, t, x)$  is the least of the indices  $j$  for which  $\bar{f}_{\bar{\omega}, t}(x) = \bar{f}_{\bar{\omega}, t}^j(x)$ , and  $\bar{i}(\bar{\omega}, t, x)$  is the least  $i$  for which

$$\bar{f}_{\bar{\omega}, t}^{\bar{j}}(x) = -\frac{a}{2} + \theta_{\omega_j^i}^{t_i}(x^i - t_{j-1}(\omega^i)) + \tau_j^i.$$

Here we have taken  $\omega_0^i = +1$ ,  $t_{-1}(\omega^i) = 0$ .

Working on the same lines as in the proof of Proposition 2 in Section 4.4.3, and using (5.18)–(5.19), we can prove that, for given  $M$ ,  $\mathcal{B}^M$ ,  $N$ , and the  $\tilde{\varepsilon}$

corresponding to these objects, we can find  $\delta > 0$  arbitrarily small, and an  $R$  arbitrarily large, such that, for a certain deterministic, regular method  $\mathcal{B}_R$  using the oracle  $\bar{\mathcal{O}}$ , the following assertion will hold. Let

$$A(R) = \{(\bar{\omega}, t) \in \mathcal{F}_R | l(\mathcal{B}_R, f_{\bar{\omega}, t}) \leq l(\mathcal{B}'^R, f_{\bar{\omega}, t}) < 8M\}$$

and

$$\tilde{\varepsilon}(\mathcal{B}'^R, f_{\bar{\omega}, t}) = \tilde{\varepsilon}(\mathcal{B}_R, f_{\bar{\omega}, t}) < 4\tilde{\varepsilon}.$$

Then

$$p(A(R)) > \frac{1}{8}. \quad (5.23)$$

We shall now assume that  $\delta$  has been chosen sufficiently small, and  $R$  sufficiently large, for the above assertion to be true.

6° It is clear from what has been said that a  $\bar{t} \in S_R$  can be chosen so that, putting

$$A_{\bar{t}} = \{\bar{\omega} | l(\mathcal{B}_R, f_{\bar{\omega}, \bar{t}}) < 8M \text{ and } \tilde{\varepsilon}(\mathcal{B}_R, f_{\bar{\omega}, \bar{t}}) < 4\tilde{\varepsilon}\} \text{ for this } \bar{t},$$

we obtain

$$r(A_{\bar{t}}) > \frac{1}{8}. \quad (5.24)$$

It is clear from the definition of the oracle  $\bar{\mathcal{O}}$  that on problems from the field  $\mathcal{F}_k = \{f_{\bar{\omega}, t} | \bar{\omega} \in T^n, t = \bar{t}\}$ , it provides exactly as much information as the oracle  $\bar{\mathcal{O}}$  which, at a point  $x$ , gives the quantities  $\bar{j} = \bar{j}(\bar{\omega}, \bar{t}, x)$ ,  $\bar{i} = \bar{i}(\bar{\omega}, \bar{t}, x)$ ,  $\omega_{\bar{j}}^{\bar{i}}$ . Therefore in considering the work of  $\mathcal{B}_R$  on  $A_{\bar{t}}$  we can suppose that  $\mathcal{B}_R$  uses the oracle  $\bar{\mathcal{O}}$ . But the answers of  $\bar{\mathcal{O}}$  to a question about any problem at any point  $x$  are elements of a fixed set of not more than  $2(N+1)n$  elements. Hence it follows that the number of different results provided by the method  $\mathcal{B}_R$  in solving problems from  $A_{\bar{t}}$  not later than the  $[8M]$ th step is not more than  $(2n(N+1))^{[8M]}$ . On the other hand, it follows from (5.20) that, if  $\delta$  is sufficiently small and if the result of applying  $\mathcal{B}_R$  to  $f_{\bar{\omega}, \bar{t}}$  does not lie in  $K_N(\bar{\omega})$ , then

$$\tilde{\varepsilon}(\mathcal{B}_R, f_{\bar{\omega}, \bar{t}}) > \frac{1}{4 \cdot 64^N \alpha_0 \dots \alpha_{N-1}}.$$

Moreover, the sets  $K_N(\bar{\omega})$  corresponding to different  $\bar{\omega}$  are disjoint.

From what has been said it is clear that, if

$$4\tilde{\varepsilon} \leq \frac{1}{4 \cdot 64^N \alpha_0 \dots \alpha_{N-1}}, \quad (5.25)$$

then the set  $A_{\bar{t}}$  cannot contain more than  $(2n(N+1))^{[8M]}$  elements. Combining this fact with (5.24) we obtain that, under the condition (5.25), the following inequality holds:

$$(2n(N+1))^{[8M]} > \frac{1}{8} 2^{nN}, \quad (5.26)$$

i.e.

$$M > \frac{nN - 3}{8 \log_2 (2n(N+1))}. \quad (5.27)$$

7° In particular, let  $\varepsilon > 0$  be given and let  $\mathcal{B}^M$  be a method such as features in the formulation of the assertion in Proposition 4.5.3 and for which, for all  $f \in A(\xi^n)$ , (5.4) does not hold. This implies that, for this method,  $\bar{\varepsilon} \leq 4\varepsilon$ , i.e. the left-hand side of (5.25) is not greater than  $16\varepsilon$ .

Suppose first that  $\varepsilon \leq 10^{-3}$ . Then, choosing  $\alpha_0, \dots, \alpha_{N-1}$  sufficiently close to 1, we see that (5.25) will be satisfied when

$$N = \left( \left[ \left( \ln \frac{1}{64\varepsilon} \right) / (\ln 64) \right] - 1 \right),$$

which, because of (5.27), gives the bound

$$[M] \geq \frac{\varphi_{n,\infty}(8\varepsilon)}{1 + [\ln \varphi_{n,\infty}(8\varepsilon)]_+},$$

when  $(d(\infty))$  is suitably chosen. Therefore, under the conditions (5.3) and with this choice of  $d(\infty)$ , the relation (5.3) is necessarily satisfied for at least one  $f \in A(\xi^n)$ .

8° To complete the proof of Proposition II, we have still to verify that it is true when  $\varepsilon > 10^{-3}$ . It is clear that it is trivially true when  $8\varepsilon \geq \frac{1}{4}$ , i.e. when  $\varepsilon \geq 1/32$ . We now have to show that, when  $1/32 > \varepsilon > 10^{-3}$  and under the condition

$$M > c \frac{n}{\ln(n+1)},$$

where  $c$  is some absolute constant, the relation (5.4) holds for some  $f \in A(\xi^n)$ . This is easily done by applying the construction in the proof of Proposition 2 in Section 4.4.3 to the set  $\xi_1, \dots, \xi_n$ . Proposition 4.5.3 has been proved.

## 4.6 PROOF OF LOWER BOUNDS FOR THE COMPLEXITY. III

We still have to prove the assertions A2, A3, A4 of Theorem 4.2.1. A4 is only of academic interest, and we shall not present its proof. From the form of the functions  $\varphi_{n,p}(\cdot)$  it follows that the assertion A2 for values of  $p$  in the range  $1 < p < 2$  is a particular case of assertion A3. So it suffices to prove A2 for  $p \geq 2$ , and A3.

### 4.6.1

We begin with the proof of A2 for  $p \geq 2$ . What we really have to prove is this: there is a right-continuous, non-decreasing function  $k_p(t)$  which takes positive integer values and is such that, under the conditions of A2 and when  $\alpha' > \alpha_{\| \cdot \|}(G)$  and  $n \geq k_p(\alpha' v)$ , the inequality

$$\tilde{N}^\delta(v) > c(p)(\alpha' v)^{-p}, \quad (6.1)$$

holds, where  $c(p) > 0$  depends only on  $p$ ,  $p \geq 2$ .

### Proof of lower bounds for the complexity. III

For the proof we observe that, if  $k$  is a natural number and if  $\dim E \geq k$  ( $E, \|\cdot\| = L_p, \|\cdot\|_p$ ), then there are  $k$  functionals  $\xi_1, \dots, \xi_k \in E^*$  and a  $k$ -dimensional subspace  $E^k \subset E$  such that

$$\left. \begin{aligned} \|\xi_i\|_* &\leq 2, & 1 \leq i \leq k, \\ \sum_{i=1}^k |\langle \xi_i | x \rangle|^p &\leq 2 \|x\|, & x \in E, \\ \min \left\{ \left\| \sum_1^j \lambda_i \xi_i \right\| \mid \sum |\lambda_i| = 1 \right\} &\geq j^{-1/p}, & 1 \leq j \leq k. \end{aligned} \right\} \quad (6.2)$$

In connection with the proof, in Section 4.6.8, of A3 we remark here that we shall obtain below the required bounds for complexity starting only from the one fact that, for a suitable  $k$ , there exists a system of  $k$  functionals on  $E$  which have the property (6.2). Moreover, the condition that  $E, \|\cdot\|$  is  $L_p, \|\cdot\|_p$  is not used explicitly below.

Now let  $\xi_1, \dots, \xi_k$  be a system of functionals having the properties (6.2), and let  $\mu$  be a probability measure concentrated on the unit ball  $V$  in the space  $E$ . Then

$$2^p \geq \int \sum_{i=1}^k |\langle \xi_i | x \rangle|^p d\mu(x).$$

Therefore, for every  $\delta, \kappa > 0$ , the number of functionals  $\xi_i$  such that

$$\mu\{x \in V \mid |\langle \xi_i | x \rangle| \geq \delta\} \geq \kappa, \quad (6.3)$$

does not exceed

$$N_p(\delta, \kappa) = 2^p \delta^{-p} \kappa^{-1}. \quad (6.4)$$

In particular, if  $k > N_p(\delta, \kappa)$ , then not less than  $k - N_p(\delta, \kappa)$  of the functionals  $\xi_i$  are such that for the set  $\Pi_\delta(\xi_i) = \{x \in V \mid |\langle \xi_i | x \rangle| \geq \delta\}$  we have  $\mu(\Pi_\delta(\xi_i)) < \kappa$ .

### 4.6.2

Now let  $r$  and  $N$  be natural numbers with  $2r \leq 2N \leq \dim E$ . Then we can take  $k = 2N$  in the argument of Section 4.6.1. We consider all possible permutations  $\sigma(i)$  of  $2N$  elements, and we form the functions

$$f_{\sigma, y}^{(r)}(x) = \max_{1 \leq i \leq r} \{ \langle \xi_{\sigma(i)} | x \rangle - iy \}, \quad (6.5)$$

where  $y > 0$  satisfies the condition  $ry \leq 1$ . Also let

$$g_{\sigma, y}^{(r)}(x) = \max \{ f_{\sigma, y}^{(r)}(x), 3\|x\| - 6 \}. \quad (6.6)$$

It is easy to see from (6.2) that  $g_{\sigma, y}^{(r)}(x)$  has the following properties:

- (i)  $g_{\sigma, y}^{(r)}$  is a Lipschitz convex function with the Lipschitz constant 3 on  $E$ ;

- (ii)  $g_{\sigma,y}^{(r)}(x) = f_{\sigma,y}^{(r)}(x)$        $\|x\| \leq 1$ ;  
 (iii)  $g_{\sigma,y}^{(r)}(x) = 3\|x\| - 6 > g_{\sigma,y}^{(r)}(0)$        $\|x\| > 6$ ;  
 (iv)  $\min \{g_{\sigma,y}^{(r)}(x) \mid \|x\| \leq 1\} < -r^{-1/p}$ .

#### 4.6.3

Now let  $\tilde{\mathcal{B}}$  be any method of minimizing convex functions on  $E$  using a local deterministic oracle  $\emptyset$ . We shall examine the work of  $\tilde{\mathcal{B}}$  on a family of functions  $\mathfrak{U}$  formed by all possible functions of the kind just described. We can suppose that  $\tilde{\mathcal{B}} = \int_0^1 \mathcal{B}' dt$  is a mixture of regular deterministic methods. We write

$$\tilde{\varepsilon}(x, g_{\sigma,y}^{(r)}) = [g_{\sigma,y}^{(r)}(x) - \inf \{g_{\sigma,y}^{(r)}(x) \mid \|x\| \leq 1\}]_+,$$

and let  $\tilde{\varepsilon}(\tilde{\mathcal{B}}, g_{\sigma,y}^{(r)})$  be obtained by averaging  $\tilde{\varepsilon}(\tilde{x}, g_{\sigma,y}^{(r)})$  over the distribution of the result  $\tilde{x}$  of applying  $\tilde{\mathcal{B}}$  to  $g$ . Let

$$\varepsilon = \sup \{\tilde{\varepsilon}(\tilde{\mathcal{B}}, g_{\sigma,y}^{(r)}) \mid g_{\sigma,y}^{(r)} \in \mathfrak{U}\}$$

and

$$l = \sup \{l(\tilde{\mathcal{B}}, g_{\sigma,y}^{(r)}) \mid g_{\sigma,y}^{(r)} \in \mathfrak{U}\}.$$

We shall prove that, if  $\varepsilon$  is small, then  $l$  cannot be small. Without worsening the characteristics  $\varepsilon$  and  $l$  of the method  $\tilde{\mathcal{B}}$  on the set of problems  $\mathfrak{U}$  we can suppose that none of the methods  $\mathcal{B}'$  poses questions (or gives results) outside the ball  $W = \{x \in E \mid \|x\| \leq 6\}$ . Indeed, by virtue of (iii) all functions in  $\mathfrak{U}$  coincide with one another outside  $W$  and are greater than their values at  $x = 0$ .

Further, it is convenient to suppose that the trajectories of all  $\mathcal{B}'$  on problems in  $\mathfrak{U}$  consist, after the moment of output of the result, in repetitions of this result and not of the symbol  $\emptyset$ . Moreover, it is, of course, always possible to suppose that the result is a point of  $E$  and is not  $*$ . Thus the trajectories of  $\mathcal{B}'$  on problems in  $\mathfrak{U}$  are sequences of points of  $W$ .

#### 4.6.4

Now let  $\frac{1}{8} > \tilde{\varepsilon} > 0$ , and let  $r = r(\tilde{\varepsilon})$  be the biggest integer such that  $r^{-1/p} > 8\tilde{\varepsilon}$ . We put  $\gamma = \gamma(\tilde{\varepsilon}) = \tilde{\varepsilon}/(2r)$ ,  $\delta = \delta(\tilde{\varepsilon}) = \gamma(\tilde{\varepsilon})/12$ , and find a natural number  $N = N_p(\tilde{\varepsilon})$  which is the least integer greater than or equal to  $r$ , and is such that  $N \geq N_p(\delta(\tilde{\varepsilon}), 1/(2r^3))$ . We assume that  $\dim E \geq 2N$ , and we consider the following  $r$ -step construction relating to the method  $\tilde{\mathcal{B}}$  of the previous section. Let  $\kappa = 1/(2r^3)$ , and let  $\xi_1, \dots, \xi_{2N}$  be the functionals in Section 4.6.1 corresponding to the choice  $k = 2N$ .

The first step in the construction is as follows. Let  $x_1(t)$  be the first point of the work of  $\mathcal{B}'$ , and  $\mu_1 = \mu^1$  be its distribution (i.e. the distribution of the values of  $x_1(t)$  induced by the Lebesgue measure on  $[0, 1]$ ). We find  $l_1$  with  $1 \leq l_1 \leq 2N$  such that  $\mu^1(\Pi_{6\delta}(\xi_{l_1})) < \kappa$  (this is possible because of the choice

of  $N$ ). We define  $\sigma^1$  as the permutation for which  $\sigma^1(1) = l_1$ , and we put  $f_1 = g_{\sigma^1,y}^{(1)}$ . The first step is ended.

Now suppose that  $i < r$  steps of the construction have already been done, and, as a result, there have been constructed the functions  $x_1(t), \dots, x_i(t)$  with values in  $W$ , the functions  $f_1, \dots, f_i \in \mathfrak{U}$ , the permutations  $\sigma^1, \dots, \sigma^i$  of  $2N$  elements, and measures  $\mu_1, \dots, \mu_i$  on  $W$ , such that

- (a)  $x_j(t)$  is the  $j$ th point of the work of  $\mathcal{B}'$  on  $f_{j-1}$ ,  $2 \leq j \leq i$ ;
- (b)  $\mu_j$  is the distribution of the values of  $x_j(t)$ ,  $1 \leq j \leq i$ ;
- (c)  $\sigma^s(f) = \sigma^t(j)$  for all  $j, s, t$  such that  $j \leq s \leq t \leq i$ ;
- (d)  $f_j = g_{\sigma^j,y}^{(j)}$ ,  $j \leq i$ ;
- (e)  $\mu_j(\Pi_{6\delta})(\xi_{\sigma^j(l)}) \leq r\kappa$  for all  $j$  and  $l$  such that  $j \leq l \leq i$ .

In the  $(i+1)$ th step the following operations are performed.  $x_{i+1}(t)$  is determined in accordance with (a), and then  $\mu_{i+1}$  in accordance with (b). The measure

$$\mu^{i+1} = \frac{1}{i+1} \sum_{j=1}^{i+1} \mu_j.$$

is constructed. Further, from among the indices  $s \in \{1, 2, \dots, 2N\}$  which are different from the  $i$  indices  $\sigma^1(1), \dots, \sigma^i(i)$  an index  $l_{i+1}$  is found such that

$$\mu^{i+1}(\Pi_{6\delta}(\xi_{l_{i+1}})) \leq \kappa$$

(such an  $l_{i+1}$  exists because of the choice of  $N$ ; see the concluding remark in Section 4.6.1). We determine  $\sigma^{i+1}$  as an (arbitrary) permutation such that

$$\sigma^{i+1}(j) = \sigma^i(j), \quad 1 \leq j \leq i, \quad \text{and} \quad \sigma^{i+1}(i+1) = l_{i+1}.$$

After this,  $f_{i+1}$  is found from (d). It is then also clear that (e) is satisfied for all  $j \leq l \leq i+1$ .

#### 4.6.5

Let the construction described have been carried out. Let  $T_i$ ,  $1 \leq i \leq r$ , denote the set  $\{t \mid x_i(t) \in \cup_{j \geq i} \Pi_{6\delta}(\xi_{\sigma^j(l)})\}$ . We estimate the Lebesgue measure  $\text{meas } T_i$  of the set  $T_i$ . Clearly

$$\text{meas } T_i \leq \sum_{j=1}^r \text{meas } \{t \mid x_i(t) \in \Pi_{6\delta}(\xi_{\sigma^j(l)})\} = \sum_{j=1}^r \mu_j(\Pi_{6\delta}(\xi_{\sigma^j(l)})) \leq r^2 \kappa$$

(we have used (e)). Therefore, putting  $T = [0, 1] \setminus \cup_{i=1}^r T_i$ , we obtain

$$\text{meas } T \geq 1 - r^3 \kappa \geq \frac{1}{2}.$$

Next we satisfy ourselves that, when  $t \in T$ , the  $x_1(t), \dots, x_r(t)$  are the first  $r$  points of the trajectory of  $\mathcal{B}'$  on  $f \equiv f_r$ . By the lemma on indistinguishability (Section 1.5), it suffices to verify that  $f_j = f_s$  in a neighbourhood of  $x_j$  when

$r \geq s \geq j$ . Using (d) we see that to do this, it is necessary only to satisfy ourselves that

$$\langle \xi_{\sigma^*(l)} | x_j(t) \rangle - l < \langle \xi_{\sigma^*(j)} | x_j(t) \rangle - jy \text{ for } j < l \leq s.$$

Since  $\sigma^*(l) = \sigma^{(0)}(l)$  when  $l \leq s$ , we can suppose that  $s = l$ . Since  $t \notin T$ ,

$$\langle \xi_{\sigma^*(l)} | x_j(t) \rangle < 6\delta \quad \text{and} \quad \langle \xi_{\sigma^*(j)} | x_j(t) \rangle > -6\delta.$$

Therefore the inequality to be proved certainly holds if  $y \geq 12\delta$ , and this is indeed so.

#### 4.6.6

We are now in a position to complete the proof of (6.1). Using the notation of Sections 4.6.3–4.6.5, let  $r > 4l$ . Then  $x_r(t)$  is the result of applying  $\mathcal{B}'$  to  $f_r$  for all  $t \in T$  except perhaps for a set  $\tilde{T}$  of values of  $t$  of measure  $\leq \frac{1}{4}$ , for, by the definition of  $l$ ,

$$\int_0^1 l(\mathcal{B}', f_r) dt \leq l.$$

But  $\bar{\varepsilon}(x_r(t), f_r) > 4\bar{\varepsilon}$  when  $t \in T$ . For, by property (iv) in Section 4.6.2 and the rule for choosing  $r$ ,

$$\inf_{\|x\| \leq 1} f_r(x) < -r^{-1/p} < -8\bar{\varepsilon},$$

while, as we have already seen, for  $t \in T$ ,

$$\langle \xi_{\sigma^*(r)} | x_r(t) \rangle \geq -6\delta$$

and

$$\langle \xi_{\sigma^*(r)} | x_r(t) \rangle - ry \leq f_r(x_r(t)).$$

Hence

$$f_r(x_r(t)) \geq -6\delta - ry > -4\bar{\varepsilon},$$

and this gives the required bound for  $\bar{\varepsilon}(x_r(t), f_r)$ . Thus

$$\bar{\varepsilon} \equiv \int_0^1 \bar{\varepsilon}(\mathcal{B}', f_r) dt > \text{meas}(T \setminus \tilde{T}) \cdot 4\bar{\varepsilon} > \bar{\varepsilon}.$$

We have obtained the following result. Let  $\mathcal{B}$  be the method in Section 4.6.3, and let  $\varepsilon, l$  be its characteristics. Suppose further that

$$r = \left[ \frac{9^{-p}}{\varepsilon^p} \right], \quad \varepsilon \leq \frac{1}{2}, \quad N = N_p(\delta(\varepsilon), \kappa(\varepsilon))$$

and

$$\dim E \geq 2N(E, \|\cdot\| = L_p, \|\cdot\|_p).$$

Then the following inequality holds:

$$l \geq \frac{1}{4} \left[ \frac{9^{-p}}{\varepsilon^p} \right]. \quad (6.7)$$

#### 4.6.7

Now let  $G$  and  $\alpha'$  be objects as in Section 4.6.1. Without loss of generality we can suppose that the class  $C_{Lip}$  considered in A2 has  $m = 0$ , and that  $G$  contains the unit ball in  $E$  and is contained in a ball of radius  $\alpha'$ . Problems of minimizing functions in  $\mathfrak{U}$  can be regarded as problems in  $C_{Lip}$ . If now  $v > 0$  and  $12\alpha'v \leq \frac{1}{3}$ , and if  $\bar{N}^\theta(v) < l$ , then for some method  $\mathcal{B}$  of solving problems in  $C_{Lip}$  we have:  $v(\mathcal{B}, f) \leq v$  and  $l(\mathcal{B}, f) \leq l$  for all  $f \in \mathfrak{U}$ . But, when  $f \in \mathfrak{U}$ ,  $L_{\|\cdot\|_p}(f) \leq 3$ , whereas  $\rho_{\|\cdot\|_p}(G) \leq 2\alpha'$ . Further,  $\{x | \|x\|_p \leq 1\} \subset G$ , and so

$$\varepsilon(\mathcal{B}, f) = L_{\|\cdot\|_p}(f) \cdot 2\rho_{\|\cdot\|_p}(G)v(\mathcal{B}, f) \leq 12\alpha'v \text{ when } f \in \mathfrak{U}.$$

Hence and from the proposition in Section 4.6.6 we obtain, because of (6.7), that

$$l \geq \frac{1}{4} \left[ \left( \frac{1}{9 \cdot 12\alpha'v} \right)^p \right],$$

i.e. also

$$\bar{N}^\theta(v) \geq \frac{1}{4} \left[ \left( \frac{1}{9 \cdot 12\alpha'v} \right)^p \right].$$

Thus, (6.1) holds when  $\alpha'v < 1/108$ . It is clear that (6.1) also holds when  $\alpha'v \geq 1/108$ . The assertion A2 has now been proved for  $p \geq 1$ .

#### 4.6.8

We now prove the assertion A3. It suffices to examine the case  $m = 0$ . Let  $\alpha' > \alpha_{2,k}(G)$ . We can suppose that  $G$  contains the ball of radius 2 with centre at 0 in  $E$ , and that it is contained in the ball of radius  $2\alpha'$  with the same centre. We put  $\bar{k}(t) = k_2(t)$ , where  $k_2(\cdot)$  is the function in A2. By Dvoretzky's theorem (see Exercise 4 in Section 4.1) there is an  $n(k) < \infty$  such that, if  $\dim E \geq n(k)$ , then  $\alpha_{2,k}(\|\cdot\|) \leq 2$ . We put  $\bar{k}(t) = n(\bar{k}(t))$ . It will be sufficient for us to verify that, if  $\dim E \geq \bar{k}(\alpha'v)$ , then

$$\bar{N}^\theta(v) \geq \frac{c}{(\alpha'v)^2} \quad (6.8)$$

for some absolute constant  $c$ .

For, let  $\dim E > \bar{k}(\alpha'v)$ . Then we also have  $\dim E^* \geq \bar{k}(\alpha'v)$ , and so  $\alpha_{2,k}(\|\cdot\|_*) \leq 2$  when  $k = \bar{k}(\alpha'v)$ . Therefore there is a  $k$ -dimensional subspace  $L \subset E^*$  and a Euclidean norm  $\|\cdot\|_2$  on  $L$  such that, when  $\xi \in L$ ,

$$\|\xi\|_2 \leq \|\xi\|_* \leq 2\|\xi\|_2. \quad (6.9)$$

Let  $\xi_1, \dots, \xi_k$  be elements of  $L$  which form an orthonormal system relative to  $\|\cdot\|_2$ . We consider the  $k$ -dimensional Euclidean space  $R^k$ ,  $|\cdot|$ , and the operator  $T: E \rightarrow R^k$  defined by the formula  $Tx = \{\langle \xi_1 | x \rangle, \dots, \langle \xi_k | x \rangle\}$ . We observe

that, when  $x \in E$ ,

$$|Tx| \leq 2\|x\|. \quad (6.10)$$

This follows immediately from (6.9) and the relation

$$|Tx| = \sup \left\{ \left\langle \sum_{i=1}^k \varphi^i \xi_i | x \right\rangle \mid \sum_{i=1}^k (\varphi^i)^2 \leq 1 \right\}.$$

It follows from (6.9) and (6.10) that the system of functionals  $\xi_1, \dots, \xi_k$  satisfies (6.2) with  $p = 2$ . Moreover,  $k \geq k_2(\alpha'v)$ . The construction in Sections 4.6.3–4.6.7 applied to these functionals yields the required lower bound for  $\tilde{N}^0(v)$  immediately. (We remark that in this construction only the fact that such a system of functionals exists for the  $k$  mentioned is used; the condition that ' $E, \|\cdot\|$  coincides with  $L_p, \|\cdot\|_p$ ' is not used.) The assertion A3, and, with it, all the lower bounds for the complexity stated in the theorems in Sections 4.1 and 4.2 have now been proved.

## 5

# Problems of convex stochastic programming

In the preceding chapters we have described methods of the first order for solving convex problems: methods which work with exact information about the problems, or, at any rate, with information little distorted in comparison with the accuracy of solution demanded.

Often more complicated situations are considered, in which exact information about the problem being solved is inaccessible, and, instead, information distorted by random noise has to be used. Such a situation can arise, for example, when the functionals in the problems represent, in content, mean values of random variables which depend on both the controls (with respect to which the optimization is being carried out) and on random factors not subject to control (with respect to which an averaging is carried out). If, here, the information about the problem, which is needed for the optimization, is obtained, during the course of the work, from a stochastic imitation model of the object, then it is clear that this information itself will have a stochastic character. Situations can also be envisaged where there is no appreciable noise physically present in the original problem, but there is noise connected with the conditions under which the necessary information about the problem is obtained. Be that as it may, the analysis of numerical methods of solving problems of stochastic programming of the widely extended type of stochastic approximation can be reduced to the study of methods of solving ordinary—deterministic—extremal problems; it is sufficient to suppose that it is only the source of information which is stochastic, i.e. the oracle which the methods can consult. It is this approach to numerical methods of stochastic programming which is adopted below.

Our purpose is to learn how to apply the methods of mirror descent to the situation under consideration. It turns out that it is indeed possible to do this; moreover, under extremely wide hypotheses, MD-methods prove to be sub-optimal—essentially not improvable as regards laboriousness.

In this chapter we introduce classes of stochastic programming problems (Section 5.1) and MD-methods of solving unconstrained problems in

these classes are constructed (Section 5.2). An extensive literature is devoted to these questions. We mention in this connection the important work of Yu. M. Ermolev [9]. In contrast to well-known works, we examine a more general situation than in Hilbert space, namely, in regular spaces. The methods described later bear the same relation to the traditional methods of stochastic approximation as the MD-methods of Chapter 3 bear to the ordinary method of gradient descent.

In the next chapter we shall consider problems with constraints, a case which has been studied much less. Such problems will be reduced to games. The development of MD-methods in connection with the problem of the approximation of saddle-points of convex-concave games will enable us to obtain practically complete results concerning (convex) conditional stochastic problems. Both chapters (so far as the new results are concerned) are based on the authors' papers [23, 24].

## 5.1 CLASSES OF CONVEX STOCHASTIC PROGRAMMING PROBLEMS

### 5.1.1

From now on, in this and the next chapter, we consider problems which are defined on convex, closed, bounded subsets of real Banach spaces. All these spaces  $E, \|\cdot\|$ , are assumed to be regular and separable. It is further assumed that the pairs,  $E, \|\cdot\|$ , considered have the following property: for any convex, closed, bounded subset  $G \subset E$ , there are Borel mappings  $\pi_G(x): E \rightarrow G$  and  $\mu_G(x): E \rightarrow E^*$  such that, for every  $x$ ,  $\pi_G(x)$  is the point in  $G$  closest to  $x$ , and  $\mu_G(x)$  is the support functional at the point  $x$  to the convex function  $\rho_{\|\cdot\|}(x, G)$ . It will always be taken that  $\mu_G(x) = 0$  when  $x \in G$ .

The properties of  $E, \|\cdot\|$ , just mentioned ensure that the rules of search in the methods to be constructed will have the Borel property; this is an essential circumstance, for we must be able to make use of probabilistic characterizations of the methods. On the other hand, the properties of  $E, \|\cdot\|$ , formulated above actually ensure the separability and reflexivity of  $E$  (indeed, under these conditions the necessary demands on  $E, \|\cdot\|$ , are satisfied automatically if  $\|\cdot\|$  is strictly convex and smooth, i.e. the unit  $\|\cdot\|$ -spherical surface contains no segments which do not reduce to a point, and the unit ball at each point of the boundary admits a unique supporting hyperplane. At the same time the indicated properties of the norm can always be secured by 'adjusting' in the necessary way (and to an arbitrarily small extent) the original norm on  $E$ ).

*Exercise 1.* Prove the above assertion.

### 5.1.2

Let  $E, \|\cdot\|$ , be a Banach space, and let  $G$  be a convex, bounded, closed subset of  $E$ . We consider the field of problems  $C_{\text{Lip}} \equiv C_{\text{Lip}}(G, E, \|\cdot\|, m)$  (see Section 2.2). Our purpose is to equip  $C_{\text{Lip}}$  with a 'stochastic oracle of the first order', and then to define classes of convex stochastic programming problems. The customary oracle for this circle of questions is a stochastic oracle  $\mathcal{C} = ((\Omega, \mathcal{F}_\omega); I; \psi(x, f, \omega))$  (see Section 1.3), for which the information space  $I$  is the same as for a deterministic oracle of the first order, i.e.

$$I = \underbrace{(\mathbb{R} \times \dots \times \mathbb{R})}_{m+1} \times \underbrace{(E^* \times \dots \times E^*)}_{m+1}.$$

In other words, the observation function  $\psi(x, f, \omega)$  has the following structure:

$$\psi(x, f, \omega) = \{\psi_0^0(x, f, \omega), \dots, \psi_m^0(x, f, \omega); \psi_0^1(x, f, \omega), \dots, \psi_m^1(x, f, \omega)\},$$

where the  $\psi_j^0(x, f, \omega)$  are scalar functions (interpreted as observations of the values of  $f_j(x)$  with oracle noise  $\omega$ ), and the  $\psi_j^1(x, f, \omega)$  take values in  $E^*$  (and are interpreted as observations of the support functionals (on  $G$ )  $f'_j(x)$  to the functions  $f_j$  at the point  $x \in G$ ). An oracle of this kind will be called a *natural oracle*.

Clearly it is impossible to construct a meaningful theory under such general hypotheses about the oracle. The customary hypothesis about an oracle is that it is unbiased, i.e.

$$\mathbf{M}_{\mathcal{F}_\omega} \psi_j^0(x, f, \omega) = f_j(x)$$

and

$$\mathbf{M}_{\mathcal{F}_\omega} \psi_j^1(x, f, \omega) \equiv f'_j(x) \in \partial_G f_j(x), \quad (1.1)$$

for all  $x \in G$  and  $f \in C_{\text{Lip}}$ . Here, as always,  $\mathbf{M}$  denotes the mean over the corresponding distribution indicated by the subscript.

But it is not enough that there be no bias. The intensity of the noise must be limited in the necessary way. There are many ways of making this limitation. For example, upper limits could be imposed on the second central moments of the random variables  $\psi_j^1(x, f, \cdot)$ ,  $0 \leq j \leq m$ . But that is not too convenient: the situation will be described by a large number of parameters, about which the absolute values of these moments will say nothing. For a sensible theory, the dimensionless ratios of these moments to the quantities characteristic of the problem must be essential.

We choose the following way—in our opinion, the simplest—of characterizing the noise. Let  $L > 0$  be a certain number. We shall say that the 'intensity of the oracle' on a problem  $f$  with  $m = 0$  does not exceed  $L$  if, for all  $x \in G$ ,

$$\left. \begin{aligned} \mathbf{M}_{\mathcal{F}_\omega} \|\psi_j^1(x, f, \omega)\|_*^2 &\leq \frac{L^2}{4\rho_{\|\cdot\|}(G)} \\ \mathbf{M}_{\mathcal{F}_\omega} |\psi_j^0(x, f, \omega) - f(x)|^2 &\leq L^2 \end{aligned} \right\} \quad (1.2)$$

and

We see that (1.2) imposes in a definite way compatible constraints on the moments of the observations of  $f(x)$  and  $f'(x)$ . Constraints of this kind will turn out to be the most natural ones.

For the general case with  $m > 0$ , the 'intensity of the oracle' could be limited on each component of the problem (by a separate constant for each component). It is convenient, however, to impose constraints which are not completely 'by components':

$$\left. \begin{aligned} \mathbf{M}_{\mathbf{F}_\omega} \|\psi_j^1(x, f, \omega)\|_*^2 &\leq \frac{L_j^2}{4\rho_{\parallel}(G)}, \quad 0 \leq j \leq m, x \in G, \\ \mathbf{M}_{\mathbf{F}_\omega} \left( \max_{0 \leq j \leq m} \left\{ \frac{|\psi_j^0(x, f, \omega) - f_j(x)|^2}{L_j^2} \right\} \right) &\leq 1. \end{aligned} \right\} \quad (1.3)$$

For a given oracle  $\mathcal{O}$  and given  $L_0, \dots, L_m > 0$  the relations (1.1) and (1.3) separate out from  $C_{\text{Lip}}$  a certain field of problems. It is convenient to equip it with the normalizing factors

$$r_0(f) = L_0, r_j(f) = \max_{x \in G} \{0, \min_{x \in G} f_j(x)\} + L_j, \quad 1 \leq j \leq m.$$

We remark that, for compatible  $f$ , we have  $r_j(f) = L_j, 1 \leq j \leq m$ . The weighted field so obtained (Section 2.1) and equipped with the oracle  $\mathcal{O}$  constitutes a certain class of problems. It is precisely classes of this sort which we shall study. The parameters  $L_0, \dots, L_m$  which specify the class, will be taken to be known *a priori*. We observe that, in such a case, instead of solving problems  $f = (f_0, \dots, f_m)$ , we can go over to solving the 'normalized' problems  $\tilde{f} = (f_0/L_0, \dots, f_m/L_m)$  and so, by this normalization, obtain the relations  $L_0 = \dots = L_m = 1$ . Thus the relations (1.2) distinguish a class which essentially is independent of the oracle noise.

It will be convenient later to consider classes with  $L_0 = \dots = L_m = L$ . As just remarked, with the approach we have adopted, the general case can be reduced to this one (even with  $L = 1$ ).

### 5.1.3

The heuristic considerations above enable us to pass on to the formal definition of the classes under investigation. We shall specify them by conditions like (1.1), (1.3), but we shall generalize the latter in two directions: we shall pass from consideration of second moments to consideration of any moments, and we shall drop the requirement for no bias in (1.1) and allow a definite (small) systematic error.

Let  $G, E, \|\cdot\|, C_{\text{Lip}}$  be the same objects as in Section 5.1.2, and let  $\mathcal{O} = ((\Omega, \mathbf{F}_\omega), I, \psi(\cdot))$  be a natural oracle for  $C_{\text{Lip}}$ . Further let  $v_0 \geq 0, L > 0$ , and  $r > 1$ . We define the field of problems  $\tilde{C}_{r,L}^{\mathcal{O},v_0} = \tilde{C}_{r,L}^{\mathcal{O},v_0}(G, E, \|\cdot\|, m)$  as the set of all problems  $f \in C_{\text{Lip}}$  which satisfy the following requirements A–C.

A. For all  $x \in G$  and all  $j, 0 \leq j \leq m$ , the affine functionals

$$g_j(y) = \mathbf{M}_{\mathbf{F}_\omega} \psi_j^0(x, f, \omega) + \langle \mathbf{M}_{\mathbf{F}_\omega} \psi_j^1(x, f, \omega) | y - x \rangle$$

satisfy the conditions

$$f_j(y) \geq g_j(y), y \in G, \text{ and } f_j(x) \leq g_j(x) + v_0 L.$$

B. For all  $j, 0 \leq j \leq m$ , and  $x \in G$

$$\mathbf{M}_{\mathbf{F}_\omega} \|\psi_j^1(x, f, \omega)\|_*^r \leq \left( \frac{L}{2\rho_{\parallel}(G)} \right)^r.$$

C. For all  $x \in G$

$$\mathbf{M}_{\mathbf{F}_\omega} \|\psi^0(x, f, \omega) - f(x)\|_\infty^r \leq L^r.$$

Here

$$\psi^0(x, f, \omega) = (\psi_0^0(x, f, \omega), \dots, \psi_m^0(x, f, \omega))$$

and  $f(x) = (f_0(x), \dots, f_m(x))$  are regarded as mappings in  $I_\infty^{(m+1)}$ . We choose

$$r_j(f) = \begin{cases} L, & j = 0, \\ L + [\min_{x \in G} f_j]_+, & j \geq 1. \end{cases}$$

as the normalizing factors for the field  $\tilde{C}_{r,L}^{\mathcal{O},v_0}$ . We equip the weighted field so obtained with the oracle  $\mathcal{O}$ , and so obtain a class of problems which we also denote by  $\tilde{C}_{r,L}^{\mathcal{O},v_0}(G, E, \|\cdot\|, m)$ .

We point out that, in contrast to the situations in the previous chapters, the problem fields of the classes now introduced depend explicitly on the oracle. We may say that, previously, the field of problems was specified in advance and the corresponding demands on the oracle were introduced when we described the methods. Here, reversely, the oracle is specified in advance, and A–C are the demands on the problem under which we undertake to solve it using the given oracle.

It is important to observe that one and the same method will be proposed for solving problems of a whole family of classes  $\tilde{C}_{r,L}^{\mathcal{O},v_0}(G, E, \|\cdot\|, m)$  which are obtained for all possible oracles  $\mathcal{O}$  for given  $G, m, L, v_0, r, E, \|\cdot\|$ . Thus we do not need to know anything *a priori* about  $\mathcal{O}$ . We need only to be satisfied that the problem to be solved lies in  $C_{\text{Lip}}(G, E, \|\cdot\|, m)$  and is related by the conditions A–C to the oracle to be used (of course,  $G, E, \|\cdot\|, r, L, m, v_0$  must be known *a priori*).

### 5.1.4

*Remark.* In the important particular case where  $v_0 = 0$  and  $m = 0$  the requirement C need not actually be imposed for the oracle. Moreover, if  $m = 0$ , and  $v_0 = 0$ , and if there is a function  $f$  satisfying the conditions A, B relative to

the oracle  $\mathcal{C}$ , then it is easy to obtain from  $\mathcal{C}$  an oracle  $\tilde{\mathcal{C}}$  which satisfies all three conditions A, B, C applied to the function  $\tilde{f}(x) = f(x) - f(x_0)$ , where  $x_0$  is the centre of  $G$ . Clearly, it is all the same whether  $f(x)$  or  $\tilde{f}(x)$  is minimized.

The oracle  $\tilde{\mathcal{C}}$  is constructed in the following way. Suppose it is required to observe the support functional and the value of  $\tilde{f}$  at a point  $x \in G$ . To do this, we consult the oracle  $\mathcal{C}$  twice—the first time at the point  $x$ , and the second time at a point  $\xi$  chosen at random (with respect to a uniform measure) in the segment  $[x_0, x]$ . Let  $\omega$  and  $\omega'$  be the noise of  $\mathcal{C}$  at each of these consultations. We take the quantity  $\psi_0^1(x, f, \omega)$  as the observation of  $\tilde{f}'(x)$  provided by  $\tilde{\mathcal{C}}$ , and the quantity

$$\langle \psi_0^1(\varepsilon, f, \omega') | x - x_0 \rangle \equiv \tilde{\psi}_0^0(x, f, \omega, \omega', \xi)$$

as the observation of  $\tilde{f}(x)$ . One answer of the oracle  $\tilde{\mathcal{C}}$  thus costs two consultations of  $\mathcal{C}$ .

*Exercise 2.* Prove that

$$\mathbf{M}\psi_0^1(x, f, \omega) \in \partial_G \tilde{f}(x)$$

and

$$\mathbf{M}\tilde{\psi}_0^0(x, f, \omega, \omega', \xi) = \tilde{f}(x).$$

Also that

$$\mathbf{M}|\tilde{\psi}_0^0(x, f, \omega, \omega', \xi) - \tilde{f}(x)|^r \leq L^r.$$

Thus  $\tilde{f}$  satisfies all the conditions A, B, C relative to  $\tilde{\mathcal{C}}$ .

### 5.1.5

Before passing on to the description of the methods, we introduce a quantity which is useful in the group of questions with which we are concerned. Let  $V$  be an  $(E, \|\cdot\|)$ -regular function,  $\omega_V(t)$  be its modulus of smoothness (Section 3.2), and let  $r > 1$ . We define  $A(r)$  as the family of all distributions  $\mathbf{F}_t$ , but on the half-line  $t \geq 0$ , such that

$$\int_0^\infty t^r d\mathbf{F}_t \leq 1.$$

We introduce the function

$$\omega_{V,r}(t) = \sup \left\{ \int_0^\infty \tau \omega_V(t\tau) d\mathbf{F}_\tau \mid \mathbf{F}_\tau \in A(r) \right\}.$$

The function  $\omega_{V,r}(t)$  has the same qualitative properties as  $\omega_V(t)$ : it is continuous and bounded for  $t \geq 0$ ,  $\omega_{V,r}(0) = 0$ , and  $\omega_{V,r}(t)$  does not decrease. We can define the function  $\gamma_{V,r}(s)$  inverse to  $\omega_{V,r}(t)$ :

$$\gamma_{V,r}(s) = \sup \{t \mid \omega_{V,r}(t) \leq s\}.$$

It is easy to see that  $\gamma_{V,r}(s)$  is a non-decreasing function of  $s$ , is positive when  $s > 0$ , takes the value  $+\infty$ , and  $\omega_{V,r}(\gamma_{V,r}(s)) \leq s$  for all  $s > 0$  (here  $\omega_{V,r}(+\infty) = \lim_{t \rightarrow \infty} \omega_{V,r}(t)$ ).

*Exercise 3.* Prove these assertions.

The rôle of the function  $\omega_{V,r}(t)$  in the following discussions is based on the next two assertions.

**Proposition.** Let  $E, \|\cdot\|, V, r, \omega_{V,r}(t)$  be the objects just described, and let  $(\Omega, \mathbf{F}_\omega)$  be a Polish space with a regular, Borel, Lebesgue-complete, probability measure  $\mathbf{F}_\omega$ . Further, let  $\omega_i, i = 1, 2, \dots, s$ , be independent random variables distributed over  $\mathbf{F}_\omega$ , let  $\omega^i = (\omega_1, \dots, \omega_s)$ , and let  $\xi_i(\omega^i): \Omega \rightarrow E^*$  be Borel functions such that

$$\mathbf{M}_{\mathbf{F}_{\omega^i}} \|\xi_i(\omega^i)\|_* \leq L, \quad L < \infty,$$

where  $\mathbf{F}_{\omega^i}$  is the joint distribution  $(\omega_1, \dots, \omega_s)$  induced by  $\mathbf{F}_\omega$ . Then

(1) for all  $\varphi \in E^*$  and all  $x \in E$  the following inequality holds for the function  $V_*(\xi) = V(\xi) - \langle \xi | x \rangle$ :

$$V_*(\varphi + \xi_1(\omega_1)) \leq V_*(\varphi) + \langle \xi_1(\omega_1) | V_*(\varphi) \rangle + r(\omega_1), \quad (1.4)$$

where  $r(\omega_1) \geq 0$  is determined from  $\xi_1$  and  $\varphi$ , and  $\mathbf{M}_{\mathbf{F}_{\omega^i}} r(\omega_1)$

(2) if  $\mathbf{M}_{\mathbf{F}_{\omega^i}} \xi_i(\omega^i) = 0$  for all  $\omega^{i-1}, 1 \leq i \leq N$ , and  $\rho_i \geq 0$ , then

$$\mathbf{M}_{\mathbf{F}_{\omega^N}} \left\| \sum_1^N \rho_i \xi_i(\omega^i) \right\|_* \leq L \sum_{i=1}^N \rho_i \omega_{V,r}(L\rho_i) + \frac{1}{2}. \quad (1.5)$$

*Exercise 4.* Prove the proposition.

<(1). By the assertion of Proposition 3.2.4 the first inequality in (1.4) is true for the choice  $r(\omega_1) = \|\xi_1(\omega_1)\|_* \omega_V(\|\xi_1(\omega_1)\|_*)$ . Putting  $\bar{\xi}_1(\omega_1) = \xi_1(\omega_1)/L$ , and letting  $\mathbf{F}_\tau$  denote the distribution of the quantity  $\|\bar{\xi}_1(\omega_1)\|_*$ , we obtain that  $\mathbf{F}_\tau \in A(r)$  and that

$$\int r(\omega_1) d\mathbf{F}_{\omega_1} = \int L\tau \omega_V(L\tau) d\mathbf{F}_\tau \leq L\omega_{V,r}(L),$$

as required by the second inequality in (1). Assertion (1) has been proved.

(2) We estimate  $\mathbf{M}\mathbf{F}_{\omega^i} V(s_i(\omega^i))$ , where

$$s_i(\omega^i) = \sum_1^i \rho_j \xi_j(\omega^j), \quad s_0(\omega^0) = 0.$$

By (1.4), when  $0 \leq i < N$ ,

$$V(s_{i+1}(\omega^{i+1})) \leq V(s_i(\omega^i)) + \rho_{i+1} \langle \xi_{i+1}(\omega^{i+1}) | V'(s_i(\omega^i)) \rangle + r_i(\omega^{i+1}).$$

Averaging both sides of this inequality first with respect to  $\omega_{i+1}$ , and then with

respect to  $\omega^i$ , and using  $\mathbf{M}_{F_{\omega^{i+1}}} \zeta_{i+1}(\omega^{i+1}) = 0$  and (1.4), we obtain

$$\mathbf{M}_{F_{\omega^{i+1}}} V(s_{i+1}(\omega^{i+1})) \leq \mathbf{M}_{F_{\omega^i}} V(s_i(\omega^i)) + \rho_{i+1} L \omega_{V,r}(\rho_{i+1} L).$$

Hence

$$\mathbf{M}_{F_{\omega^N}} V(s_N(\omega^N)) \leq \sum_{i=1}^N \rho_i L \omega_{V,r}(\rho_i L).$$

To obtain (1.5) from this inequality it suffices to note that, by the definition of  $V$ , we have

$$\|\varphi\|_* \leq V(\varphi) + \frac{1}{2} \text{ for all } \varphi \in E^*. \quad >$$

### 5.1.6

In conclusion we give lower bounds for the quantities  $\gamma_{V,r}(\cdot)$  for the standard  $L_p$ -regular functions. In the following,  $c(p, r) > 0$  is a number depending only on  $r$  and  $p$ , and  $r > 1$ .

For  $1 < p < \infty$ ,  $V = V_p$  (see Section 3.2):

$$\gamma_{V,r}(s) \equiv \gamma_{p,r}(s) \geq c(p, r) s^{\frac{1}{p-1}}, \text{ where } \bar{r} = \min \left\{ 2, r, \frac{p}{p-1} \right\} \quad (1.6)$$

For  $p = 1$ ,  $V = V_{1,n}$  (see Section 3.2):

$$\gamma_{V_{1,n},r}(s) \equiv \gamma_{1,n,r}(s) \geq \frac{c(1, r) s^{\frac{1}{p-1}}}{\ln(n+1)}, \quad (1.7)$$

where  $\bar{r} = \min \{2, r\}$ . We note that, when  $r = 2$ , the bounds given for the functions  $\gamma_{p,r}$  and  $\gamma_{1,n,r}$  coincide, as regards order of dependence on  $v$  (resp.  $v$  and  $n$ ), with the similar estimates for  $\gamma_p$  and  $\gamma_{1,n}$  in Section 3.2. For  $p > 2$  this circumstance also holds when  $r \geq p/(p-1)$ .

*Exercise 5.* Using the bounds for  $\gamma_p$  and  $\gamma_{1,n}$  in Section 3.2, prove (1.6)–(1.7).

*Hint.* In deriving (1.7) bear in mind that  $\sup_{t \geq 0} \omega_{V,n}(t) \leq a < \infty$ , and that  $a$  does not depend on  $n$ . (see Section 3.5).

We are now prepared for constructing MD-methods of solving stochastic programming problems.

## 5.2 UNCONSTRAINED OPTIMIZATION

In this section methods are described for solving problems in the classes  $\tilde{\mathcal{C}}_{L,v_0}^r(G, E, \|\cdot\|, 0)$ , i.e. problems of the form

$$f(x) \rightarrow \min |x \in G \subset E.$$

It turns out that for solving these problems we can adapt the method of mirror descent; the modified method is basically similar to the MD-method of solving

### Unconstrained optimization

Lipschitz-convex problems with an exact, first-order oracle (Section 3.3). Let  $V(\cdot)$  be an  $(E, \|\cdot\|)$ -regular function. We shall denote the method associated with  $V$  of solving problems in the above class by  $\widetilde{\text{MD}}_V$ . For brevity we write  $\tilde{\mathcal{C}}$  for the class  $\tilde{\mathcal{C}}_{L,v_0}^r(G, E, \|\cdot\|, 0)$ . In describing the method we shall suppose, as usual, that  $G$  is normalized by the condition: 0 is the centre of  $G$  and  $\rho_{1,1}(G) = 1$ . It is clear that such a normalization can be achieved by simple homothetic transformations.

#### 5.2.1

We start by describing ‘convergent’, ‘infinite-step’ versions of the method. The work of the method is determined by the choice of a sequence of displacements  $\{\rho_i\}_{i=0}^\infty$ ,  $\rho_i > 0$ , and consists in the construction of sequences  $\{\varphi_i \in E^*\}_{i=0}^\infty$  and  $\{x_i \in E\}_{i=0}^\infty$ ,  $\{x_i \in G\}_{i=1}^\infty$ . These sequences are constructed in accordance with the following recursive rules, in which  $f \in \tilde{\mathcal{C}}$  is the problem being solved,  $\psi(x, f, \omega) = (\psi_0^0(x, f, \omega), \psi_0^1(x, f, \omega))$  is the observation function of the oracle  $\mathcal{O}$ , and  $\omega_i$  is the noise of the oracle at the  $i$ th step, and  $\omega^i = (\omega_1, \dots, \omega_i)$ :

$$\varphi_0 = 0, \quad (2.1)$$

$$x_i \equiv x_i(\omega^{i-1}) = V'(\varphi_{i-1}(\omega^{i-1})); \bar{x}_i \equiv \bar{x}_i(\omega^{i-1}) = \pi_G(x_i), \quad (2.2)$$

$$\varphi_i \equiv \varphi_i(\omega^i) = \varphi_{i-1}(\omega^{i-1}) - \rho_i \{\psi_0^1(\bar{x}_i, f, \omega_i) + \|\psi_0^1(x_i, f, \omega_i)\|_* \mu_G(x_i)\}. \quad (2.3)$$

If the oracle  $\mathcal{O}$  were an exact deterministic oracle of the first order, and if the displacements  $\rho_i$  satisfied the natural conditions  $\rho_i \rightarrow 0$  as  $i \rightarrow \infty$ ,  $\sum_i \rho_i = +\infty$ , then the trajectory  $\{\bar{x}_i\}$  would converge to the point of a minimum of  $f$  on  $G$  in the following sense. Let  $x^i$  be the ‘best’ (as regards the value of  $f$ ) among the points  $x_1, \dots, x_i$ . Then  $f(x^i) \rightarrow \inf_G f$  as  $i \rightarrow \infty$  (cf. the results for output of the results of the methods in chapters 2 and 3). In the stochastic case, as may easily be seen, the same thing is also true (after averaging over the random realizations of the methods, of course). Thus the rule for output of the result after  $i$  steps of the type indicated above is capable of ensuring the convergence of the method, whereas the trajectory  $\{\bar{x}_i\}$  itself may not indeed, without additional hypotheses, converge (even in the mean over the functional). However, we are not in a position to realize such a rule. For, the choice of the ‘best’ among the points  $x_1, \dots, x_i$  presupposes exact knowledge of the numbers  $f(x_1), \dots, f(x_i)$ , and we do not have this.

It turns out (and this consideration lies at the basis of the method) that the difficulty mentioned can easily be overcome: the trajectory  $\{\bar{x}_i\}_{i=1}^\infty$  always converges to the optimum of  $f$  (in the mean over the functional) in the Césaro sense. Let us consider the trajectory of convex combinations  $\{x^i\}$  of points  $\bar{x}_j$ , of the form

$$x^i \equiv x^i(\omega^{i-1}) = \left( \sum_{j=1}^i \rho_j \right)^{-1} \sum_{j=1}^i \rho_j \bar{x}_j(\omega^{j-1}) \quad (2.4)$$

(on the right the  $\omega^{j-1}$  are the initial fragments of  $\omega^{i-1}$ ). It turns out that under the requirements on  $\rho_i$  formulated above, the trajectory  $\{x^i\}$  converges to the point of the optimum of  $f$  (in the mean over the functional), and, moreover, the rate of convergence can be estimated explicitly.

**Theorem.** Let the displacements  $\rho_i$  satisfy the conditions

$$\rho_i > 0, \sum_i \rho_i = \infty, \lim_{i \rightarrow \infty} \rho_i = 0,$$

and let  $f \in \mathcal{C}$ . Then for the trajectory of the process (2.1)–(2.4) the following relation holds:

$$\bar{v}_i(f) \equiv \mathbf{M}_{F_{\omega^{i-1}}} v(x^i(\omega^{i-1}), f) \leq v_0 + \frac{1}{2L \sum_{j=1}^i \rho_j} + \frac{\sum_{j=1}^i \rho_j \omega_{V,r}(L\rho_j)}{\sum_{j=1}^i \rho_j}. \quad (2.5)$$

In particular,

$$\overline{\lim_{i \rightarrow \infty} v_i} \leq v_0.$$

*Proof.* It is clear that, when  $i \rightarrow \infty$ , the right-hand side of (2.5) tends to  $v_0$ , and so the second assertion of the theorem follows immediately from the first, to the proof of which we now turn. Let  $x^*$  be the point of the minimum of  $f$  on  $G$ . We put  $V_*(\varphi) = V(\varphi) - \langle \varphi | x^* \rangle$ , and we investigate the variation of  $V_*$  along the trajectory  $\{\varphi_i\}$ . We put

$$\Delta_i(\omega^i) = \psi_0^1(\bar{x}_i(\omega^{i-1}), f, \omega_i) + \|\psi_0^1(\bar{x}_i(\omega^{i-1}), f, \omega_i)\|_* \mu_G(x_i(\omega^{i-1}))$$

and

$$\delta_i(\omega^{i-1}) = \mathbf{M}_{F_{\omega_i}} \Delta_i(\omega^i) \equiv \psi_f(\bar{x}_i(\omega^{i-1})) + b_i(\omega^{i-1}) \mu_G(x_i(\omega^{i-1})),$$

where

$$\psi_f(x) = \mathbf{M}_{F_\omega} \psi_0^1(x, f, \omega)$$

and

$$b_i(\omega^{i-1}) = \mathbf{M}_{F_{\omega_i}} \|\psi_0^1(\bar{x}_i(\omega^{i-1}), f, \omega_i)\|_*.$$

It is clear that

$$b_i(\omega^{i-1}) \geq \|\psi_f(\bar{x}_i(\omega^{i-1}))\|_* \quad (2.6)$$

and

$$\mathbf{M}_{F_{\omega_i}} \|\Delta_i(\omega^i)\|_* \leq \frac{2^r L^r}{[2\rho_{\|\cdot\|}(G)]^r} = L^r \quad (2.7)$$

(in the last estimate the property B of the oracle  $\mathcal{O}$  has to be used). Because of (1.4) we have

$$V_*(\varphi_j(\omega^j)) \leq V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j \langle \Delta_i(\omega^i) | -V'_*(\varphi_{j-1}(\omega^{j-1})) \rangle + r_j(\omega^j),$$

or

$$V_*(\varphi_j(\omega^j)) \leq V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j \langle \Delta_i(\omega^i) | x^* - x_j(\omega^{j-1}) \rangle + r_j(\omega^j), \quad (2.8)$$

where, because of (2.7),

$$\mathbf{M}_{F_{\omega_j}} r_j(\omega^j) \leq L \rho_j \omega_{V,r}(L\rho_j). \quad (2.9)$$

Averaging both sides of (2.8) over  $\omega_j$ , we obtain

$$\begin{aligned} \mathbf{M}_{F_{\omega_j}} V_*(\varphi_j(\omega^j)) &\leq V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j \langle \psi_f(\bar{x}_j(\omega^{j-1})) \\ &\quad + b_j(\omega^{j-1}) \mu_G(x_j(\omega^{j-1})) | x^* - x_j(\omega^{j-1}) \rangle + L \rho_j \omega_{V,r}(L\rho_j). \end{aligned}$$

Because of (2.6) the right-hand side of the last inequality is not greater than

$$V_*(\varphi_{j-1}) + \rho_j \langle \psi_f(\bar{x}_j) | x^* - \bar{x}_j \rangle + L \rho_j \omega_{V,r}(L\rho_j)$$

(cf. the corresponding stage in Section 3.3.5. Thus

$$\begin{aligned} \mathbf{M}_{F_{\omega_j}} V_*(\varphi_j(\omega^j)) &\leq V_*(\varphi_{j-1}(\omega^{j-1})) \\ &\quad + \rho_j \langle \psi_f(\bar{x}_j(\omega^{j-1})) | x^* - \bar{x}_j(\omega^{j-1}) \rangle + L \rho_j \omega_{V,r}(L\rho_j). \end{aligned} \quad (2.10)$$

By property A of the oracle  $\mathcal{O}$ , the middle term of the sum on the right-hand side of (2.10) does not exceed  $f(x^*) - f(\bar{x}_j(\omega^{j-1})) + v_0 L$ , i.e. (2.10) gives

$$\begin{aligned} \mathbf{M}_{F_{\omega_j}} V_*(\varphi_j(\omega^j)) &\leq V_*(\varphi_{j-1}(\omega^{j-1})) \\ &\quad + \rho_j v_0 L + \rho_j (f(x^*) - f(\bar{x}_j(\omega^{j-1}))) + \rho_j L \omega_{V,r}(\rho_j L). \end{aligned} \quad (2.11)$$

Averaging both sides of (2.11) over  $\omega^{j-1}$ , we obtain

$$\begin{aligned} \mathbf{M}_{F_{\omega_j}} V_*(\varphi_j(\omega^j)) &\leq \mathbf{M}_{F_{\omega^{j-1}}} V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j v_0 L \\ &\quad + \rho_j (f(x^*) - \mathbf{M}_{F_{\omega^{j-1}}} f(\bar{x}_j(\omega^{j-1}))) + \rho_j L \omega_{V,r}(\rho_j L). \end{aligned} \quad (2.12)$$

Summing (2.12) over  $j = 1, 2, \dots, i$ , and noting that  $V_*(\varphi_0(\omega^0)) = 0$ , we find

$$\begin{aligned} \sum_{j=1}^i \rho_j \{ \mathbf{M}_{F_{\omega^{j-1}}} f(\bar{x}_j(\omega^{j-1})) \} - \left( \sum_{j=1}^i \rho_j \right) f(x^*) &\leq \left( \sum_{j=1}^i \rho_j \right) v_0 L \\ &\quad + \sum_{j=1}^i \rho_j L \omega_{V,r}(\rho_j L) - \mathbf{M}_{F_{\omega^i}} V_*(\varphi_i(\omega^i)). \end{aligned} \quad (2.13)$$

Further, because  $f$  is convex, we have

$$\sum_{j=1}^i \rho_j f(\bar{x}_j(\omega^{j-1})) \geq \left( \sum_{j=1}^i \rho_j \right) f(x^i(\omega^{i-1})),$$

and by the properties of  $V$

$$V_*(\varphi) = V(\varphi) - \langle \varphi | x^* \rangle \geq V(\varphi) - \|\varphi\|_* \geq -\frac{1}{2}$$

(we have used the fact that  $\|x^*\| \leq 1$  as a result of the normalization conditions

on  $G$ ). Using these inequalities in (2.13), we obtain

$$\begin{aligned} & \left( \sum_{j=1}^i \rho_j \right) (\mathbf{M}_{\mathbf{F}_{\omega^{i-1}}} (f(x^i(\omega^{i-1})) - f(x^*))) \\ & \leq \left( \sum_{j=1}^i \rho_j \right) v_0 L + \frac{1}{2} + \sum_{j=1}^i \rho_j L \omega_{V,r}(\rho_j, L), \end{aligned} \quad (2.14)$$

and this gives the required inequality (2.5). The theorem has been proved.

### 5.2.2

Let us discuss the result obtained.

#### 5.2.2.1

First of all we remark that when  $E, \|\cdot\|$  is a Hilbert space, for  $t \geq 2$ , the method described, corresponding to  $V(\cdot) = V_2(\cdot)$  (see Section 3.2.3), can be somewhat simplified (similarly to what was done in Section 3.3). After this simplification, the method assumes the following form ( $\varphi_{t-1} = x_i = \bar{x}_i$ ):

$$x_1 = 0; \quad x_{i+1} = \pi_G(x_i - \rho_i \psi_0^1(x_i, f, \omega_i)), \quad (2.15)$$

$$x^i = \left( \sum_{j=1}^i \rho_j \right)^{-1} \sum_{j=1}^i \rho_j x_j. \quad (2.16)$$

The bound (2.5) can then be replaced by

$$\bar{v}_i \leq v_0 + \frac{1}{2L} \sum_{j=1}^i \frac{\rho_j^2}{\sum_{j=1}^i \rho_j} + \frac{8}{\sum_{j=1}^i \rho_j}. \quad (2.17)$$

All these facts are proved in the same way as in the general case; in the proof the variation of the function  $\bar{V}_*(x) = \frac{1}{2}(x - x^*)^2$  has to be considered.

*Exercise 1.* Prove (2.17).

We point out that the equations (2.15) describe the well-known process of stochastic approximation due to Yu. M. Ermol'ev. Thus, in the case of a Hilbert space, the proposed method differs from the customary method only in the rule for forming the result (2.16). However, this difference is very substantial from the point of view of the analysis of the method. For the trajectory itself (2.16) convergence (in some probabilistic sense or other), and estimates of the (2.17) type, have been proved only (so far as the authors are aware) under additional assumptions about the displacements  $\rho_i$  and the functions  $f$ . We see that a small (and, from the practical point of view, costing

nothing) change in the method leads to a substantial simplification of the situation.

#### 5.2.2.2

Let us discuss the question of a sensible choice of the displacements  $\rho_i$ . Suppose the function  $\omega_{V,r}(t)$  admits a bound,  $\omega_{V,r}(t) \leq ct^\sigma$  with  $\sigma \leq 1$ . We shall be concerned with qualitative results, and so we take  $L = 1$ . We restrict our attention to rules for choice of displacements of the form  $\rho_i \sim i^{-x}$ ,  $0 < x \leq 1$ . It is easy to deduce from (2.5) that we should choose  $x = 1/(1+\sigma)$ . The right-hand side of (2.5) will then decrease like  $v_0 + i^{-\sigma/(1+\sigma)} \ln i$  as  $i$  increases. In the simplest case—a Hilbert space—when  $r \geq 2$ , we obtain a rule for choice of the displacements which is of the form  $\rho_i \sim 1/\sqrt{i}$ . The right-hand side of (2.5) then behaves like  $v_0 + O(\ln i/\sqrt{i})$ . We observe that the rule for choice of displacements can be made more precise, if we relinquish choosing displacements of the form  $O(i^{-x})$ ; for example, if we take

$$\rho_i \sim i^{-\frac{1}{1+\sigma}} (\ln i)^{-\frac{1}{1+\sigma}},$$

then the bound for the right-hand side of (2.5) is of the form

$$v_0 + O(i^{-\frac{2}{1+\sigma}} (\ln i)^{\frac{1}{1+\sigma}} \ln \ln i).$$

We shall not proceed further in this direction, in view of the fact that, a little later, we shall be considering ‘constructive’, i.e. finite-step, versions of the method described, in which similar complications regarding the choice of step do not arise.

#### 5.2.2.3

We shall say a few more words about the recommended rule for choice of the step. We restrict ourselves to the simplest case:  $E, \|\cdot\|$  is a Hilbert space,  $r = 2$ ,  $v_0 = 0$ . In this case we recommend choosing steps  $\rho_i = O(1/\sqrt{i})$ , whereas the ‘traditional’ rules for choice of step indicate  $\rho_i = O(1/i)$ . From our point of view the second choice is extremely bad; for it, the right-hand side of (2.5) behaves like  $O(1/\ln i)$ , i.e. it converges extremely slowly to 0 when  $i \rightarrow \infty$ .

The origin of the rule  $\rho_i = O(1/i)$  can be explained in the following way. Roughly speaking, the greater the displacement  $\rho_i$ , the quicker the method would converge when there is exact information, at any rate, initially. On the other hand, larger displacements lead to a strong influence of the noise on the trajectory, and so it is desirable to make small displacements. Traditionally, a compromise is effected between these two tendencies on the assumption that the function  $f$  is ‘good’, i.e. smooth and convex. From the point of view of the

limiting behaviour of the trajectories (traditionally, of the random index of the work of the method), such an assumption about  $f$  is justified to a sufficient extent. Moreover, with such an approach to the matter the choice  $\rho_i = O(1/i)$  is indeed sensible, for here precisely the noise does determine the rate of convergence.

Let us consider a model example: the minimization of a function on the axis,  $f_a(x) = \frac{1}{2}(x - a)^2$ . It is assumed that at the  $i$ th moment of time an observation  $a_i = a + \omega_i$  of the quantity  $a$  is accessible, where the noises  $\omega_i$  are scalars, independent from one step to the next, and say, normal with the mean = 0 and the dispersion = 1. From the point of view under discussion the oracle communicates realizations of the random (not biased) estimate

$$\psi_0^1(x, f_a, \omega) = x - a - \omega,$$

where  $\omega \in \mathbb{R}$  is distributed in the way indicated earlier. The process (2.15) (we suppose, in the form of exclusion of  $G = \mathbb{R}$ ) has in this case the form

$$x_{i+1} = x_i - \rho_i(x_i - a - \omega_i),$$

or, putting  $\Delta_i = x_i - a$ ,

$$\Delta_{i+1} = \Delta_i - \rho_i \Delta_i + \rho_i \omega_i = (1 - \rho_i) \Delta_i + \rho_i \omega_i.$$

Hence, under the condition  $\rho_1 = 1$  (which causes no loss of generality), it follows that  $\Delta_i$ ,  $i \geq 2$ , is a combination of the random variables  $\omega_1, \dots, \omega_i$  with the sum of the coefficients equal to 1, such that the minimal dispersion of  $\Delta_i$  (i.e. the mean value of  $2(f_a(x_i) - \min_x f_a(x))$ ) corresponds to the case where all the coefficients of this combination are equal to one another. The latter, in turn, is equivalent to the condition  $\rho_i = 1/i$ ,  $i \geq 1$ .

Thus the choice  $\rho_i = 1/i$  is indeed optimal for the example considered. This same example can, with a well-known justification, be regarded as a suitable model for the 'remote-from-the-origin' part of the trajectory of the method of stochastic approximation in the general case too. At the same time this rule proves to be unacceptably bad on the initial parts of the trajectory if the function  $f$  is not assumed to be smooth and strictly convex.

For, let  $f_\epsilon$  be a convex function on the axis, whose minimum is attained at  $x = 0$ , and let  $f_\epsilon(x) = 4\epsilon x$  when  $x > 0$ . We assume that there is no noise, and we start the motion from the point  $x_1 = 1$ , and the displacements are of the form  $\rho_i = 1/i$ . After  $i$  steps we arrive at the point

$$x_i = 1 - 4\epsilon \sum_{j=1}^{i-1} \rho_j \geq 1 - 4\epsilon \ln i.$$

To ensure an (absolute) accuracy  $\epsilon$  requires a hit on a point  $x$  with  $|x| \leq \frac{1}{4}$ , i.e. the number of steps  $N_\epsilon$  needed to do this satisfies the estimate

$$4\epsilon \ln N_\epsilon \geq \frac{3}{4}, \text{ i.e. } N_\epsilon \geq \exp(3/(16\epsilon)).$$

For small  $\epsilon$  the number of steps obtained is fantastically large. The point here is that in the model example considered above, at the expense of the choice  $\rho_1 = 1$ , we immediately obtain a point  $x$  with  $Mx_1 = a$  (in the case of exact information, an immediate hit at the point of the minimum of  $f$ ), and all the subsequent behaviour of the trajectory represents the oscillations about this point due to the noise. There is nothing similar in the general case to these effects in the quadratic situation, as the second example shows.

The rule recommended here for choice of the displacements  $\rho_i = 1/\sqrt{i}$  ensures an incomparably quicker (than the 'classical' one) emergence of the trajectory  $x_i$  at the optimum point (for the function  $f_\epsilon$  described above we achieve it after a time  $O(1/\epsilon^2)$  instead of  $\exp(O(1/\epsilon))$ ). On the other hand, this method is very 'sensitive' to the noise; because of it, the trajectory  $x_i$  oscillates strongly about the minimum. Here the second peculiarity of the proposed method comes into action, a feature which we have not yet mentioned; this is the fact that we average this oscillating trajectory (we go from  $x_i$  to  $x^i$  in accordance with (2.16)), and we consider precisely this trajectory of the means as the resulting trajectory. The averaging, which is carried out not during the course of the search (roughly speaking, this is just what steps  $\rho_i = 1/i$  do) but independently of it, makes it possible, simultaneously and acceptably quickly, to get into the neighbourhood of the optimum, independently of whether the function being minimized is smooth or not, and to smooth out the fluctuations of the trajectory arising from the presence of noise.

### 5.2.3

We now pass on to 'constructive', finite-step versions of the method, supposing that the (relative) accuracy  $v$ ,  $v_0 < v < 1$ , demanded of the solution is specified in advance.

In this situation we can consider the  $\widetilde{MD}_v(v)$ -method, which consists in constructing the initial fragment of length  $N$  of the trajectory (2.1)–(2.3) and putting out the  $N$ th point of the trajectory (2.4) as the result. It is necessary only to choose the displacements  $\rho_i$ ,  $1 \leq i \leq N$ , and the number  $N$  itself in the appropriate way. From the point of view of optimizing the laboriousness, all the  $\rho_i$  should be chosen equal to one another in accordance with the relation

$$\rho_i = \frac{1}{L} \rho_{v,r} (v - v_0),$$

where

$$\rho_{v,r}(v) = \gamma_{v,r} \left( \frac{v}{2} \right). \quad (2.18)$$

The number of steps should be chosen according to the formula

$$N = N_{v,r}(v - v_0),$$

where

$$N_{V,r}(v) = \left\lceil \frac{1}{v\rho_{V,r}(v)} \right\rceil. \quad (2.19)$$

With such a choice of the parameters  $\rho$  and  $N$ , the right-hand side of (2.5) when  $i = N$  does not exceed  $v$ . Thus the constructed method ensures an accuracy  $v$  in the solution of problems of the class  $\mathcal{C}$  with a laboriousness  $N_{V,r}(v - v_0)$  (and not  $N_{V,r}(\cdot) + 1$ ; the point being that the construction of  $x^N$  requires not  $N$  but  $N - 1$  questions to be put concerning the problem under solution).

In the Hilbert-space case when  $r = 2$  we can use (2.15)–(2.17) instead of (2.1)–(2.4) with the choice of parameters

$$\rho_i = \frac{4}{L(v - v_0)}, \quad N = \left\lceil \frac{1}{4(v - v_0)^2} \right\rceil.$$

#### 5.2.4

Let us see what bounds for laboriousness the  $\widetilde{MD}_V(v)$ -method provides. Let  $\omega_{V,r}(t) \leq ct^\sigma$ , as in Section 5.2.2. Then

$$N_{V,r}(v) = O\left(\left(\frac{1}{v}\right)^{\frac{\sigma}{1+\sigma}}\right),$$

i.e. in setting up the method for an accuracy specified in advance, the latter can be ensured rather more quickly (after

$$O\left(\left(\frac{1}{v - v_0}\right)^{\frac{\sigma}{1+\sigma}}\right)$$

steps instead of after

$$O\left(\left(\frac{1}{v - v_0}\right)^{\frac{\sigma}{1+\sigma}} \ln\left(\frac{1}{v - v_0}\right)\right)$$

steps) than with the rational choice for  $\rho_i$  in the ‘infinite-step’ version in Section 5.2.1.

We enumerate the bounds corresponding to  $\widetilde{MD}^{p,r}$ -methods associated with functions  $V_p(\cdot)$  on the spaces  $(E, \|\cdot\|) = (L_p(T, \mu), \|\cdot\|_p)$ . The bounds given below for the laboriousness follow from (2.18)–(2.19) and the bounds for  $\gamma_{V_p,r}(\cdot)$  in (1.6)–(1.7).

First let  $1 < p < \infty$ . Then the laboriousness of  $\widetilde{MD}^{p,r}(v)$  when  $v_0 < v < 1$  does not exceed the value  $M_{p,r}(v - v_0)$ , where

$$M_{p,r}(v) = D(p, r) \left(\frac{1}{v}\right)^{\frac{r}{p-1}}, \quad (2.20)$$

$$\bar{r} = \min \left\{ 2, r, \frac{p}{(p-1)} \right\}.$$

But if  $p = 1$  and if  $(E, \|\cdot\|) = (l^{(n)}, \|\cdot\|_1)$ ,  $1 < n < \infty$ , then the laboriousness of the  $\widetilde{MD}^{1,r}$  associated with  $V = V_{1,n}(\cdot)$ , when  $v_0 < v < 1$  does not exceed  $M_{1,n,r}(v - v_0)$ , where

$$M_{1,n,r}(v) = D(1, r) \left(\frac{1}{v}\right)^{\frac{r}{r+1}} \ln n, \quad \bar{r} = \min \{2, r\}. \quad (2.21)$$

In (2.20)–(2.21),  $D(p, r)$  is a finite function of both arguments for all  $r > 1$ .

We point out that, when  $r \geq 2$  (in fact, even when  $r \geq \min\{2, p/(p-1)\}$ ), the bound (2.20) is the same as regards the character of its dependence on  $v$  (and the bound (2.21) is the same as regards the character of its dependence on  $v$  and  $n$ ) as the corresponding bounds for laboriousness of MD-methods of solving Lipschitz-convex problems with an exact oracle of the first order (see Section 3.3.4). Thus, in the most typical case  $r = 2$ , the laboriousness of the stochastic-programming methods presented is, roughly speaking, the same as for MD-methods in ordinary convex programming. In other words, the MD-methods scarcely react to noise in the presented information.

### 5.3 THE COMPLEXITY OF STOCHASTIC PROGRAMMING PROBLEMS

In this section lower bounds for the complexity of the classes of stochastic programming problems introduced in Section 5.1 are constructed. These estimates, on the one hand, enable us to judge to what extent the methods constructed are efficient, and, on the other hand, they indicate what should be strived for in the general case when there are constraints and when values and their derivatives are observed with noise present.

Turning to the construction of lower bounds for the complexity of the classes with which we are concerned, we point out the following fact. In chapter 4 bounds were constructed, effective for any (local deterministic or even any local) oracles. In considering stochastic oracles, it is impossible to obtain such universal results. For, the oracle considered may turn out by chance to be an exact, deterministic oracle, and so a ‘universal’ *lower* bound for the complexity would necessarily have to ignore the fact of the oracle’s being stochastic. Accordingly, we shall, roughly speaking, concern ourselves with the *upper* bound for the complexity functions of a given field of problems relative to all possible ways of equipping this field with oracles of this type.

#### 5.3.1

We start by considering problems without constraints. Here the following simple result holds.

**Theorem.** There is a function  $c(r) > 0$  having the following property. For any  $r > 1$ , any Banach space  $E$ ,  $\|\cdot\|$ , and any bounded, convex, closed set  $G \subseteq E$

which does not reduce to a point, there is a natural (in the sense of Section 5.1) oracle  $\mathcal{O}$  for the field of problems  $C_{\text{Lip}}(G, E, \|\cdot\|, 0)$  such that the stochastic complexity  $\tilde{N}(v)$  of the class of problems  $\tilde{\mathcal{C}} = \tilde{\mathcal{C}}_{r,L}^{e,0}(G, E, \|\cdot\|, 0)$  satisfies the estimate

$$\tilde{N}(v) \geq c(r) \left(\frac{1}{v}\right)^{\frac{r}{r-1}}$$

where

$$\bar{r} = \min\{2, r\}. \quad (3.1)$$

(Note: This bound already obtains on the sub-class of the simplest linear problems.)

*Proof:* Without loss of generality we can suppose that  $L = 1$ . We can further suppose that  $\rho_{\|\cdot\|}(G) = 1$ , and that  $G$  contains a pair of points  $h, -h$  with  $\|h\| > \frac{1}{4}$ . Then  $G$  contains 0, and is contained in a ball of radius 2 with centre at 0. Let  $\varphi \in E^*$  be such that  $\langle \varphi | h \rangle = \|h\|$  and  $\|\varphi\|_* = 1$ . For all  $t$  with  $|t| < \frac{1}{4}$ , we take  $f_t$  to be the problem generated by the function  $f_t(x) = t \langle \varphi | x \rangle$ .

1°. First let  $r \leq 2$ . We define the oracle  $\mathcal{O}$  as follows. On all problems distinct from problems of the family  $\{f_t\}_{|t| \leq \frac{1}{4}}$ ,  $\mathcal{O}$  is an exact, first-order oracle. The answer of  $\mathcal{O}$  to a question about the problem  $f_t$  at a point  $x \in G$  is a random variable with values in  $R \times E^*$  which takes the value  $(0, 0)$  with probability  $1 - p(r, t)$ , and the value  $(p(r, t))^{-1} (f_t(x), f'_t(x))$  with probability  $p(r, t)$ . Here

$$p(r, t) = (4|t|)^{\frac{r}{r-1}}. \quad (3.2)$$

Direct calculation shows that, when  $|t| \leq \frac{1}{4}$ ,  $f_t \in \tilde{\mathcal{C}} \equiv \tilde{\mathcal{C}}_{r,1}^{e,0}(G, E, \|\cdot\|, 0)$ . Now let  $\mathcal{B}_t$  be any method of solving problems of the class  $\tilde{\mathcal{C}}$  with (means) accuracy  $v$  and laboriousness  $M$  on this class. We consider a subclass  $\tilde{\mathcal{C}}_t$  of the class  $\tilde{\mathcal{C}}$  formed by the pair of problems  $f_{+|t|}$  and  $f_{-|t|}$ ,  $|t| \leq \frac{1}{4}$ , and we equip  $\tilde{\mathcal{C}}_t$  with a probability distribution. By Theorem 1.5.2 there is a *deterministic* method  $\mathcal{B}_t$  of solving problems of the class  $\tilde{\mathcal{C}}$  which is such that

$$l(\mathcal{B}_t, f_{+|t|}) + l(\mathcal{B}_t, f_{-|t|}) \leq 4M$$

and

$$v(\mathcal{B}_t, f_{+|t|}) + v(\mathcal{B}_t, f_{-|t|}) \leq 4v. \quad (3.3)$$

Let us examine what the method  $\mathcal{B}_t$  can do in the case where, in the first  $]4M[ = M_1$  steps of its work it always receives zero information. We observe that, in solving each of the problems  $f_{+|t|}$  and  $f_{-|t|}$  this receipt of zero information occurs with probability  $p_M = (1 - p(r, t))^{M_1}$ .

It is possible that, with such information,  $\mathcal{B}_t$  puts out the result  $\bar{x}$  not later than the  $M_1$ th step. It is clear that, either for  $f = f_{+|t|}$  or for  $f = f_{-|t|}$ , we have  $v(\bar{x}, f) > |t|/4$  (it is important to note that, by definition,  $\bar{x}$  does not depend on which particular problem is being solved). By the second inequality in (3.3) we

obtain  $|t| p_M / 4 < 4v$ , i.e.

$$p_M < 16v/|t|. \quad (3.4)$$

It is also possible that  $\mathcal{B}_t$  does not put out the result in the first  $M_1$  steps of the work of the method, when zero information is received all the time. Then both  $l(\mathcal{B}_t, f_{+|t|})$  and  $l(\mathcal{B}_t, f_{-|t|})$  are greater than  $p_M M_1$ , i.e. from the first inequality in (3.3) we have  $p_M |4M| < 2M$ , or  $p_M < \frac{1}{2}$ . We now choose  $t = 32v$ , and we shall suppose that  $v \leq 1/256$ . Then it follows from the above argument that

$$p_M \equiv (1 - p(r, t))^{14M} < \frac{1}{2},$$

which immediately gives

$$]4M[ > \frac{c}{p(r, t)},$$

where  $c > 0$  is an absolute constant. Recalling the definition of  $p(r, t)$  and using the fact that  $M$  is always  $\geq 1$ , we obtain the bound

$$M \geq c(r) \left(\frac{1}{v}\right)^{\frac{r}{r-1}},$$

in which  $c > 0$  depends only on  $r$ . Since  $\mathcal{B}_t$  was an arbitrary method of solving problems in  $\tilde{\mathcal{C}}$  with accuracy  $v$ , we have

$$\tilde{N}(v) \geq c(r) \left(\frac{1}{v}\right)^{\frac{r}{r-1}-1},$$

as required.

2°. Now let  $r \geq 2$ . We change the oracle  $\mathcal{O}$  constructed above, on problems  $f_t$  with  $|t| \leq \frac{1}{4}$ , in the following way. The answer of  $\mathcal{O}$  to a question about  $f_t$  at a point  $x \in G$  is, with probability  $\frac{1}{2} + 2t$ , the answer of the exact, first-order oracle to a question about  $f_{1/4}$  at the point  $x$ , or, with complementary probability  $\frac{1}{2} - 2t$ , it is the answer of the same oracle to a question about the problem  $f_{-1/4}$  at  $x$ . It is easily seen that, with this definition of  $\mathcal{O}$ , for all  $t$  with  $|t| \leq \frac{1}{4}$ , we have  $f_t \in \tilde{\mathcal{C}} \equiv \tilde{\mathcal{C}}_{r,1}^{e,0}(G, E, \|\cdot\|, 0)$ .

Now let the objects  $\mathcal{B}_t, v, M, \tilde{\mathcal{C}}_t, \mathcal{B}_t$  be defined as in 1°, let  $M_1 = ]16M[$ , let  $\Omega$  be the space of oracle noises in the first  $M_1$  steps of the work of the method  $\mathcal{B}_t$ , and let  $\hat{\Omega} = \Omega \times \tilde{\mathcal{C}}_t = \{(\bar{\omega}, \kappa), \bar{\omega} \in \Omega, \kappa = \pm 1\}$ ;  $\hat{\Omega}$  is regarded as a product of probability spaces. Here  $\kappa = +1$  corresponds to the function  $f_{+|t|}$ , and  $\kappa = -1$  to the function  $f_{-|t|}$ . We define an ordered pair of random variables  $z$  and  $w$ —functions on the space  $\hat{\Omega}$ ; namely,  $z(\bar{\omega}, \kappa) = +1$  if the method  $\mathcal{B}_t$ , working on the problem  $\kappa$  when oracle noise  $\bar{\omega}$  is present puts out the result in the first  $M_1$  steps of its work and if this result  $\bar{x}$  is such that  $\langle \varphi | \bar{x} \rangle \leq 0$ , or if the method  $\mathcal{B}_t$ , under the same conditions, does not put out a result in the first  $M_1$  steps. In the remaining cases,  $z(\bar{\omega}, \kappa) = -1$ . Further, let  $w(\bar{\omega}, \kappa) = \kappa$ .

We assume that

$$v < 1/256 \text{ and } t = 64v. \quad (3.5)$$

We prove that  $z = w$  with a probability  $\geq \frac{3}{4}$ . For, if  $z(\bar{\omega}, \kappa) \neq \bar{\omega}, \kappa$ , then either the laboriousness of the corresponding  $\bar{\omega}, \kappa$ -realization of  $\mathcal{B}_t$  is greater than  $16M$ , or the error of its result regarded as a solution of the corresponding problem is greater than  $16v$  (this follows immediately from the choice of  $t$  and the definition of  $z$ ). By (3.3) the probabilities of each of the events mentioned does not exceed  $\frac{1}{8}$ , as required.

Thus,  $\Pr\{z = w\} \geq \frac{3}{4}$ . Moreover, it is clear that

$$\Pr\{w = 1\} = \Pr\{w = -1\} = \frac{1}{2}$$

(the probabilities are calculated from the distribution over  $\hat{\Omega}$ ). From these considerations it is easy to deduce that the Shannon mutual information  $I(z, w)$  of the random variables  $z$  and  $w$  is not too small:

$$I(z, w) \geq c_0 > 0, \quad (3.6)$$

where  $c_0$  is an absolute constant.

On the other hand, it is clear that the answers of the oracle  $\mathcal{O}$  about the problems  $f_{+|I|}$  and  $f_{-|I|}$  at any point  $x$  are obtained by a univalent transformation (depending on the parameter and on  $x$ , but not depending on which of the problems ( $f_{+|I|}$  or  $f_{-|I|}$ ) is concerned or on what noise is realized in the oracle) from the oracle's answers about the same problem at a point  $h$ , say. Therefore the information supplied by the oracle  $\mathcal{O}$  to the method  $\mathcal{B}_t$  is obtained by a univalent transformation (independent of whichever problem  $f \in \mathcal{C}_t$  is being solved) from the sequence of realizations of its answers (corresponding to the given series of oracle noises) to a question about  $f$  at the point  $h$ . By the same token, the random variable  $z$  is also a univalent function of the series  $\zeta$  of answers of the oracle to a question at the point  $h$  about the problem under solution. Since  $z$  is a univalent function of  $\zeta$ ,  $I(z, w) \leq I(\zeta, w)$ . The last quantity does not depend at all on  $\mathcal{B}_t$ , but is completely determined by the structure of the oracle  $\mathcal{O}$ ;  $I(\zeta, w)$  can be calculated explicitly. The calculation, which we omit, leads, under the conditions (3.5), to the bound

$$I(\zeta, w) \leq ((1 + cv^2)^M - 1),$$

with  $c > 0$  an absolute constant. Comparing this estimate with the inequality  $I(\zeta, w) \geq I(z, w)$ , we obtain, because of (3.6) and the definition of  $M_1$ , that, when (3.5) holds,

$$M \geq c_2/v^2, \quad (3.7)$$

where  $c_2 > 0$  is an absolute constant.

Hence, as in 1°, it follows that the bound to be proved, (3.1), is valid. The theorem has been proved.

### 5.3.2

The assertion of Theorem 5.3.1 enables us to delineate the domain of sub-optimality of an MD-method of solving the unconstrained stochastic-programming problems in Section 5.2. First of all we see that, if we are speaking about solving problems of the class  $\mathcal{C} = \mathcal{C}_{r,L}^{e,0}(G, E, \|\cdot\|, 0)$  on a regular space  $E$ ,  $\|\cdot\|$ , for which it is possible by a suitable choice of  $V(\cdot)$  to ensure that  $\omega_v(t) \leq c_v t$ , then, in principle, the laboriousness of  $\widetilde{MD}_v$ —for certain oracles  $\mathcal{O}$ —cannot be reduced by more than  $d_{v,r}$  times, where  $d_{v,r}$  is a constant depending only on  $V$  and  $r$ . For, the laboriousness of the method  $\widetilde{MD}_v$  constructed for an accuracy  $v$  in this case does not, by (2.18)–(2.19), exceed the value  $b_v(1/v)^{r/(r-1)}$ , where  $\bar{r} = \min\{r, 2\}$ ; and the lower bound (3.1) has the same order of dependence on  $v$ . The situation just mentioned holds, in particular, when  $(E, \|\cdot\|) = (L_p, \|\cdot\|_p)$  with  $1 < p \leq 2$ . In fact, in the latter case, because of (2.20), the laboriousness of  $\widetilde{MD}_v^{p,r}$  is not reduced by more than  $b(p, r)$  times also when  $p > 2$  but  $p/(p-1) \geq r$ .

We continue the analysis of the question of optimality of  $\widetilde{MD}_v^{p,r}$ -methods. We are, of course, taking  $(E, \|\cdot\|) = (L_p, \|\cdot\|_p)$ . We still have to analyze the case  $p/(p-1) < r$ , and also the case  $p = 1$ . If  $p/(p-1) < r$ , then, as noted in Section 5.2.4 (and as follows directly from (2.20)), the laboriousness of the  $\widetilde{MD}_v^{p,r}$  method—with accuracy up to a factor depending only on  $p$  and  $r$ —is the same as for a method, using an exact, first-order oracle, of solving Lipschitz-convex problems in Section 3.3. Therefore, with the  $p$  and  $r$  mentioned, the  $\widetilde{MD}_v^{p,r}$ -methods are automatically sub-optimal in those situations in which  $MD_p$ -methods (with an exact oracle) are sub-optimal, namely, for problems of high dimension in domains  $G$  of the type of  $L_p$ -balls.

Finally, if  $p = 1$  and if  $\dim E = n < \infty$ , then, as follows from a comparison of (2.21) and (3.1), if an  $\widetilde{MD}_v^{1,n,r}$ -method can be improved at all as regards laboriousness, it is improvable only by  $b(r) \ln(n+1)$  times.

We emphasize that in these discussions the clauses like 'the laboriousness of this or that method of solving problems of the class  $\mathcal{C}_{r,L}^{e,0}(G, E, \|\cdot\|, 0)$  is, in principle, reducible by not more than so-and-so number of times' has the following precise meaning. For the given  $G, E, \|\cdot\|, r, L$  a natural oracle  $\mathcal{O}$  can be found such that the stated assertion is true for the class of problems corresponding to  $\mathcal{O}$ .

### 5.3.3

We shall now construct a lower bound for the complexity for classes of stochastic programming problems with constraints.

**Theorem.** There is a function  $c(r)$  ( $r > 1$ ) and a number  $\bar{v}_0 > 0$  such that the following assertion holds. For any  $r > 1$ , any Banach space  $E$ ,  $\|\cdot\|$ , and any bounded, convex, closed set  $G \subset E$ , not reducing to a point, there is a

natural (in the sense of Section 5.1) oracle  $\mathcal{O}$  for the field of problems  $C_{\text{Lip}} = C_{\text{Lip}}(G, E, \|\cdot\|, m)$  such that the stochastic complexity  $\tilde{N}(v)$  of the class of problems  $\tilde{\mathcal{C}} = \tilde{\mathcal{C}}_{r,L}^{e,0}(G, E, \|\cdot\|, m)$  satisfies, when  $v < v_0$ , the bound

$$\tilde{N}(v) \geq c(r) \ln(m+2)(1/v)^{\bar{r}/(\bar{r}-1)}, \bar{r} = \min\{r, 2\}. \quad (3.8)$$

When  $m > 0$ , the bound (3.8) holds for the subclass of problems  $\tilde{\mathcal{C}}$  formed by all compatible problems  $f$  from  $\tilde{\mathcal{C}}$  with  $f_0 \equiv 0$ .

*Proof.* In view of Theorem 5.3.1 it suffices to prove the second assertion of this theorem. When  $m = 1$  it is proved by a simple modification of the construction used in proving Theorem 5.3.1. We shall not describe this modification, but leave the reader to effect it. Now let  $m > 1$ . Without loss of generality, we can suppose that  $m$  is even and equal to  $2k$ . Suppose, as in Section 5.3.1, that  $\rho_{\parallel 1}(G) = 1$ , that  $G$  contains the points  $h$  and  $-h$  with  $\|h\| > \frac{1}{4}$ , and that  $\varphi \in E^*$  is such that  $\|\varphi\|_* = 1$  and  $\langle \varphi | h \rangle = \|h\|$ . We construct an oracle  $\mathcal{O}$  in the following way. For a given  $t$  with  $0 < t < \frac{1}{4}$ , We consider  $m+1$  problems  $f^{t,s}$ ,  $0 \leq s \leq m$ , of the form

$$f_0^{t,s} \equiv 0, f_j^{t,s}(x) = \begin{cases} -t - \langle \varphi | x \rangle + 2\delta_{js}t, & 1 \leq j \leq k, \\ -t + \langle \varphi | x \rangle + 2\delta_{js}t, & k+1 \leq j \leq m, \end{cases}$$

where

$$\delta_{js} = \begin{cases} 0, & j \neq s, \\ 1, & j = s \end{cases}$$

is the Kronecker symbol. Let  $\bar{h} = h/\|h\|$ .

It is clear that  $f^{t,s} \in C_{\text{Lip}}(G, E, \|\cdot\|, m)$ , and when  $s > 0$  the problem  $f^{t,s}$  has admissible feasible points, namely, when  $s \leq k$  the points  $\{x \in G | \langle \varphi | x \rangle = t\}$  are admissible feasible points of  $f^{t,s}$ , and when  $s > k$  so are the points  $\{x \in G | \langle \varphi | x \rangle = -t\}$ .

Now let the oracle  $\mathcal{O}$  on all problems except problems of the family  $\{f^{t,s}\}_{1 \leq s \leq m, 0 < t < 1/4}$  be an exact oracle of the first order, and let  $\mathcal{O}$ 's answers to questions about a problem  $f^{t,s}$  be constructed in the following way.  $\mathcal{O}$  communicates exactly the support functionals to the components  $f_j(x)$  of the problem at  $x$ . Moreover,  $\mathcal{O}$ 's answer about the value of the vector  $f^{t,s}(x)$ ,  $s \geq 1$ , is of the form  $f^{t,0}(x) + \xi^{s,t}$ , where the random vector  $\xi^{s,t}$  is distributed in the following way:

- (i) with probability  $1 - p(t, r)$ ,  $\xi^{s,t} = 0$ ;
- (ii) the conditional distribution of the vector  $\xi^{s,t}$  with the condition  $\xi^{s,t} \neq 0$  is a distribution with independent co-ordinates; the zeroth component takes the value 0, and the  $j$ th component,  $1 \leq j \leq m$ , takes the value  $a(t, r)$  with probability  $\frac{1}{2} + \delta_{js}\Delta(t, r)$ , and takes the value  $(-a(t, r))$  with the

probability  $\frac{1}{2} - \delta_{js}\Delta(t, r)$ . Here

$$\left. \begin{aligned} p(t, r) &= \begin{cases} (2t)^{\frac{r}{r-1}}, & 1 < r < 2, \\ 1, & r \geq 2; \end{cases} \\ a(t, r) &= \begin{cases} (1/2t)^{\frac{1}{r-1}}, & 1 < r < 2, \\ 1, & r \geq 2; \end{cases} \\ \Delta(t, r) &= \begin{cases} \frac{1}{2}, & 1 < r < 2, \\ t, & r \geq 2. \end{cases} \end{aligned} \right\} \quad (3.9)$$

Direct verification shows that, with this definition of  $\mathcal{C}$ , the problems  $f^{t,s}$ ,  $0 < t < \frac{1}{4}$ ,  $s = 1, \dots, m$ , fall into the class  $\tilde{\mathcal{C}}_{r,L}^{e,0}(G, E, \|\cdot\|, m)$  with  $L = 2$ . Without loss of generality, we may suppose that it is precisely with this class that the assertion of the theorem is concerned. Let  $\tilde{\mathcal{C}}$  be the corresponding subclass. Then  $f^{t,s} \in \tilde{\mathcal{C}}$ ,  $0 < t < \frac{1}{4}$ ,  $1 \leq s \leq m$ . Now let  $v_0 = 10^{-2}$ ,  $v < v_0$ , and let  $\mathcal{B}$  be a method of solving problems of the class  $\tilde{\mathcal{C}}$  with accuracy  $v$ , and with laboriousness  $M$  on this class. We consider the subclass  $\tilde{\mathcal{C}}_t$  of the class  $\tilde{\mathcal{C}}$  formed by the  $m$  problems  $f^{t,s}$ ,  $1 \leq s \leq m$ , and we equip it with an equiprobable distribution of probabilities  $p(s)$  (a problem  $f^{t,s} \in \tilde{\mathcal{C}}_t$  is identified with its 'index'  $s$ ). As in the proof of Theorem 5.3.1, there is a deterministic method  $\mathcal{B}_t$  for which

$$\int_{\tilde{\mathcal{C}}_t} l(\mathcal{B}_t, f^{t,s}) dp(s) \leq 2M \text{ and } \int_{\tilde{\mathcal{C}}_t} v(\mathcal{B}_t, f^{t,s}) dp(s) \leq 2v. \quad (3.10)$$

Let  $\bar{\Omega} = \{\bar{\omega}\}$  be the space of sets of oracle noises on the first  $\lceil 16M \rceil$  steps of the method, and let  $\hat{\Omega} = \bar{\Omega} \times \tilde{\mathcal{C}}_t$  (the right-hand side is a product of probability spaces). We consider two random variables  $z, w$ —functions on  $\hat{\Omega}$  which are defined in the following way.  $z(\bar{\omega}, s) = +1$  if the method  $\mathcal{B}_t$ , in solving a problem  $f^{t,s}$  with a set of oracle noises  $\bar{\omega}$  on the first  $16M$  steps, does not put out the result in the first  $16M$  steps or if it puts out the result  $\bar{x}$  with  $\langle \varphi | \bar{x} \rangle > 0$ ; otherwise,  $z(\bar{\omega}, s) = -1$ . Further,

$$w(\bar{\omega}, s) = \begin{cases} +1, & 1 \leq s \leq k, \\ -1, & k+1 \leq s \leq m. \end{cases}$$

Let  $t = 32v$ . Then, as in part 2° of the proof of Theorem 5.3.1, it follows from (3.10) that  $\Pr\{z = w\} \geq \frac{3}{4}$ , and consequently the mutual information  $I(z, w)$  of the random variables satisfies the estimate

$$I(z, w) \geq c_0 > 0, \quad (3.11)$$

where  $c_0$  is an absolute constant. On the other hand, it is clear that  $z$  is a univalent function of the  $\lceil 16M \rceil$  realizations of the random variable  $\xi^{s,t}$  (i.e. for some function  $\varphi_t(\cdot)$  we have  $z(\bar{\omega}, s) = \varphi_t(\xi^t(\bar{\omega}, s))$ , where  $\xi^t(\bar{\omega}, s)$  is a series

of  $16]M$  successive realizations of the variable  $\xi^{s,t}$  corresponding to the noise  $\bar{\omega}$ ). Therefore  $I(\xi^t, w) \geq I(z, w) \geq c_0$ . But the quantity  $I(\xi^t, w)$  does not depend on the method  $\mathcal{B}$ , but is determined only by the structure of the oracle  $\mathcal{O}$ . Its estimation, which we leave to the reader, gives

$$I(\xi^t, w) \leq \frac{c}{2m} \{ (1 + 4\Delta^2(t, r) p(t, r))^{16]M} - 1 \},$$

with  $c$  an absolute constant. The inequality  $I(\xi^t, w) \geq c_0$  together with this estimate and the relations (3.9) immediately leads to (3.8). The theorem has been proved.

### 5.3.4

We see that the complexity of classes of stochastic programming problems cannot fail to depend on the number of constraints, although this dependence is quite weak. We shall prove in the next chapter that, under quite general hypotheses, the *lower bound* obtained for the complexity is actually exact. We further observe that the lower bounds for the complexity provided by Theorems 5.3.1 and 5.3.3 already obtain on the simplest (linear) problems, and so no restriction of the classes considered, achieved at the cost of increasing the demands regarding the smoothness of the problems constituting the classes, will lead, in the stochastic case, to a substantial reduction of the laboriousness (in the case of a deterministic oracle, this is not true; see Chapter 7). But if we impose on the problems of the class, not only smoothness restrictions, but restrictions of such a type as strong convexity (taking  $E, \|\cdot\|$  to be a Hilbert space), then, for unconstrained problems, the laboriousness can sometimes be reduced, but it is impossible to do so for problems with constraints ( $m \geq 2$ ). The second statement follows immediately from a slight modification of the proof of Theorem 5.3.3 (the addition of a quadratic term  $x^2/20$  to the formulae for  $f_j^{t,s}$  changes nothing essential in the derivation of (3.8)). The possibility of reducing the complexity of stochastic programming *without constraints* at the expense of assumptions about the smoothness and strong convexity of the function to be optimized is well-known (see, for example, [9]). This is illustrated by the result formulated in the following exercise.

*Exercise.* Let  $E, \|\cdot\|$  be a Hilbert space, let  $G \subset E$  be a convex, closed, bounded body, let  $\rho_{\|\cdot\|}(G) = 1$ , and let  $\mathcal{C}_{c_1, c_2}$  be a subclass of the class  $\mathcal{C}_{2,1}^{c,0}(G, E, \|\cdot\|, 0)$  formed by functions  $f$  such that, if  $x^*$  is the point where  $f$  has a minimum on  $G$ , then, for  $y \in G$ ,

$$f(x^*) + c_2 \|y - x^*\|^2 \geq f(y) \geq f(x^*) + c_1 \|y - x^*\|^2.$$

Prove that the method (2.15)–(2.16), with a suitable choice of the number of

steps and of the displacements  $\rho_i$ , is capable of solving problems in  $\mathcal{C}_{c_1, c_2}$  with a (mean) accuracy  $v_0$  and with a laboriousness  $\leq a(c_1, c_2)/v$  (instead of  $O(1/v^2)$ ) in the general case. Take the point  $x_i$  after  $i$  steps to be the result of the method. By considering the simplest quadratic problem,  $(x - a)^2 \rightarrow \min$ , verify that the laboriousness  $O(1/v)$  is the best that can be expected in the situation described (irrespective of which method is used for the solution).

# Solution of convex-concave games and constrained stochastic programming problems

In this chapter the application of MD-methods to solving convex-concave games, observed both exactly and with noise, is described. We show how constrained stochastic-programming problems can be reduced to problems of solving games. Methods of solving constrained stochastic problems are constructed. The results of the previous chapter enable us to prove the sub-optimality of the methods mentioned. The following treatment is based on the authors' papers [23], [24].

## 6.1 CLASSES OF CONVEX-CONCAVE GAME PROBLEMS

In this section we describe the classes of problems we shall be dealing with throughout the chapter. We shall be concerned with problems of approximating the solutions of convex-concave games, or what is the same thing, of saddle-points of convex-concave functions.

### 6.1.1

Let  $E$ ,  $\|\cdot\|$  and  $E_l$ ,  $\|\cdot\|_l$  be regular spaces satisfying the conditions formulated at the beginning of Section 5.1, and let  $G \subset E$  and  $G_l \subset E_l$  be bounded, convex, closed, non-void sets (more general situations will not arise in our discussions). Let  $F(x, l): G \times G_l \rightarrow \mathbb{R}$  be a continuous function. The function  $F$  is said to be *convex-concave* if, for each  $l \in G_l$ ,  $F$  regarded as a function of  $x \in G$  is convex on  $G$ , and if, for each  $x \in G$ ,  $F$  regarded as a function of  $l \in G_l$  is concave with respect to  $l \in G_l$ . A point  $(x^*, l^*) \in G \times G_l$  is called a *saddle-point* of  $F$  if at this point  $F$ , as a function of  $x$ , attains a minimum, and as a function of  $l$ , it attains a

maximum:

$$F(x, l^*) \geq F(x^*, l^*) \geq F(x^*, l) \quad \text{for all } (x, l) \in G \times G_l. \quad (1.1)$$

It is convenient to interpret the idea of a saddle-point in terms of the theory of games. We associate with the function  $F$  a 2-person, zero-sum *game*  $F$ , in which the first player's strategies are identified with points  $x \in G$ , and the second player's strategies with points  $l \in G_l$ , and the outcome of the play  $(x, l)$ , in which the first player chose the strategy  $x$  and the second the strategy  $l$ , consists in the payment by the first player to the second player of the sum  $F(x, l)$ . In such an interpretation of a saddle-point the function  $F$  is called the *solution of the game*  $F$ ; this is precisely an ordered pair of players' strategies which has the property that departure of either player from the corresponding strategy brings him no advantage if the other player continues to hold to his strategy.

The same thing can be formulated as follows. Let  $\bar{F}(x) = \sup_{l \in G_l} F(x, l)$  be the guaranteed pay-out by the first player, and  $F(l) = \inf_{x \in G} F(x, l)$  be the guaranteed income of the second. If the function  $F$  has a saddle-point, then

$$\inf_{x \in G} \bar{F}(x) = \min_{x \in G} \bar{F}(x) = \sup_{l \in G_l} F(l) = \max_{l \in G_l} F(l). \quad (1.2)$$

The set of saddle-points of  $F$  is  $X^*(F) \times L^*(F)$ , where

$$X^*(F) = \{x \in G | \bar{F}(x) = \inf_{y \in G} \bar{F}(y)\}$$

and

$$L^*(F) = \{l \in G_l | F(l) = \sup_{s \in G_l} F(s)\}. \quad (1.3)$$

Conversely, if (1.2) holds, then the function  $F$  has a saddle-point. The question arises whether a convex-concave, continuous function  $F$  always has a saddle-point. In our situation ( $E, E_l$  reflexive,  $G, G_l$  convex, closed, and bounded), this is always the case, by virtue of a fundamental theorem due to von Neumann [29].

### 6.1.2

Our purpose is to construct first-order methods for finding approximate solutions of convex-concave games. In order to talk about such methods, we need to fix the following objects:

- (i) the field of problems, i.e. the set of games, to be solved;
- (ii) a measure of the error of a point  $(x, l) \in G \times G_l$  regarded as an approximate solution of a game;
- (iii) the source of information, accessible to the method, about the game being solved, i.e. the oracle.

We shall describe these objects in the next sub-section.

**6.1.2.1**

The field of problems which we shall consider is the set of all games with convex-concave Lipschitz pay-off functions  $F: G \times G_I \rightarrow \mathbb{R}$ . (We shall later use the concepts ‘game with pay-off function  $F$ ’, ‘function  $F$ ’, ‘problem of finding a solution of a game with pay-off function  $F$ ’ as synonyms, and we simply speak of ‘a problem  $F$ ’.) We denote the set of such games with given  $G$  and  $G_I$  by

$$\mathcal{D} \equiv \mathcal{D}(G \times G_I, E, E_I, \|\cdot\|, \|\cdot\|_I).$$

**6.1.2.2**

The error of a point  $(x, l) \in G \times G_I$  regarded as an approximate solution of a game  $F \in \mathcal{D}$  is defined as follows. An exact solution of a game is an ordered pair of solutions of the extremal problems

$$\bar{F}(x) \rightarrow \min |x \in G \text{ and } \underline{F}(l) \rightarrow \max |l \in G_I. \quad (1.4)$$

It is natural to measure the error of a point  $(x, l) \in G \times G_I$  regarded as an approximate solution of a game  $F$  by half the sum of the errors of its components, each regarded as an approximate solution of its ‘own’ extremal problem, i.e. by the number

$$\varepsilon(x, l; F) = \frac{1}{2}\{\varepsilon(x, F) + \varepsilon_l(l, F)\},$$

where

$$\varepsilon(x, F) = \bar{F}(x) - \inf_G \bar{F}(x),$$

$$\varepsilon_l(l, F) = -\underline{F}(l) + \sup_{G_I} \underline{F}(l).$$

As well as these absolute measures of errors it is useful also to consider the relative errors

$$v(x, F) \equiv \frac{\varepsilon(x, F)}{r(F)}, \quad v_l(l, F) \equiv \frac{\varepsilon_l(l, F)}{r(F)},$$

$$v(x, l; F) = \frac{\varepsilon(x, l; F)}{r(F)} = \frac{1}{2}\{v(x, F) + v_l(l, F)\}.$$

In the case of a deterministic oracle (see below), it is convenient to take the normalizing factor  $r(F)$  to be

$$r(F) = 2 \max \{\mathbf{L}_{\|\cdot\|}(F) \rho_{\|\cdot\|}(G), \mathbf{L}_{\|\cdot\|_I}(F) \rho_{\|\cdot\|_I}(G_I)\},$$

where  $\mathbf{L}_{\|\cdot\|}(F)$  is the upper bound (over  $l \in G_I$ ) of the Lipschitz constants of the functions  $F(\cdot, l)$  obtained from  $F$  by fixing  $l$ , relative to  $\|\cdot\|$ , and  $\mathbf{L}_{\|\cdot\|_I}(F)$  is defined in the ‘symmetric’ way. It is clear that  $v(x, l; F) \geq 0$ , and

$$v(x, l; F) = 0 \Leftrightarrow \{(x, l) \text{ is a solution of the game } F\}.$$

With  $r(F)$  chosen as above,  $2r(F)$  is the natural upper bound for the variation of  $F$  on  $G \times G_I$ , and so in this case,

$$v(x, l; F) \leq 1 \quad \text{for all } (x, l) \in G \text{ and all } F \in \mathcal{D}$$

Thus  $v(x, l; F)$  has indeed the meaning of relative error.

**6.1.2.3**

We shall consider two types of oracle—deterministic and stochastic oracles of the first order. As in mathematical programming, a deterministic oracle is a particular case of a stochastic oracle. The latter, in the case of interest to us, is a set

$$\mathcal{O} = ((\Omega, \mathbf{F}_\omega); \psi_x(x, l; F), \psi_l(x, l; F)),$$

where  $(\Omega, \mathbf{F}_\omega)$  is a Polish space with a regular, Borel, Lebesgue-complete, probability measure;

$$\psi_x: (G \times G_I) \times \mathcal{D} \rightarrow E^* \text{ and } \psi_l: (G \times G_I) \times \mathcal{D} \rightarrow E_l^*$$

are the observation functions which associate to the ordered pair  $(x, l) \in G \times G_I$  and to the problem  $F \in \mathcal{D}$  the observations of the support functionals to  $F$  as a convex function of  $x$  and as a concave function of  $l$  respectively. These observations, for given  $x, l, F$ , are random variables—functions on the space of elementary events  $\Omega$ . The requirements that  $\psi_x, \psi_l$  be Borel with respect to  $x, l, \omega$  form part of the definition.

As in optimization methods, it is assumed that at the  $i$ th step the method can put a question to the oracle at any point (chosen by the method)  $(x_i, l_i) \in G \times G_I$ , and it obtains as the answers the values of the functions  $\psi_x(x_i, l_i, \omega_i)$  and  $\psi_l(x_i, l_i, \omega_i)$ , where  $\omega_i$  is the random noise occurring in the oracle at the  $i$ th step. The noises of distinct steps are assumed to be independent in the aggregate and to have the same distribution.

On the pattern of Sections 1.2–1.4 it is possible to define the concepts of: ‘a method of solving problems from the field  $\mathcal{D}$  using the oracle  $\mathcal{O}$ ’, ‘the trajectory of a method’, ‘the results of the work of a method’, ‘the mean laboriousness and error (relative and absolute)’. We shall not write out these definitions, but leave this task to the reader.

Having indicated the way to define the concepts with which we shall be operating, we return to the definition of the two types of oracle which we shall consider in connection with games. The first is a deterministic oracle with a specified (relative) accuracy  $v_0$ . For such an oracle,  $\Omega$  is a 1-point set. Also, by definition, the following conditions are satisfied

$$\left. \begin{aligned} \|\psi_x(x, l; F)\|_* &\leq \mathbf{L}_{\|\cdot\|}(F), & \|\psi_l(x, l; F)\|_{l*} &\leq \mathbf{L}_{\|\cdot\|_I}(F), \\ F(y, l) - F(x, l) &\geq \langle \psi_x(x, l; F) | y - x \rangle - v_0 r(F), \\ F(x, s) - F(x, l) &\leq \langle \psi_l(x, l; F) | s - l \rangle + v_0 r(F) \end{aligned} \right\} \quad (1.5)$$

for all  $F \in D$ ,  $x, y \in G$ ,  $s, l \in G_I$ . We shall call the classes of problems obtained by equipping  $D$  (regarded as a field of problems together with a fixed measure of the error  $v(\cdot, \cdot)$  corresponding to the choice of  $r(F)$  made earlier) with oracles of this type *classes of type  $D^{v_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$* .

The second type of oracle is a stochastic oracle, unbiased, and of accuracy  $v_0$ . Its definition is similar to that of the stochastic, first-order oracle in Section 5.1. Namely, let  $\mathcal{C}((\Omega, \mathbf{F}_\omega); \psi_x(\cdot), \psi_l(\cdot))$  be of the type defined above. Further, let  $L > 0$  and  $r > 1$ . We associate with the oracle  $\mathcal{O}$  the field of problems  $\tilde{D}_{r,L}^{v_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$  formed by all problems  $F \in D$  such that, for all  $x, y \in G$  and all  $l, s \in G_I$  the following relations hold:

$$(A) \begin{cases} F(y, l) - F(x, l) \geq \langle \mathbf{M}_{F_\omega} \psi_x(x, l; F, \omega) | y - x \rangle - v_0 L, \\ F(x, s) - F(x, l) \leq \langle \mathbf{M}_{F_\omega} \psi_l(x, l; F, \omega) | s - l \rangle + v_0 L, \end{cases}$$

$$(B) \begin{cases} \mathbf{M}_{F_\omega} \|\psi_x(x, l; F, \omega)\|_*^r \leq \left( \frac{L}{2\rho_{\|\cdot\|}(G)} \right)^r, \\ \mathbf{M}_{F_\omega} \|\psi_l(x, l; F, \omega)\|_{l*}^r \leq \left( \frac{L}{2\rho_{\|\cdot\|_I}(G_I)} \right)^r. \end{cases}$$

The field of problems  $\tilde{D}_{r,L}^{v_0}$ , equipped with the normalizing factors  $r(F) \equiv L$  (i.e. with a measure of the relative error  $v(x, l; F) = (1/L)v(x, l; F)$ ) and with the oracle  $\mathcal{O}$ , forms a class of problems denoted, like the original field, by  $\tilde{D}_{r,L}^{v_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$ .

We remark that, if one of the sets  $G, G_I$ , let us say  $G_I$ , is a point, then the class of problems of solving games of the form  $\tilde{D}_{r,L}^{v_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$  actually becomes the class  $\tilde{\mathcal{C}}_{r,L}^{v_0}(G, E, \|\cdot\|)$  of problems of stochastic programming without constraints.

## 6.2 MD-METHODS OF SOLVING GAMES: THE CASE OF A DETERMINISTIC ORACLE

In this section we describe the application of MD-methods to solving games with exact information (or, more correctly, information weakly distorted by noise)—to solving problems of the classes  $D^{v_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$ .

### 6.2.1

We start by picking out the special features of the situation with games as compared with the simpler problem of solving convex extremal problems without constraints. We shall watch out for what sort of difficulties are lying in wait for us in the simplest—the Hilbert—situation ( $E, \|\cdot\|$  and  $E_I, \|\cdot\|_I$ , Hilbert spaces). In this case MD-methods must turn into some analogue of the gradient method.

Gradient methods of solving convex-concave games are well-known. The simplest of them—the Arrow-Hurwicz method [2]—consists in constructing successive approximations  $(x_i, l_i)$  to the required solution from the formulae

$$\begin{aligned} x_{i+1} &= \pi_G(x_i - \rho_i \nabla_x F(x_i, l_i)), \\ l_{i+1} &= \pi_{G_I}(l_i + \rho_i \nabla_l F(x_i, l_i)). \end{aligned} \quad (2.1)$$

Here  $\nabla_x, \nabla_l$  denote the sub-gradients with respect to  $x$  and  $l$  respectively, and the  $\rho_i > 0$  are suitably chosen displacements. Unfortunately, this version does not, generally speaking, converge, however the displacements  $\rho_i$  may be chosen (a very simple example is:  $F(x, l) = xl$ ,  $|x| \leq 1$ ,  $|l| \leq 1$ ). To ensure the convergence of the method for suitable  $\rho_i$ , additional constraints have to be imposed on  $F$  such as strong convexity-concavity. There are many ways of modifying the method so that it becomes applicable to any  $F \in D$  (regularization of  $F$ , choosing different displacements along the  $x$ - and  $l$ -components, etc.). So far as the authors are aware, all such modifications lead to methods whose laboriousness on  $D$  is of a higher order than for convex-optimization methods of the gradient type.

It turns out that the difficulties mentioned can easily be overcome; under extremely general hypotheses about the  $\rho_i$  ( $\rho_i \rightarrow 0$  as  $i \rightarrow \infty$ ,  $\sum_i \rho_i = \infty$ ), the procedure (2.1) always converges to a saddle-point of  $F$ , although not in the ordinary sense, but in the Cesaro sense. Namely, the sequence

$$(x^i, l^i) = \frac{1}{i} \sum_{j=1}^i \rho_j (x_j, l_j) \quad (2.2)$$

converges to a saddle-point (cf. with the situation in Section 5.2). The rate of convergence of the trajectory (2.2) turns out to be the same as for gradient methods of solving Lipschitz-convex extremal problems, if the displacements  $\rho_i$  are chosen rationally.

We remark that an averaging of the (2.2) type in the Hilbert situation was apparently first proposed by R. Bruck in [4] (applied to the problem of solving variational inequalities with monotonic operators, which is close to the games problem), and, a little later but independently, by the authors in [24, 23] (where the general, regular situation is discussed).

The methods described below are obtained from the Arrow-Hurwicz method by replacing gradient descent by mirror descent, and with the addition of the averaging (2.2). We proceed to describe them.

### 6.2.2

We fix on an  $(E, \|\cdot\|)$ -regular function  $V(\varphi)$  and an  $(E_I, \|\cdot\|_I)$ -regular function  $V_I(\lambda)$ . We describe the ‘convergent’, infinite-step method, corresponding to  $V$ ,

$V_I$ , of solving games of the class  $D^{v_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$ . In describing the method we shall take  $G$  and  $G_I$  to be normalized by the conditions  $0 \in G, 0 \in G_I$ ,  $\rho_{\|\cdot\|}(G) = \rho_{\|\cdot\|_I}(G_I) = 1$  (these conditions on  $G$  and  $G_I$  are assumed to hold throughout this section).

The method (it will be denoted by  $MD_{V, V_I}$ ) consists in constructing sequences  $\{\varphi_i \in E^*\}_{i=0}^\infty$ ,  $\{\lambda_i \in E_I^*\}_{i=0}^\infty$  and their ‘images’  $\{x_i \in E, \bar{x}_i \in G\}_{i=1}^\infty$ ,  $\{l_i \in E_I, \bar{l}_i \in G_I\}_{i=1}^\infty$ , and also the sequence of results  $\{(x^i, l^i) \in G \times G_I\}_{i=1}^\infty$  according to the following recursive formulae (in which  $\rho_i > 0$  are certain displacements,  $\psi_x$  and  $\psi_l$  are the observation functions,  $F \in D^{v_0}$  is the problem being solved):

$$\left. \begin{aligned} \varphi_0 = 0, \lambda_0 = 0; \quad x_i = V'(\varphi_{i-1}), \quad \bar{x}_i = \pi_G(x_i), \\ l_i = V'_I(\lambda_{i-1}), \quad \bar{l}_i = \pi_{G_I}(l_i), \end{aligned} \right\} \quad (2.3)$$

$$\left. \begin{aligned} \varphi_i = \varphi_{i-1} - \rho_i(\psi_x(\bar{x}_i, \bar{l}_i; F) + \|\psi_x(\bar{x}_i, \bar{l}_i; F)\|_* \mu_G(x_i)), \\ \lambda_i = \lambda_{i-1} + \rho_i(\psi_l(\bar{x}_i, \bar{l}_i; F) - \|\psi_l(\bar{x}_i, \bar{l}_i; F)\|_{I*} \mu_{G_I}(l_i)); \end{aligned} \right\} \quad (2.4)$$

$$(x^i, l^i) = \frac{1}{i} \sum_{j=1}^i \rho_j (\bar{x}_j, \bar{l}_j). \quad (2.5)$$

**Theorem.** Let the displacements  $\rho_i$  satisfy the conditions  $\rho_i > 0$ ,  $\rho_i \rightarrow 0$  as  $i \rightarrow \infty$ ,  $\sum_i \rho_i = \infty$ . Then the process (2.3)–(2.5) converges on any problem  $F \in D^{v_0}$  in the sense that  $\lim_{i \rightarrow \infty} v(x^i, l^i; F) \leq v_0$ .

Moreover, for all  $i$ , the bound

$$v(x^i, l^i; F) \leq v_0 + \frac{1 + \sum_{j=1}^i \rho_j r(F)(\omega_V(\rho_j r(F)) + \omega_{V_I}(\rho_j r(F)))}{2r(F) \sum_{j=1}^i \rho_j} \quad (2.6)$$

is satisfied.

*Proof.* The first assertion of the theorem follows, as may easily be seen, from the second, and it is the latter which we shall now prove. Let  $x \in G$  and  $l \in G_I$ . We form the functions

$$V_x(\varphi) = V(\varphi) - \langle x | \varphi \rangle \text{ and } V_I(\lambda) = V_I(\lambda) - \langle \lambda | l \rangle,$$

and we shall follow the variation of the function  $\bar{V}_{x,I}(\varphi, \lambda) = V_x(\varphi) + V_I(\lambda)$  along the trajectory  $\{(\varphi, \lambda)\}$ . We note that because of the normalization conditions on  $G$  and  $G_I$  and the properties of the oracle (1.5), we have

$$\|\psi_x(x, l; F)\|_* \leq \frac{1}{2} r(F)$$

and

$$\|\psi_l(x, l; F)\|_{I*} \leq \frac{1}{2} r(F).$$

By the standard arguments for proving MD-methods (see Section 3.3) we have

$$\begin{aligned} V_x(\varphi_{i+1}) &\leq V_x(\varphi_i) + \rho_i \langle \psi_x(\bar{x}_i, \bar{l}_i; F) \\ &\quad + \|\psi_x(\bar{x}_i, \bar{l}_i; F)\|_* \mu_G(x_i) | x - x_i \rangle + \rho_i r(F) \omega_V(\rho_i r(F)) \\ &\leq V_x(\varphi_i) + \rho_i \langle \psi_x(\bar{x}_i, \bar{l}_i; F) | x - \bar{x}_i \rangle + \rho_i r(F) \omega_V(\rho_i r(F)) \\ &\leq V_x(\varphi_i) + \rho_i F(x, \bar{l}_i) - \rho_i F(\bar{x}_i, \bar{l}_i) + v_0 \rho_i r(F) + \rho_i r(F) \omega_V(\rho_i r(F)) \end{aligned}$$

(we have used the relation (1.5)). Similarly,

$$V_I(\lambda_{i+1}) \leq V_I(\lambda_i) + \rho_i F(\bar{x}_i, \bar{l}_i) - \rho_i F(\bar{x}_i, l) + v_0 \rho_i r(F) + \rho_i r(F) \omega_{V_I}(\rho_i r(F)).$$

Hence

$$\begin{aligned} V_{x,I}(\varphi_{j+1}, \lambda_{j+1}) &\leq V_{x,I}(\varphi_j, \lambda_j) + \rho_j F(x, \bar{l}_j) \\ &\quad - \rho_j F(\bar{x}_j, l) + \rho_j r(F)(\omega_V(\rho_j r(F)) + \omega_{V_I}(\rho_j r(F))) + 2v_0 r(F) \rho_j. \end{aligned} \quad (2.7)$$

Summing the relations (2.7) over  $j = 1, 2, \dots, i$ , and using the fact that

$$V_{x,I}(\varphi, \lambda) \geq -1, \quad V_{x,I}(\varphi_0, \lambda_0) = 0,$$

we obtain

$$\begin{aligned} \sum_{j=1}^i (F(\bar{x}_j, l) - F(x, \bar{l}_j)) \rho_j \\ \leq 1 + \sum_{j=1}^i \rho_j r(F)(\omega_V(\rho_j r(F)) + \omega_{V_I}(\rho_j r(F))) + 2v_0 \sum_{j=1}^i \rho_j r(F). \end{aligned} \quad (2.8)$$

Since  $F$  is convex relative to  $x$  and concave relative to  $l$ , the left-hand side of (2.8) is not less than

$$\left( \sum_{j=1}^i \rho_j \right) (F(x^i, l) - F(x, l^i)),$$

and so (2.8) gives

$$F(x^i, l) - F(x, l^i)$$

$$\leq \frac{1 + \sum_{j=1}^i \rho_j r(F)(\omega_V(\rho_j r(F)) + \omega_{V_I}(\rho_j r(F)))}{\sum_{j=1}^i \rho_j} + 2v_0 r(F). \quad (2.9)$$

The maximum of the left-hand side of (2.9) over  $x \in G, l \in G_I$  is equal to

$$\bar{F}(x^i) - \underline{F}(l^i) = \bar{F}(x^i) - \inf_G \bar{F} + \sup_{G_I} \bar{F} - \underline{F}(l^i) = \varepsilon(x^i, F) + \varepsilon_I(l^i, F),$$

and so (2.6) follows immediately from (2.9). The theorem has been proved.

**6.2.3**

We discuss the result obtained.

**6.2.3.1**

We start with the case where  $E$ ,  $\|\cdot\|$  and  $E_I$ ,  $\|\cdot\|_I$  are Hilbert spaces. Here the method can be simplified. The resulting method is defined by the relations

$$\begin{aligned} x_1 &= 0; \quad x_i = \pi_G(x_{i-1} - \rho_{i-1} \psi_x(x_{i-1}, l_{i-1}; F)), \\ l_1 &= 0; \quad l_i = \pi_{G_I}(l_{i-1} + \rho_{i-1} \psi_I(x_{i-1}, l_{i-1}; F)); \end{aligned} \quad (2.10)$$

$$(x^i, l^i) = \frac{1}{\sum_{j=1}^i \rho_j} \sum_{j=1}^i \rho_j (x_j, l_j). \quad (2.11)$$

This is precisely the Arrow-Hurwicz method supplemented by the averaging (2.11). For it Theorem 6.2.2 is also true, and it gives the bound

$$v(x^i, l^i, F) \leq v_0 + \frac{+\frac{1}{4}r^2(F) \sum_{j=1}^i \rho_j^2}{2r(F) \sum_{j=1}^i \rho_j}, \quad (2.12)$$

which is better than the one given in this theorem.

**6.2.3.2**

The question of choosing the displacements  $\rho_i$  rationally is solved in exactly the same way as in Section 5.2.2.2. Namely, if  $\omega_V(t) + \omega_{V_I}(t) \leq ct^\sigma$ ,  $\sigma \leq 1$ , then in the class of families of the form  $\rho_i = i^{-\sigma}$ , the best choice is  $x = 1/(1+\sigma)$ . And, with that choice, the right-hand side of (2.6) behaves like

$$v_0 + O(i^{-\sigma/(1+\sigma)}) \ln i$$

as  $i$  increases. In particular, in the Hilbert case, one should take  $\rho_i = O(1/\sqrt{i})$ ; and then the right-hand side of (2.6) decreases like  $v_0 + O((\ln i)/\sqrt{i})$ .

**6.2.4**

We now describe a ‘constructive’ version  $MD_{V,V_I}(v)$  of the  $MD_{V,V_I}$ -method capable of ensuring a specified accuracy  $v$ ,  $v_0 < v < 1$ , in the solution of problems in the  $D^{v_0}$  class. This version is obtained from the ‘infinite-step’ method by adding rules for the ‘automatic’ choice of the displacements  $\rho_i$  (like the rule MD 1.6’ in Section 3.3) and of the rule for stopping.

In solving a problem  $F \in D^{v_0}$ , the method  $MD_{V,V_I}(v)$  constructs finite sequences

$$\{\varphi_i \in E^*, \lambda_i \in E_I^*\}_{i=0}^{M_F}, \quad \{x_i \in E, \bar{x}_i \in G, l_i \in E_I, \bar{l}_i \in G_I\}_{i=1}^{M_F}$$

and the result  $(\bar{x}, \bar{e})$ . Here  $M_F$  is the number (formed by the method) of questions after putting which it solves the problem  $F$ .

The work of the method  $MD_{V,V_I}$  is described by the following rules.

$MD_{V,V_I} 0$ .

Put  $i = 1$ ,  $\varphi_0 = 0$ ,  $\lambda_0 = 0$ . Go to  $MD_{V,V_I} 1$ .

$MD_{V,V_I} I$ . *i*th step.

For the *i*th step  $\varphi_{i-1} \in E^*$ ,  $\lambda_{i-1} \in E_I^*$  are available. In the *i*th step the following operations are carried out.

$MD_{V,V_I} I.1$ . Put

$$x_i = V'(\varphi_{i-1}), \quad \bar{x}_i = \pi_G(x_i), \quad l_i = V'_I(\lambda_{i-1}), \quad \bar{l}_i = \pi_{G_I}(l_i).$$

$MD_{V,V_I} I.2$ . At the point  $(\bar{x}, \bar{l})$  put a question to the oracle about the problem  $F$  under solution. Let  $\eta_i = \psi_x(\bar{x}_i, \bar{l}_i; F)$  and  $\eta_{II} = \psi_I(\bar{x}_i, \bar{l}_i; F)$  be the answers communicated by the oracle. Put  $r(i) = 2 \max \{\|\eta_i\|_*, \|\eta_{II}\|_{I*}\}$ . If  $r(i) = 0$ , go to  $MD_{V,V_I} 2$ ; otherwise go to  $MD_{V,V_I} 1.3$ .

$MD_{V,V_I} I.3$ . Put

$$\bar{\eta}_i = \eta_i + \|\eta_i\|_* \mu_G(x_i), \quad \bar{\eta}_{II} = \eta_{II} - \|\eta_{II}\|_{I*} \mu_{G_I}(l_i).$$

$MD_{V,V_I} I.4$ . Let  $\rho_i$  be the upper bound of those  $\rho$  such that

$$\begin{aligned} &\{V(\varphi_{i-1} - \rho \bar{\eta}_i) - V(\varphi_{i-1}) + \langle \bar{\eta}_i | x_i \rangle \rho\} \\ &+ \{V_I(\lambda_{i-1} + \rho \bar{\eta}_{II}) - V_I(\lambda_{i-1}) - \langle \bar{\eta}_{II} | l_i \rangle \rho\} \leq (v - v_0) \rho r(i). \end{aligned} \quad (2.13)$$

If  $\rho_i = \infty$ , go to  $MD_{V,V_I} 2$ ; otherwise go to  $MD_{V,V_I} 1.5$ .

$MD_{V,V_I} I.5$ . Put

$$\varphi_i = \varphi_{i-1} - \rho_i \bar{\eta}_i, \quad \lambda_i = \lambda_{i-1} + \rho_i \bar{\eta}_{II}.$$

If  $\sum_{j=1}^i \rho_j > 0$  and

$$\|\varphi_i\|_* + \|\lambda_i\|_{I*} - V(\varphi_i) - V_I(\lambda_i) \leq (\max_{1 \leq j \leq i} r(j))(v - v_0) \sum_{j=1}^i \rho_j, \quad (2.14)$$

go to  $MD_{V,V_I}$  2; otherwise, increase  $i$  by 1 and go to the next step, i.e. to  $MD_{V,V_I}$  1.1.

$MD_{V,V_I}$  2. Rule for output of the result.

Let  $M_F$  be the moment of reversion to  $MD_{V,V_I}$  2. At this moment the work of the method stops, with output of the result

$$(\bar{x}, \bar{l}) = \begin{cases} (\bar{x}_{M_F}, \bar{l}_{M_F}), r(M_F) = 0 \text{ or } \rho_{M_F} = +\infty, \\ \frac{1}{\sum_{j=1}^i \rho_j} \sum_{j=1}^i \rho_j (\bar{x}_j, \bar{l}_j) \text{ in the remaining cases.} \end{cases}$$

**Theorem.** The method described solves every problem  $F \in D^{v_0}$  with a relative error  $v$ ,  $v_0 < v < 1$ , and with a laboriousness  $M_F + 1$  not exceeding  $M_{V,V_I}(v - v_0)$ , where

$$\begin{aligned} M_{V,V_I}(v) &\equiv \max \left\{ \left[ \frac{1}{v\gamma_V(v/2)} \right] + 2, \left[ \frac{1}{v\gamma_{V_I}(v/2)} \right] + 2 \right\} \\ &= \max \{ M_V(2v), M_{V_I}(2v) \} \end{aligned}$$

(The functions  $M_V(\cdot)$  were introduced in Section 3.3 in connection with the bounds for the laboriousness of MD-methods of solving Lipschitz-convex problems.)

*Proof.*

1°. *Bound for the laboriousness.* Suppose the method, when solving the problem  $F$ , did not stop after  $N$ th step. Then  $r(i) > 0$ ,  $\rho_i < \infty$  in the first  $N$  steps. In accordance with inequality (2.6) in Chapter 3, the left-hand side of (2.13) does not exceed  $\omega_V(r(i)\rho) + \omega_{V_I}(r(i)\rho)$ . Therefore (2.13) holds when

$$\rho = \frac{1}{r(i)} \min \left\{ \gamma_V \left( \frac{v - v_0}{2} \right), \gamma_{V_I} \left( \frac{v - v_0}{2} \right) \right\} \equiv \frac{\bar{\rho}}{r(i)}.$$

Thus  $\rho_i \geq \bar{\rho}/r(i)$ . But then the right-hand side of (2.14) when  $i = N$  is not less than

$$(\max r(j)) \sum_{j=1}^N \rho_j (v - v_0) \geq \bar{\rho} N (v - v_0).$$

But the left-hand side of (2.14) does not exceed 1. When  $i = N$  the relation (2.14) does not hold, by definition of  $N$ , and this leads to  $\bar{\rho} N (v - v_0) < 1$ . This gives the required bound for  $N$  and hence for  $M_F$ .

2°. *The error of the method.* First let  $r(M_F) = 0$ . In this case, by the properties of the oracle,

$$F(x, l_{M_F}) \geq F(\bar{x}_{M_F}, \bar{l}_{M_F}) - v_0 r(F) \text{ and } F(\bar{x}_{M_F}, l) - F(\bar{x}_{M_F}, \bar{l}_{M_F}) \leq v_0 r(F)$$

for all  $(x, l) \in G \times G_I$ , i.e.

$$F(\bar{l}_{M_F}) - F(\bar{x}_{M_F}, \bar{l}_{M_F}) \geq -v_0 r(F)$$

and

$$F(\bar{x}_{M_F}) - F(\bar{x}_{M_F}, \bar{l}_{M_F}) \leq v_0 r(F).$$

Hence and as in the proof of Theorem 6.2.2,  $\varepsilon(\bar{x}_{M_F}, \bar{l}_{M_F}; F) \leq v_0 r(F)$ , which proves the inequality  $v(\bar{x}_{M_F}, \bar{l}_{M_F}; F) \leq v_0 < v$ .

Now let  $r(M_F) > 0$ . Let  $x \in G$ ,  $V_x(\varphi) = V(\varphi) - \langle \varphi | x \rangle$ , and let  $l \in G$ ,  $V_{x,l}(\lambda) = V_l(\lambda) - \langle \lambda | l \rangle$ . When  $j \leq M_F$  and for every  $\rho > 0$  such that (2.13) holds, we have

$$\begin{aligned} V_x(\varphi_{j-1} - \rho \bar{\eta}_j) \\ = \{ V(\varphi_{j-1} - \rho \bar{\eta}_j) - V(\varphi_{j-1}) + \rho \langle \bar{\eta}_j | x_j \rangle \} + V_x(\varphi_{j-1}) + \rho \langle \bar{\eta}_j | x - x_j \rangle \end{aligned}$$

and

$$\begin{aligned} V_{x,l}(\lambda_{j-1} + \rho \bar{\eta}_{lj}) \\ = \{ V_l(\lambda_{j-1} + \rho \bar{\eta}_{lj}) - V_l(\lambda_{j-1}) - \rho \langle \bar{\eta}_{lj} | l_j \rangle \} + V_{x,l}(\lambda_{j-1}) + \rho \langle \bar{\eta}_{lj} | l_j - l \rangle. \end{aligned}$$

Adding these inequalities and using the fact that  $\rho$  satisfies (2.13), and putting  $V_{x,l}(\varphi, \lambda) \equiv V_x(\varphi) + V_{x,l}(\lambda)$ , we obtain

$$\begin{aligned} V_{x,l}(\varphi_{j-1} - \rho \bar{\eta}_j, \lambda_{j-1} + \rho \bar{\eta}_{lj}) \\ \leq (v - v_0)r(j)\rho + V_{x,l}(\varphi_{j-1}, \lambda_{j-1}) \\ + \rho \langle \eta_j | x - x_j \rangle + \rho \langle \bar{\eta}_{lj} | l_j - l \rangle. \end{aligned} \quad (2.15)$$

As always, the right-hand side of (2.15) is not greater than

$$(v - v_0)r(j)\rho + V_{x,l}(\varphi_{j-1}, \lambda_{j-1}) + \rho \{ F(x, l_j) - F(\bar{x}_j, l) + 2v_0 r(F) \},$$

i.e. (2.15) gives

$$\begin{aligned} V_{x,l}(\varphi_{j-1} - \rho \bar{\eta}_j, \lambda_{j-1} + \rho \bar{\eta}_{lj}) \\ \leq (v - v_0)r(j)\rho + V_{x,l}(\varphi_{j-1}, \lambda_{j-1}) + \rho \{ F(x, l_j) - F(\bar{x}_j, l) + 2v_0 r(F) \}. \end{aligned} \quad (2.16)$$

When  $j < M_F$  the relation (2.16) holds for  $\rho = \rho_j$ . It is also true when  $j = M_F$  if  $\rho_{M_F} < \infty$ . Adding the inequalities (2.16) corresponding to the choice  $\rho = \rho_j$  over  $j = 1, 2, \dots, M_F - 1$ , and adding (2.16) with  $j = M_F$  with any  $\rho$  which satisfies (2.13) for this  $j$  to the result, we obtain, as in the proof of Theorem 6.2.2 that

$$\begin{aligned} \sum_{j=1}^{M_F-1} \rho_j (F(\bar{x}_j, l_j) - F(x, l_j)) + \rho (F(\bar{x}_{M_F}, l) - F(x, l_{M_F})) \\ \leq (v - v_0) \left( \sum_{j=1}^{M_F-1} r(j)\rho_j + r(M_F)\rho \right) - V_{x,l}(\varphi_{M_F-1} - \rho \bar{\eta}_{M_F}, \lambda_{M_F-1} + \rho \bar{\eta}_{lM_F}) \\ + 2v_0 r(F) \left( \sum_{j=1}^{M_F-1} \rho_j + \rho \right). \end{aligned} \quad (2.17)$$

If  $\rho_{M_F} = +\infty$ , then (2.17) holds for a certain sequence of values of  $\rho$  tending to infinity. Dividing both sides of (2.17) by  $\rho$  and passing to the limit through this sequence of values of  $\rho$ , we obtain that

$$F(\bar{x}, l) - F(x, \bar{l}) \leq 2v_0 r(F) + (v - v_0) r(M_F)$$

for all  $(x, l) \in G \times G_I$ . From the properties of the oracle it is clear that  $r(j) \leq r(F)$  for all  $j$ . Thus, in the case considered,

$$F(\bar{x}, l) - F(x, \bar{l}) \leq (2v_0 + (v - v_0)) r(F) \leq 2vr(F),$$

and hence, as also when  $r(M_F) = 0$ , we have  $v(\bar{x}, \bar{l}; F) \leq v$ .

Now let  $\rho_{M_F} \neq +\infty$ . Then we can take  $\rho = \rho_{M_F}$  in (2.17). Also, as in the proof of Theorem 6.2.2, the left-hand side of (2.17) is not less than  $(\sum_{j=1}^{M_F} \rho_j)(F(\bar{x}, l) - F(x, \bar{l}))$ , and the right-hand side, in view of (2.14), is not greater than

$$\begin{aligned} & (v - v_0) \sum_{j=1}^{M_F} r(j) \rho_j + \| \varphi_{M_F} \|_* + \| \lambda_{M_F} \|_{I*} - V(\varphi_{M_F}) - V_I(\lambda_{M_F}) + 2v_0 r(F) \sum_{j=1}^{M_F} \rho_j \\ & \leq ((v - v_0) + (v - v_0) + 2v_0) r(F) \sum_{j=1}^{M_F} \rho_j. \end{aligned}$$

Hence  $F(\bar{x}, l) - F(x, \bar{l}) \leq 2vr(F)$ ,  $(x, l) \in G \times G_I$ , which, as above, gives  $v(\bar{x}, \bar{l}; F) \leq v$ . The theorem has been proved.

## 6.2.5

Let us discuss the results obtained.

### 6.2.5.1

We start with the Hilbert situation. As always, in this case the method can be simplified. The resulting version is obtained from the original one by making the following changes:

- (i) the rule  $MD_{V, V_I}$  1.1 is replaced by  $\varphi_{i-1} = x_i = \bar{x}_i$ ,  $\lambda_{i-1} = l_i = \bar{l}_i$ ;
- (ii) the rule  $MD_{V, V_I}$  1.3 is replaced by  $\bar{\eta}_i = \eta_i$ ,  $\bar{\eta}_{ii} = \eta_{ii}$ ;
- (iii) the rule  $MD_{V, V_I}$  1.4 is replaced by

$$\rho_i = \frac{2(v - v_0)r(i)}{\eta_i^2 + \eta_{ii}^2}; \quad (2.18)$$

- (iv) the rule for constructing  $\varphi_i$ ,  $\lambda_i$  is replaced by

$$\varphi_i = \pi_G(\varphi_{i-1} - \rho_i \bar{\eta}_i), \quad \lambda_i = \pi_{G_I}(\lambda_{i-1} + \rho_i \bar{\eta}_{ii}),$$

and inequality (2.14) by

$$\| \varphi_i \| + \| \lambda_i \| - \frac{1}{2} \| \varphi_i \|^2 - \frac{1}{2} \| \lambda_i \|^2 \leq (\max_{1 \leq j \leq i} (r(j))(v - v_0) \sum_{j=1}^i \rho_j). \quad (2.19)$$

The method obtained solves every problem  $f \in D^{v_0}$  with an accuracy  $v$ ,  $v_0 < v < 1$ , and with a laboriousness not exceeding

$$\left[ \frac{1}{4(v - v_0)^2} \right] + 2 \equiv \bar{M}(v - v_0).$$

The proof of this assertion is carried out in the same way as in the proof of Theorem 6.2.2, the part of  $V_{x, l}$  being played by

$$\frac{1}{2}(\varphi - x)^2 + \frac{1}{2}(\lambda - l)^2.$$

The reader is recommended to work through all the necessary calculations. We remark that the method just described is the same as the method (2.10)–(2.11) with a special choice of the displacements and of the moment of stopping.

### 6.2.5.2

Up to now we have presented versions of the methods which are intended to solve problems with a given relative error. It is possible to modify these versions so as to achieve a specified absolute error  $\varepsilon$ . To do this it suffices to replace the right-hand side of (2.13) by  $\varepsilon\rho$ , and the right-hand side of (2.14) by  $\varepsilon \sum_{j=1}^i \rho_j$ . (For the Hilbert case, (2.18) has to be replaced by  $\rho_i = 2\varepsilon/(\eta_i^2 + \eta_{ii}^2)$ , and the right-hand side of (2.19) by  $\varepsilon \sum_{j=1}^i \rho_j$ .) The methods so obtained solve problems  $F \in D^{v_0}$  with an absolute error  $v(\bar{x}, \bar{l}; F) \leq \varepsilon + v_0 r(F)$  (i.e. one not exceeding the specified error—with accuracy up to the non-removable error of the oracle). Moreover, their laboriousness does not exceed  $M_{V, V_I}(v(F, \varepsilon))$  and  $\bar{M}(v(F, \varepsilon))$  respectively, where  $v(F, \varepsilon) = \varepsilon/r(F)$  is the (maximum) relative error which ensures the specified absolute error. The proof of these facts is similar to the previous one, and we leave it to the reader.

### 6.2.5.3

We see that solving convex-concave games on  $E \times E_I$  is ‘no more complicated’ than solving, by MD-methods, convex extremal problems on ‘the worser’ of the spaces  $E, E_I$ .

## 6.3 MD-METHODS OF SOLVING GAMES: THE CASE OF A STOCHASTIC ORACLE

### 6.3.1

In this section we prove that the methods in Section 6.2 are also suitable for solving games when the oracle is stochastic, for solving, that is, problems in the classes  $\tilde{D}_{r, L}^{v, v_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$ . In describing the methods we shall

assume  $G$  and  $G_i$  to be normalized by the conditions stated at the beginning of Section 6.2.2. Let  $V(\cdot)$  and  $V_i(\cdot)$  be the same as in Section 6.2.2, let  $\omega_i$  be the noise of the oracle

$$\mathcal{O} = ((\Omega, \mathbf{F}_\omega), \psi_x(x, l; F, \omega) \psi_i(x, l; F, \omega))$$

in the  $i$ th step, and let  $\omega^i = (\omega_1, \dots, \omega_i)$ .

We start by describing the ‘infinite-step’ version of the method—we denote it by  $\overline{\text{MD}}_{V, V_i}$ . The method, in solving a problem  $F \in \tilde{\mathbf{D}}_{r, L}^{e, v_0}$  constructs the random sequences

$$\begin{aligned} &\{\varphi_i = \varphi_i(\omega^i) \in E^*, \quad \lambda_i = \lambda_i(\omega^i) \in E_l^*\}, \\ &\{x_i = x_i(\omega^{i-1}) \in E, \quad \bar{x}_i = \bar{x}_i(\omega^{i-1}) \in G, \\ &x^i = x^i(\omega^{i-1}) \in G; \quad l_i = l_i(\omega^{i-1}) \in E_l, \\ &\bar{l}_i = \bar{l}_i(\omega^{i-1}) \in G_l, \quad l^i = l^i(\omega^{i-1}) \in G_l\} \end{aligned}$$

according to rules similar to (2.4)–(2.5):

$$\begin{aligned} &\varphi_0 = 0, \quad \lambda_0 = 0; \\ &x_i = V'(\varphi_{i-1}); \quad \bar{x}_i = \pi_G(x_i); \quad l_i = V'_i(\lambda_{i-1}); \quad \bar{l}_i = \pi_{G_l}(l_i), \\ &\varphi_i = \varphi_{i-1} - \rho_i \{\psi_x(\bar{x}_i, \bar{l}_i; F, \omega_i) + \|\psi_x(\bar{x}_i, \bar{l}_i; F, \omega_i)\|_* \mu_G(x_i)\}, \\ &\lambda_i = \lambda_{i-1} + \rho_i \{\psi_i(\bar{x}_i, \bar{l}_i; F, \omega_i) - \|\psi_i(\bar{x}_i, \bar{l}_i; F, \omega_i)\|_* \mu_{G_l}(l_i)\}, \\ &(x^i, l^i) = \frac{1}{\sum_{j=1}^i \rho_j} \sum_{j=1}^i \rho_j (\bar{x}_j, \bar{l}_j). \end{aligned} \quad (3.1)$$

In (3.1), (3.2),  $\{\rho_i\}$  ( $\rho_i > 0$ ) is a sequence of (deterministic) displacements.

**Theorem.** Under the conditions  $\rho_i > 0$ ,  $\lim_{i \rightarrow \infty} \rho_i = 0$ ,  $\sum_i \rho_i = \infty$ , the process (3.1)–(3.2) converges (in the mean) on any problem  $F \in \tilde{\mathbf{D}}_{r, L}^{e, v_0}$  in the following sense:

$$\lim_{i \rightarrow \infty} \mathbf{M}_{\mathbf{F}_{\omega^{i-1}}} v(x^i, l^i; F) \leq v_0.$$

Moreover, the following bound is satisfied:

$$\mathbf{M}_{\mathbf{F}_{\omega^{i-1}}} v(x^i, l^i; F) \leq v_0 + \frac{2 + 3L \left[ \sum_{j=1}^i \rho_j (\omega_{V, r}(2L\rho_j) + \omega_{V_l, r}(2L\rho_j)) \right]}{2L \sum_{j=1}^i \rho_j}. \quad (3.3)$$

*Proof.* Let  $F$  be the problem being solved. As always, it suffices to prove (3.3). We put

$$\begin{aligned} d_i(\omega^i) &= \psi_x(\bar{x}_i(\omega^{i-1}), \bar{l}_i(\omega^{i-1}); F, \omega_i) + \|\psi_x(\bar{x}_i, \bar{l}_i; F, \omega_i)\|_* \mu_G(x_i), \\ \tilde{d}_i(\omega^{i-1}) &= \mathbf{M}_{\mathbf{F}_{\omega_i}} d_i(\omega^i) = \bar{\psi}_x(\bar{x}_i, \bar{l}_i; F) + a_i(\omega^{i-1}) \mu_G(x_i), \end{aligned}$$

where

$$\bar{\psi}_x(x, l; F) = \mathbf{M}_{\mathbf{F}_\omega} \psi_x(x, l; F, \omega).$$

Further let  $\tilde{d}_i(\omega^i) = d_i(\omega^i) - \tilde{d}_i(\omega^{i-1})$ . Then, because of property (B) of the oracle  $\mathcal{O}$  (see Section 6.1.2.3) and the definition of  $a_i(\omega^{i-1})$ , we have

$$\left. \begin{aligned} \mathbf{M}_{\mathbf{F}_{\omega_i}} \|d_i(\omega^i)\|_*^r &\leq L^r, \quad \mathbf{M}_{\mathbf{F}_{\omega_i}} \|\tilde{d}_i(\omega^i)\|_*^r \leq (2L)^r, \\ \mathbf{M}_{\mathbf{F}_{\omega_i}} \tilde{d}_i(\omega^i) &= 0, \quad a_i(\omega^{i-1}) \geq \|\bar{\psi}_x(\bar{x}_i, \bar{l}_i; F)\|_*. \end{aligned} \right\} \quad (3.4)$$

Similarly we define

$$\begin{aligned} \delta_i(\omega^i) &= \psi_i(\bar{x}_i, \bar{l}_i; F, \omega_i) - \|\psi_i(\bar{x}_i, \bar{l}_i; F, \omega_i)\|_* \mu_{G_l}(l_i), \\ \tilde{\delta}_i(\omega^{i-1}) &= \mathbf{M}_{\mathbf{F}_{\omega_i}} \delta_i(\omega^i), \quad -\tilde{\delta}_i(\omega^i) = \delta_i(\omega^i) - \tilde{\delta}_i(\omega^{i-1}). \end{aligned}$$

Then

$$\left. \begin{aligned} \mathbf{M}_{\mathbf{F}_{\omega_i}} \|\delta_i(\omega^i)\|_{l*}^r &\leq L^r, \quad \mathbf{M}_{\mathbf{F}_{\omega_i}} \|\tilde{\delta}_i(\omega^i)\|_{l*}^r \leq (2L)^r, \\ \mathbf{M}_{\mathbf{F}_{\omega_i}} \tilde{\delta}_i(\omega^i) &= 0. \end{aligned} \right\} \quad (3.5)$$

Now let

$$\begin{aligned} x \in G, \quad l \in G_l, \quad V_x(\varphi) &= V(\varphi) - \langle \varphi | x \rangle, \\ V_H(\lambda) &= V_I(\lambda) - \langle \lambda | l \rangle, \quad V_{x, l}(\varphi, \lambda) = V_x(\varphi) + V_H(\lambda). \end{aligned}$$

Proceeding exactly as in the proof of Theorem 5.2.1, we obtain

$$V_x(\varphi_j(\omega^j)) \leq V_x(\varphi_{j-1}(\omega^{j-1})) + \rho_j \langle d_j(\omega^j) | x - x_j \rangle + b_j(\omega^j), \quad (3.6)$$

where, because of (3.4),

$$\mathbf{M}_{\mathbf{F}_{\omega_i}} b_i(\omega^i) \leq \rho_i L \omega_{V, r}(\rho_i L). \quad (3.7)$$

Moreover,

$$\begin{aligned} \langle d_i(\omega^i) | x - x_i \rangle &= \langle \tilde{d}_i(\omega^{i-1}) | x - x_i \rangle + \langle \tilde{d}_i(\omega^i) | x \rangle - \langle \tilde{d}_i(\omega^i) | x_i \rangle \\ &\leq \langle \bar{\psi}_x(\bar{x}_i, \bar{l}_i; F) | x - x_i \rangle + \langle \tilde{d}_i(\omega^i) | x \rangle - \langle \tilde{d}_i(\omega^i) | x_i \rangle. \end{aligned}$$

Further,

$$\langle \psi_x(\bar{x}_i, \bar{l}_i; F) | x - x_i \rangle \leq v_0 L + F(x, \bar{l}_i) - F(\bar{x}_i, \bar{l}_i).$$

Using these inequalities in (3.6), we obtain

$$\begin{aligned} V_x(\varphi_j(\omega^j)) &\leq V_x(\varphi_{j-1}(\omega^{j-1})) + \rho_j (v_0 L + F(x, \bar{l}_j) - F(\bar{x}_j, \bar{l}_j)) \\ &\quad + \rho_j \langle \tilde{d}_j(\omega^j) | x \rangle + c_j(\omega^j), \end{aligned} \quad (3.8)$$

where we have put

$$c_j(\omega^j) = b_j(\omega^j) - \rho_j \langle \tilde{d}_j(\omega^j) | x_j(\omega^{j-1}) \rangle$$

and because of (3.4) and (3.7)

$$\mathbf{M}_{\mathbf{F}_{\omega_j}} c_j(\omega^j) \leq \rho_j L \omega_{V, r}(\rho_j L), \quad (3.9)$$

Similarly,

$$\begin{aligned} V_l(\lambda_j(\omega^j)) &\leq V_l(\lambda_{j-1}(\omega^{j-1})) + \rho_j(v_0 L + F(\bar{x}_j, \bar{l}_j) - F(\bar{x}_j, l)) \\ &\quad + \rho_j \langle \delta_j(\omega^j) | l \rangle + c_{lj}(\omega^j), \end{aligned} \quad (3.10)$$

where

$$\mathbf{M}_{F_{\omega^j}} c_{lj}(\omega^j) \leq \rho_j L \omega_{V_l, r}(\rho_j L). \quad (3.11)$$

It is of essential importance that  $c_j$  and  $c_{lj}$  do not depend on the parameters  $x$  and  $l$ . Adding (3.8) and (3.10) and summing the inequalities obtained over  $j = 1, 2, \dots, i$ , we obtain, as in the proof of Theorem 6.2.2,

$$\begin{aligned} &\left( \sum_{j=1}^i \rho_j \right) \{F(x^i(\omega^{i-1}), l) - F(x, l^i(\omega^{i-1}))\} \\ &\leq 1 + 2 \left( \sum_{j=1}^i \rho_j \right) v_0 L + \left\| \sum_{j=1}^i \rho_j \delta_j(\omega^j) \right\|_{I_*} + \left\| \sum_{j=1}^i \rho_j \tilde{\delta}_j(\omega^j) \right\|_* \\ &\quad + \sum_{j=1}^i (c_j(\omega^j) + c_{lj}(\omega^j)) \equiv A(\omega^i). \end{aligned} \quad (3.12)$$

Here  $A(\omega^i)$  does not depend on  $x$  and  $l$ . Taking the maximum of both sides of (3.12) over  $(x, l) \in G \times G_l$ , we obtain

$$\begin{aligned} &\left( \sum_{j=1}^i \rho_j \right) \varepsilon(x^i(\omega^{i-1}), l^i(\omega^{i-1}); F) \leq \frac{1}{2} + \left( \sum_{j=1}^i \rho_j \right) v_0 L \\ &+ \frac{1}{2} \left\| \sum_{j=1}^i \rho_j \delta_j(\omega^j) \right\|_{I_*} + \frac{1}{2} \left\| \sum_{j=1}^i \rho_j \tilde{\delta}_j(\omega^j) \right\|_* + \frac{1}{2} \sum_{j=1}^i (c_j(\omega^j) + c_{lj}(\omega^j)). \end{aligned} \quad (3.13)$$

Averaging (3.13) over  $\omega^i$  and taking (3.9) and (3.11) into account, we obtain  $L \mathbf{M}_{F_{\omega^{i-1}}} v(x^i, l^i; F)$

$$\begin{aligned} &\leq \frac{1}{2 \sum_{j=1}^i \rho_j} + v_0 L + \frac{1}{2 \sum_{j=1}^i \rho_j} \mathbf{M}_{F_{\omega^i}} \left\| \sum_{j=1}^i \rho_j \delta_j(\omega^j) \right\|_{I_*} \\ &+ \frac{1}{2 \sum_{j=1}^i \rho_j} \mathbf{M}_{F_{\omega^i}} \left\| \sum_{j=1}^i \rho_j \tilde{\delta}_j(\omega^j) \right\|_* \\ &+ \frac{1}{2 \sum_{j=1}^i \rho_j} \left( \sum_{j=1}^i \rho_j L (\omega_{V_r, r}(\rho_j L) + \omega_{V_l, r}(\rho_j L)) \right). \end{aligned} \quad (3.14)$$

In view of (3.4), for estimating  $\mathbf{M}_{F_{\omega^i}} \left\| \sum_{j=1}^i \rho_j \tilde{\delta}_j(\omega^j) \right\|_*$  we can apply Proposition 5.1.5 and this gives the bound

$$\mathbf{M}_{F_{\omega^i}} \left\| \sum_{j=1}^i \rho_j \tilde{\delta}_j(\omega^j) \right\|_* \leq \sum_{j=1}^i 2 L \rho_j \omega_{V_r, r}(2 L \rho_j) + \frac{1}{2}.$$

$\mathbf{M}_{F_{\omega^i}} \left\| \sum_{j=1}^i \rho_j \delta_j(\omega^j) \right\|_{I_*}$  is estimated similarly. Substituting these estimates in (3.14), we obtain the required bound (3.3). The theorem has been proved.

**6.3.2** Let us discuss the results obtained.

### 6.3.2.1

In the Hilbert case when  $r \geq 2$ , simplifications are possible, as always. The resulting method has the form

$$\left. \begin{aligned} x_1 &= 0, & l_1 &= 0, \\ x_{i+1} &= \pi_G(x_i - \rho_i \psi_x(x_i, l_i; F, \omega_i)); \\ l_{i+1} &= \pi_{G_l}(l_i + \rho_i \psi_l(x_i, l_i; F, \omega_i)), \end{aligned} \right\} \quad (3.15)$$

$$(x^i, l^i) = \frac{1}{i} \sum_{j=1}^i \rho_j (x_j, l_j). \quad (3.16)$$

A theorem like 6.3.1 holds for the method obtained; the bound (3.3) simplifies and takes the form

$$\mathbf{M}_{F_{\omega^{i-1}}} v(x^i, l^i; F) \leq v_0 + \frac{1 + \frac{1}{4} L^2 \sum_{j=1}^i \rho_j^2}{L \sum_{j=1}^i \rho_j}. \quad (3.17)$$

### 6.3.2.2

The question of a rational choice of the displacements  $\rho_j$  is solved as in Section 5.2.2.2.

### 6.3.3

We now describe a ‘constructive’ version  $\widetilde{\text{MD}}_{V_r, V_l}(v)$  of the method  $\widetilde{\text{MD}}_{V_r, V_l}$ , intended to solve problems from  $\widetilde{\mathcal{D}}_{r, L}^{v, v_0}$  with a specified (averaged over realizations of the method) relative error  $v$ ,  $v_0 < v < 1$ . A method  $\widetilde{\text{MD}}_{V_r, V_l}(v)$  consists in the realization of  $N$  steps of the process (3.1)-(3.2) in which all the displacements are equal to some value  $\rho > 0$ . The parameters  $N$  and  $\rho$  have to be chosen as functions of the required accuracy. The result of applying the method is a point  $(x^N, l^N)$ . The laboriousness of such a method (which is equal to the number of questions put to the oracle plus one for the output of the result) is equal to  $N$ , since the construction of  $(x^N, l^N)$  requires  $N - 1$  questions to be put to the oracle.

By virtue of the estimate (3.3), to ensure an accuracy  $v > v_0$  we can choose  $\rho$  and  $N$  to be of the form

$$\rho = \frac{1}{2L} \rho_{V, V_I, r}(v - v_0),$$

where

$$\rho_{V, V_I, r}(v) = \min \left\{ \gamma_{V, r}\left(\frac{v}{6}\right), \gamma_{V_I, r}\left(\frac{v}{6}\right) \right\},$$

and

$$N = N_{V, V_I, r}(v - v_0),$$

where

$$N_{V, V_I, r}(v) = \left\lceil \frac{4}{v \rho_{V, V_I, r}(v)} \right\rceil. \quad (3.18)$$

In the Hilbert case when  $r \geq 2$ , one should similarly apply the process (3.15)–(3.16), taking  $\rho$  and  $N$  to be of the form

$$\rho = 2v/L \text{ and } N = \lceil 1/v^2 \rceil. \quad (3.19)$$

### 6.3.4

Comparing (3.18)–(3.19) with the bound for the laboriousness of MD-methods of solving stochastic, convex, unconstrained extremal problems (Section 5.2), we see that solving games with a stochastic oracle is not more complicated, roughly speaking, than solving stochastic extremal problems on ‘the worser’ of the spaces  $E$  and  $E_I$ . On the other hand, problems of the latter type are simple particular cases of solving games when either  $G$  or  $G_I$  degenerates to a point.

We give for reference the bounds for the laboriousness of  $\widetilde{MD}_{V, V_I, r}(v)$  (i.e. for the function  $N_{V, V_I, r}(v)$ ) for the case where  $(E, \|\cdot\|) = (L_p, \|\cdot\|_p)$  and  $(E_I, \|\cdot\|) = (L_{p_I}, \|\cdot\|_{p_I})$ . These bounds are obtained by comparing (3.18) and the relations (1.6), (1.7) in Chapter 5. Let  $p > 1$ ,  $p_I > 1$ . Then with the standard choice of  $V = V_p(\cdot)$ ,  $V_I = V_{p_I}(\cdot)$  (see Section 3.2), we obtain (we assume  $v < 1$ )

$$N_{V, V_I, r}(v) \leq \max \left\{ D(p, r) \left( \frac{1}{v} \right)^{\bar{r}(p)}, D(p_I, r) \left( \frac{1}{v} \right)^{\bar{r}(p_I)} \right\}, \quad (3.20)$$

where  $\bar{r}(q) = \min \{2, r, q/(q-1)\}$ , and  $D(q, r) < \infty$  depends only on  $q > 1$  and  $r > 1$ . If  $p = 1$  and  $\dim E \equiv n < \infty$  (and similarly if  $p_I = 1$  and  $\dim E_I \equiv n_I < \infty$ ), then  $D(p, r)$  in (3.20) should be replaced by  $D(1, r) \ln(n+1)$  (and similarly  $D(p_I, r)$  should be replaced by  $D(1, r) \ln(n_I+1)$ ).

It is worth noting that, if  $r \geq \min \{2, p/(p-1)\}$  and  $r \geq \min \{2, p_I/p_I-1\}$ , then the laboriousness of the ‘stochastic’ version of the MD-method of solving games on  $L_p \times L_{p_I}$  is the same, as regards the order of its dependence on  $v$  and  $\dim E$  and  $\dim E_I$ , as for the version using an exact, first-order oracle.

## 6.4 SOLUTION OF CONVEX OPERATOR INEQUALITIES

Starting with this section and to the end of the chapter we shall be dealing with the application of MD-methods of solving games to the solution of various kinds of convex stochastic problems. The questions studied below could also be posed in the deterministic case, but there is no reason in the deterministic situation to consider them specially and separately (some of the questions have already been considered in Chapters 2 and 3, and the rest of them can be reduced immediately to the formulations in those chapters).

In this section we consider problems of solving operator inequalities. These problems are interesting in themselves, and because conditional problems of convex stochastic programming will be reduced to them in Section 6.5.

### 6.4.1

We shall be discussing the approximate solution of problems of the form

$$(H) \quad \text{find } x \in G \text{ such that } H(x) \leq 0. \quad (4.1)$$

Here  $G$  is a closed, convex, bounded subset of a Banach space  $E$ ,  $\|\cdot\|$ ,  $H: G \rightarrow E'$  is a mapping of  $G$  into a Banach space  $E'$ ,  $\|\cdot\|'$ , and  $E'$  is equipped with the structure of an ordered linear space, and so the property  $H(x) \leq 0$  is meaningful.

The ordering of  $E'$  with which we shall be concerned is equivalent to specifying a *non-negative cone*  $K \subset E'$ , i.e. a closed, non-void subset of  $E'$  such that

$$(x \in K, y \in K) \Rightarrow x + y \in K \text{ and } \lambda x \in K \text{ for all } \lambda \geq 0.$$

Such a set is necessarily convex and it contains 0. Fixing on a non-negative cone  $K$  enables us to define an ordering relation in  $E'$  (more precisely, a *pre-ordering relation*)  $\geq_K$ . Namely, we write  $x \geq_K y$  if  $x - y \in K$ . This relation has the natural ordering properties:

$$\begin{aligned} x &\geq_K x \text{ for all } x; \\ (x &\geq_K y, y \geq_K z) \Rightarrow x \geq_K z; \\ x &\geq_K y \Rightarrow \lambda x \geq_K \lambda y \text{ for all } \lambda \geq 0 \\ (x &\geq_K y \text{ and } z \geq_K w) \Rightarrow x + z \geq_K y + w. \end{aligned}$$

We shall also write  $x \leq_K y$  if  $y \geq_K x$ . Generally speaking,  $x \geq_K y$  and  $y \leq_K x$  do not imply  $x = y$ . The latter is true only under the additional assumption that  $K \cap (-K) = \{0\}$ .

We give a standard example of such an ordering. Let  $E' = R^m$  and let  $K = \{x = (x^1, \dots, x^m) \in R^m \mid x^i \geq 0\}$ . The ordering  $\geq_K$  is the usual ordering  $x \geq y \Leftrightarrow (\text{for all } i \leq m, x^i \geq y^i)$ . If  $E'$  in (4.1) has this structure, then (4.1) is an ordinary problem of solving a system of scalar inequalities.

Another example is  $E' = L_p(T, \mu)$  with  $K = \{x = x(t) \mid x(t) \geq 0\}$  is  $\mu$ -almost certain}. Here the inequality  $x \geq_K y$  is the usual (true  $\mu$ -almost everywhere) inequality between the functions  $x(t)$  and  $y(t)$ .

We need to introduce some further concepts. Let  $K$  be a non-negative cone in  $E$ , and let  $(E')^*$  be the dual space to  $E$ . In  $(E')^*$  there is a cone  $K^*$  dual to  $K$ :

$$K^* = \{\varphi \in (E')^* \mid \langle \varphi | x \rangle \geq 0 \text{ for all } x \in K\}.$$

It is easy to see that  $K^*$  actually is a cone. It specifies an ordering  $\geq_{K^*}$  in  $(E')^*$ .

Further, let  $S: E \rightarrow E'$  be a bounded linear operator. With it we can associate a linear operator (likewise bounded)  $S^*: (E')^* \rightarrow E^*$  operating on an element  $\varphi \in (E')^*$  according to the rule:  $S^*\varphi$  is a functional on  $E$  whose value on  $x \in E$  is  $\langle \varphi | Sx \rangle$ .

#### 6.4.2

Having explained the formulation of the problem, we pass on to the description of the classes of problem (4.1) which we shall be considering. These will be Lipschitz-convex problems of the (4.1) type. We make these ideas precise.

##### 6.4.2.1

A mapping  $H: G \rightarrow E'$  is said to be *Lipschitz with constant  $L$*  if, for some  $L < \infty$ , we have  $\|H(x) - H(x')\|' \leq L \|x - x'\|$  for all  $x, x' \in G$ . Such a mapping is said to be *convex* if

$$H\left(\frac{x+y}{2}\right) \leq_K \frac{1}{2}(H(x) + H(y)) \text{ for all } x, y \in G.$$

For example, if  $E'$  is  $E^n$  with the standard ordering (indicated above), then a convex Lipschitz mapping of  $G$  into  $E'$  is simply an  $n$ -dimensional vector-function on  $G$ , which is Lipschitz and convex in each of its co-ordinates. We denote the set of Lipschitz-convex mappings of  $G$  into  $E'$  by  $U \equiv U(G, E, \|\cdot\|, E', \|\cdot\|', K)$ . Each of the mappings  $H$  will be identified with the problem (4.1) corresponding to it.

##### 6.4.2.2

We now define a measure of the error of a point  $x \in G$  regarded as a solution of a problem  $H \in U$ . The exact solution is a point  $x \in G$  such that  $H(x) \in -K$ . It is now natural to measure the (absolute) error of a point  $x \in G$  regarded as a solution of the problem  $H$  by the distance from  $x$  to  $(-K)$ , i.e. by the quantity

$$\varepsilon(x, H) = \rho_{\|\cdot\|'}(H(x), -K). \quad (4.2)$$

For example, if  $E'$  is  $R^m$  with the standard ordering, and if  $\|\cdot\|'$  is  $\|\cdot\|_\infty$  (so that  $E'$  is simply  $I_\infty^{(m)}$  with the standard ordering and norm), then for  $H = (H^1(x), \dots, H^m(x))$ , where  $H^i(x)$  is scalar,  $\varepsilon(x, H) = \max_i H_+^i(x)$ , where  $H_+^i = \max \{0, H^i\}$ . Thus in this case the error is simply the maximal residual over the inequalities of the system at the point considered. We shall write

$$y \leq_K \varepsilon + z, \quad y, z \in E', \quad \varepsilon \geq 0, \quad \text{if } \rho_{\|\cdot\|'}(y - z, -K) \leq \varepsilon.$$

##### 6.4.2.3

We now describe the informational basis of the process of solving problems. Let  $\mathcal{O} = ((\Omega, \mathbf{F}_\omega); \psi(H, x, \omega))$  be the oracle with  $(\Omega, \mathbf{F}_\omega)$  as the space of noises, and with the observation functions  $\psi(H, x, \omega) = \{\psi^0(H, x, \omega); \psi^1(H, x, \omega)\}$ . The first component  $\psi^0(H, x, \omega)$  takes values in  $E'$  and serves as an estimate of the quantity  $H(x)$ , this estimate depending on the realization  $\omega$  of the oracle noise; and the second component  $\psi^1(H, x, \omega)$  takes values in the space  $L(E, E')$  of bounded Lipschitz operators acting from  $E$  into  $E'$ , and serves, roughly speaking, as an estimate of the differential of  $H$  at the point  $x$ . As in Section 6.1, the methods under consideration are able, at each stage of their work, to put a question to the oracle, about the problem  $H$  being solved, at a point  $x_i \in G$  and to obtain as answer the values of the estimate  $\psi(H, x_i, \omega_i)$  distorted by the noise  $\omega_i$  actually occurring at the  $i$ th step. The noises at different steps are assumed to be independent in the aggregate, and to be distributed according to the measure  $\mathbf{F}_\omega$ . It is, of course, assumed that  $(\Omega, \mathbf{F}_\omega)$  is a Polish space with a regular, Borel, Lebesgue-complete, probability measure, that  $E$  and  $E'$  are separable, and that the  $\psi^i(H, x, \omega)$  are Borel functions with respect to  $x$  and  $\omega$  for all  $H \in U$ . The Borel property of  $\psi^1$  is understood to mean that all functions of the form  $\langle \eta | \psi^1(H, x, \omega)e \rangle, e \in E, \eta \in (E')^*$  are Borel functions with respect to  $x$  and  $\omega$ .

##### 6.4.2.4

Let  $r > 1$ ,  $L > 0$ ,  $v_0 > 0$ . We associate with the parameters  $r$ ,  $L$ ,  $v_0$ , and the given oracle  $\mathcal{O}$  the class of problems  $U_{r, L}^{v_0} \equiv U_{r, L}^{v_0}(G, E, \|\cdot\|, E', \|\cdot\|', K)$  described as follows.

The field of problems of the class  $U_{r, L}^{v_0}$  consists of all problems  $H \in U$  such that the following two relations UA and UB hold:

$$\left. \begin{aligned} &\| \mathbf{M}_{\mathbf{F}_\omega} \psi^0(H, x, \omega) - H(x) \|' \leq \frac{v_0 L}{2} \\ &\mathbf{M}_{\mathbf{F}_\omega} \{ \psi^1(H, x, \omega)(y - x) \} \leq H(y) - H(x) + v_0 L, \end{aligned} \right\} \begin{array}{l} \text{UA} \\ \text{UB} \end{array}$$

for all  $x, y \in G$ . The meaning of the first of these relations UA is clear. The second means, roughly speaking, that  $\mathbf{M}_{F_\omega} \psi^1(H, x, \omega)$  (with accuracy  $v_0 L$ ) is a support linear mapping to the convex mapping  $H(x)$  (cf. with the case  $E' = \mathbb{R}$ );

$$\mathbf{M}_{F_\omega} (\|\psi^0(H, x, \omega)\|')' \leq \left(\frac{L}{4}\right)' \text{ for all } x \in G \quad \text{UB}$$

and

$$\mathbf{M}_{F_\omega} \|[\psi^1(H, x, \omega)]^* l\|_*' \leq \left(\frac{L}{2\rho_{\|\cdot\|}(G)}\right)',$$

for all

$$x \in G \text{ and } l \in K_* \text{ such that } \|l\|_*' \leq 1.$$

The (relative) error of a point  $x \in G$  regarded as an approximate solution of the problem  $H$  is defined as  $\varepsilon(x, H) = (1/L)\varepsilon(x, H)$ .

The oracle of the class is  $\emptyset$ .

We shall be concerned below with solving problems in the classes  $U_{r, L}^{e, v_0}$ .

**Exercise 1.** Let  $E' = \mathbb{I}_\infty^{(m)}$  with the natural ordering, so that the operators  $H \in U$  are vector-functions  $H^1(x), \dots, H^m(x)$ . Prove that the 'translation' of the definition of the class  $U_{r, L}^{e, v_0}(G, E, \|\cdot\|, \mathbb{I}_\infty^{(m)})$  into 'scalar' language gives the following definition:

- (i) problems in the class  $U_{r, L}^{e, v_0}$  are problems of solving systems of convex Lipschitz inequalities  $H^1(x) \leq 0, \dots, H^m(x) \leq 0, x \in G$ , and the observation functions of the oracle  $\emptyset$  have the form  $(\psi_1^0(H, x, \omega), \dots, \psi_m^0(H, x, \omega); \psi_1^1(H, x, \omega), \dots, \psi_m^1(H, x, \omega))$ , where the  $\psi_j^0$  are scalar observations of  $H^j(x)$ , and the  $\psi_j^1$  are observations of the support functionals to  $H^j$  at the point  $x$  (i.e. are functions with values in  $E'^*$ ), which are Borel functions with respect to  $x$  and  $\omega$ . Here the fact that  $\psi_j^1$  are Borel functions means, by definition, that all the functions  $\langle \psi_j^1(H, x, \omega) | e \rangle, e \in E$  are Borel functions with respect to  $x$  and  $\omega$ ;
- (ii) the class  $U_{r, L}^{e, v_0}$  consists precisely of those functions of the type mentioned for which the following four conditions hold:

$$|\mathbf{M}_{F_\omega} \psi_j^0(H, x, \omega) - H_j(x)| \leq \frac{1}{2} v_0 L, \quad x \in G, \quad 1 \leq j \leq m;$$

$$H^j(y) - H^j(x) \geq \mathbf{M}_{F_\omega} \{ \langle \psi_j^1(H, x, \omega) | y - x \rangle - v_0 L \}, \quad x, y \in G, \quad 1 \leq j \leq m;$$

$$\mathbf{M}_{F_\omega} \max_j \{ |\psi_j^0(H, x, \omega)| \}' \leq \left(\frac{L}{4}\right)', \quad x \in G;$$

$$\mathbf{M}_{F_\omega} \|\psi_j^1(H, x, \omega)\|_*' \leq \left(\frac{L}{2\rho_{\|\cdot\|}(G)}\right)', \quad x \in G, \quad 1 \leq j \leq m;$$

- (iii) the relative error of a point  $x \in G$  regarded as a solution of the problem  $H$  is defined as  $(1/L) \max \{ \max_{1 \leq j \leq m} H^j(x), 0 \}$ .

### 6.4.3

We now describe a way of solving operator inequalities in the class  $U_{r, L}^{e, v_0}$ . Let  $E, \|\cdot\|$  and  $(E')^*, \|\cdot\|_*$  be regular spaces satisfying the conditions in Section 5.1.1. Under these conditions we reduce the problems being solved to games.

#### 6.4.3.1

Let  $y \in E'$  and  $-y \notin K$ . By the Hahn–Banach theorem there is a functional  $\varphi \in (E')^*$  with norm equal to 1 such that

$$\langle \varphi | y \rangle \geq \sup_{z \in -K} \langle \varphi | z \rangle + \rho_{\|\cdot\|'}(y, -K).$$

In particular,  $\sup_{z \in -K} \langle \varphi | z \rangle$  is finite, and so  $\varphi \geq_{\kappa^*} 0$ . But then  $\sup_{z \in K^*} \langle \varphi | z \rangle = 0$ . Thus, for every  $y \notin -K$  there is a  $\varphi \in G_I = \{\varphi \in (E')^*: \|\varphi\|_*' \leq 1, \varphi \geq_{\kappa^*} 0\}$  such that  $\langle \varphi | y \rangle \geq \rho_{\|\cdot\|'}(y, -K)$ . On the other hand, it is clear that  $\langle \varphi | y \rangle \leq \rho_{\|\cdot\|'}(y, -K)$  when  $\varphi \in G_I$ . When  $y \in -K$  it is clear that  $\sup \{\langle \varphi | y \rangle | \varphi \in G_I\} = 0$ . So

$$s(y) \equiv \sup \{\langle \varphi | y \rangle | \varphi \in G_I\} = \rho_{\|\cdot\|'}(y, -K) \text{ for all } y. \quad (4.3)$$

#### 6.4.3.2

Having fixed  $G_I$ , we associate with the problem  $H$  the problem

$$(\bar{H}) \quad \bar{H}(x) = s(H(x)) \rightarrow \min | x \in G.$$

It is clear that the problem  $\bar{H}$  is equivalent to the problem  $H$  in the following precise sense:  $H$  is soluble if and only if  $\bar{H}_* \equiv \inf_{x \in G} \bar{H}(x) \leq 0$ . If this condition holds, then every solution of the problem  $\bar{H}$  is a solution of the problem  $H$ . Moreover, if we put  $\varepsilon(x, \bar{H}) = \bar{H}(x) - \inf_G \bar{H}$ , then, for every  $H$ , we have

$$\varepsilon(x, H) \leq \varepsilon(x, \bar{H}) + \bar{H}_*. \quad (4.4)$$

#### 6.4.3.3

It remains to learn how to solve the problem  $\bar{H}$ . To do this, we observe that, by definition of  $s(y)$ , we have

$$\bar{H}(x) = \sup_{l \in G_I} \langle l | H(x) \rangle. \quad (4.5)$$

We consider the function  $F_H(x, l) = \langle l | H(x) \rangle$ . It is easy to see that when  $H \in U$ , this function is Lipschitz with respect to  $x$  and  $l$  and convex with respect to  $x$  (because  $H$  is convex, and  $\varphi \geq_{\kappa^*} 0$  when  $\varphi \in G_I$ ) and linear (and also concave) with respect to  $l \in G_I$ . In view of the relation (4.5) the solutions of the problem  $\bar{H}$  are precisely the first components of the saddle-points of  $F_H$  on

$G \times G_I$ , and

$$\varepsilon(x, H) = \varepsilon(x, F_H). \quad (4.6)$$

Let us satisfy ourselves that we can construct by means of the oracle  $\mathcal{O}$  an oracle  $\bar{\mathcal{O}}$  of the type, described in Section 6.1.2.3, which is needed by the methods of solving games in Section 6.3. We define  $\bar{\mathcal{O}}$  as an oracle whose answers are constructed in the following way from the answers of  $\mathcal{O}$ :

$$\left. \begin{aligned} \psi_x(x, l; F_H, \omega) &= [\psi^1(H, x, \omega)]^* l, \\ \psi_l(x, l; F_H, \omega) &= \psi^0(H, x, \omega). \end{aligned} \right\} \quad (4.7)$$

In the second of the relations (4.7)  $\psi^0(H, x, \omega)$  is interpreted as a mapping with values in  $((E')^*)^*$  (the latter space is canonically identified with  $E'$ ).

We prove that, with such a definition of the oracle  $\bar{\mathcal{O}}$ , the games  $F_H$  generated by problems  $H \in U_{r, L}^{\varepsilon, v_0}(\cdot)$  fall into the class

$$\bar{D}_{r, L}^{\varepsilon, v_0} \equiv \bar{D}_{r, L}^{\varepsilon, v_0}(G \times G_I, E, \|\cdot\|, (E')^*, \|\cdot\|'_*).$$

By the same token, a solution of a problem  $\bar{H}$  with (mean absolute) error  $\varepsilon > 2v_0L$  can be obtained in the following way. Imitating the answers of the oracle  $\bar{\mathcal{O}}$  about the game  $F_H$  by means of the answers of the oracle  $\mathcal{O}$  about  $H$ , we solve the problem  $F_H$  with the relative accuracy  $\varepsilon/(2L)$  by the method in Section 6.3. Then the first component of the approximation obtained to the saddle-point of  $F_H$  is regarded as an approximate solution of the problem  $\bar{H}$ . The method just described solves the problem  $\bar{H}$  with (mean absolute) error  $\varepsilon$ .

#### 6.4.3.4

We now prove that the game  $F_H$  does indeed lie in  $\bar{D}_{r, L}^{\varepsilon, v_0}$ . We have to verify that the conditions A and B in Section 6.1.2.3 hold.

Let  $x, y \in G$  and  $l', l \in G_I$ . We have

$$\begin{aligned} F_H(y, l) - F_H(x, l) &= \langle l | H(y) - H(x) \rangle \\ &= \langle l | \mathbf{M}_{F_\omega} \{ \psi^1(H, x, \omega)(y - x) \} \rangle \\ &\quad + \langle l | \{ H(y) - H(x) - \mathbf{M}_{F_\omega} \psi^1(H, x, \omega)(y - x) \} \rangle. \end{aligned}$$

Because  $l \in K^*$  and  $\|l\|'_* \leq 1$  and also in view of the second relation in UA, the second term is not less than  $(-v_0L)$ . Thus

$$\begin{aligned} F_H(y, l) - F_H(x, l) &\geq \langle l | \mathbf{M}_{F_\omega} \psi^1(H, x, \omega)(y - x) \rangle - v_0L \\ &= \langle \mathbf{M}_{F_\omega} [\psi^1(H, x, \omega)]^* l | y - x \rangle - v_0L. \end{aligned}$$

This is the first of the inequalities in A. Further,

$$\begin{aligned} F_H(x, l') - F_H(x, l) &= \langle l' - l | H(x) \rangle \leq \|l' - l\|'_* \|H(x) - \mathbf{M}_{F_\omega} \psi^0(H, x, \omega)\|'_* \\ &\quad + \langle l' - l | \mathbf{M}_{F_\omega} \psi^0(H, x, \omega) \rangle \leq v_0L + \langle l' - l | \mathbf{M}_{F_\omega} \psi_l(x, l; F_H, \omega) \rangle \end{aligned}$$

(we have used the fact that  $\|l' - l\|'_* \leq 2$  because  $l, l' \in G_I$ , and also the first relation in UA). So A has now been proved.

Further,

$$\mathbf{M}_{F_\omega} \|\psi_x(x, l; F_H, \omega)\|'_* = \mathbf{M}_{F_\omega} \|[\psi^1(H, x, \omega)]^* l\|'_* \leq \left( \frac{L}{2\rho_{\|\cdot\|_1}(G)} \right)^r$$

because of UB and the fact that  $l \in G_I$ . Finally, in view of UB,

$$\mathbf{M}_{F_\omega} \|\psi_l(x, l, F_H, \omega)\|'_* = \mathbf{M}_{F_\omega} \{ \|\psi^0(H, x, \omega)\|' \}^r \leq \left( \frac{L}{2} \right)^r \leq \left( \frac{L}{2\rho_{\|\cdot\|_1}(G_I)} \right)^r.$$

It remains to observe that a solution of the problem  $\bar{H}$  with an absolute accuracy  $\varepsilon > 2Lv_0$  can be reduced to the solution of the problem  $F_H$  with the relative accuracy  $\varepsilon/(2L)$ . On the other hand, if  $H$  is compatible, then a solution of  $\bar{H}$  with an absolute accuracy  $\varepsilon$  is at the same time also a solution of  $H$  with the relative accuracy  $\varepsilon/L$ . Therefore we can solve compatible problems  $H \in U_{r, L}^{\varepsilon, v_0}$  with a relative accuracy  $v > 2v_0$ , and solution of such problems reduces to solving the game  $F_H$  with relative accuracy  $v/2$ .

Everywhere in the preceding we have, of course, been speaking of *mean* values of the accuracies.

What is to be done if the problem being solved is incompatible? Roughly speaking, the procedure described does the most that can be done in this case—the result  $\bar{x}$  of applying it is (in the mean with relative accuracy  $v$ ) such that  $H(\bar{x})$  realizes the minimum (over  $x \in G$ ) possible distance from  $H(x)$  to  $-K$ . Thus we may say that the method described is also suitable for analyzing insoluble problems.

For the purpose of our further exposition, an additional demand is imposed on the construction described: namely, in solving a problem  $H$ , to keep track of the quantity  $\bar{H}_*$ . This demand can be satisfied in the following way. Let  $N$  be the number of steps in the method which is being applied to  $F_H$ , of solving games as in Section 6.3.3; let  $\omega^i = (\omega_1, \dots, \omega_i)$  be the set of noises of the oracle in the first  $i \leq N$  steps of the work of the method, let  $(x^i, l^i)$  be the points at which the method puts questions to the oracle about the problem being solved, and let  $(\psi_l(x_i(\omega^{i-1}), l_i(\omega^{i-1}), F_H, \omega_i)) \equiv \bar{H}_i(\omega^i)$  be the observations of  $H(\bar{x}_i(\omega^{i-1}))$  communicated by the oracle  $\bar{\mathcal{O}}$ . We form the random variables

$$\bar{H}(\omega^N) = \frac{1}{N} \sum_{i=1}^N \bar{H}_i(\omega^i) \text{ and } \bar{a}(\omega^N) = s(\bar{H}(\omega^N)).$$

We remark that, after a realization of the work of the method corresponding to the oracle noise  $\omega^N$ , the quantity  $\bar{a}(\omega^N)$  becomes known to us. It turns out that  $\bar{a}(\omega^N)$  is quite a good estimate of  $\bar{H}_*$ ; if  $\tilde{v} > v_0$  is the accuracy to which the method under examination of solving games is constructed, then

$$\mathbf{F}_{\omega^N} \{ |\bar{a}(\omega^N) - \bar{H}_*| > 12\tilde{v}L \} \leq \frac{1}{4}. \quad (4.8)$$

To verify this it suffices to follow out the proof of Theorem 6.3.1 applied to a game  $F_H$  which is linear with respect to  $l$ . One has to operate in exactly the same way with regard to the Hilbert version in Section 6.3.3.

*Exercise 2. Prove (4.8).*

<Adding inequalities (3.8) and (3.10) and then summing the resulting inequality over  $j = 1, 2, \dots, N$ , we have, taking,  $\rho_j = \rho$ ,  $1 \leq j \leq N$ , that

$$\sum_{j=1}^N \rho(F_H(\bar{x}_j(\omega^{j-1}), l) - F_H(x, \bar{l}_j(\omega^{j-1}))) \leq A(\omega^N), \quad (4.9)$$

where  $A(\omega^N)$  does not depend on  $x$  and  $l$ , and

$$\mathbf{M}_{F_{\omega^N}} A(\omega^N) \leq 2\tilde{v}LN\rho. \quad (4.10)$$

We can take the right-hand side of (3.13) as  $A(\omega^N)$ , and then (4.10) will follow from the rules for the choice of  $\rho$  and  $N$  for the method under consideration.

Substituting for  $x$  in (4.9) any component  $x^*$  of the saddle-point of  $F_H$ , and using the fact that we then have, for all  $l' \in G_l$ ,

$$F_H(x^*, l') \leq \bar{F}_H(x^*) = \inf_{x \in G} \bar{F}_H(x) = \bar{H}_*,$$

we obtain from (4.9)

$$\sum_{j=1}^N \rho(F_H(\bar{x}_j(\omega^{j-1}), l) - \bar{H}_*) \leq A(\omega^N).$$

Hence

$$\frac{1}{N} \sum_{j=1}^N F_H(\bar{x}_j(\omega^{j-1}), l) \leq \bar{H}_* + \frac{1}{N\rho} A(\omega^N). \quad (4.11)$$

Recalling the definition of  $F_H$ , we hence obtain that, for all  $l \in G_l$ ,

$$\left\langle \frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1})) | l \right\rangle \leq \bar{H}_* + \frac{1}{N\rho} A(\omega^N),$$

or

$$s\left(\frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1}))\right) \leq \bar{H}_* + \frac{1}{N\rho} A(\omega^N).$$

At the same time, because  $H(x)$  is convex, we have

$$s\left(\frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1}))\right) \geq s\left(H\left(\frac{1}{N} \sum_{j=1}^N \bar{x}_j(\omega^{j-1})\right)\right) \geq \bar{H}_*.$$

Thus, for all  $\omega^N$

$$\bar{H}_* \leq s\left(\frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1}))\right) \leq \bar{H}_* + \frac{1}{N\rho} A(\omega^N). \quad (4.12)$$

Further, by the properties of the oracle  $\mathcal{O}$  and by assertion (2) of the

proposition in Section 5.1.5, we have

$$\mathbf{M}_{F_{\omega^N}} \left\| \sum_{j=1}^N \rho(H(\bar{x}_j(\omega^{j-1})) - \tilde{H}_j(\omega^j)) \right\|' \leq \frac{\rho}{2} v_0 LN + \frac{N}{2} \rho L \omega_{V_l, r} \left( \frac{\rho L}{2} \right) + \frac{1}{2},$$

and hence, because  $s$  is Lipschitz with constant 1 relative to  $\|\cdot\|'$ ,

$$\begin{aligned} \mathbf{M}_{F_{\omega^N}} \left| s\left(\frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1}))\right) - s\left(\frac{1}{N} \sum_{j=1}^N \tilde{H}_j(\omega^j)\right) \right| \\ \leq \frac{v_0 L}{2} + \frac{L}{2} \omega_{V_l, r} \left( \frac{\rho L}{2} \right) + \frac{1}{2N\rho} \leq \tilde{v}L \end{aligned} \quad (4.13)$$

(we have taken into account the rule in Section 5.3.3 for choosing  $\rho$  and  $N$ ). Combining (4.12) with (4.13), and using (4.10), we obtain

$$\mathbf{M}_{F_{\omega^N}} \{ |\bar{H}_* - \tilde{a}(\omega^N)| \} \leq 3\tilde{v}L. \quad (4.14)$$

It is clear that (4.8) is an immediate consequence of (4.14). >

*Remark.* The constant on the left-hand side of (4.8) is somewhat high. We could reduce it by working more accurately, but we shall not do this in order not to make the exposition lengthier. We shall operate in a similar way also in the following. The point is that in this and the next sections we are concerned primarily with the principles of the business and not with the immediate construction of practical recommendations. If necessary, the reader himself will be able to make the constant suitably more precise and at this cost reduce the laboriousness of the methods described.

## 6.5 SOLUTION OF EXTREMAL PROBLEMS WITH OPERATOR CONSTRAINTS

### 6.5.1

We shall consider problems which are more general than (4.1), namely, convex extremal problems with operator constraints of the form

$$f \equiv (f_0, H_f); f_0(x) \rightarrow \min | x \in G, H_f(x) \leq 0. \quad (5.1)$$

Here  $H_f: G \rightarrow E'$  is assumed to be a mapping belonging to a set  $U(G, E, \|\cdot\|, E', \|\cdot\|', K) \equiv \bar{U}$ , and  $f_0$  is a convex, Lipschitz, real-valued function defined on  $G$ . As regards  $E$ ,  $\|\cdot\|$  and  $E'$ ,  $\|\cdot\|'$  we shall assume the conditions stated at the beginning of Section 6.4.3 to be satisfied; as usual,  $G$  is assumed to be a non-empty, convex, closed, bounded subset of  $E$ . Convex conditionally extremal problems constitute a particular case of the problems (5.1) obtained when  $E' = \ell_\infty^{(m)}$  with the natural ordering of this space. We shall denote the field

of all problems of the type described, by

$$\mathcal{W} \equiv \mathcal{W}(G, E, \|\cdot\|, E', \|\cdot\|', K).$$

### 6.5.2

Our purpose is to learn how to solve problems in  $\mathcal{W}$ , using a stochastic, first-order oracle. Roughly speaking, this oracle will communicate observations (distorted by noise) of the values and differentials of the target function  $f_0$  and of the constraint  $H$ . We shall consider an oracle  $\mathcal{O} = ((\Omega, \mathbf{F}_\omega), \varphi^0(x, f, \omega), \varphi^1(x, f, \omega); \psi^0(x, f, \omega), \psi^1(x, f, \omega))$ , in which  $\varphi^0$  is scalar,  $\varphi^1$  takes values in  $E^*$ ,  $\psi^0$  takes values in  $E'$ , and  $\psi^1$  in  $L(E, E')$ . Here  $\varphi^0$  is an estimate of  $f_0$ ,  $\varphi^1$  is an estimate of the support functional to  $f_0$ ,  $\psi^0$  is an estimate of  $H_f$ , and  $\psi^1$  an estimate of the differential of  $H_f$ . As always, in the  $i$ th step of the work the method can put a question to the oracle  $\mathcal{O}$  about the problem being solved, at a point  $x_i \in G$ , and it obtains in answer the values of

$$\varphi^0(x_i, f, \omega_i), \varphi^1(x_i, f, \omega_i), \psi^0(x_i, f, \omega_i), \psi^1(x_i, f, \omega_i).$$

Here  $\omega_i$  are the noises in the oracle; they are independent and distributed according to the measure  $\mathbf{F}_\omega$ . Naturally,  $\Omega, \mathbf{F}_\omega$  is assumed to be a Polish space with a regular, Borel, Lebesgue-complete, probability measure, and  $\varphi^0, \varphi^1, \psi^0, \psi^1$  are Borel functions relative to  $x$  and  $\omega$ .

We now describe the classes of problems we shall study. We fix  $r > 1$  and  $L > 0$ , and also  $\omega_0 > 0$ , and we denote by

$$\tilde{\mathcal{W}}_{r,L}^{\epsilon, v_0} \equiv \tilde{\mathcal{W}}_{r,L}^{\epsilon, v_0}(G, E, \|\cdot\|, E', \|\cdot\|', K)$$

the class of problems of the form (5.1) described as follows.

The field of problems of the class  $\tilde{\mathcal{W}}_{r,L}^{\epsilon, v_0}$  consists of all problems

$$f = (f_0, H_f) \in \mathcal{W}(G, E, \|\cdot\|, E', \|\cdot\|', K)$$

for which

(1) the mapping  $H_f$  falls into the class  $U_{r,L}^{\epsilon, v_0}(G, E, \|\cdot\|, E', \|\cdot\|', K)$  (see Section 6.4.2.3 where  $\tilde{\mathcal{O}}$  is the oracle induced by the oracle  $\mathcal{O}((\Omega, \mathbf{F}_\omega); \psi^0, \psi^1)$ .

(2)

$$\left. \begin{aligned} |\mathbf{M}_{\mathbf{F}_\omega} \varphi^0(x, f, \omega) - f_0(x)| &\leq \frac{v_0 L}{2}; \\ \mathbf{M}_{\mathbf{F}_\omega} |\varphi^0(x, f, \omega)|^r &\leq \left(\frac{L}{4}\right)^r, \end{aligned} \right\} \quad (5.2)$$

$$\left. \begin{aligned} f_0(y) - f_0(x) &\geq \langle \mathbf{M}_{\mathbf{F}_\omega} \varphi^1(x, f, \omega) | y - x \rangle - v_0 L; \\ \mathbf{M}_{\mathbf{F}_\omega} \|\varphi^1(x, f, \omega)\|_*^r &\leq \left(\frac{L}{2\rho_{\|\cdot\|}(G)}\right)^r \end{aligned} \right\} \quad (5.3)$$

for all  $x, y \in G$ .

The conditions (1)–(2) are the conditions, of the standard type for our considerations, on the possible bias and moments of the stochastic information provided by the oracle.

The *oracle* of the class is  $\mathcal{O}$ .

The *measure of the absolute error* of a point  $x \in G$  regarded as a solution of a problem  $f$  is defined as

$$\varepsilon(x, f) = \max \{f_0(x) - f_*; s(H_f(x))\}.$$

Here

$$f_* = \begin{cases} +\infty & \text{for all } x \in G \quad H(x) \leq 0, \\ \inf \{f(x) | x \in G, H(x) \leq 0\} & \text{otherwise,} \end{cases}$$

is the optimal value of the objective functional of the problem  $f$ , and  $s(x)$  is the function introduced in Section 6.4.3.

The *measure of the relative error* of a point  $x \in G \cup \{*\}$  regarded as an approximate solution of a problem  $f$  (as usual,  $*$  is interpreted as the answer that  $f$  is incompatible) is defined by the formula

$$v(x, f) = \begin{cases} 0, x = *, & f \text{ is incompatible,} \\ 1, x = *, & f \text{ is compatible,} \\ \frac{1}{L} \varepsilon(x, f) & \text{in the case } x \in G. \end{cases}$$

### 6.5.3

We fix the class of problems

$$\tilde{\mathcal{W}}_{r,L}^{\epsilon, v_0} \equiv \tilde{\mathcal{W}}_{r,L}^{\epsilon, v_0}(G, E, \|\cdot\|, E', \|\cdot\|', K)$$

and we shall construct a method of solving problems in this class with a relative accuracy  $v$ . We shall assume that the error  $v_0$  of the oracle is quite small compared with  $v$ , let us say

$$800v_0 < v. \quad (5.4)$$

The idea of the construction effected below is to reduce the optimization problem (5.1) to the solution of a series of operator inequalities. We shall explain how this is done. In the description it is convenient to assume that  $\rho_{\|\cdot\|}(G) = 1$ . From now on we make this assumption.

#### 6.5.3.1

We begin with a very simple case: it is required to solve a problem  $f \in \tilde{\mathcal{W}}_{r,L}^{\epsilon, v_0}$ , and it is known *a priori* that  $f_* = 0$ . In this case, solving the problem  $f$  reduces immediately to solving an operator inequality, namely,

$$f_0(x) \leq 0, \quad H_f(x) \leq 0, \quad x \in G. \quad (5.5)$$

(5.5) can, in turn, be written as

$$H^f(x) \leq 0, \quad x \in G, \quad (5.6)$$

where  $H^f = (f_0, H_f)$  is a mapping of  $G$  into  $\tilde{E} = \mathbb{R} \times E'$ , and the ordering in  $\tilde{E}$  is specified by a cone  $\tilde{K} = \{(t, y) | t \geq 0, y \in K\}$ . Moreover, it is clear that

$$H^f \in U(G, E, \|\cdot\|, \tilde{E}, \|\cdot\|^\sim, \tilde{K}) \equiv U;$$

let us agree to take  $\|t, y\|^\sim = \max\{|t|, \|y\|\}$ .

The problem  $H^f$  is equivalent to the problem  $f$  in the following sense:  $H^f$  is compatible whenever  $f$  is, and, for every  $x \in G$ , we have  $\varepsilon(x, H^f) = \varepsilon(x, f)$ . Thus, to solve the problem  $f$  with a given mean (absolute) accuracy  $\varepsilon$  is just the same as solving the problem  $H^f$  in the same sense.

We now examine whether we have at our disposition a source of information about the problem  $H^f$  such as is required by the method in Section 6.4. The answer to this question is, of course, affirmative. For, the components  $\varphi^0, \psi^0$  of the observation function of the oracle  $\mathcal{O}$  together form the necessary estimate of  $H^f(x)$ , and the components  $\varphi^1, \psi^1$  form the necessary estimate for the differential of  $H^f$ . Thus the oracle  $\mathcal{O}$  can be interpreted as the oracle for the field of problems  $U$  in Section 6.4.2.3. Moreover, comparison of the definitions of the class of problems  $\tilde{U}_{r,L}^{e,v_0}$  and the class of problems  $\tilde{W}_{r,L}^{e,v_0}$  leads to the conclusion that  $H^f \in \tilde{U}_{r,L}^{e,v_0}$ . Moreover,  $\tilde{E}$  is regular whenever  $(E')^*$  is; and so the method in Section 6.4 can be applied to  $H^f$ . By solving  $H^f$  by this method, constructed for a mean relative accuracy  $v/2$ , we obtain a solution of the problem  $f$  with a mean relative accuracy  $v$ .

It is clear that, if it is known *a priori* that  $f_* < \infty$  (but not necessarily equal to 0), then we can proceed as in the case  $f_* = 0$  but with the problem  $f = \{f_0 - f_*, H_f\}$ . The only difference from the preceding case is that (under the proviso  $v_0 < \frac{1}{2}$ , which is of no consequence for the sequel) the corresponding operator inequality, as may easily be seen, falls into the class  $\tilde{U}_{r,4L}^{e,v_0}$  and not into  $\tilde{U}_{r,2L}^{e,v_0}$  as before, and so solution of the problem  $f$  with relative accuracy  $v$  requires the application to  $H^f$  of a method constructed to give an accuracy  $v/4$ .

### 6.5.3.2

We now consider the general case when  $f_*$  is not known *a priori*. One can first try to find  $f_*$ , and then proceed according to the recommendations in Section 6.5.3.1. We describe a procedure for finding  $f_*$ . Let  $f \in \tilde{W}_{r,L}^{e,v_0}$ , and, for every  $t \in \mathbb{R}$ , let  $f^t$  be the problem  $(f_0 - t, H_f)$ . Clearly,  $f$  is compatible if and only if, for some  $t$ , the problem  $H^{f^t}$  is compatible; and then

$$f_* = \max\{t | H^{f^t} \text{ is compatible}\}.$$

We observe that, when  $2|t| \leq L$ , the problem  $H^{f^t}$  falls into the class  $\tilde{U}_{r,L}^{e,v_0}$ . If we now apply to  $H^{f^t}$  the method described in Section 6.4 of solving problems

of this class, constructed for a mean relative accuracy  $\tilde{v} > 2v_0$  (we denote this method by  $\mathcal{B}_{\tilde{v}}$ ), then, in accordance with what was said in Section 6.4, we obtain a (random) result  $\tilde{x}_t \in G$  such that

$$\mathbf{M}\varepsilon(\tilde{x}_t, H^{f^t}) \leq 4\tilde{v}L + \overline{H_*^{f^t}}, \quad (5.7)$$

and also a (random) estimate  $\tilde{a}_t$  of the number  $\overline{H_*^{f^t}}$  such that

$$\mathbf{Pr}\{|\tilde{a}_t - \overline{H_*^{f^t}}| > 48\tilde{v}L\} \leq \frac{1}{4}. \quad (5.8)$$

Here  $\mathbf{M}$  denotes the mean, and  $\mathbf{Pr}$  the probability calculated from the distribution of the realizations of  $\mathcal{B}_{\tilde{v}}$  on the method  $H^{f^t}$  (see (4.4), (4.8)).

Further, let  $g(t) = H_*^{f^t}$ . From the definition of  $\varepsilon(x, f)$  we have

$$\begin{aligned} \varepsilon(x, f) &= \max\{f_0(x) - f_*, s(H_f(x))\} \\ &\leq [t - f_*]_+ + \max\{f_0(x) - t, s(H_f(x))\} \\ &\leq [t - f_*]_+ + \varepsilon(x, H^{f^t}). \end{aligned}$$

Hence and from (5.7) we obtain

$$\mathbf{M}\varepsilon(\tilde{x}_t, f) \leq [t - f_*]_+ + g(t) + 4\tilde{v}L, \quad (5.9)$$

or

$$\mathbf{M}\varepsilon(\tilde{x}_t, f) \leq \frac{[t - f_*]_+}{L} + \frac{g(t)}{L} + 4\tilde{v}. \quad (5.10)$$

Hence it is clear that to solve the problem  $f$  with (mean relative) accuracy  $v$  it is sufficient to seek an approximation  $t^*$  of the quantity  $f_*$  such that

$$\frac{1}{L}(t^* - f_*) = O(v) \quad \text{and} \quad \frac{1}{L}g(t^*) = O(v),$$

after which we apply to  $H^{f^{t^*}}$  a method  $\mathcal{B}_{\tilde{v}}$  with  $\tilde{v} = O(v)$ .

We have still to see how to find the required  $t^*$ . We note that, by definition,  $g(t)$  is a non-decreasing function which is Lipschitz with constant 1. Moreover,

$(f \text{ is compatible}) \Leftrightarrow (\text{the equation } g(t) = 0 \text{ is solvable}),$

and in the latter case  $t^*$  is the least root of  $g(t) = 0$ . From what has been said it is clear that we may take as  $t^*$  a sufficiently exact approximation to the least root of the equation  $g(t) = 0$ . The latter can be obtained by the 'repeated bisection of the segment' method, and to determine the values of  $g(t)$  we can make use of the estimate  $\tilde{a}_t$  provided by applying  $\mathcal{B}_{\tilde{v}}$  (with a suitable  $\tilde{v}$ ) to  $H^{f^t}$ .

### 6.5.3.3

We now implement the projected plan. We describe the procedure  $\mathcal{B}^v$  of solving problems in  $\tilde{W}_{r,L}^{e,v_0}$  with a specified (mean relative) accuracy  $v$  satisfying (5.4). We note that, from the definition of the class, it follows that, for

$f \in \tilde{W}_{r,L}^{v_0}$ , the variation of the functions  $f_0$  and the maximum of the function  $s(H_j)$  on  $G$  do not exceed  $(1+2v_0)L/2$ , and so the only case of interest is  $v < 1+v_0$ .

Thus, we assume that

$$800v_0 < v < 1+v_0. \quad (5.11)$$

Under this assumption the procedure  $\mathcal{B}^*$  as applied to a problem  $f \in \tilde{W}_{r,L}^{v_0}$  consists of  $N_1 \equiv N_1(v)$  search stages and one concluding stage (the form of the function  $N_1(v)$  will be shown later). At the  $i$ th stage of the search we shall know the points  $\underline{t}_{i-1}, \bar{t}_{i-1}$ , the ends of the interval localizing  $t^*$  at that stage. Initially,  $\underline{t}_0 = -L/2, \bar{t}_0 = L/2$ .

In the  $i$ th search stage the point  $t_i = (\underline{t}_{i-1} + \bar{t}_{i-1})/2$  is constructed. The procedure  $\mathcal{B}_\tilde{v}$  for solving the operator inequalities is applied  $N_2 \equiv N_2(v)$  times to the problem  $H^{J^i}$  with a certain  $\tilde{v} > v_0$ . The form of the function  $N_2(v)$  and the value of  $\tilde{v}$  will be shown later.

In the  $j$ th application of  $\mathcal{B}_\tilde{v}$  to  $H^{J^i}$  the observed realization  $\tilde{a}_{t_i}^j$  of the bound  $\tilde{a}$  for the quantity  $H_*^{J^i} = g(t_i)$  is fixed. After the numbers  $\tilde{a}_{t_i}^1, \dots, \tilde{a}_{t_i}^{N_2}$  have been obtained, the number

$$\gamma_i = \begin{cases} 1 & \text{if the bigger half of } \tilde{a}_{t_i}^j \text{ is greater than } 50\tilde{v}L, \\ 0 & \text{otherwise,} \end{cases}$$

is constructed. The numbers

$$\underline{t}_i = \begin{cases} \underline{t}_{i-1}, & \gamma_i = 0, \\ t_i, & \gamma_i = 1, \end{cases} \quad \bar{t}_i = \begin{cases} t_i, & \gamma_i = 0, \\ \bar{t}_{i-1}, & \gamma_i = 1 \end{cases}$$

are then constructed. This ends the step.

When  $i < N_1$  we go to the next search stage, and when  $i = N_1$  to the concluding stage. In the latter case we put  $t^* = \underline{t}_{N_1}$ .

In the concluding stage it is tested whether  $\gamma_i = 1$  all the time (i.e. for  $1 \leq i \leq N_1$ ). If 'yes', then the result of the work of  $\mathcal{B}^*$  on  $f$  is equal to  $f_*$ . If 'no', we apply  $\mathcal{B}_\tilde{v}$  to  $H^{J^*}$ , and we call the result obtained the result of applying  $\mathcal{B}^*$  to  $f$ .

#### 6.5.3.4

We now analyse the work of the method  $\mathcal{B}^*$ , and we prove that, with an appropriate choice of  $N_1(v)$  and  $N_2(v)$ , the specified mean accuracy in solving any problem  $f \in \tilde{W}_{r,L}^{v_0}$  will indeed be secured. We fix on an  $f$ .

1°. We shall say that there was *success* in the  $i$ th search stage if neither of the following events occurred:

- (i)  $g(t_i)$  was  $\leq 2\tilde{v}L$  and  $\gamma_i$  was  $= 1$ ;
- (ii)  $g(t_i)$  was  $> 100\tilde{v}L$  and  $\gamma_i$  was  $= 0$ .

Otherwise the  $i$ th stage will be said to be *unsuccessful*.

2°. We fix a realization of the oracle noises on all  $N_2$  search stages. Suppose it is such that the following condition is satisfied:

(A) each of the  $N_2$  search stages is successful.

We shall prove that, if  $N_1$  and  $\tilde{v}$  satisfy the relations

$$2\tilde{v} - 2^{-N_1} \geq v_0, \quad 100\tilde{v} + 2^{-N_1} + 4\tilde{v} \leq \frac{1}{2}v, \quad (5.12)$$

then, under the condition (A), the mean (over realizations of the oracle noise in the concluding stage of the work of the method) error of the result does not exceed  $v/2$ .

For, suppose the condition (A) holds. It is possible that we might then have  $\gamma^i = 1$ ,  $1 \leq i \leq N_1$ . Under condition (A) this means that, for all  $i$ , we would have  $g(t_i) > 2\tilde{v}L$ , and then

$$0 < \frac{1}{2}L - t_{N_1} \leq L \cdot 2^{-N_1}.$$

Since  $g(t)$  is Lipschitz with constant 1, we would then have

$$g(L/2) > 2\tilde{v}L - L \cdot 2^{-N_1} \geq v_0L.$$

(we have used the first relation in (5.12)). It is easy to see that the last relation is impossible for compatible  $f$  (one has to use the fact that, by the definition of the class,  $|f_0| \leq (1+2v_0)L/4$  on  $G$ ). Thus, in the case considered where (A) holds and  $\gamma_i \equiv 1$ , the mean error of the result of the work of the method is equal to 0.

Now suppose (A) is satisfied, but not all the  $\gamma_i = 1$ . Then  $t^*$  is well-defined, and from its definition and the satisfaction of (A) it follows that

- (i) there is a certain  $i_0$  such that  $t_{i_0} \geq t^*$  and  $\gamma_{i_0} = 0$ . Then  $t_{i_0} - t^* < L \cdot 2^{-N_1}$ ;
- (ii) either there is an  $i_1$  such that  $t_{i_1} = t^*$  and  $\gamma_{i_1} = 1$  (case 1), or  $t^* = -L/2$  (case 2).

By the first of these remarks and condition (A), we have  $g(t_{i_0}) \leq 100\tilde{v}L$ . But then

$$g(t^*) \leq g(t_{i_0}) + t_{i_0} - t^* \leq 100\tilde{v}L + L \cdot 2^{-N_1}. \quad (5.13)$$

Moreover, in case 1,  $g(t^*) > 0$ , i.e.  $t^* \leq f_*$ . In case 2, it is clear that  $t^* \leq f_*$  (we have used the fact that

$$f_* \geq -L(1+2v_0)/4, \quad v_0 < \frac{1}{4}.$$

So, in all cases,

$$t^* - f_* \leq 0. \quad (5.14)$$

Combining (5.13), (5.14), and (5.10), we see that, under the conditions (5.12), when (A) is satisfied the mean error of the result produced by the method does indeed not exceed  $v/2$ .

We observe that the inequalities (5.12) are satisfied if we take

$$\tilde{v} = c_0v \quad \text{and} \quad N_1(v) = c_1 \ln(2/v) \quad (5.15)$$

with suitable absolute constants  $c_0$  and  $c_1$  (and here  $c_0 \geq 1/200$ , so that

$\tilde{v} > 4v_0$ ). We could write out these constants, but we shall not pause to do so, in order not to encumber the exposition.

3°. Now let  $\Pr(A) = p$ . Then the mean relative error  $v(\mathcal{B}^v, f)$  of the method  $\mathcal{B}^v$  on the problem  $f$  under the conditions (5.15) does not exceed  $v/2 + (1 + v_0)(1 - p)$ . For, in this case, the mean error of the method (for a fixed noise in the search stages) does not exceed  $v/2$  under the condition (A), or  $(1 + v_0)$  in all other cases. In order to construct a method of the specified accuracy, it suffices to supplement the conditions (5.15) with a rule for choosing  $N_2(v)$  such that, under it, the condition

$$1 - p \leq \frac{v}{2(1 + v_0)} \left( \geq \frac{v}{4} \right) \quad (5.16)$$

is satisfied (we have used the fact that it follows from (5.4) and the condition  $v < v_0 + 1$  that  $v_0 < 1$ ).

We shall prove that (5.16) is ensured by the choice

$$N_2(v) = c_2 \ln(2/v) \quad (5.17)$$

with a suitable absolute constant  $c_2$ . Thus the choice of  $\tilde{v}, N_1, N_2$  according to (5.15)–(5.17) enables us to obtain a method of solving problems in  $\tilde{\mathcal{W}}_{r,L}^{v,v_0}$  with accuracy  $v$ . We estimate  $1 - p$  in the following way. We look at the  $i$ th search stage. We consider the conditional distribution of the series of estimates  $\{\tilde{a}_{t_i}^j\}_{j=1}^{N_2}$  obtained when the ‘previous history’ of the solution process up to the  $i$ th stage is fixed. It is clear that under these conditions the  $\tilde{a}_{t_i}^j$  are independent as an aggregate and that each of the events  $A_j = \{|\tilde{a}_{t_i}^j - g(t_i)| > 48\tilde{v}L\}$  has a probability  $\leq \frac{1}{4}$ . Then the simultaneous occurrence of  $N_2/2$  or more of these events is possible with a probability  $\leq \exp(-\kappa N_2)$ , where  $\kappa > 0$  is an absolute constant. It is clear that if less than half of the events  $A_j$  occur, then the  $i$ th stage is successful. It follows from what has been said that the probability of simultaneous success of all  $N_1$  of the search stages (i.e. the probability of the event A) is  $\geq 1 - N_1 \exp(-\kappa N_2)$ , and this provides the estimate  $1 - p \leq N_1 \exp(-\kappa N_2)$ . Recalling the definition of  $N_1(v)$ , we find that, with a suitable choice of  $c_2$ , (5.17) implies (5.16).

#### 6.5.4

Later, when speaking of a method of solving problems in  $\tilde{\mathcal{W}}_{r,L}^{v,v_0}$  with accuracy  $v$ , we shall have in mind the method  $\mathcal{B}^v$ , provided by the construction described above, with the parameters (5.15)–(5.17). The laboriousness of this method, under the conditions (5.11), is bounded above by the function

$$M_w(v) = c \lceil \ln^2(1/v) \rceil \tilde{M}(v/d),$$

where  $c$  and  $d$  are absolute constants, and  $\tilde{M}(v)$  is the laboriousness of an

MD-method, constructed for accuracy  $v$ , of solving games in a class of the  $\tilde{\mathcal{D}}_{r,L}^{v,0}(G \times G; E, \|\cdot\|, \tilde{E}, \|\cdot\|)$  type.

## 6.6 SOLUTION OF CONSTRAINED STOCHASTIC PROBLEMS

### 6.6.1

The results of the previous section are applicable almost immediately to solving constrained extremal, convex, stochastic problems (problems in the classes  $\tilde{\mathcal{C}}_{r,L}^{v,v_0}(G, E, \|\cdot\|, m)$ ) introduced in Section 5.1. For, every such problem

$$f: \{f_0(x) \rightarrow \min |x \in G \subset E, f_j(x) \leq 0, 1 \leq j \leq m\}$$

can be regarded as a problem with operator constraints of the form

$$f: f_0(x) \rightarrow \min |x \in G, H_f(x) \leq 0,$$

where  $H_f(x) = (f_1(x), \dots, f_m(x))$  is a mapping of  $E$  into  $E' = I_\infty^{(m)}$  ( $I_\infty^{(m)}$  is provided with its natural ordering). The answers of the oracle  $\mathcal{C}$  about the problem  $f$  can be interpreted as the answers of the oracle considered in Section 6.5 about the problem  $(f_0, H_f)$ . In other words, the field of problems  $\tilde{\mathcal{C}}_{\text{Lip}}(G, E, \|\cdot\|, m)$  is identified directly with the field of problems  $\mathcal{W} \equiv \mathcal{W}(G, E, \|\cdot\|, I_\infty^{(m)}, \|\cdot\|_\infty, K_\infty)$ , where  $K_\infty$  is the standard cone in  $I_\infty^{(m)}$ , and the oracle  $\mathcal{C}$  can be regarded as the oracle considered in Section 6.5.2 for the field of problems  $\mathcal{W}$ .

The absolute error of a point  $x \in G$  regarded as a solution of the problem  $f \in \tilde{\mathcal{C}}_{\text{Lip}}$ —the value of  $\max \{f_0(x) - f_*, f_1(x), \dots, f_m(x)\} \equiv \epsilon(x, f)$ —is at the same time the absolute error of the point  $x$  as a solution of the same problem  $f$  regarded as an element of  $\mathcal{W}$ . Thus, for solving problems in the classes  $\tilde{\mathcal{C}}_{r,L}^{v,v_0}$  it would seem possible to apply methods of solving problems in the classes  $\tilde{\mathcal{W}}_{r,L}^{v,v_0}$ .

However, without special additional transformations of the original problem it is impossible to effect this reduction. The point is that, in general, not only does a problem  $f \in \tilde{\mathcal{C}}_{r,L}^{v,v_0}$  not fall into the class  $\tilde{\mathcal{W}}_{r,L}^{v,v_0}$ , it does not even fall into a class  $\tilde{\mathcal{W}}_{r,L}^{v,v_0}$  where  $\tilde{v}_0$  and  $\tilde{L}$  are an absolute constant times greater than  $v_0$  and  $L$  respectively. In the definition of the class  $\tilde{\mathcal{C}}_{r,L}^{v,v_0}$  the constraints on the observations of the values of  $f_i(x)$  concerned only the deviations of these observations from the true values, whereas in the definition of the classes  $\tilde{\mathcal{W}}_{r,L}^{v,v_0}$  the constraints on these same observations concerned not only the deviations but the values of the observations themselves.

*Exercise 1.* Indicate where in Sections 6.4–6.5 the *a priori* limitations on the magnitudes of the observations of values of the components in the problem were used in the definitions of the corresponding classes.

*Answer.* It was necessary to have the possibility of reducing the business to a

game in which the observations of the support functionals for both components were effectively estimated. >

### 6.6.2

The difficulty mentioned above is easily overcome. To do this, we use the preparatory procedure described later. As everywhere in this chapter, it is assumed that  $G, E, \|\cdot\|$  satisfy the conditions stated at the beginning of Section 5.1.

Suppose it is required to solve a problem  $f \in \tilde{C}_{r,L}^{\epsilon,v_0}(G, E, \|\cdot\|, m)$ . Let  $\bar{x}$  denote the 'centre' of the set  $G$ . Suppose we have available a family of statistical estimates  $a_\delta, \delta > 0$ , of the vector  $f(\bar{x})$  such that

$$\Pr\{\|a_\delta - f(\bar{x})\|_\infty > L\} \leq \delta. \quad (6.1)$$

It is assumed that the estimate  $a_\delta$  is constructed from  $R(\delta)$  observations of the vector  $f(\bar{x})$  provided by the oracle  $\mathcal{O}$  (to which it is necessary to address  $R(\delta)$  questions about the problem  $f$  at the point  $\bar{x}$ ). We fix a  $\delta > 0$  and consider a problem  $\tilde{f}$  which is constructed from the realized value  $a = (a^0, \dots, a^m)$  of the estimate  $a_\delta$ :

$$\tilde{f} = (\tilde{f}_0, \dots, \tilde{f}_m) : \tilde{f}_0(x) = f_0(x) - a^0,$$

and, for  $1 \leq j \leq m$ ,

$$\tilde{f}_j(x) = \begin{cases} L, & \text{if } a^j > 3L, \\ 0, & \text{if } a^j < -3L, \\ f_j(x), & \text{if } -3L \leq a^j \leq 3L. \end{cases}$$

We point out that the problem  $\tilde{f}$  is a random transform of the problem  $f$ , determined by the realization  $a$  of the estimate  $a_\delta$ . We shall call this the *preparatory transformation*. We shall say that the *preparation was successful* if the condition  $\|a - f(\bar{x})\|_\infty \leq L$  is satisfied.

*Exercise 2.* Suppose that the preparation of a problem  $f \in \tilde{C}_{r,L}^{\epsilon,v_0}(G, E, \|\cdot\|, m)$  took place successfully. Prove that, when  $v_0 < \frac{1}{2}$ ,

- (i) if, for some  $j \geq 1$ ,  $a^j$  were  $> 3L$ , then  $f$  and  $\tilde{f}$  would both be incompatible; and, independently of the value of the  $a^j$ , compatibility of the one problem implies compatibility of the other;
- (ii) if, for all  $j \geq 1$  we have  $a^j > 3L$ , then, for all  $x \in G$ ,  $\varepsilon(x, f) \leq \varepsilon(x, \tilde{f})$ . In particular, a method which solves the problem  $\tilde{f}$  with a mean absolute error  $\leq \varepsilon$  will at the same time also solve the problem  $f$  with the same accuracy.
- (iii) The information about  $f$  provided by the oracle  $\mathcal{O}$  is converted in an obvious way into information about  $\tilde{f}$ . Prove that the oracle  $\bar{\mathcal{O}}$  obtained in

this way ensures the inclusion

$$\tilde{f} \in \tilde{W}_{r,40L}^{\epsilon,v_0}(G, E, \|\cdot\|, I_\infty^{(m)}, \|\cdot\|_\infty, K_\infty) \equiv \tilde{W}_{r,40L}^{\epsilon,v_0}.$$

<From the definition of the class  $\tilde{C}_{r,L}^{\epsilon,v_0}$  it is clear that the variation of each of the functions  $f_j$  on  $G$  does not exceed  $(1 + v_0)L \leq 2L$ . Hence follow immediately properties (i) and (ii). Further, it is clear that if the preparation is successful, then  $|\tilde{f}_j(x)| \leq 5L$  when  $x \in G$ . Hence and from the relations A-B in Section 6.1.2.3 we have (iii) (see Exercise 1, Section 6.4.5). >

### 6.6.3

We now consider the following method of solving problems in the class  $\tilde{C}_{r,L}^{\epsilon,v_0}(G, E, \|\cdot\|, m)$  with a specified mean relative accuracy  $v$  such that

$$cv_0 < v < 1, \quad (6.2)$$

where  $c > 12$  is a sufficiently large absolute constant (chosen so that (5.11) shall hold for  $\tilde{v} = v/80$ . As regards the values of the absolute constants in these and the following estimates, see the final remark in Section 6.4).

We take  $\delta = v/4$  and apply to the problem  $f$  being solved the preparatory procedure with the parameter  $\delta$ . If we then obtain  $a^j > 3L$  for a certain  $j \geq 1$ , then we stop the solution, and put out the answer that  $f$  is incompatible. Otherwise, we apply to the problem  $f$  (regarded as an element of the class  $\tilde{W}_{r,40}^{\epsilon,v_0}$ ) the method  $\mathcal{B}^{v/80}$  from the previous section. The result  $\tilde{x}$  of its application to  $\tilde{f}$  will be regarded as the result of the work of the method being described on  $f$ . This method will solve  $f$  with a (mean relative) error  $\leq v$ . Fr, by the properties of  $\mathcal{B}^{v/80}$ , the mean, conditional (under the condition of success of the preparation) error of  $\tilde{x}$  as a solution of the problem  $f$  does not (by the assertions (i)-(iii) in Exercise 2) exceed  $v/2$ . The relative error of any point of  $G \cup \{\star\}$  as a soluton of  $f$  clearly does not exceed  $1 + v_0 \leq 2$ . Thus, the error of the method described of solving problems in the class  $\tilde{C}_{r,L}^{\epsilon,v_0}$ , under the conditions (6.2), does not exceed  $(v/2) + 2\delta \leq v$ . Its laboriousness does not exceed  $R(v/4) + M_\omega(v/80)$ .

### 6.6.4

We have still to learn how to construct the estimates  $a_\delta$  with  $R(\delta)$  as small as possible. We point out a special characteristic of this problem. The accuracy which is demanded for the estimates is low; it is of the order of the  $r$ th moment of the noise (to the power  $1/r$ ). On the other hand, the reliability demanded of the estimate may be very high (the estimate is permitted not to have the required accuracy only with probability  $\leq v/4$ ).

We now expound the general method of constructing such estimates; the parts of the text enclosed in square brackets will give the particularization of the general scheme to the needs of this section.

Let  $(X, \Sigma)$  be a space with a  $\sigma$ -algebra of subsets, let  $\mathcal{F} = \{\mathbf{F}\}$  be a certain family of probability distributions on  $(X, \Sigma)$ , let  $Q$  be a metric space with the metric  $\rho(\cdot, \cdot)$ , and let  $\sigma(\mathbf{F}) : \mathcal{F} \rightarrow Q$  be a mapping setting up a correspondence between a distribution  $\mathbf{F}$  and its parameter  $\sigma(\mathbf{F})$ . [In the case with which we are concerned,  $(X, \Sigma)$  is  $I_\infty^{(m+1)}$  with a  $\sigma$ -algebra of Borel subsets of  $I_\infty^{(m+1)}$ ,  $\mathcal{F}$  is a family of distributions of the values of the random variables  $\psi^0(x, f, \omega)$  induced by the measure  $\mathbf{F}_\omega$  on  $\Omega$ . Here  $\ell = \{(\Omega, \mathbf{F}_\omega); \psi^0(x, f, \omega), \psi^1(x, f, \omega)\}$  is the oracle considered,  $Q = I_\infty^{(m+1)}$  with the metric defined by the norm  $\|\cdot\|_\infty$ , and  $\sigma(\mathbf{F}) = \int_{I_\infty^{(m+1)}} y d\mathbf{F}(y)$ .]

The function  $a(x_1, \dots, x_n)$  on the sample space of length  $n$ , i.e. on the space

$$X^n = \underbrace{X \times X \times \dots \times X}_{n \text{ times}}$$

with values in  $Q$ , the function being measurable in the sense that the pre-image of any ball in  $Q$  lies in the least  $\sigma$ -algebra of  $X^n$  which contains all sets of the form  $\{x^n | x_i \in A\}$ , where  $A \in \Sigma$ , will be called an *estimate with the observation time n*. We shall say that the *estimate a has accuracy r and reliability  $\alpha$  on a distribution  $\mathbf{F} \in \mathcal{F}$*  if the probability  $\Pr_{\mathbf{F}}\{\rho(a(x^n), \sigma(\mathbf{F})) > r\}$  of the event shown in the braces, calculated over the distribution on  $X^n$  obtained as the  $n$ th power  $\mathbf{F}^n$  of the distribution  $\mathbf{F}$  (i.e. over the distribution on  $X^n$  with the coordinates  $x_i$  independently distributed according to  $\mathbf{F}$ ) is not greater than  $\alpha$ , that is,  $\Pr_{\mathbf{F}}\{\rho(a(x^n), \sigma(\mathbf{F})) > r\} \leq \alpha$ . We shall say that *a has accuracy r and reliability  $\alpha$  on the class  $\mathcal{F}$*  if it has this property on every distribution  $\mathbf{F} \in \mathcal{F}$ . [In our case the estimate with the observation time  $n$  is constructed from the  $n$  answers of the oracle to a question about the value of  $f$  at the point  $\bar{x}$ ; and any estimate of accuracy  $L/2$  and reliability  $\delta$  can be taken as the estimate  $a_\delta$ .]

It turns out that by means of an estimate of accuracy of order  $r$  (namely,  $r/3$ ) and of a comparatively low reliability (any value less than  $\frac{1}{2}$ ) we can obtain an estimate of accuracy  $r$  and of a reliability specified in advance.

**Proposition.** Let  $a(x^n)$  be an estimate with observation time  $n$  which has on  $\mathcal{F}$  an accuracy  $r/3$  and a reliability  $p < \frac{1}{2}$ . Then, for some integer  $c(p)$  depending only on  $p$  (but not on  $\mathcal{F}$ ,  $a(\cdot)$ ,  $r$ , and  $n$ ) and every  $\alpha < p$ , an estimate  $a^\alpha$  can be found of accuracy  $r$  and reliability  $\alpha$  and with an observation time

$$n^* = nc(p) \left\lceil \ln \frac{1}{\alpha} \right\rceil. \quad (6.3)$$

The estimate  $a^\alpha$  is constructed in the following way. We divide the sample  $x^{n^*}$  into  $n$  successive blocks  $x^{(1)}, \dots, x^{(s)}$ , each of length  $n$ ;  $s = c(p) \lceil \ln 1/\alpha \rceil$ . To each of the blocks  $x^{(i)}$  we apply the estimate  $a$ . We obtain  $s$  points  $y_1, \dots, y_s$  of the space  $Q$ . It is possible that, for a certain  $i$ , the inequality  $\rho(y_i, y_j) \leq 2r/3$  holds for more than  $s/2$  of the indices  $j$ . In this case, let  $a^{(s)}(x^{ns}) \equiv a^\alpha(x^{n^*})$  be the first among the  $y_i$  which have the property mentioned. If there is no such  $i$ , we put  $a^\alpha(x^{n^*}) = y_1$ .

*Proof of the proposition.* We fix  $\mathbf{F} \in \mathcal{F}$ . We consider a natural number  $k$  and the distribution induced by  $\mathbf{F}$  on  $X^{nk}$ . Let  $x^{(1)}, \dots, x^{(k)}$  be successive blocks of  $x^{nk}$  each of length  $n$ , and let  $y_i = a(x^{(i)})$ . The quantities  $y_i$  are independent in aggregate, and the probability for each of them not to fall into the ball  $W_0 = \{y \in Q | \rho(\sigma(\mathbf{F}), y) \leq r/3\}$  is not greater than  $p < \frac{1}{2}$ , by the definition of  $a(\cdot)$ . Let  $A$  be the event consisting in the simultaneous occurrence of more than  $k/2$  of the events  $\{y_i \in W_0\}$ ,  $i = 1, \dots, k$ . It is easy to see that the probability of occurrence of  $A$  is not less than  $1 - \exp(-\kappa(p)k)$ , where  $\kappa(p) > 0$  depends only on  $p$ . Suppose the event  $A$  took place. Then more than  $k/2$  of the points  $y_i$  fell into  $W_0$ . In particular, they all fell into a ball of radius  $2r/3$  with centre at one of the points  $y_i$ ,  $i = 1, 2, \dots, k$  (we can take as the centre of this ball any of the points  $y_i \in W_0$ ). We now construct  $a^{(k)}(x^{nk})$  as described in the proposition. The value of  $a^{(k)}$ , by virtue of this description, determines the centre  $y$  of a ball  $W_1$  of radius  $2r/3$  which contains more than half of the points  $y_1, \dots, y_k$ . It is clear that  $W_1 \cap W_0 \neq \emptyset$ , and this gives  $\rho(\sigma(\mathbf{F}), y) \leq r$ . Thus  $\rho(\sigma(\mathbf{F}), a^{(k)}(x^{nk})) \leq r$  when  $A$  occurs. Therefore,  $a^{(k)}$  has a reliability not worse than  $\exp(-\kappa(p)k)$  for an accuracy  $r$ . From this the assertion of the proposition follows immediately.

*Remark.* When  $p = \frac{1}{4}$  we can take

$$\lceil \ln 1/\alpha \rceil \cdot c(p) = \lceil 2 \ln 3/2\alpha / \ln 4/3 \rceil.$$

We apply the general result proved to our problem of a suitable estimate of  $f(\bar{x})$ . For this it suffices to learn how to construct an estimate  $a$  of the quantity  $\sigma(\mathbf{F}) = \int y d\mathbf{F}(y)$  of accuracy  $L/6$  and of reliability  $< \frac{1}{2}$ , let us say of reliability  $\frac{1}{4}$ . About the distribution  $\mathbf{F}$  it is known from the properties of the oracle that

$$\mathbf{M}_{\mathbf{F}} \|y - \sigma(\mathbf{F})\|_\infty' \leq [L(1 + v_0)]'.$$

We construct the estimate  $a$  in the form

$$a(y^n) = \frac{1}{n} \sum_{i=1}^n y_i$$

and choose the necessary  $n$ . By virtue of the inequality (1.5) in Chapter 5 applied to the sum

$$\sum_1^n \frac{1}{n} \xi_i, \text{ with } \xi_i = 36y_i/(L(1 + v_0)),$$

we have

$$\begin{aligned} \mathbf{M}_{\mathbf{F}^n} \|a(y^n) - \sigma(\mathbf{F})\|_\infty &\leq \frac{L(1 + v_0)}{36} \left[ \frac{1}{2} + 36\omega_{V_{1,n+1};r} \left( \frac{36}{n} \right) \right] \\ &\leq L \left[ \frac{1}{36} + 2\omega_{V_{1,n+1};r} \left( \frac{36}{n} \right) \right]. \end{aligned}$$

It is clear that the estimate being used will have accuracy  $L/6$  and reliability  $\frac{1}{4}$  if the right-hand side of the last inequality is not greater than  $L/24$ . Thus  $n$  must

be chosen from the condition

$$\frac{1}{36} + 2\omega_{V_{1,m+1},r} \left( \frac{36}{n} \right) \leq \frac{1}{24}, \quad (6.4)$$

which, because of (1.7) in Chapter 5, gives

$$n = \lceil \ln(m+2) c(r) \rceil, \quad (6.5)$$

where  $c(r)$  depends only on  $r$ .

Thus, we known how to construct the estimate  $a_\delta$  required in Section 6.6.2 with a function  $R(\delta)$  of the form

$$R(\delta) = d(r) \ln(m+2) \ln(2/\delta) \quad (6.6)$$

where  $d(r)$  depends only on  $r$ . Applying this estimate to the method in Section 6.6.3, we obtain a method of solving problems in the class  $\tilde{C}_{r,L}^{e,v_0}(G, E, \|\cdot\|, m)$  with accuracy  $v$ .

### 6.6.5

Thus we have reduced conditionally extremal problems to extremal problems with operator constraints (in doing this, we had to apply a preparatory procedure, and also had to modify somewhat our use of words).

The solution of extremal problems with operator constraints was, in its turn, reduced to solving a series of  $O(\ln^2(1/v))$  operator inequalities. Each inequality of this kind was reduced to convex-concave games, and the latter were solved, without further reductions, by the method described in Section 6.3.

In the final analysis, problems from  $\tilde{C}_{r,L}^{e,v_0}$  are reduced to games on  $E \times I_1^{(m+1)}$ , or, more correctly, on a product of bounded subsets of each of the components. For the solution of these games we recommend the use of the MD-methods in Section 6.3.3 associated with a pair of functions  $V, V_p$ , where  $V$  is an  $(E, \|\cdot\|)$ -regular function, and  $V_p$  is an  $(I_1^{(m+1)}, \|\cdot\|_1)$ -regular function, which it is sensible to take to be  $V_{1,m+1}$  (see Section 3.2.5).

*Exercise 3.* It is well-known that, under very general hypotheses, solving extremal convex problems  $f$  can be reduced to solving games with a pay-off function

$$F(x, \lambda) = f_0(x) + \sum_{j=1}^m \lambda_j f_j(x)$$

on  $G \times \{\lambda | \lambda_j \geq 0\}$ . Would it be immediately possible to reduce the problem  $f$  to such a game?

〈No, at least not in the sense of the reduction described above. The point is that a game  $F(a, \lambda)$  is defined on an unbounded set, and this affords no possibility of solving it ‘constructively’ by MD-methods. Cutting off the ‘very large’ values of  $\lambda$  would also not lead to success since the size of the ‘curtailed’ domain increases as the accuracy required for the solution increases.〉

### 6.6.6

We shall now discuss the question of the laboriousness of the method solving problems in the class  $\tilde{C}_{r,L}^{e,v_0}$  which implements the scheme set out in Sections 6.4–6.6. The bound for the laboriousness  $\hat{M}(v)$  under the conditions (6.2) and with  $V_1 = V_{1,m+1}$  is of the form

$$\begin{aligned} \hat{M}(v) &\leq c(r) \ln(m+2) \ln 2/v + d(r) \left[ \ln^2 \frac{1}{v} \left[ \max \left\{ \frac{1}{v\gamma_{V,r}(v/c)}, \frac{\ln(m+2)}{v^{\frac{r}{r-1}}} \right\} \right. \right. \\ &\quad \left. \left. \leq D(r) \left[ \ln^2 \frac{1}{v} \left[ \max \left\{ \frac{1}{v\gamma_{V,r}(v/c)}, \frac{\ln(m+2)}{v^{\frac{r}{r-1}}} \right\} \right] \right] \right], \end{aligned} \quad (6.7)$$

Here  $\bar{r} = \min(r, 2)$ ,  $c$  is an absolute constant, and  $D(r)$  depends only on  $r$ . In particular, if  $(E, \|\cdot\|) = (L_p, \|\cdot\|_p)$ ,  $1 < p < \infty$ , the laboriousness of the method constructed admits the bound

$$D(r, p) \left[ \ln^2 \frac{1}{v} \left[ \max \left\{ \left( \frac{1}{v} \right)^{\frac{r}{r-1}}, \ln(m+2) \left( \frac{1}{v} \right)^{\frac{r}{r-1}} \right\} \right] \right],$$

where  $\tilde{r} = \min\{\bar{r}, p/(p-1)\}$ .

We see, in particular, that in the case  $\tilde{r} \leq p/(p-1)$  (which automatically holds when  $1 < p \leq 2$ ) the laboriousness of the constructed method cannot, in principle, be reduced by more than  $\sim_{p,r} \ln^2 1/v$  times provided that  $v < \bar{v}$ , where  $\bar{v} > 0$  is an absolute constant (cf. the lower bound for the stochastic complexity of classes  $\tilde{C}_{r,L}^{e,v_0}$  in Section 5.3). But if  $\tilde{r} > p/(p-1)$ , then, for all sufficiently small  $v$  ( $v \leq \bar{v}(p, \tilde{r}, m)$ ), the laboriousness is the same, up to a factor of order  $a(p, r) \ln^2 1/v$ , as for the MD-method of solving problems in the class  $C_{Lip}$  on  $L_p$  with exact information (see Section 3.3). Therefore in this case the recommended method cannot be substantially improved regarding laboriousness (by not more than, say,  $a_1(p, r) \ln^2 1/v$  times) where the MD-methods using exact information about the method are sub-optimal (i.e. in the case where  $G$  is a body like an  $I_p$ -ball of high dimension, see Chapter 4).

When  $(E, \|\cdot\|) = (I_1^{(n)}, \|\cdot\|_1)$ ,  $1 < n < \infty$ , the bound for the laboriousness of the constructed method is

$$D(1, r) \left( \frac{1}{v} \right)^{\frac{r}{r-1}} \max \{ \ln n, \ln(m+2) \},$$

and when  $v < \bar{v}$  its laboriousness cannot in principle be reduced by more than

$$a(1, r) \left[ \ln^2 \frac{1}{v} \left[ \max \left\{ \frac{\ln n}{\ln m}, 1 \right\} \right] \right]$$

times. We observe further that if it is a matter of solving problems  $f \in \tilde{C}_{r,L}^{e,v_0}$  with

the optimal value of the objective functional  $f_*$  known in advance, then such problems can be reduced to the single solution of an operator inequality (see Section 6.5.3.1), and so in this case the upper bound for the laboriousness of the resulting method and the potential upper limits for reducing its laboriousness are reduced by  $O(\ln^2 1/\nu)$  times.

In conclusion we remark that if we are concerned only with problems in the class  $\tilde{C}_{r,L}^{e,v_0}$ , then the actual description given in the last two sections of the method of solving them could be considerably simplified and made more perspicuous (we emphasize that we are speaking now of the description, and not of the actual structure of the method). We chose the more involved, but on the other hand more general, way of explaining the business.

## 6.7 ‘COMPLEX’ STOCHASTIC PROGRAMMING PROBLEMS

### 6.7.1

In this section we consider stochastic problems of the form

$$\Phi(x) \equiv \Phi_m(\Phi_{m-1}(\dots(\Phi_0(x))\dots)) \rightarrow \min |x \in G_0, \quad (7.1)$$

consisting in the minimization of a composition of  $m+1$  functions  $\Phi_0, \dots, \Phi_m$ . It will be supposed that here information about  $\Phi$  is not immediately accessible, although all the  $\Phi_i$  can be observed (by means of first-order oracles). If the information about the  $\Phi_i$  is exact, then, under certain quite general hypotheses about the  $\Phi_i$ , it is possible to derive from the oracles' answers the usual first-order information about  $\Phi$  (the values and the support functionals). Thus, when there is exact observation of all the  $\Phi_i$ , there is no necessity at all to distinguish problems of the form (7.1) from the whole family of extremal problems.

The situation changes when the  $\Phi_i$  are observed by means of stochastic oracles. In that case it is impossible to deduce (directly, at any rate) from the information about the  $\Phi_i$  the information about  $\Phi$  needed for solving this problem by first-order methods of the sort described above. In this case problems of the form (7.1) merit special examination. We shall prove that, under certain hypotheses regarding the structure of the  $\Phi_i$ , the problem  $\bar{\Phi}$  can be reduced to solving a convex-concave game.

### 6.7.2

We start by describing the situation. Suppose there are  $m+1$  Banach spaces  $(E_0, \|\cdot\|_0), \dots, (E_m, \|\cdot\|_m)$ . Suppose further that  $G_i$  are bounded, convex, closed subsets with  $G_i \subset E_i$ . For uniformity we let  $E_{m+1}, \|\cdot\|_{m+1}$ , denote the real line with the usual norm. We consider the mappings

$$\Phi_0: G_0 \rightarrow G_1, \Phi_1: G_1 \rightarrow G_2, \dots, \Phi_m: G_m \rightarrow E_{m+1} \equiv \mathbb{R},$$

and we define the composition  $\bar{\Phi}(x) = \Phi_m(\Phi_{m-1}(\dots(\Phi_0(x))\dots)): G_0 \rightarrow \mathbb{R}$ . The composition  $\bar{\Phi}(x)$  defines the problem (7.1).

Naturally, we shall study only the convex case where all the mappings  $\Phi_i$  and their composition  $\bar{\Phi}$  are convex. More precisely, we assume that each of the spaces  $E_i, i > 0$ , is equipped with a convex, closed cone  $K_i$  which specifies an ordering  $\geq_i$  in  $E_i$  (the ordering in  $E_{m+1} = \mathbb{R}$  is the usual one). We assume that all the mappings  $\Phi_i: G_i \rightarrow E_{i+1}$ ,  $i = 0, 1, \dots, m$ , are Lipschitz and convex (relative to the ordering  $\geq_{i+1}$  in  $E_{i+1}$ ), and also that they are all monotonic (i.e.  $x, y \in G_i$  and  $x \leq_i y$  imply  $\Phi_i(x) \leq_i \Phi_i(y)$ ). Under these hypotheses the composition  $\bar{\Phi}$  turns out to be a Lipschitz-convex function on  $G_0$  (verify this!).

We further assume that the spaces  $E_i, \|\cdot\|_i$ ,  $0 \leq i \leq m$ , are separable and regular, and that so are the spaces  $E_i^*, \|\cdot\|_i^*$ ,  $1 \leq i \leq m$ , as well. We also assume that each set  $G_i$ ,  $1 \leq i \leq m$ , has a non-empty interior.

### 6.7.3

We now formulate a hypothesis regarding the possibility of observing the components  $\Phi_i$  of the problem  $\bar{\Phi}$ . Roughly speaking, we suppose that each of the mappings  $\Phi_i$  can be observed by means of a stochastic oracle of the type described in Section 6.4. For simplicity we shall suppose that this oracle is unbiased and that the situation is normalized. That is, we assume that  $\rho_{\Omega, \Omega}(G_i) = 1$  and that the centre of  $G_i$  is 0. Further, let  $x^m = (x_0, \dots, x_m)$  be a point of  $G^m = G_0 \times \dots \times G_m$ . We assume that we have at our disposition an oracle with a noise space  $(\Omega, \mathcal{F}_\omega)$  and an observation function having  $2(m+1)$  components. The first  $m+1$  components  $\psi_0^i(x^m, \omega): G^m \times \Omega \rightarrow E_{i+1}$  (we omit the index  $\bar{\Phi}$  from the notation for these) are observations, accessible to the method, of the values of  $\Phi_i(x_i)$ , and the last  $m+1$  components  $\psi_1^i(x^m, \omega): G^m \times \Omega \rightarrow L(E_i, E_{i+1})$  are the observations of the differentials of the  $\Phi_i$ . Here  $L(E_i, E_{i+1})$  is the space of continuous linear operators from  $E_i$  into  $E_{i+1}$ . As always  $(\Omega, \mathcal{F}_\omega)$  is assumed to be a Polish space with a regular, Borel measure, and  $\psi_0^i, \psi_1^i$  are taken to be Borel functions with respect to  $x^m$  and  $\omega$  ( $\psi_1^i$  in the sense of Section 6.4.2.3). It is also assumed that, for all  $x^m \in G^m$  and for all  $i$ ,  $0 \leq i \leq m$ ,

$$\begin{aligned} \mathbf{M}_{\mathcal{F}_\omega} \psi_0^i(x^m, \omega) &= \Phi_i(x_i); \quad \mathbf{M}_{\mathcal{F}_\omega} \{\psi_1^i(x^m, \omega)(y - x_i)\} \\ &\leq \Phi_i(y) - \Phi_i(x_i), \quad y \in G_i; \quad \mathbf{M}_{\mathcal{F}_\omega} \|\psi_0^i(x^m, \omega)\|_{i+1}^* \leq 1; \\ \mathbf{M}_{\mathcal{F}_\omega} \{[\psi_1^i(x^m, \omega)]^* l\}_{i,*}^r &\leq 1 \text{ for all } l \geq 0 \text{ with } \|l\|_{i+1,*} \leq 1. \end{aligned}$$

Here  $\geq_{i+1,*}$  denotes the ordering in  $E_{i+1}^*$  induced by the cone  $K_{i+1}^*$  dual to  $K_{i+1}$ ,  $[\psi_1^i]^*$  is the operator dual to  $\psi_1^i$  (see Section 6.4.1), and  $r > 1$  is a parameter.

## 6.7.4

We shall prove  $\bar{\Phi}$  can be reduced to a game; and that the information needed for solving this game can be obtained from the oracle  $\mathcal{C}$ . To be exact, let

$$G_i^* = \{l \in E_i^* \mid l \geq 0, \|l\|_{i,*} \leq 1\}, \quad 1 \leq i \leq m,$$

and

$$G^{*m} = G_1^* \times \dots \times G_m^* = \{l^m \mid l^m = (l_1, \dots, l_m), l_i \in G_i^*\}.$$

We consider the game  $F_{\Phi}$  defined on  $G^m \times G^{*m}$  by

$$\begin{aligned} F_{\Phi}(x^m, l^m) &= \Phi_m(x_m) + \langle l_m | \Phi_{m-1}(x_{m-1}) - x_m \rangle \\ &\quad + \langle l_{m-1} | \Phi_{m-2}(x_{m-2}) - x_{m-1} \rangle + \dots + \langle l_1 | \Phi_0(x_0) - x_1 \rangle. \end{aligned}$$

We shall show later that the game  $F_{\Phi}$  is closely connected with the problem  $\bar{\Phi}$ . Namely, let

$$\bar{F}_{\Phi}(x^m) = \sup_{l^m \in G^{*m}} F_{\Phi}(x^m, l^m)$$

be the payoff function of the first player. Then, for all  $x_0 \in G_0$ ,

$$\bar{\Phi}(x_0) = \inf_{x_1 \in G_1, \dots, x_m \in G_m} \bar{F}_{\Phi}(x_0, x_1, \dots, x_m), \quad (7.2)$$

i.e.  $\bar{\Phi}(x_0)$  is the pay-out made by the first player after the best of his moves with first component  $x_0$ . Hence it is clear that, having solved the game  $F_{\Phi}$  with a (mean) absolute error  $\varepsilon/2$ , and having taken the zeroth component of  $x^m$ —the components of the solution obtained for  $F_{\Phi}$ —as an approximate solution of the problem of minimizing  $\bar{\Phi}$ , we obtain a (random) point  $\tilde{x}_0 \in G_0$  for which

$$\mathbf{M}\bar{\Phi}(\tilde{x}_0) \leq \min_{G_0} \bar{\Phi}(x_0) + \varepsilon \quad (7.3)$$

( $\mathbf{M}$  denotes the mean over the distribution of  $\tilde{x}_0$ ).

## 6.7.5

We have still to learn how to solve the game  $F_{\Phi}$ . To do this we use the MD-method in Section 6.3.3. Its use is possible. For, in view of the conditions imposed on  $E_i$  and  $E_i^*$  earlier, the spaces  $E \equiv E_0 \times \dots \times E_m$  with the norm  $\|x^m\| = \max_i \|x_i\|_i$  and  $E_l \equiv E_1^* \times \dots \times E_m^*$  with the norm  $\|l^m\|_i = \max_i \|l_i\|_{i,*}$  are regular and separable. Further, the oracle  $\mathcal{C}$  does indeed provide the information necessary for this method. For, it is easy to see that  $F_{\Phi}$  is a Lipschitz convex-concave game. Moreover, the function

$$\psi_x(x^m, l^m, \omega) = \{[\psi_1^m(x^m, \omega)]^* \mathbb{1}, [\psi_1^{m-1}(x^m, \omega)]^* l_m, \dots, [\psi_1^0(x^m, \omega)]^* l_1\}$$

with values in  $E_0^* \times \dots \times E_m^*$  ( $\mathbb{1}$  is the standard linear functional  $\langle \mathbb{1} | t \rangle = t$  on

$E_{m+1} = \mathbb{R}$ ) has the property that  $\mathbf{M}_{F_{\Phi}}\psi_x(x^m, l^m, \omega)$  is a support functional to  $F_{\Phi}$  as a function of  $x^m$ . Similarly, the function

$$\psi_l(x^m, l^m, \omega) = \{\Phi_{m-1}(x_{m-1}) - x_m, \dots, \Phi_0(x_0) - x_1\}$$

with values in  $E_1 \times \dots \times E_m = [E_1^* \times \dots \times E_m^*]^*$  is such that  $\mathbf{M}_{F_{\Phi}}\psi_l(x^m, l^m, \omega)$  is a support functional to  $F_{\Phi}$  as a function of  $l^m$ .

Thus by means of the oracle  $\mathcal{C}$  unbiased estimates  $\psi_x, \psi_l$  of the support functionals to  $F_{\Phi}$  with respect to the  $x^m$  components and the  $l^m$  components can be constructed. From the properties of the oracle  $\mathcal{C}$  it is easy to deduce that, with the method described of observing  $F_{\Phi}$ , this game falls into a class of the form  $\bar{D}_{r,L}^{\varepsilon,0}(G^m \times G^{*m}, E, \|\cdot\|, E_l, \|\cdot\|_l)$  with  $L = 4(m+1)$ . The parameter  $L$  is thus written out explicitly, and the application to  $F_{\Phi}$  of the method in Section 6.3.3 presents no difficulties in principle.

## 6.7.6

It remains to prove (7.2). For this we need a simple lemma.

**Lemma.** Let  $\tilde{\Phi}_i: G_i \rightarrow \mathbb{R}$  be a monotone, convex mapping, Lipschitz with constant 1,  $i \geq 1$ . Then, for all  $x \in G_i$ ,

$$\tilde{\Phi}_i(x) = \sup_{l \in G_i^*} \inf_{y \in G_i} \{\tilde{\Phi}_i(y) - \langle l | y - x \rangle\}. \quad (7.4)$$

*Proof.* By von Neumann's lemma

$$\sup_{l \in G_i^*} \inf_{y \in G_i} \{\tilde{\Phi}_i(y) - \langle l | y - x \rangle\} = \inf_{y \in G_i} \sup_{l \in G_i^*} \{\tilde{\Phi}_i(y) - \langle l | y - x \rangle\}. \quad (7.5)$$

It is clear that the right-hand side of (7.5), for a given  $x \in G$ , is not greater than  $\tilde{\Phi}_i(x)$  (one can take  $y = x$ ). Denoting the right-hand side of (7.4) by  $H(x)$ , we obtain

$$H(x) \leq \tilde{\Phi}_i(x), \quad x \in G. \quad (7.6)$$

But, by hypothesis,  $\text{int } G_i \neq \emptyset$ . Let  $x \in \text{int } G_i$ . Then because the function  $\tilde{\Phi}_i$  is monotone and Lipschitz with constant 1, it follows that the support functional  $l_x$  to it at the point  $x$  lies in  $G_i^*$  (why?). The function  $\tilde{\Phi}_i(y) - \langle l_x | y - x \rangle$  attains a minimum over  $y$  when  $y = x$ , and this minimum is  $\tilde{\Phi}_i(x)$ . Therefore, when  $x \in \text{int } G$ , the left-hand side of (7.5) is not less than  $\tilde{\Phi}_i(x)$ . Hence and from (7.6),  $H(x) = \tilde{\Phi}_i(x)$  when  $x \in \text{int } G$ . Both sides of this equality are continuous on  $G_i$  (why?), and this proves (7.4).

*Exercise 1.* Is the condition that  $\text{int } G_i \neq \emptyset$  essential?

We are now in a position to prove (7.2). Indeed, from the definition of  $F_{\Phi}$ ,

$$\begin{aligned} \bar{F}_{\Phi}(x_0, \dots, x_m) &= \max_{l^m \in G^{*m}} \{\Phi_m(x_m) + \langle l_m | \Phi_{m-1}(x_{m-1}) - x_m \rangle \\ &\quad + \dots + \langle l_1 | \Phi_0(x_0) - x_1 \rangle\}. \end{aligned} \quad (7.7)$$

Hence, in view of von Neumann's lemma,

$$\begin{aligned} & \inf_{x_m \in G_m} \bar{F}_\Phi(x_0, \dots, x_m) \\ &= \max_{l_1, \dots, l_{m-1}, l_j \in G_j^*} \left\{ \max_{l_m \in G_m^*} \min_{x_m \in G_m} \{ \Phi_m(x_m) - \langle l_m | x_m - \Phi_{m-1}(x_{m-1}) \rangle \} \right. \\ & \quad \left. + \langle l_{m-1} | \Phi_{m-2}(x_{m-2}) - x_{m-1} \rangle + \dots + \langle l_1 | \Phi_0(x_0) - x_1 \rangle \right\}. \end{aligned}$$

By the lemma, the expression in the inner braces is  $\Phi_m(\Phi_{m-1}(x_{m-1}))$ . (The lemma is actually applicable to  $\Phi_m$ ; the only one of its hypotheses which needs verifying is that  $\Phi_m$  is Lipschitz with constant 1, and this follows at once from the oracle's properties, in view of which  $\mathbf{M}_{F_\omega} \psi_0^m(x^m, \omega)$  is a support functional to  $\Phi_m(x_m)$  with norm  $\leq 1$ ). Thus

$$\begin{aligned} & \min_{x_m \in G_m} \bar{F}_\Phi(x_0, \dots, x_{m-1}; x_m) \\ &= \max_{l_1, \dots, l_{m-1}, l_j \in G_j^*} \{ \Phi_m(\Phi_{m-1}(x_{m-1})) + \langle l_{m-1} | \Phi_{m-2}(x_{m-2}) - x_{m-1} \rangle \\ & \quad + \dots + \langle l_1 | \Phi_0(x_0) - x_1 \rangle \}. \end{aligned} \tag{7.8}$$

The expression (7.8) is constructed exactly like (7.7) except that  $m$  is replaced by  $m-1$ , and the function  $\Phi_m$  is replaced by  $\Phi_m(\Phi_{m-1}(\cdot))$ . If we knew that  $\Phi_m(\Phi_{m-1}(\cdot))$ , like  $\Phi_m(\cdot)$ , was a convex, monotone mapping, Lipschitz with constant 1, then it would be possible to pass from (7.8) to a similar relation for  $\min_{x_m \in G_m, x_{m-1} \in G_{m-1}} \bar{F}_\Phi(x_0, \dots, x_m)$ , and then, continuing in the same way, to arrive at (7.2).

From what we have said, it is clear that, to prove (7.2), it suffices to show that, if  $j > 1$  and  $H_j: G_j \rightarrow \mathbb{R}$  is a monotone convex mapping, Lipschitz with constant 1, then the same is also true of the mapping  $H_j(\Phi_{j-1}(x_{j-1})) = \tilde{H}_{j-1}(x_{j-1}): G_{j-1} \rightarrow \mathbb{R}$ . The monotonicity and convexity of  $\tilde{H}_{j-1}$  are obvious.

Further,  $\text{int } G_j \neq \emptyset$ , and at each point  $x_j \in \text{int } G_j$ , the function  $H_j$  has a support functional belonging to  $G_j^*$ . By a standard argument, compactness is also true for  $x_j \in G_j$ . Now let  $\tilde{x}_{j-1} \in G_{j-1}$ , and let  $\tilde{x}_j = \Phi_{j-1}(\tilde{x}_{j-1})$ . Let  $l \in G_j^*$  be the support functional to  $H_j$  at  $\tilde{x}_j$ . It is clear that  $\tilde{l} = \mathbf{M}_{F_\omega}[[\psi_1^{j-1}(\tilde{x}^m, \omega)]^* l]$  is a support functional to  $\tilde{H}_{j-1}$  at  $\tilde{x}_{j-1}$  (here  $x^m$  is any point the  $(j-1)$ th co-ordinate of which is equal to  $\tilde{x}_{j-1}$ ). From the properties of the oracle,  $\|\tilde{l}\|_{j-1,*} \leq 1$ . Thus, at each point of  $G_{j-1}$  the function  $\tilde{H}_{j-1}$  has a support functional with norm  $\leq 1$ . So  $\tilde{H}_{j-1}$  is Lipschitz with constant 1 on  $G_{j-1}$ , as required. The relation (7.2) has been proved completely.

### 6.7.7

*Remark.* Actually, we have already used the construction described above, namely in Section 6.4. Indeed, solution of the operator inequality  $H(x) \leq 0$  was

there reduced to solving a ‘complex’ stochastic problem  $s(H(x)) \rightarrow \min |x \in G$  (see Section 6.4.3.1), and the latter was solved by reduction to a game of the type  $F_\Phi$ . The specific feature of the situation in Section 6.4 is that the function  $s(y)$  is known *a priori* and does not require observation; furthermore, it is homogeneous and can be expressed in the form  $\max_{l \in G_1^*} \langle l | y \rangle$  ( $G_1^*$  here is the same as  $G_1$  in Section 6.4.3.1). This special representation of  $s(y)$  (which, of course, is also obtainable from (7.4) by taking the lower bound over  $y$ ) is used in Section 6.4 instead of the representation (7.4), on which the whole plan of the argument in this section is based.

# Strongly convex problems

Up to now we have been studying first-order methods of solving ‘arbitrary’ convex problems. Now we pass on to the consideration of special classes of such problems, which are distinguished by certain conditions of smoothness and strict convexity. It should be remarked that most of the standard numerical methods of optimization are oriented precisely towards ‘good’ problems of this sort.

We start in Section 7.1 with a description of the classes of problems to be considered, and in Section 7.2 we obtain lower bounds for their complexity. In Sections 7.3, 7.4, and 7.5 methods are constructed which ‘almost’ implement this lower bound for complexity. The results in this chapter are based mainly on the authors’ paper [32].

## 7.1 CLASSES OF SMOOTH AND STRONGLY CONVEX EXTREMAL PROBLEMS

### 7.1.1

Let  $E$  be a real, separable Hilbert space (this understanding holds throughout the chapter). The norm in  $E$  will be denoted simply by  $|\cdot|$ . A function  $f: E \rightarrow \mathbb{R}$  is said to be *L-smooth* if

- (1)  $f$  is continuously differentiable;
- (2) for all  $x, y \in E$  we have

$$|\langle f'(x) - f'(y), x - y \rangle| \leq L|x - y|^2.$$

The condition (2) is equivalent, in fact, to

$$(2') \quad |f'(x) - f'(y)| \leq L|x - y| \text{ for all } x, y \in E,$$

i.e. an *L*-smooth function is simply a differentiable function whose first derivatives are Lipschitz, with Lipschitz constant *L*.

A function  $f: E \rightarrow \mathbb{R}$  is said to be *(l, Q)-strongly convex*,  $l > 0$ ,  $Q \geq 1$ , if

- (1)  $f$  is continuously differentiable
- (2) for all  $x, y \in E$

$$l|x - y|^2 \leq \langle f'(x) - f'(y), x - y \rangle \leq Ql|x - y|^2.$$

Let us explain this definition. Suppose  $f$  is twice differentiable. Its second differential (its Hessian) is a quadratic form on  $E$ . The function  $f$  is *(l, Q)-convex* if and only if the least eigenvalue of the Hessian is everywhere not less than *l* and the greatest eigenvalue is everywhere not greater than *Ql*. (In the infinite-dimensional case, the eigenvalues are to be replaced by the lower and upper bounds of the spectrum.)

A function  $f$  which is *L*-smooth for arbitrary  $L < \infty$  is said to be *smooth*. A function which is *(l, Q)-strongly convex* for arbitrary  $l > 0$  and  $Q > 1$  is said to be *strongly convex*.

Let  $f$  be smooth. Then there is a minimal  $L = L_f$  such that  $f$  is *L*-smooth. Similarly, if  $f$  is strongly convex, then there are  $l_f > 0$  and  $Q_f \geq 1$  such that *(l, Q)-convexity* of  $f$  implies  $l \leq l_f$  and  $Q \geq Q_f$ ; moreover,  $Q_f = L_f/l_f$ . Suppose  $f$  is *(l, Q)-strongly convex*. Then the following inequalities hold:

$$\left. \begin{aligned} f(x + h) &\geq f(x) + \langle f'(x), h \rangle + \frac{l}{2}h^2, \\ f(x + h) &\leq f(x) + \langle f'(x), h \rangle + \frac{lQ}{2}h^2; \end{aligned} \right\} \quad (1.1)$$

we shall be constantly using these later. Conversely, if  $f$  is continuously differentiable and has the property (1.1), then  $f$  is *(l, Q)-strongly convex*.

*Exercise 1.* Prove the assertions just made.

### 7.1.2

Let  $G \subset E$  be a convex, closed, non-void (but not necessarily bounded) set, and let  $x_1 \in G$ . We shall consider classes of problems of the form

$$f_0(x) \rightarrow \min_{x \in G}, f_i(x) \leq 0, 1 \leq i \leq m, \quad (1.2)$$

determined by strongly convex (and smooth) functions  $f_i$ . More precisely, let  $l_i > 0$ ,  $Q_i \geq 1$ . Let  $H = H_{x_1}(G, m; l_0, \dots, l_m, Q_0, \dots, Q_m)$  denote the following class of problems (we shall call it *the class of strongly convex problems*):

- (i) the set of problems in this class is precisely the set of all problems (1.2) (i.e. of functions  $f(x) = f_0(x), \dots, f_m(x))$  such that the  $f_i$  are  $(l_i, Q_i)$ -strongly convex,  $i = 0, 1, \dots, m$ ;
- (ii) the oracle is an exact, first-order oracle with  $E$  as its domain of questions. In other words, a question can be put to the oracle at any point of  $E$ , and the values of  $f(x)$  and  $f'(x)$  are obtained as the answer;

(iii) the normalizing factors  $r(f)$  are defined in the following way. For every  $x \in G$  let

$$V_f(x) = \max\{f_0(x) - f_*, f_1(x), \dots, f_m(x)\}$$

( $f_*$  is the optimal value of the objective functional of the problem  $f$ ; if  $f$  is incompatible, then  $f_* = +\infty$ ). Then

$$r_i(f) \equiv r(f) \equiv V_f(x_1).$$

The meaning of the normalization is clear.  $V_f(x)$  measures the residual of  $x$  regarded as an approximate solution of  $f$ , i.e. the maximum deviation of the functionals in the problem  $f$  at the point  $x$  from their 'nominal' values (i.e. from the values required for an exact solution). The residual at a fixed point  $x_1 \in G$  is taken as  $r_i(f)$ . The clause, 'the method has an accuracy  $v$ ', now means that 'the residual of the result of its work is at least  $1/v$  times smaller than the residual at  $x_1$ '. As we shall see later, the complexity of the class  $H$  is basically determined by the quantity  $Q = \max_i Q_i$ .  $Q$  is called the *modulus of strong convexity of the class  $H$* .

As well as the classes  $H_{x_1}$  we shall consider the classes  $H_{x_1}^v(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$ . These classes consist of all the problems  $f \in H_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$  which are either incompatible or such that

$$V_f(x_1) \equiv \max\{|f_0(x_1) - f_*|, f_1(x_1), \dots, f_m(x_1)\} \leq V.$$

The oracle for this class is the same as for the class  $H_{x_1}$ , and the normalizing factors for all problems in the class are equal to  $V$ .

Let us look at the differences between the definitions of the classes  $H_{x_1}$  and  $H_{x_1}^v$ . Clearly, application of the  $H_{x_1}^v$ -method to a given problem requires rather more *a priori* information than the application of the  $H_{x_1}$ -method to it. In both cases we need to estimate the parameters of strong convexity of the functionals in the problem  $f$ . But in the first case we also need to have *a priori* an upper bound for  $\bar{V}_f(x_1)$  in order to know in which of the classes  $H_{x_1}^v f$  would fall. Moreover, an  $H_{x_1}$ -method with error  $v$  will solve  $f$  with an (absolute) error  $\leq V_f(x_1)v$ , but the  $H_{x_1}^v$ -method of the same accuracy will solve it with an error  $\leq Vv$ . Since  $V \geq \bar{V}_f(x_1) \geq V_f(x_1)$  for compatible  $f$ , the second bound is always worse than the first. Thus, from the point of view of application,  $H_{x_1}^v$ -methods are less convenient than  $H_{x_1}$ -methods.

However, the loss here is largely theoretical. The point is that it is not difficult to find an upper bound for  $\bar{V}_f(x_1)$ : it suffices to know the answer to a question about  $f$  at the point  $x_1$  and then to use (1.1). There is still a defect relating to the 'coarser' measurement of the error by  $H_{x_1}^v$ -methods than by  $H_{x_1}$ -methods. But the dependence of the laboriousness of  $H_{x_1}^v$ -methods on the accuracy is, as we shall prove, logarithmic. Therefore, in all 'reasonable' situations, this dependence is little changed if we go over from the  $H_{x_1}$ -scale for error to the  $H_{x_1}^v$ -scale.

The reason for introducing the  $H_{x_1}^v$  classes is that, in certain cases, we are unable to construct for  $H_{x_1}$  such effective methods (as regards the bounds) as for  $H_{x_1}^v$ .

### 7.1.3

Let  $G$  be, as before, a convex, closed, and non-void, but now a necessarily bounded, subset of  $E$ , and let  $V > 0$  and  $x_1 \in G$ . Let  $CS_{x_1}^v(G, m)$  denote the class of all problems of the form (1.2) in which the  $f_i$  are convex and the  $L_i(f)$  are smooth functions on  $E$  such that

$$\max_i (\frac{1}{2}L_i(f)\rho_{G, x_1}^2 + \rho_{G, x_1}|f'_i(x_1)|) \leq V;$$

here

$$\rho_{G, x_1} = \max_{y \in G} |y - x_1|$$

(on the left we have written down the natural upper bounds for the quantities  $\sup_G f_i(x) - f_i(x_1)$ ). The oracle for  $CS_{x_1}^v$  is the same as for  $H_{x_1}^v$ , and the normalizing factor is  $V$ . The classes  $CS_{x_1}^v$  are called *classes of smooth convex problems*.

We observe immediately that solving smooth convex problems can be reduced to solving strongly convex problems. For, let  $v > 0$  be the required accuracy of solution of problems in the class  $CS_{x_1}^v(G, m)$ . With each problem  $f \in CS_{x_1}^v(G, m)$  we associate its *regularization*  $f^v$ .

$$f_i^v(x) = f_i(x) + \frac{vV(x - x_1)^2}{2\rho_{G, x_1}^2} - \frac{vV}{2}.$$

it is clear that

$$f^v \in H_{x_1}^{(1+\frac{v}{2})+}(G, m; \frac{vV}{\rho_{G, x_1}^2}, \dots, \frac{vV}{\rho_{G, x_1}^2}; \frac{2+v}{v}, \dots, \frac{2+v}{v}).$$

Suppose now that there is a deterministic method  $\mathcal{B}_v$  of solving problems of the last class with an accuracy  $v/3$ . We construct from  $\mathcal{B}_v$  a method  $\mathcal{B}'_v$  of solving problems in  $CS_{x_1}^v$  by the following obvious device: we apply the method  $\mathcal{B}_v$  to the problem  $f^v$ , and we put out the result  $x_f$  of this application as the result of the work of  $\mathcal{B}'_v$  on  $f$ .

Clearly, we have actually described a  $CS_{x_1}^v$ -method (information about  $f^v$  is calculated in an obvious way from the information about  $f$ ). We shall prove that it has the required accuracy  $v$ , provided that  $v < 1$  (this is the only case which is not trivial).

It is possible that  $x_f \neq x_*$ . Then

$$f_i^v(x_f) \leq \left(1 + \frac{v}{2}\right) \frac{v}{3} V, 1 \leq i \leq m,$$

and

$$f_i(x_f) \leq f_i^*(x_f) + \frac{vV}{2} \leq vV, \quad 1 \leq i \leq m.$$

Therefore if  $f$  is incompatible, we have  $v(x_f, f) \leq v$ .

Now let  $f$  be compatible. The  $f^*$  is also compatible, and so, from the properties of  $\mathcal{B}_v$ , we have  $x_f \neq *$ . Moreover,  $(f^*)_* \leq f_*$ , and, from the properties of  $\mathcal{B}_v$ ,

$$\begin{aligned} f_0^*(x_f) &\leq (f^*)_* + \frac{v}{3}V\left(1 + \frac{v}{2}\right) \leq f_* + \frac{v}{3}V\left(1 + \frac{v}{2}\right), \\ f_0(x_f) &\leq f_0^*(x_f) + \frac{vV}{2} \leq f_* + vV. \end{aligned}$$

This inequality together with the previous one shows that  $x_f$  is a solution of the problem  $f$  with an error  $\leq v$ . Thus  $\mathcal{B}'_v$  solves problems in  $CS_{x_1}^V(G, m)$  with accuracy  $v$ ; and its laboriousness is equal to that of  $\mathcal{B}_v$ .

In view of what has been said, we shall pay main attention to classes of strongly convex problems. The methods of solving them also induce methods of ‘smooth convex’ optimization.

#### 7.1.4

We begin with a simple—‘simple’, at any rate, so far as the idea is concerned—method of solving strongly convex problems, which is a natural generalization of the classical gradient method. Suppose we have a class of problems  $H = H_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$ . The method being described constructs on a problem  $f \in H$  a sequence of points  $x_1, x_2, \dots$  such that, for all  $i$  for which  $x_i \in G$ , we have

$$\mathbf{V}_f(x_i) \leq \left(1 - \frac{1}{Q}\right)\mathbf{V}_f(x_{i-1}) \quad (1.3)$$

( $Q$  is the modulus of strong convexity of the class  $H$ ). But if  $x_i \notin G$ , then  $x_{i+s} = *$ ,  $s \geq 1$ , and also  $v(x_i, f) = 0$  (i.e.  $f$  is incompatible). In particular, to ensure an accuracy  $v$  it suffices to make

$$\left\lceil \frac{\ln 1/v}{\ln Q/(Q-1)} \right\rceil \leq \left\lceil Q \ln \frac{1}{v} \right\rceil \text{ steps.}$$

We describe the  $(i-1)$ th step of the method. Suppose that at this step we already have a point  $x_{i-1} \in G \cup \{*\}$ . If  $x_{i-1} = *$ , then we also put  $x_i = *$ . Otherwise let

$$f_{j,x}(y) = f_j(x) + \langle f'_j(x)|y - x \rangle + \frac{1}{2}Ql_j(y - x)^2.$$

We solve the problem

$$(P_{i-1}): \quad f_{0,x_{i-1}}(y) \rightarrow \min |y \in G, \\ f_{j,x_{i-1}}(y) \leq \left(1 - \frac{1}{Q}\right)f_{j,x_{i-1}}(x_{i-1}), \quad 1 \leq j \leq m.$$

To form the problem  $P_{i-1}$  it suffices to put to the oracle a question about  $f$  at the point  $x_{i-1}$ . We take the solution of  $P_i$  as  $x_i$  (if  $P_{i-1}$  is incompatible,  $x_i = *$ ).

Let us satisfy ourselves that the method described really does have the property formulated in (1.3) of reducing the residual after the step. Suppose first that  $x_i \in G$ . In view of (1.1), when  $j \geq 1$  we have

$$f_j(x_i) \leq f_{j,x_{i-1}}(x_i) \leq \left(1 - \frac{1}{Q}\right)f_{j,x_{i-1}}(x_{i-1}) = f_j(x_{i-1})\left(1 - \frac{1}{Q}\right). \quad (1.4)$$

Hence it is clear that if  $f$  is incompatible, then

$$\mathbf{V}_f(x_i) \equiv \max_{j \geq 1} f_j(x_i) \leq \left(1 - \frac{1}{Q}\right)\max_{j \leq 1} f_j(x_{i-1}) = \left(1 - \frac{1}{Q}\right)\mathbf{V}_f(x_{i-1}),$$

as is required in (1.3).

Now let  $f$  be compatible and let  $x^*$  be its solution. Then, by (1.1), for all  $j \geq 0$ ,

$$f_j(x^*) \geq f_j(x_{i-1}) + \langle f'_j(x_{i-1})|x^* - x_{i-1} \rangle + \frac{l_j}{2}(x^* - x_{i-1})^2. \quad (1.5)$$

Putting

$$y = x_{i-1} + \frac{1}{Q}(x^* - x_{i-1}),$$

we find that  $y \in G$  and

$$\begin{aligned} f_{j,x_{i-1}}(y) &= f_{j,x_{i-1}}(x_{i-1}) + \frac{1}{Q} \left[ \langle f'_j(x_{i-1})|x^* - x_{i-1} \rangle + \frac{1}{2}l_j(x^* - x_{i-1})^2 \right] \\ &\leq f_{j,x_{i-1}}(x_{i-1}) + \frac{1}{Q}(f_j(x^*) - f_j(x_{i-1})) \end{aligned}$$

(we have used (1.5)). Thus

$$f_{j,x_{i-1}}(y) \leq \frac{1}{Q}f_j(x^*) + \left(1 - \frac{1}{Q}\right)f_{j,x_{i-1}}(x_{i-1}). \quad (1.6)$$

Hence it is clear that  $P_{i-1}$  is compatible ( $y$  is a feasible point for it), and so, when  $f$  is compatible,  $x_i \in G$ . Moreover, putting  $j = 0$  in (1.6), we find

$$f_0(x_i) \leq f_{0,x_{i-1}}(x_i) \leq f_{0,x_{i-1}}(y) \leq \frac{1}{Q}f_j(x^*) + \left(1 - \frac{1}{Q}\right)f_0(x_{i-1}),$$

i.e.

$$f_0(x_{i-1}) - f_* \leq \left(1 - \frac{1}{Q}\right)(f_0(x_{i-1}) - f_*). \quad (1.7)$$

We see that, when  $f$  is compatible, (1.7) holds as well as (1.4), and hence

$$\begin{aligned}\mathbf{V}_f(x_i) &= \max \{f_0(x_i) - f_*, f_1(x_i), \dots, f_m(x_i)\} \\ &\leq \left(1 - \frac{1}{Q}\right) \max \{f_0(x_{i-1}) - f_*, f_1(x_{i-1}), \dots, f_m(x_{i-1})\} \\ &= \left(1 - \frac{1}{Q}\right) \mathbf{V}_f(x_{i-1}),\end{aligned}$$

as required. The proof is completed.

*Example 1.* Let  $G = E$ ,  $m = 0$ . The method becomes the classical method of gradient descent with a constant step:

$$x_i = x_{i-1} - \frac{1}{Q_0 l_0} f'_0(x_{i-1}).$$

*Example 2.* Let  $m = 0$  and  $G$  be arbitrary. The function  $f_{i,x}(y)$  attains a global minimum (i.e. over all  $y \in E$ ) at the point

$$y_x = x - \frac{1}{Q_0 l_0} f'_0(x).$$

As we move away from  $y_x$ , the function increases proportionally to  $|y - y_x|^2$ . Therefore its minimum on  $G$  is the point of  $G$  closest to  $y_x$ , i.e.  $\pi_G(y_x)$ . The method thus is of the form

$$x_i = \pi_G \left( x_{i-1} - \frac{1}{Q_0 l_0} f'_0(x_{i-1}) \right).$$

We obtain the standard gradient projection method with a constant step.

Apart from these two cases, the method described is acceptable in practice, it would appear, only when  $G = E$  and for small  $m$ . For, in this case, a step consists in solving a conditional quadratic problem over the whole space. Moreover, all the quadratic terms have the standard form  $\lambda x^2$ . This problem can be solved by going over to the dual problem, the objective function of which can be written out explicitly. Thus, a step reduces to solving a convex problem of small dimension ( $m$  is small, by hypothesis).

In other cases use of the method is limited because of the high computational complexity of a step.

### 7.1.5

We see that the complexity of the class  $H_{x_1}$  admits an estimate in terms of its modulus of strong convexity (apart from its dependence on the other parameters of the situation):

$$N(v) \leq \left[ Q \ln_+ \frac{1}{v} \right]. \quad (1.8)$$

It is not clear, however, to what extent this bound is accurate. Running ahead slightly, we shall say that it is 'correct as regards  $v$ ', but 'too much dependent' as regards  $Q$ . The proper bound is  $O(\sqrt{Q} \ln 1/v)$ , provided the dimension of  $E$  is sufficiently high. Obtaining lower and upper bounds of this sort and constructing a method which realizes the upper bound are our objectives in the next section.

## 7.2 QUADRATIC PROGRAMMING AND LOWER BOUNDS OF THE COMPLEXITY OF STRONGLY CONVEX CLASSES

### 7.2.1

In this section we study the complexity of the simplest model object of strongly convex programming, namely, the class of quadratic problems. We shall deal with problems of the form

$$f_{A,b,c}(x) = \frac{1}{2} \langle Ax | x \rangle - \langle b | x \rangle + c \rightarrow \min | x \in E, \quad (2.1)$$

where  $A$  is a bounded, non-negative, Hermitian operator. (A reader who is not familiar with these concepts may consider the finite-dimensional case;  $A$  is then a non-negative definite, symmetric matrix.) Also,  $b \in E$  in (2.1). The problem  $f_{A,b,c}$  is  $(l_A, Q_A)$ -strongly convex,  $l_A$  and  $l_A Q_A$  being the minimum point and maximum point in the spectrum of  $A$  (in the finite-dimensional case, the least and greatest eigenvalues of  $A$ ). Let  $H(E)$  denote the set of all such problems. If  $l_A = 0$ , then  $f_{A,b,c}$  is  $L_A$ -smooth,  $L_A$  being the maximum point in the spectrum of  $A$ .

For solving problems of this kind there is the well-known conjugate-gradients method (CGM). Its work, if one is not interested in the algorithmic aspect, looks like this. Let  $E_0 = \{0\}$  and, for  $i \geq 1$ , let  $E_i(A, b) = \mathfrak{L}(b, Ab, \dots, A^{i-1}b)$ . Here, and later,  $\mathfrak{L}$  denotes the closed linear hull of the vectors listed in the brackets. Then the  $i$ th point of the trajectory of CGM on  $f_{A,b,c}$  is the point where this function has a minimum on  $E_{i-1}(A, b)$ ,  $i \geq 1$ .

It turns out that the CGM is, in a certain precise sense, the optimal method of solving quadratic problems. We shall first prove this fact, and then estimate the laboriousness of the CGM.

### 7.2.2

Let  $H'(E)$  be a certain subclass of problems of the form (2.1) such that  $\min_x f_{A,b,c}(x) = 0$ . We shall say that  $H'(E)$  *withstands rotations* if

$$f_{A,b,c} \in H'(E) \text{ implies } f_{U^*AU, U^*b, c} \in H'(E),$$

where  $U$  is any orthogonal operator. In other words,  $H'(E)$  necessarily contains all ‘rotations’ of  $f$  whenever it contains  $f$ .

For  $f \in H'(E)$  and  $x \in E$  we define the error of  $x$  regarded as a solution of  $f$  thus: the *relative error* is

$$v(x, f) = \frac{f(x)}{f(0)}$$

and the *absolute error* is

$$\bar{v}(x, f) = f(x).$$

Suppose that a deterministic first-order method (i.e. the oracle gives only the values of  $f(x)$  and  $f'(x)$ ) is used for solving problems from  $H'(E)$ . Let  $I(v, H'(E))$  and  $J(v, H'(E))$  denote the complexity functions of the class  $H'(E)$  in relation to the arsenal of methods mentioned, corresponding respectively to the definitions of relative error and absolute error. Also let  $I_0(v, H'(E))$  resp.  $J_0(v, H'(E))$  be the least  $N$  for which a CGM with  $N$  steps still ensures a relative error resp. absolute error  $\leq v$  in the solution of all problems from  $H'(E)$ .

The sub-optimality mentioned earlier of the CGM consists in the following result.

**Theorem.** Let  $H'(E)$  withstand rotations. Then

$$I_0(v, H'(E)) \leq 2I(v, H'(E)); J_0(v, H'(E)) \leq 2J(v, H'(E))$$

(i.e. no method can be more than twice better than the CGM).

*Proof.* We shall prove the first inequality. Suppose it is false. This means that for a certain  $v > 0$  and a certain natural number  $M$  there is a regular method  $\mathcal{B}$  of solving problems from  $H'(E)$  with  $M$  steps which solves all the problems with accuracy  $v$ , whereas the CGM with  $2M$  steps does not have this capability. In particular,  $\dim E \geq 2M$ , because the CGM with  $\dim E + 1$  steps solves exactly any convex quadratic problem, (Why?)

Let  $f_{A,b,c} \equiv f$  be that problem from  $H'(E)$  which the CGM with  $2M$  steps does not solve with accuracy  $v$ . We consider the chain of subspaces  $E_i = E_i(A, b)$ . If, for a given  $i$ ,  $E_i = E_{i+1}$ , then  $E_{i+j} = E_i$  for all  $j \geq i$ . It is easy to see (and we shall not prove this fact in linear algebra) that the point  $x^*$  of a minimum of  $f_{A,b,c}$  (i.e.  $A^{-1}b$ ) lies in  $\mathfrak{L}(b, Ab, A^2b, \dots)$ . Therefore, for such an  $i$ ,  $x^* \in E_i$ . Therefore  $i \geq 2M$ , otherwise the CGM with  $2M$  steps would solve  $f_{A,b,c}$  exactly.

Thus  $E_i \neq E_{i+1}$  for  $0 \leq i \leq 2M - 1$ . Now let  $g_i$  be a unit vector lying in  $E_{i-1}$  and orthogonal to  $E_i$ ,  $i = 0, 1, \dots, 2M - 1$ . We proceed in the following way. Let  $\mathcal{B}$  put its first question at the point  $x_1$ . We find an  $e_0$  with  $|e_0| = 1$  and such that  $x_1 \in \mathfrak{L}(e_0)$ , and choose as  $e_1$  any unit vector orthogonal to  $e_0$ . Let  $U_1$  be any orthogonal operator such that  $U_1 e_0 = g_0$ ,  $U_1 e_1 = g_1$ .

Now let  $x_2$  denote the second point at which the method  $\mathcal{B}$  puts a question about the problem  $f_{U_1^*AU_1, U_1^*b, c}$ . We continue the construction in the same way.

Suppose that, up to the  $n$ th step ( $1 \leq n \leq M$ ), the orthonormal vectors  $e_0, e_1, \dots, e_{2n-4}, e_{2n-3}$  and the orthogonal operator,  $U_{n-1}$  have already been constructed such that  $U_{n-1} e_j = g_j$ ,  $0 \leq j \leq 2n-3$ . Let  $x_n$  be the  $n$ th point at which  $\mathcal{B}$  ‘puts a question’ about the problem  $f_{U_{n-1}^*AU_{n-1}, U_{n-1}^*b, c}$  (when  $n = M$  it puts out the result). We choose  $e_{2n-2}, e_{2n-1}$  so that  $x_n \in \mathfrak{L}(e_0, \dots, e_{2n-2})$  and so that  $e_0, \dots, e_{2n-1}$  forms an orthonormal system. Such a construction is possible if  $2n < \dim E$ , which is automatically so when  $n \leq M$  (we have already proved this). We now also define an orthogonal operator  $U_n$  by the condition  $U_n e_j = g_j$ ,  $0 \leq j \leq 2n-1$ . This completes the step.

After  $M$  steps of the procedure described, the vectors  $e_0, \dots, e_{2M-1}$  and an orthogonal operator  $U_M$  will have been constructed such that  $U_M e_j = g_j$ ,  $0 \leq j \leq 2M-1$ . We show that  $x_1, \dots, x_M$  is the trajectory of  $\mathcal{B}$  on the problem  $f_{U_M^*AU_M, U_M^*b, c}$ . By the lemma on indistinguishability (see Section 1.5) it suffices to verify that the information supplied by the oracle about the problems  $f_{U_M^*AU_M, U_M^*b, c}$  and  $f_{U_J^*AU_J, U_J^*b, c}$  coincide at each of the points  $x_1, x_2, \dots, x_j$ . By construction,  $x_1, \dots, x_j \in \mathfrak{L}(e_0, \dots, e_{2j-2})$ . Therefore, when  $j \geq 1$  and  $t \leq j$ ,

$$\begin{aligned} f_{U_J^*AU_J, U_J^*b, c}(x_t) &= \frac{1}{2} \langle AU_J x_t | U_J x_t \rangle - \langle b | U_J x_t \rangle + c \\ &= \frac{1}{2} \langle AU_M x_t | U_M x_t \rangle - \langle b | U_M x_t \rangle + c, \end{aligned}$$

because  $U_J$  and  $U_M$  coincide on  $\mathfrak{L}(e_0, \dots, e_{2j-2})$ . Further,

$$f'_{U_J^*AU_J, U_J^*b, c}(x_t) = U_J^* AU_J x_t - U_J^* b$$

and

$$f'_{U_M^*AU_M, U_M^*b, c}(x_t) = U_M^* AU_M x_t - U_M^* b,$$

$U_M$  coincides with  $U_J$  on  $\mathfrak{L}(e_0, \dots, e_{2j-2}) \ni x_t$  and carries this subspace into  $E_{2j-1}$ .  $A$  carries  $E_{2j-1}$  into  $E_{2j}$ , and it remains to verify that  $U_M^* = U_J^*$  on  $E_{2j}$ . This is so, since  $U_J^* = U_J^{-1}$  and  $U_M^* = U_M^{-1}$ , and  $U_J^{-1} g_s = e_s = U_M^{-1} g_s$  when  $s \leq 2j-1$ .

Similarly,  $U_J^{-1} b = U_M^{-1} b$ , because  $b \in E_1 \subset E_{2j}$  for all  $j \geq 1$ .

Thus  $x_M$  is the result of applying  $\mathcal{B}$  to  $f_{U_M^*AU_M, U_M^*b, c}$ . Since  $H'(E)$  withstands rotation, from the definition of  $\mathcal{B}$  we have

$$v \geq \frac{f_{U_M^*AU_M, U_M^*b, c}(x_M)}{f_{U_M^*AU_M, U_M^*b, c}(0)} = \frac{f_{A,b,c}(U_M x_M)}{f_{A,b,c}(0)} \geq \frac{\inf_{x \in E_{2M-1}} f_{A,b,c}(x)}{f_{A,b,c}(0)}$$

(we have used the fact that  $U_M x_M \in \mathfrak{L}(g_0, \dots, g_{2M-2}) = E_{2M-1}$ ). But the last ratio is  $> v$  (it is equal to the error of the CGM with  $2M$  steps on the problem  $f_{A,b,c}$ ). The contradiction thus obtained proves the first inequality in the theorem. The second is proved in exactly the same way. So the theorem has now been proved.

## 7.2.3

We shall apply the result obtained to estimate the complexity of classes of smooth convex and strongly convex problems. Let  $\mathbf{H}(E, Q)$  denote the set of all problems  $f_{A, b, c} \in \mathbf{H}'(E)$  such that the spectrum of  $A$  lies in the interval  $[1/Q, 1]$  and  $|b| = 1$ , and let  $\bar{\mathbf{H}}(E)$  denote the set of all problems  $f_{A, b, c} \in \mathbf{H}'(E)$  whose solutions have a norm  $\leq 1$ , and such that  $|b| = 1$ ,  $\|A\| < 1$ . The functions  $I_0, J_0$  corresponding to these classes clearly depend only on the linear dimension of  $E$  (on whether this dimension is  $+\infty$  or not—and if not, on what its value is). Accordingly, we let  $I_0(v, n, Q)$  and  $J_0(v, n)$  denote the functions  $I_0(v, \mathbf{H}(E, Q))$ ,  $\dim E = n$ , and  $J_0(v, \bar{\mathbf{H}}(E))$ ,  $\dim E = n$ . Here  $n = 1, 2, \dots, \infty$ . Then the following theorem holds.

**Theorem.** The following assertions hold.

$$(1) \quad I_0(v, n, Q) = \min \{n+1, I_0(v, \infty, Q)\}, \\ J_0(v, n) = \min \{n+1, J_0(v, \infty)\}.$$

(2) Let  $x_1 \in \text{int } G$  and let  $\rho_{|\cdot|}(x_1, \partial G) = r \leq \infty$ . Then the complexity  $N(v)$  of the class  $\mathbf{H}_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$  with  $Q = \max_i Q_i$  admits the lower bound

$$N(v) \geq \frac{1}{2} I_0(v, \dim E, Q) - 1.$$

(3) Let  $x_1 \in \text{int } G$  and let  $\rho_{|\cdot|}(x_1, \partial G) = r \leq \infty$ . Then the complexity  $N(v)$  of the class  $\mathbf{H}_{x_1}^v(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$  admits the lower bound

$$N(v) \geq \frac{1}{2} \max_{0 \leq i \leq m} I_0(\kappa_i v, \dim E, Q_i) - 1,$$

where

$$\kappa_i = \max \left[ 1, \frac{2V}{l_i r^2}, \frac{l_i}{l_0} \right].$$

In particular, if for some  $i_0$  with  $Q_{i_0} = Q$  we have  $l_{i_0} \leq l_0$  and  $V \leq \frac{1}{2} l_{i_0} r^2$ , then

$$N(v) \geq \frac{1}{2} I_0(v, \dim E, Q) - 1.$$

(4) Let  $G$  be a ball with centre at  $x_1$ . Then the complexity  $N(v)$  of the class  $\mathbf{CS}_{x_1}^v(G, m)$  admits the bound

$$N(v) \geq \frac{1}{2} J_0(2v, \dim E) - 1.$$

**Exercise 1.** Prove this theorem.

This theorem reduces the problem of finding a lower bound for the complexity of the classes considered to the computation of  $I_0(v, \infty, Q)$  and  $J_0(v, \infty)$ , i.e. to estimating the laboriousness of the CGM on a certain class of infinite-dimensional quadratic problems. We now study this matter.

It is more convenient to estimate, not  $I_0(v, \mathbf{H}'(E))$  for the corresponding  $\mathbf{H}'(E)$ , but the inverse function

$$v(n, \mathbf{H}'(E)) = \inf \{v \mid \text{the CGM with } n \text{ steps solves all problems from } \mathbf{H}'(E) \text{ with a relative error } \leq v\}.$$

$\bar{v}(n, \mathbf{H}'(E))$  is defined similarly, with the relative error replaced by the absolute error. When  $\dim E = \infty$  we let

$$v(n, \mathbf{H}'(E, Q)) = v(n, Q), \quad \bar{v}(n, \bar{\mathbf{H}}(E)) = \bar{v}(n)$$

(this is a proper definition in the sense that the left-hand side does not depend on the actual realization of  $E$  with  $\dim E = \infty$ ).

The investigation carried out in the next two sub-sections is close in character to that carried out in [20]. The resulting lower bounds of the type (2.7), (2.12) were first obtained, it would seem, in [25–26].

## 7.2.4

Let  $\Delta = [a, 1]$ ,  $0 \leq a \leq 1$ , and let  $\mu$  be a non-negative Borel measure on  $\Delta$  with  $\mu(\Delta) \leq 1$ . We put  $E = L_2(\Delta, \mu)$  and consider the quadratic problem

$$f[x] = \frac{1}{2} \int_{\Delta} t x^2(t) d\mu(t) - \int_{\Delta} x(t) d\mu(t) + \frac{1}{2} \int_{\Delta} \frac{1}{t} d\mu(t) \rightarrow \min |x \in E$$

(we suppose that  $\mu$  is such that

$$\int \frac{d\mu(t)}{t^2} < \infty \Big).$$

We examine how the CGM works on  $f$ . By definition, the  $i$ th point of the work of the CGM on  $f$  is the point of a minimum of this function on the subspace spanned by the functions  $1, t, t^2, \dots, t^{i-2}$ , i.e. on the subspace  $\mathcal{P}_{i-2}$  of polynomials of degree  $\leq i-2$ . When  $1/t \in L_2(\Delta, \mu)$  we have

$$f[x] = \frac{1}{2} \int_{\Delta} t \left( x(t) - \frac{1}{t} \right)^2 d\mu(t) \quad (2.2)$$

and  $\inf_x f[x] = 0$ . Let  $v(n, f)$ , resp.  $\bar{v}(n, f)$ , denote the relative, resp. absolute, error of the CGM with  $n$  steps on  $f$ . We find that

$$\left. \begin{aligned} v(n, f) &= \frac{\min_{p \in \mathcal{P}_{n-2}} \int_{\Delta} t (p(t) - 1/t)^2 d\mu(t)}{\int_{\Delta} 1/t d\mu(t)}, \\ \bar{v}(n, f) &= \min_{p \in \mathcal{P}_{n-2}} \int_{\Delta} t \left( p(t) - \frac{1}{t} \right)^2 d\mu(t). \end{aligned} \right\} \quad (2.3)$$

Let  $a = 1/Q$ , and let  $\mu$  be any probability measure. We introduce the measure  $\bar{\mu}(t)$ :

$$d\bar{\mu}(t) = \frac{1}{t} d\mu(t).$$

We have

$$v(n, f) = \min_{p \in \mathcal{P}_{n-2}} \frac{\int_{\Delta} (tp(t) - 1)^2 d\bar{\mu}(t)}{\int_{\Delta} d\bar{\mu}(t)}.$$

Further,  $f \in H(E, Q)$  and so

$$v(n, Q) \geq \min_{p \in \mathcal{P}_{n-2}} \frac{\int_{\Delta} (tp(t) - 1)^2 d\bar{\mu}(t)}{\int_{\Delta} d\bar{\mu}(t)} = \min_{p \in \Pi_{n-1}} \int_{\Delta} p^2(t) d\bar{\mu}(t),$$

where  $\Pi_{n-1} = \{p \in \mathcal{P}_{n-1} \mid p(0) = 1\}$ , and  $\bar{\mu}$  is the probability normalization of  $\bar{\mu}$ . When  $\mu$  changes, the measure  $\bar{\mu}$  traverses the set  $\mathcal{F}_{\Delta}$  of all Borel probability measures on  $\Delta$ . Therefore

$$\begin{aligned} v(n, Q) &\geq \max_{\gamma \in \mathcal{F}_{\Delta}} \min_{p \in \Pi_{n-1}} \int_{\Delta} p^2(t) d\gamma(t) \\ &= \min_{p \in \mathcal{P}_{n-1}} \max_{\gamma \in \mathcal{F}_{\Delta}} \int_{\Delta} p^2(t) d\gamma(t) = \min_{p \in \Pi_{n-1}} \max_{t \in \Delta} p^2(t) \end{aligned}$$

(in interchanging the order of taking the maximum and minimum we have used von Neumann's lemma). Thus

$$v(n, Q) \geq \min_{p \in \Pi_{n-1}} \max_{t \in \Delta} p^2(t). \quad (2.4)$$

(2.4) is, in fact, an exact equality, but we shall not need to use this fact.

To estimate  $\bar{v}(n)$  we proceed similarly. Let  $a = 0$ . We observe that, if

$$\int \frac{1}{t^2} d\mu(t) \leq 1,$$

then  $f \in \bar{H}(E)$ . In other words, let  $d\mu(t) = t^2 d\gamma(t)$ , where  $\gamma$  is a probability measure on  $[0, 1] = \Delta$  which is not concentrated at 0. Then  $\mu(\Delta) \leq 1$ , and so the function  $f$  from (2.2) lies in  $\bar{H}(E)$ . Therefore, in view of (2.3),

$$\bar{v}(n, f) = \min_{p \in \mathcal{P}_{n-2}} \frac{1}{2} \int_{\Delta} t(tp(t) - 1)^2 d\gamma(t),$$

i.e.

$$\bar{v}(n) \geq \min_{p \in \Pi_{n-1}} \frac{1}{2} \int_{\Delta} tp^2(t) d\gamma(t)$$

for every probability measure  $\gamma$ . Therefore, as above,

$$\bar{v}(n) \geq \max_{\gamma \in \mathcal{F}_{[0, 1]}} \min_{p \in \Pi_{n-1}} \frac{1}{2} \int tp^2(t) d\gamma(t) = \min_{p \in \Pi_{n-1}} \frac{1}{2} \max_{0 \leq t \leq 1} tp^2(t).$$

Thus

$$\bar{v}(n) \geq \frac{1}{2} \min_{p \in \Pi_{n-1}} \max_{0 \leq t \leq 1} tp^2(t). \quad (2.5)$$

The relation (2.5) is, in fact, an exact equality, but we shall not need to use this fact.

### 7.2.5

We are now in a position to estimate  $v(n, Q)$  and  $\bar{v}(n)$ . We begin with the first. From the definition of  $\Pi_{n-1}$  it is clear that

$$\begin{aligned} \min_{p \in \Pi_{n-1}} \max_{t \in \Delta} p^2(t) &= \left\{ \max_{\substack{p \in \mathcal{P}_{n-1} \\ |p(t)| \leq 1 \text{ on } \Delta}} |p(0)| \right\}^{-2} \\ &= \left[ \max_{\substack{p \in \mathcal{P}_{n-1} \\ |p(t)| \leq 1 \text{ on } [-1, 1]}} \left| p\left(1 + \frac{2}{Q-1}\right) \right| \right]^{-2} \end{aligned}$$

(in the last transformation we have extended the interval  $\Delta$  into  $[-1, 1]$ ). Thus the matter reduces to seeking a polynomial of degree  $n-1$  whose value at the point  $\bar{t} = 1 + 2/(Q-1)$  is maximal, under the condition that on  $[-1, 1]$  this polynomial does not deviate from 0 by more than 1.

The solution of this problem is given by Markov's theorem [1]. The appropriate polynomial is the Chebyshëv polynomial  $T_{n-1}(x)$  of degree  $n-1$ , independent of the choice  $\bar{t} \notin [-1, 1]$ . Moreover, for  $t \geq 1$ , we have the representation

$$T_{n-1}(t) = \cosh \{ (n-1) \cosh^{-1} t \} \quad (2.6)$$

Thus

$$v(n, Q) \geq [\cosh \{ (n-1) \cosh^{-1} (1 + 2/(Q-1)) \}]^{-2},$$

which, after an elementary estimation, gives

$$I_0(v, \infty, Q) \geq \sqrt{\frac{Q-1}{16}} \ln \frac{1}{v} + 1. \quad (2.7)$$

The calculations for  $\bar{v}(n)$  are more complicated. Let

$$g(n) = \min_{\substack{p \in \mathcal{P}_{2n-1} \\ p(0) = 0, p'(0) = 1}} \max_{t \in [0, 1]} |p(t)|. \quad (2.8)$$

It then follows from (2.5) that

$$\bar{v}(n) \geq \frac{1}{2} g(n). \quad (2.9)$$

The function  $g(t)$  can easily be calculated. Namely, let  $T_{2n-1}(x)$  be the Chebyshëv polynomial of degree  $2n-1$ , and let  $x_n$  be its zero which is furthest to the right on  $[-1, 1]$ . It turns out that

$$g(n) = \frac{1}{|T'_{2n-1}(x_n)|(1+x_n)}. \quad (2.10)$$

*Exercise 2.* Prove (2.10).

Using the fact that  $T_n(x) = \cos(n \cos^{-1} x)$ , we obtain

$$g(n) \geq \frac{\sin\left(\frac{\pi}{4(2n-1)}\right)}{2n-1}.$$

Thus

$$\bar{v}(n) \geq \frac{\pi}{2(2n-1)} > \frac{\pi\sqrt{2}}{16(2n-1)^2}, \quad (2.11)$$

which gives

$$J_0(v, \infty) \geq \frac{1}{4\sqrt{v}}. \quad (2.12)$$

*Exercise 3.* The class  $\bar{H}(E)$ , whose complexity is, in fact,  $J_0(v, \dim E)$  contains problems which are arbitrarily badly conditioned (i.e. problems not lying in  $H(E, Q)$  with arbitrarily large  $Q$ ). Explain how to reconcile the unbounded growth of the right-hand side of (2.7) when  $Q \rightarrow \infty$  with the bound (2.12).

〈The point is that the normalizations of the error in the calculation of  $I_0$  and  $J_0$  were different: in the first case, we were studying the relative error, in the second case, the absolute error.〉

## 7.2.6

Taking into account the assertions of Theorem 7.2.3, the bounds (2.7) and (2.12), we arrive at the following:

**Theorem.** For certain absolute constants  $c_1, c_2 > 0$ , the following lower bounds hold for the deterministic complexity  $N(v)$  of the classes of strongly convex and smooth convex problems on an  $n$ -dimensional Hilbert space.

A. For the class  $H_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$  with  $\max_i Q_i = Q$ , when  $x_1 \in \text{int } G$

$$N(v) \geq c_1 \min\left\{n, \sqrt{Q-1} \ln \frac{1}{v}\right\} \text{ when } v < 1. \quad (2.13)$$

B. For the class  $H_{x_1}^V(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$  with  $\max_i Q_i = Q$ , when  $x_1 \in \text{int } G$

$$N(v) \geq c_1 \min\left\{n, \sqrt{Q-1} \ln \frac{1}{v}\right\} \text{ when } v < v_0(H_{x_1}^V). \quad (2.14)$$

Here  $v_0(H_{x_1}^V) > 0$ , if, for some  $i_0$ , we have

$$Q_{i_0} = Q, l_{i_0} \leq l_0, \frac{2V}{l_{i_0} \rho_{|\cdot|}^2(x_1, \partial G)} \leq 1,$$

then we can take  $v_0(H_{x_1}^V) = 1$ .

C. For the class  $CS_{x_1}^V(G, m)$  in the case where  $G$  is a ball with centre at  $x_1$ ,

$$N(v) \geq c_2 \min\left\{n, \frac{1}{\sqrt{v}}\right\} \text{ when } v \leq 1. \quad (2.15)$$

*Remark.* The bounds provided by Theorem 7.2.6 were established on the assumption that the classes in question are equipped with an exact oracle of the first order. It can be proved that the same bounds also hold for the deterministic complexity of the classes obtained from those defined above by replacing this oracle by any local deterministic oracle. We shall not give the rather tedious proof of this assertion.

## 7.2.7

The estimates in Theorem 7.2.6 cannot describe correctly the behaviour of the complexity when  $v \rightarrow 0$  for the case where  $n \equiv \dim E < \infty$ , since their right-hand sides are bounded by a quantity of the order  $O(n)$ . This fact is connected with a special feature of the model classes of quadratic problems, the complexity of which is essentially the right-hand side of the estimates (2.13)–(2.15). In fact, any quadratic problem can be solved exactly in  $n$  steps. It is clear that these bounds are automatically under-estimated for those accuracies  $v$  for which the ‘dimensional effect’ comes into play (for (2.13)–(2.14) these are those  $v$  for which  $\sqrt{Q-1} \ln 1/v \gg n$ , and for (2.15) those  $v$  for which  $1/\sqrt{v} \gg n$ ).

*A priori* it might be found that (2.13)–(2.15) are essentially under-estimated for all  $v$ . But in the following sections it will be proved that actually this is not so, at least for the classes  $H_{x_1}$  with  $m = 0$ , for the classes  $H_{x_1}^V$ , and for the classes  $CS_{x_1}^V$ . We shall learn how to solve problems from these classes with a laboriousness respectively

$$\sim \sqrt{Q} \left[ \ln \frac{1}{v} \right] \ln Q \left[ \sim \sqrt{Q} \right] \ln^2 Q \left[ \sim \ln \frac{1}{v} \right], \text{ and } \sim \frac{1}{\sqrt{v}} \left[ \ln^3 \frac{1}{v} \right].$$

Comparison of these upper bounds of the complexity with the lower bounds (2.13)–(2.15) shows that the latter describe ‘almost correctly’ (i.e. up to a

logarithmic factor) the behaviour of the complexity  $N(v)$  of the above classes in the range of variation of  $v$  and  $Q$  in which the ‘dimensional effect’ does not operate. It can be said that the estimates (2.13)–(2.15) isolate ‘in pure form’ the effect of strong convexity on the complexity; they show what happens asymptotically as  $\dim E \rightarrow \infty$ . When  $\dim E$  is finite, various kinds of ‘side-effects’ can arise for the classes considered. For example, for solving problems  $f \in H_{x_1}^V, f \in H_{x_1}^V, f \in CS_{x_1}^V$ , we can use the methods for solving general convex problems, ignoring in practice the specific character of the classes under consideration. Analysis of this possibility shows that, when  $\dim E$  is finite, the bounds (2.13)–(2.15) cannot be exact for all values of  $v$ .

We now sharpen the lower bounds for the complexity of classes of strongly convex and smooth convex problems which are effective over the whole range of values of  $v > 0$ .

**Theorem.** There are absolute constants  $v_0, c_1, c_2 > 0, Q_0$ , and functions  $n(Q, v) < \infty, n(v) < \infty$ , such that the deterministic and stochastic complexities of the classes

$$\begin{aligned} H_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m), H_{x_1}^V(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m), \\ CS_{x_1}^V(G, m) \end{aligned}$$

satisfy the following lower bounds; here  $n = \dim E$ .

A. For the class  $H_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$  with  $\max_i Q_i = Q$ , when  $x_1 \in \text{int } G$  and  $Q \geq 2$ ,

$$N(v) > c_1 \frac{\min(n, \sqrt{Q})}{\ln \min(n, \sqrt{Q})} \ln \frac{1}{v} \text{ when } v < 1. \quad (2.16)$$

Under the conditions  $n \geq (Q, v), Q \geq Q_0$  and  $v \leq v_0$  the bound

$$\tilde{N}(v) \geq c_1 \sqrt{Q} \ln \frac{1}{v} \quad (2.17)$$

also holds.

B. For the class  $H_{x_1}^V(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$  with  $\max_i Q_i = Q$  when  $x_1 \in \text{int } G$  and  $Q \geq 2$

$$N(v) \geq c_1 \frac{\min(n, \sqrt{Q})}{\ln \min(n, \sqrt{Q})} \ln \frac{1}{v} \text{ for } v \leq v_0(H_{x_1}^V). \quad (2.18)$$

Here  $v_0(H_{x_1}^V) > 0$ . If, for some  $i_0$ , we have

$$Q_{i_0} = Q, l_{i_0} \leq l_0, \frac{2V}{l_{i_0} \rho_{|\cdot|}^2(x_1, \partial G)} \leq 1,$$

then it is possible to take  $v_0(H_{x_1}^V) = 1$ . Under the conditions  $n \geq n(Q, v), Q > Q_0$ , and  $v < \min\{v_0, v_0(H_{x_1}^V)\}$ , the following bound also holds:

$$\tilde{N}(v) \geq c_1 \sqrt{Q} \ln \frac{1}{v}. \quad (2.19)$$

C. For the class  $CS_{x_1}^V(G, m)$  in the case where  $G$  is a ball with centre at  $x_1$ ,

$$N(v) \geq c_2 \min \left\{ \frac{1}{\sqrt{v}}, \frac{n}{\ln n} \ln \frac{1}{v} \right\} \text{ for } v < v_0. \quad (2.20)$$

If also  $n > n(v)$ , then

$$\tilde{N}(v) \geq \frac{c_2}{\sqrt{v}} \text{ for } v < 1. \quad (2.21)$$

We point out that the above bounds for  $N(v)$  and  $\tilde{N}(v)$  also hold for the classes obtained from the classes above by replacing the exact first-order oracle by any local deterministic oracle (for  $\tilde{N}(v)$ —by any local oracle).

The proof of Theorem 7.2.7 is very tedious, and we omit it.

## 7.2.8

The results obtained show what should be striven for. We see that it is impossible to solve strongly convex problems by methods which converge more rapidly than linearly (i.e. the asymptotic behaviour of the complexity as  $v \rightarrow +0$  cannot be of a better order than  $O(\ln 1/v)$ ). On the other hand, the gradient method in Section 7.1.4 provides this asymptotic behaviour. However, the complexity of classes of strongly convex programming problems is determined not merely by the required accuracy, but also by the ‘conditionality’ of the classes considered—i.e. by their modulus of strong convexity  $Q$ . The lower bound for the complexity here (in the case  $\dim E = \infty$ ) is  $O(\sqrt{Q} \ln 1/v)$ , and the bound for the laboriousness of the gradient method is  $O(Q \ln 1/v)$ .

The lower bound is, as already mentioned, exact, and so the gradient method is ‘too sensitive’ to the conditionality parameter  $Q$  of the class. But, in fact, badly conditioned problems (with large  $Q$ ) are encountered rather often, and therefore it is desirable to modify the gradient method so as to reduce its sensitivity to  $Q$  down to the theoretically possible level. This problem is solved practically in the next three sections.

We further remark that the results of Theorem 7.2.7 show that the stochastic and deterministic complexities (or rather, their lower bounds) of the classes in question behave in the same way asymptotically as  $\dim E \rightarrow \infty$ , and so here too, as in the case of general convex problems, deterministic methods which ‘almost’ realize the lower bound of  $N(v)$  (and such methods will be constructed) ‘almost’ realize the lower bound of  $N(v)$  as well; that is, their randomization cannot produce a ‘radical’ effect, at any rate not for large  $\dim E$ .

### 7.3 STRONGLY CONVEX PROGRAMMING; UNCONSTRAINED PROBLEMS

#### 7.3.1

In this section a method will be constructed for solving strongly convex problems of the form  $f_0(x) \rightarrow \min | E$ , (i.e. in our notation, problems of the class  $H_{x_1}(E, 0; l_0; Q_0)$ ) which have a lower bound for laboriousness of the order  $O(\sqrt{Q} \ln Q \ln 1/\nu)$ , which, by Theorem 7.2.6, is the same (up to the logarithmic factor  $O(\ln Q)$ ) as the lower bound for the complexity of the classes (at least, when  $\dim E$  is large). We start by describing the idea of the method. To construct a method with a laboriousness  $O(\sqrt{Q} \ln 1/\nu)$  is the same as learning how to decrease the initial residual by a factor of 2 after  $O(\sqrt{Q})$  steps. The CGM is able to do this on quadratic problems. It is natural to try to extend this method to the general case, still preserving this property.

Let us establish the ‘true’ properties of the CGM—as far as possible, the simplest ones—by virtue of which it is able to halve the residual so quickly. After that we can try to ‘stretch’ these properties to meet the requirements dictated by the general case.

It turns out that the CGM halves the residual after  $O(\sqrt{Q})$  steps because of some simple geometrical circumstances. Let  $x_1 = 0, x_2, \dots, x_M$  be the first  $M$  points of the trajectory of the CGM on the quadratic problem  $f = f_{A, b, c} \in H(E, Q)$ , and let  $p_i = f'(x_i)$ . Then it is clear from the definition of the CGM that

- (1)  $f(x_{i+1}) \leq f(x_i) - \frac{1}{2} p_i^2$ ;
- (2)  $p_i$  is orthogonal to  $p_1 + \dots + p_{i-1}$ ;
- (3)  $p_i$  is orthogonal to  $x_i - x_1$ ;
- (4)  $\langle p_i | x^* - x_i \rangle \leq f(x^*) - f(x_i)$ , where  $x^*$  is the minimizing point of  $f$ ;
- (5) if  $x^*$  is the minimizing point of  $f$ , then

$$V_f(x_1) = f(x_1) - f_* \geq \frac{1}{2Q} (x^* - x_1)^2.$$

For,  $x_{i+1}$  is the minimizing point of  $f$  on  $E_i(A, b)$ , which contains  $x_i$  and  $p_i$ . Therefore

$$f(x_{i+1}) \leq f(x_i - p_i) \leq f(x_i) - p_i^2 + \frac{1}{2} p_i^2 = f(x_i) - \frac{1}{2} p_i^2$$

(we have used (1.1)). This proves (1). Assertion (2) is obvious, because  $p_1, \dots, p_{i-1} \in E_{i-1}(A, b)$ , whereas  $p_i$  is orthogonal to  $E_{i-1}(A, b)$  (since  $x_i$  is the minimizing point of  $f$  on  $E_{i-1}(A, b)$ ). Finally,  $x_i \in E_{i-1}$ , and so  $p_i$  is orthogonal to  $x_i$ , which proves (3). The inequality (5) follows immediately from (1.1) when  $x = x^*, y = x_1$ . The relation (4) is obvious because of the convexity of  $f$ .

It turns out that it follows immediately from (1)–(5) that the CGM rapidly halves the residual. For, let us suppose that

$$V_f(x_M) > \frac{1}{2} V_f(x_1).$$

Then, by (1)

$$\sum_{i=1}^{M-1} p_i^2 \leq V_f(x_1). \quad (3.1)$$

Moreover, by (4)

$$f(x^*) \geq f(x_i) + \langle p_i | x^* - x_i \rangle,$$

i.e. using (3) we have

$$\langle p_i | x^* - x_1 \rangle \leq -f(x_i) + f(x^*) = -V_f(x_i) \leq -V_f(x_M) < -\frac{V_f(x_1)}{2}.$$

Adding these inequalities over  $i = 1, \dots, M-1$ , we obtain

$$-\left| \sum_{i=1}^{M-1} p_i \right| \cdot |x^* - x_1| \leq \left\langle \sum_{i=1}^{M-1} p_i | x^* - x_1 \right\rangle \leq -\frac{(M-1)V_f(x_1)}{2},$$

i.e.

$$|x^* - x_1| \left| \sum_{i=1}^{M-1} p_i \right| \geq \frac{(M-1)V_f(x_1)}{2}.$$

Because of (5) this inequality gives

$$\sqrt{2Q} V_f(x_1) \left| \sum_{i=1}^{M-1} p_i \right| \geq \frac{(M-1)V_f(x_1)}{2}. \quad (3.2)$$

In view of (2)

$$\left| \sum_{i=1}^{M-1} p_i \right| = \sqrt{\sum_{i=1}^{M-1} p_i^2} \leq \sqrt{V_f(x_1)}$$

(the last because of (3.1)). Substituting the last estimate into (3.2), we obtain

$$\sqrt{2Q} V_f(x_1) \geq \frac{(M-1)V_f(x_1)}{2}, \text{ or } M-1 \leq 2\sqrt{2Q}.$$

Thus, at the step with iteration number  $2\sqrt{2Q}$  the CGM has certainly halved the original residual.

In this argument only the facts (1)–(5) have been used. Let us see whether it is possible to ensure their validity in the general situation where  $f$  is any  $[1/Q, Q]$ -strongly convex function and, as before,  $p_i = f'(x_i)$ . Here the relations (4) and (5) hold automatically: the fact that  $f$  was quadratic was not needed in deriving them from (1.1). To prove (1) it suffices to shift from  $x_i$  to the point  $\bar{x}_i = x_i - p_i$  (in doing this,  $f$  decreases by not less than  $\frac{1}{2} p_i^2$ , because of (1.1)), and then to take as  $x_{i+1}$  any point such that  $f(x_{i+1}) \leq f(\bar{x}_i)$ .

We have only to think out how to ensure that (2) and (3) hold, i.e. to ensure that  $p_{i+1}$  is orthogonal to the pair of vectors  $x_{i+1} - x_i$  and  $\sum_1^i p_j = q_i$ . For this it is sufficient that  $x_{i+1}$  be the minimizing point of  $f$  on some affine subspace  $E^i \subset E$  which contains the point  $x_i$  and the point  $x_i + q_i$ . It is still necessary to ensure that the inequality  $f(x_{i+1}) \leq f(\bar{x}_i)$  shall hold, and for this it suffices that  $E^i$  shall contain  $\bar{x}_i$ .

Thus  $E^i$  must contain  $x_i$ ,  $x_i + q_i$ , and  $\bar{x}_i$ ; so  $E^i$  can be regarded as a 2-dimensional (affine) plane passing through these three points. The construction of  $x_{i+1}$  thus reduces to the minimization of  $f$  on a 2-dimensional plane. But we know how to solve 2-dimensional problems quickly (by the MCG or MMCG) irrespective of whether  $f$  is strictly convex or just convex. Of course, we cannot solve the problem of minimizing  $f$  on  $E^i$  exactly, and so, in the implementation of this approach, points (1)–(3) above will be satisfied approximately.

It is not difficult, however, to estimate the accuracy with which the 2-dimensional problems must be solved, and it turns out that to solve them with this accuracy requires a number of steps of the order  $O(\ln Q)$ . Thus the construction of  $x_{i+1}$  will ‘cost’  $O(\ln Q)$  questions about  $f$ —with accuracy up to the cost of the 2-dimensional optimization of  $f$ . It suffices to construct altogether  $O(\sqrt{Q})$  points  $x_i$ ; if (1)–(5) hold with the necessary accuracy, then the residual is halved after this time. Thus the residual can be halved after a time  $O(\sqrt{Q} \ln Q)$ , losing to the quadratic CGM by  $O(\ln Q)$  times altogether. Having halved the residual, we take the last of the points constructed to be the new value of  $x_1$ , and then repeat the process, and so onwards. (In ‘analytical’ analogues of the CGM, this strategem is known as ‘re-initialization’ (a word due to Polak: see Section 8.2.1, and [27], pp. 46, 55, 259).

### 7.3.2

We now give a precise description of the method. The method is applicable to problems of the class  $H_{x_1}(E, 0; l_0; Q) \equiv H_{x_1}$ . It is made up of *stages*, a stage—of *periods*, and a period—of *steps*.

1°. Let  $v < 1$  be the required accuracy. The method constructed to this accuracy consists of

$$M_0(v) = \left\lceil \log_3 \frac{1}{v} \right\rceil$$

stages; the structure of a stage does not depend on  $v$ . The  $i$ th stage ( $i = 1, \dots, M_0(v)$ ) has an *input*  $\bar{x}_i$ , and an *output*  $\hat{x}_{i+1}$ , which (when  $i < M_0(v)$ ) is the input of the next stage. The output of the last stage,  $\hat{x}_{M_0(v)+1}$ , is the result of the application of the method to the problem under solution. The input of the first stage is the point  $\bar{x}_1 = x_1$ .

2°. All stages have identical construction (up to the value of the input). We shall describe the stage which has an input  $\bar{x}$ . Let  $f$  be the problem under solution. Let  $\bar{Q} = \max\{Q, 2\}$ . The stage consists of  $M_1(Q) = \lceil 2\sqrt{\bar{Q}} \rceil$  periods; the  $i$ th period ( $1 \leq i \leq M_1(Q)$ ) has the input  $(x_i, q_i)$ ,  $x_i, q_i \in E$ , and the output  $(x_{i+1}, q_{i+1})$ , which—when  $i < M_1(Q)$ —is the input of the  $(i+1)$ th period. Here  $x_1 = \bar{x}$ ,  $q_1 = 0$ ;  $x_{M_1(Q)+1}$  is the output of the stage.

3°. The  $i$ th period is constructed as follows. Let  $(x_i, q_i)$  be its input. Let  $E^i$  denote the affine subspace of  $E$  spanned by the points  $\bar{x}$ ,  $x_i$ ,  $\bar{x} + q_i$ , and let  $V_i$  be the ball with radius  $r = (2/l_0)|f'(x)|$  with centre at  $\bar{x}$ . In the  $i$ th period the following operations are carried out.

- (1) By means of MCG or MMCG constructed for a relative accuracy of  $\delta = (1/(96\bar{Q}))^2$ , the problem of minimizing  $f$  on  $E^i \cap V_i$  is solved. Let  $M_2(Q)$  be the number of steps in the corresponding method. Then, by the results of Chapter 2,  $M_2(Q) \leq c \ln \bar{Q}$ , where  $c$  is an absolute constant.
- (2) Let  $\hat{x}_i$  be the result of the work of the ‘2-dimensional’ method. If  $f(\hat{x}_i) > f(x_i)$ , put  $\hat{x}_i = x_i$ ; otherwise put  $\hat{x}_i = \bar{x}_i$ .

At  $\hat{x}_i$  a question about the problem  $f$  is put (in fact, this has already been done—because MCG and MMCG are so constructed). Let  $p_i$  be the projection of  $f'(\hat{x}_i)$  on to the orthogonal complement to  $E^i - \bar{x}$ . Put

$$x_{i+1} = \hat{x}_i - f'(\hat{x}_i)/l_0 Q, \quad q_{i+1} = q_i + p_i.$$

This ends the period.

### 7.3.3

**Theorem.** The method described, constructed for an accuracy  $v < 1$ , solves all problems from  $H_{x_1}(E, 0; l_0; Q_0)$  with an error  $\leq v$  and with laboriousness

$$M(v) = M_0(v)M_1(Q)M_2(Q) \leq c_1 \ln \frac{1}{v} [\sqrt{Q} \ln \bar{Q}],$$

where  $c_1 > 0$  is an absolute constant and  $\bar{Q} = \max(Q, 2)$ .

*Proof.* The bound for the laboriousness is obvious from the description. We prove that the method ensures the required accuracy. Let  $f$  be the problem under solution. It suffices to prove that at the stage with input  $\bar{x}$  and output  $\bar{y}$  we have

$$V_f(y) \leq \frac{3}{4} V_f(\bar{x}). \quad (3.3)$$

Suppose the contrary. By construction

$$f(\bar{x}) \geq f(x_1) \geq f(\hat{x}_1) \geq f(x_2) \geq \dots \geq f(\hat{x}_{M_1}) \geq f(x_{M_1+1}) = f(\bar{y}). \quad (3.4)$$

Here  $M_1 = M_1(Q)$ . Therefore  $V_f(\bar{y}) \leq V_f(\bar{x})$ , and therefore, since (3.3) is false,

$f'(\bar{x}) \neq 0$  and  $\mathbf{V}_f(\bar{x}) \neq 0$ . Further, by (1.1),

$$f(x+t) \leq f(x) + \langle f'(x)|t \rangle + \frac{1}{2}l_0Qt^2,$$

whence

$$f(x_{i+1}) \leq f(\hat{x}_i) - \frac{p_i^2}{2l_0Q} \leq f(x_i) - \frac{p_i^2}{2l_0Q},$$

which, by (3.4) and the falsity of (3.3), leads to

$$\sum_{i=1}^{M_1} p_i^2 \leq 2l_0Q(\mathbf{V}_f(\bar{x}) - \mathbf{V}_f(\bar{y})) < \frac{1}{2}l_0Q\mathbf{V}_f(\bar{x}). \quad (3.5)$$

Moreover, because of (1.1),

$$f(\bar{x}+h) \geq f(\bar{x}) - |f'(\bar{x})||h| + \frac{1}{2}l_0|h|^2,$$

and so, outside  $V_i$ , we have  $f(x) \geq f(\bar{x})$ . Therefore  $\min_{V_i \cap E^i} f = \min_{E^i} f$ . But the variation of  $f$  on  $V_i$  clearly does not exceed

$$\begin{aligned} 2r|f'(\bar{x})| + \frac{1}{2}l_0Qr^2 &= \frac{4}{l_0}|f'(\bar{x})|^2 + \frac{2}{l_0}Q|f'(\bar{x})|^2 \\ &= \frac{4+2Q}{l_0}|f'(\bar{x})|^2 \leq \frac{4\bar{Q}}{l_0}|f'(\bar{x})|^2. \end{aligned}$$

Therefore

$$f(\hat{x}_i) - \inf_{E^i} f \leq \delta \frac{4\bar{Q}}{l_0}|f'(\bar{x})|^2. \quad (3.6)$$

Now let  $e \in E^i - \bar{x}$  be any unit vector. Then

$$f(\hat{x}_i + te) \leq f(\hat{x}_i) + t \langle f'(\hat{x}_i)|e \rangle + \frac{l_0Q}{2}t^2,$$

whence, by (3.6),

$$\frac{\langle f'(\hat{x}_i)|e \rangle^2}{2l_0Q} \leq \delta \frac{4\bar{Q}}{l_0}|f'(\bar{x})|^2,$$

i.e.

$$|P_i f'(\hat{x}_i)| \leq 3\sqrt{\delta\bar{Q}}|f'(\bar{x})|, \quad (3.7)$$

where  $P_i$  is the orthoprojector on to  $E^i - \bar{x}$ . Since  $f(\hat{x}_i) \leq f(\bar{x}_i)$ , we have  $|x_i - \bar{x}| \leq r$ . Similarly, for the minimizing point  $x^*$  of  $f$  on  $E$ , we have  $|x^* - \bar{x}| \leq r$ . But

$$f(x^*) \geq f(\hat{x}_i) + \langle f'(\hat{x}_i)|x^* - \hat{x}_i \rangle,$$

or

$$\begin{aligned} \langle p_i|x^* - \bar{x} \rangle &= \langle p_i|x^* - \hat{x}_i \rangle < f(x^*) - f(\hat{x}_i) + |p_i - f'(\hat{x}_i)||x^* - \hat{x}_i| \\ &< -\frac{3}{4}\mathbf{V}_f(\bar{x}) + 6\sqrt{\delta\bar{Q}}|f'(\bar{x})|r \\ &= -\frac{3}{4}\mathbf{V}_f(\bar{x}) + \frac{12\bar{Q}}{l_0}\sqrt{\delta}|f'(\bar{x})|^2. \end{aligned} \quad (3.8)$$

We further observe that  $\mathbf{V}_f(\bar{x}) \geq f(\bar{x}) - f\left(\bar{x} - \frac{f'(\bar{x})}{l_0Q}\right) \geq |f'(\bar{x})|^2/2l_0Q$  (we have used (1.1)). Therefore

$$\frac{12\bar{Q}}{l_0}\sqrt{\delta}|f'(\bar{x})|^2 \leq 24Q\bar{Q}\sqrt{\delta}\mathbf{V}_f(\bar{x}) \leq \frac{1}{4}\mathbf{V}_f(\bar{x})$$

from the choice of  $\delta$ . So we obtain from (3.8)

$$\langle p_i|x^* - \bar{x} \rangle < -\mathbf{V}_f(\bar{x})/2.$$

Summing over  $i = 1, 2, \dots, M_1$ , we obtain

$$\langle q_{M_1+1}|x^* - \bar{x} \rangle < -\frac{1}{2}M_1\mathbf{V}_f(\bar{x}). \quad (3.9)$$

But  $p_i$  is orthogonal to  $q_i$  and so, in view of (3.5),

$$|q_{M_1+1}| \leq \sqrt{\sum_{i=1}^{M_1} p_i^2} \leq \sqrt{\frac{l_0Q}{2}\mathbf{V}_f(\bar{x})}.$$

Moreover, it is clear from (1.1) that  $\mathbf{V}_f(\bar{x}) \geq \frac{1}{2}l_0|x^* - \bar{x}|^2$ , or

$$|x^* - \bar{x}| \leq \sqrt{\frac{2}{l_0}\mathbf{V}_f(\bar{x})}.$$

Using these bounds and (3.9) we obtain

$$-\sqrt{\frac{1}{2}l_0Q\mathbf{V}_f(\bar{x})}\sqrt{\frac{2}{l_0}\mathbf{V}_f(\bar{x})} < -\frac{1}{2}M_1\mathbf{V}_f(\bar{x}),$$

i.e.  $M_1 < 2\sqrt{Q}$ , contrary to the definition of  $M_1$ . The theorem has been proved.

*Remark.* It would be possible to adjust the stage to decrease the residual, not by  $\frac{4}{3}$  times, as we have just done, but to decrease it by some number of times  $\theta$  with  $\theta > 1$ . By choosing  $\theta$  suitably it would be possible to optimize the absolute constant  $c_1$  in the bound for the laboriousness of the method, but we shall not carry out this tedious analysis.

## 7.4 MINIMAX PROBLEMS

### 7.4.1

We now extend the method of the previous section to the case of constrained problems. We shall do this in two stages. First we learn how to solve minimax problems, i.e. problems of the form

$$\bar{f}(x) = \max_{0 \leq i \leq m} f_i(x) \rightarrow \min |x \in G, \quad (4.1)$$

where

$$f(x) = (f_0(x), \dots, f_m(x)) \in H_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m).$$

Then later, in the next section, we shall reduce constrained problems to minimax problems.

We observe that it is useful in itself to know how to solve problems like (4.1) well. For, firstly, the solution of systems of objective inequalities, i.e. finding  $x \in G$  such that  $f_0(x) \leq a_0, \dots, f_m(x) \leq a_m$ , where the  $a_i$  are specified numbers, can be reduced to minimax problems. For, finding such an  $x$  can be reduced to the problem of minimizing the function  $\max_{0 \leq i \leq m} (f_i(x) - a_i)$ . Moreover, solving a conditionally extremal problem with the optimal value of the objective function known *a priori* reduces to solving a system of objective inequalities, i.e. ultimately to a problem like (4.1). Finally, when  $m = 0$ , the problem (4.1) is the problem of minimizing  $f$  in the domain  $G$ . When  $G \neq E$  we do not as yet know how to solve this sort of problem well.

#### 7.4.2

We define the error of a point  $x \in G$  regarded as a solution of the problem (4.1) by

$$v(x, f) = \frac{\mathbf{V}_{\bar{f}}(x)}{\mathbf{V}_{\bar{f}}(x_1)}, \quad (4.2)$$

where  $\mathbf{V}_{\bar{f}}(x) = \bar{f}(x) - \min_G \bar{f}$ .

It is required to construct a method of solving a minimax problem  $\bar{f}$  associated with functions  $f \in H_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$ , using a first-order oracle and solving all such problems with a specified accuracy  $v$  with the least possible laboriousness—preferably of order  $O(\sqrt{Q} \ln(1/v))$ , where  $Q = \max_{0 \leq i \leq m} Q_i$  is the modulus of convexity of  $f$  (the results of Section 7.2 show that it is impossible to demand more, in the infinite-dimensional situation at least). In fact this method will have the bound  $O(\sqrt{Q} \ln^2 Q \ln 1/v)$  for laboriousness.

Let us sketch the idea of the method. In setting out the heuristic considerations we shall refer to a certain number of facts needed in the proof of the method. We suggest that the reader should verify these. A full proof will, of course, be given when we justify the method.

As in the case of unconstrained problems, it is enough to know how to halve the residual  $\mathbf{V}_{\bar{f}}(x)$  quickly. To do this, we shall start from the same considerations as in Section 7.3, namely, we propose a method which generates sequences  $\{x_i\}$  and  $\{p_i\}$  satisfying (1)–(5) in Section 7.3.1. Suppose, for simplicity, that  $l_0 = \dots = l_m = 1/Q$  and that  $Q_0 = \dots = Q_m = Q$ . In Section 7.3 the vectors  $f'(x_i)$  were chosen as the vectors  $p_i$ , and the property (1) was ensured by the fact that in a shift along the anti-gradient  $f_0$  decreased in the necessary manner. The conditions (2) and (3) were achieved owing to the fact

that  $x_i$  was the precise minimizing point of  $f_0$  on the plane passing through  $\bar{x}$  which contained the direction  $\sum_{j=1}^{i-1} p_j$ . The situation has now become more complicated: the function  $\bar{f}$  is in no way bound to decrease under a displacement along its anti-gradient—because it is not smooth. Therefore the choice  $p_i = \bar{f}'(x_i)$  is not convenient to us.

In order to understand what is ‘natural’ from the point of view of the demands presented by the definition of the ‘gradient’ of  $\bar{f}$ , we observe the following fact. When  $m = 0$  the gradient  $f'(x)$  of a  $[1/Q, Q]$ -strongly convex function  $f$  can be defined thus: we consider a natural majorant of the function  $f(y)$  and we find its minimizing point  $\bar{x}$ . Then the gradient is  $x - \bar{x}$ . The natural generalization of this definition to the case of the function  $\bar{f}(x) = \max_{0 \leq i \leq m} f_i(x)$  with the  $f_i$   $[1/Q, Q]$ -strictly convex is this: for a given  $x$ , let

$$f_x(y) = \max_{0 \leq i \leq m} \left\{ f_i(x) + \langle f'_i(x) | y - x \rangle + \frac{(y - x)^2}{2} \right\}$$

be the natural majorant  $\bar{f}(y)$ , and let  $Tx$  be the minimizing point of  $f_x(y)$ . Then we call the vector  $p(x) = x - Tx$  the ‘gradient’ of  $\bar{f}$  at the point  $x$ . The vector  $p(x)$  is not a support functional to  $\bar{f}$  at the point  $x$ . However, it can be shown that it has the following properties:

- (1)  $\bar{f}(Tx) \leq \bar{f}(x) - \frac{1}{2}p^2(x);$
- (2)  $\langle p(x) | x^* - x \rangle \leq \bar{f}(x^*) - \bar{f}(Tx),$

where  $x^*$  is the minimizing point of  $\bar{f}$ . Hence it is clear that, with the choice  $p_i = p(x_i)$  and with  $x_{i+1}$  constructed so that  $\bar{f}(x_{i+1}) \leq \bar{f}(Tx_i)$ , the relation (1) in Section 7.3.1 will be satisfied. The relation (5) is automatically satisfied. It remains to ensure that (2), (3), and (4) hold. It is just this which presents the greatest difficulty.

In the situation of Section 7.3 (i.e. when  $m = 0$  and  $p(x) = f'_0(x)$ ) to ensure that (2) and (3) hold,  $x_{i+1}$  is chosen as a point in the plane  $E^i$  passing through  $x_1, Tx_1$ , and  $x_1 + \sum_1^i p_j$  such that  $p(x_{i+1})$  would be orthogonal to  $E^i - x_1$ . With such a choice we would have automatically  $f_0(x_{i+1}) \leq f_0(Tx_i)$ . On the other hand, the choice mentioned reduces to a simple problem of minimizing  $f_0$  on  $E^i$ .

In the general case with  $m > 0$ , choice of  $x_{i+1}$  according to the rule  $p(x_{i+1}) \perp E^i - x_1, x_{i+1} \in E^i$  would also satisfy us: for then we would automatically have  $\bar{f}(x_{i+1}) \leq \bar{f}(Tx_i)$ , and (1)–(5) would hold. Unfortunately,  $p(x)$  is not the gradient field, and so it is not clear how to effect the choice of  $x_{i+1}$  just mentioned. We therefore have to proceed more cunningly.

Let us suppose that we know how to construct a family of points  $x_i^j \in E^i$ ,  $j = 1, \dots, M$ , such that

- (a)  $\bar{f}(Tx_i) - \bar{f}(Tx_i^j) \geq \frac{1}{2}p^2(x_i^j);$
- (b)  $\langle p(x_i^j) | Tx_i - x_i^j \rangle = 0;$

(c) a certain convex combination of the vectors  $p(x_i^j)$ , namely,  $\sum_j \mu_j p(x_i^j)$  is orthogonal to  $E^i - x_1$ .

We shall prove that it is possible to take  $p_{i+1} = \sum_j \mu_j p(x_i^j)$  and to choose as  $x_{i+1}$  the best (as regards the value of  $\bar{f}$ ) among the points  $\{Tx_i^j\}_{j=1}^M$ .

For, with such a choice of  $p_{i+1}$ , we shall have, because of (a),

$$\begin{aligned}\frac{1}{2}p_{i+1}^2 &\leq \frac{1}{2} \sum_j p^2(x_i^j) \mu_j \leq \sum_j \mu_j (\bar{f}(Tx_i) - \bar{f}(Tx_i^j)) \\ &\leq \sum_j \mu_j (\bar{f}(Tx_i) - \bar{f}(x_{i+1})) \leq \bar{f}(x_i) - \bar{f}(x_{i+1}),\end{aligned}$$

and so (1) is satisfied.

Further,  $p_{i+1}$  is orthogonal to  $E^i - x_1$  by hypothesis, and so (2) and (3) hold. It remains to prove that (4) is satisfied, or rather, that version (4') below is sufficient for the argument in Section 7.3.1. But by condition (2)

$$\langle p(x_i^j) | x^* - x_i^j \rangle \leq \bar{f}(x^*) - \bar{f}(Tx_i^j) \leq \bar{f}(x^*) - \bar{f}(x_{i+1}).$$

Because of (b) the left-hand side is equal to  $\langle p(x_i^j) | x^* - Tx_i \rangle$ . Thus

$$\langle p(x_i^j) | x^* - Tx_i \rangle \leq \bar{f}(x^*) - \bar{f}(x_{i+1}),$$

Hence

$$\langle p_{i+1} | x^* - Tx_i \rangle \leq \bar{f}(x^*) - \bar{f}(x_{i+1}),$$

or, since  $p_{i+1} \perp E^i - x_1 \ni Tx_i - x_1$ ,

$$\langle p_{i+1} | x^* - x_1 \rangle \leq \bar{f}(x^*) - \bar{f}(x_{i+1}).$$

This last is the version (4'). It is easy to see that (4') successfully replaces (4) in the argument of Section 7.3.1.

We have still to learn how to construct the points  $x_i^j$ . It may be noted that (a) follows from (b), and so we need only bother about (b) and (c). For every  $e \in E^i - Tx_i$ , there is a point  $x(e)$  on the line  $\{Tx_i + te | t \in \mathbb{R}\}$  such that  $\langle p(x(e)) | e \rangle = 0$ . This follows from the fact that  $\langle p(Tx_i + te) | e \rangle$  changes sign as  $t$  goes from  $-\infty$  to  $+\infty$ . It is easy to find the point  $x(e)$  with any required degree of accuracy, since to do so reduces to finding the root of a function which changes sign, and can be done by 'repeated bisection'. Thus, for every  $e \in E^i - x_1$  we can find  $x(e)$  such that (b) will be satisfied when  $x(e)$  is substituted for  $x_i^j$  in it.

We still have to devise a sequential procedure for choosing that  $e_j \in E^i - x_1$  so that, for some  $M$ , we can satisfy (e) as well (not exactly, of course, but sufficiently accurately). We want to ensure that

$$|\sum P_i \sum_j \mu_j p(x(e_j))| < \varepsilon$$

with  $\varepsilon$  sufficiently small. Here  $P_i$  is the orthoprojector on to  $E^i - x_1$ , and

$\sum \mu_j = 1$ ,  $\mu_j \geq 0$ . We observe that, by von Neumann's lemma,

$$\inf_{\mu_j \geq 0, \sum_j \mu_j = 1} |\sum P_i \sum_j \mu_j p(x(e_j))| = \sup_{\substack{e \in E^i - x_1 \\ |e| \leq 1}} \min_{j \leq m} \langle e | p(x(e_j)) \rangle. \quad (4.3)$$

Therefore it is sufficient to organize the choice of the  $e_j$  so that the right-hand side of (4.3) shall be less than  $\varepsilon$  for a suitable  $M$ , but an  $M$  which is as small as possible. For a given  $s$  the set

$$K_s = \left\{ e \in E^i - x_1 \mid |e| = 1, \frac{\varepsilon}{2} \leq \langle e | p(x(e_j)) \rangle \text{ for } j = 1, 2, \dots, s \right\}$$

forms an arc. If we choose as  $e_{s+1}$  a unit vector directed towards the mid-point of this arc, then the arc  $K_{s+1}$  will be at least twice shorter than  $K_s$  (because  $K_{s+1} \subset K_s$ , and  $K_{s+1}$  does not contain a neighbourhood of the mid-point of  $K_s$ , since  $\langle e_{s+1} | p(x(e_{s+1})) \rangle = 0$ ). Thus, with this system of choosing  $e_{s+1}$ , the arcs  $K_s$  will be decreased each time by more than twice. Hence it is easily deduced that, after a time  $M$  roughly proportional to  $\ln(1/\varepsilon)$ ,  $K_M$  will become void, i.e. for such an  $M$  the left-hand side of (4.3) will become less than  $\varepsilon$ . Then property (c) will be satisfied 'to accuracy  $\varepsilon$ ' for a suitable choice of  $\mu_j$ .

Thus, finding  $p_{i+1}$  and  $x_{i+1}$  requires a certain process of a 'double division into halves'. The 'cost' (in number of questions about  $f$ ) of this process turns out to be of the order  $O(\ln^2 Q)$ , since  $O(\ln Q)$  is the 'cost' of finding  $x_{i+1}$  with suitable accuracy, and  $O(\ln Q)$  of such points have to be found.

Implementation of these ideas encounters well-known difficulties which arise both from the presence of the domain  $G$  and from the fact that the  $l_i$ ,  $0 \leq i \leq m$ , are not all equal. At the cost of certain additional dodges, these difficulties can be overcome.

#### 7.4.3

We proceed to a precise description of the method. We fix the class  $H_{x_1} \equiv H_{x_1}(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$ , and put  $Q = \max_{0 \leq i \leq m} Q_i$ . We introduce the following notation.

1°. For  $f \in H_{x_1}$  and  $x \in E$ , let

$$f_{i,x}(y) = f_i(x) + \langle f'_i(x) | y - x \rangle + \frac{1}{2} l_i Q(y - x)^2$$

and

$$f_x(y) = \max_{0 \leq i \leq m} f_{i,x}(y).$$

2°. Let  $Tx$  be the point where  $f_x(y)$  has its minimum on  $G$ . Since  $f_x(y)$  is strictly convex,  $Tx$  is uniquely determined. Put  $p(x) = x - Tx$ . We note that the calculation of  $Tx$  and  $f_x(Tx)$  requires one question about  $f$  to be put at the point  $x$ .

**Description of the method**

1°. The method constructed for an accuracy  $v < 1$  consists of

$$M_0(v) = \left\lceil \ln_{\frac{1}{3}} \frac{1}{v} \right\rceil$$

stages. A stage consists of  $M_1(Q)$  periods, a period consists of  $\leq M_2(Q)$  procedures, and a procedure consists of  $M_3(Q)$  steps. Here

$$\left. \begin{aligned} M_1(Q) &= \lceil \sqrt{Q} \rceil, \\ M_2(Q) &= 4 + \lceil \frac{1}{2} M_3(Q) \rceil, \\ M_3(Q) &= \lceil 8 \log_2(100\bar{Q}) \rceil, \quad \bar{Q} = \max\{Q, 2\}. \end{aligned} \right\} \quad (4.6)$$

2°. The  $i$ th stage has the input  $\bar{x}_i$  and output  $\bar{x}_{i+1}$ ,  $i = 1, 2, \dots, M_0(v)$ . Here  $\bar{x}_1 = x_1$ , and  $\bar{x}_i \in G$ . The output of the  $i$ th stage is the input of the  $(i+1)$ th stage (if  $i < M_0(v)$ ) or is the result of the work of the method (if  $i = M_0(v)$ ). All stages are constructed in the same way, with accuracy up to the value of the output. We shall describe a stage with input  $\bar{x} \in G$ .

3°. A stage with input  $\bar{x}$  consists of periods; the  $i$ th period ( $i = 1, 2, \dots$ ) has the input  $(x_i, q_i)$ , and output  $(x_{i+1}, q_{i+1})$ ; the latter (when  $i < M_1(Q)$ ) is the input of the next period. Here  $x_i \in G$ ,  $q_i \in E$ ,  $x_1 = \bar{x}$ , and  $q_1 = 0$ . The output of the stage is a point  $x_{M_1(Q)+1}$ . In the  $i$ th period the following operations are carried out.

- A. A 2-dimensional affine subspace  $E^i$  of the space  $E$  containing the points  $\bar{x}$ ,  $x_i$ , and  $\bar{x} + q_i$  is constructed. Let  $E_0^i$  denote the 2-dimensional linear subspace  $E^i - \bar{x}$ .
- B. At the point  $x_i$  a question is put to the oracle about the problem  $f$ . The number  $L_i = (\sqrt{5} + 1)\bar{Q}|p(x_i)|$ ,  $\bar{Q} = \max\{Q, 2\}$  is constructed. If  $L_i = 0$ , then the work of the method stops, with output of the result  $x_i$ . Otherwise, a number  $\leq M_2(Q)$  of procedures are executed in accordance with rule C.
- C. C.1. The  $j$ th procedure of the  $i$ th period has the input  $e_i^{j-1} \in E_0^i$ ,  $|e_i^{j-1}| = 1$ . Here  $e_i^0$  is chosen as an arbitrary unit vector in  $E_0^i$ .
- C.2. Let  $\varphi_{ij}(t) = \langle p(x_i + te_i^{j-1}) | e_i^{j-1} \rangle$ . We note that the calculation of  $\varphi_{ij}$  at a point  $t$  specified in advance requires one question about  $f$ .

In the  $j$ th procedure the following operations are carried out.

C.2.1.  $\varphi_{ij}(-L_i)\varphi_{ij}(L_i)$  is calculated. If  $\varphi_{ij}(-L_i)\varphi_{ij}(L_i) > 0$ , then the stage ends, with output of  $\bar{y} = Tx_i$ . Otherwise, go to C.2.2.

C.2.2. If  $\varphi_{ij}(-L_i)\varphi_{ij}(L_i) \leq 0$ , then make  $M_3(Q) + 1$  iterations of solving the equation

$$\varphi_{ij}(t) = 0, \quad |t| \leq L_i,$$

by the ‘bisection’ method.

C.2.3. Let  $\Delta_i^j = [t_{ij}, \bar{t}_{ij}]$  be the last interval of localization of the root obtained after C.2.2. Put  $t_{ij} = 0$  if  $\Delta_i^j \ni 0$ . Otherwise, take  $t_{ij} = t_{ij}$  or  $\bar{t}_{ij}$  such that  $t_{ij}\varphi_{ij}(t_{ij}) \leq 0$  (this can be done, because  $\varphi_{ij}(t_{ij})\varphi_{ij}(\bar{t}_{ij}) \leq 0$ ). Then put

$$p_{ij} = p(x_i + t_{ij}e_i^{j-1}), \quad x_{ij} = x_i + t_{ij}e_i^{j-1}.$$

The  $j$ th procedure is now completed.

C.3. After finishing the  $j$ th procedure, put

$$K_j = \{e \in E_0^i \mid \langle e | p_{is} \rangle > \frac{16L_i|e|}{2^{1/2}M_3(Q)}, \quad 1 \leq s \leq j\}.$$

If  $K_j = \{\emptyset\}$  or  $j = M_2(Q)$ , then go to D. Otherwise  $K_j$  is an acute angle on the plane  $E_0^i$ . Let  $e_i^j$  be the unit vector along its bisector. Go to procedure  $j+1$ .

D. Let  $M_i \geq 1$  be the number of vectors constructed according to the rules in C. Let  $P_i$  be the orthoprojector on to the plane  $E_0^i$ . Solve the problem

$$\left| \sum_{j=1}^{M_i} \mu_j P_i p_{ij} \right| \rightarrow \min \quad \mu_j \geq 0, \quad \sum_j \mu_j = 1.$$

Let  $\{\mu_j\}$  be its solution. Then proceed as follows.

D.1. If, for some  $j$ ,  $p_{ij} = 0$ , then stop the solution process and put out  $x_{ij}$  as the result. If this is not so, then go to D.2.

D.2. Put

$$y_{ij} = T(Tx_{ij}), \quad s_{ij} = \frac{2(\bar{f}(x_i) - \bar{f}(y_{ij}))}{p_{ij}^2},$$

and let  $x_{i+1}$  be the best (according to the value of  $\bar{f}(\cdot)$ ) of the points  $y_{ij}$ ,  $T(Tx_i)$ . We note that the construction of  $y_{ij}$ ,  $s_{ij}$  requires  $2(M_2(Q) + 1)$  questions about  $f$ .

D.3. Further, put

$$\theta_i^2 = \sum_{j=1}^{M_i} \mu_j \frac{1}{s_{ij}}, \quad \theta_i > 0$$

(it will be proved that  $s_{ij} > 0$ , and so this definition makes sense). Put

$$\eta_i = \sum_{j=1}^{M_i} \mu_j p_{ij}, \quad \eta'_i = \eta_i - P_i \eta_i \text{ and } \xi_i = \frac{\eta'_i}{\theta_i}.$$

Then put  $q_{i+1} = q_i + \xi_i$ . The period is completed.

**Theorem.** The method just described solves the minimax problem  $\bar{f}$  associated with any problem  $f$  from  $H_{x_1}$  with accuracy  $v$  and with laboriousness

$$\begin{aligned} M_Q(v) &= M_0(v)M_1(Q)[M_2(Q)(M_3(Q) + 3) + 2(M_2(Q) + 1)] \\ &\leq c \sqrt{Q} \ln^2 \bar{Q} \left[ \ln \frac{1}{v} \right]. \end{aligned}$$

Here  $c$  is an absolute constant,  $v < 1$ , and  $\bar{Q} = \max\{Q, 2\}$ .

*Proof.* The bound for laboriousness is clear from the description of the method. The proof of the assertion about the accuracy is given in Section 7.6.

#### 7.4.4

We make a few remarks about the practical implementation of the proposed method. It is clear from its description that two of the operations are not trivial:

- (i) calculation for a given  $x$  of the point  $Tx$ , i.e. solution of the minimax problem generated by quadratic functions with the simplest quadratic part (proportional to a scalar square);
- (ii) implementation of the rules D, i.e. choice of a convex combination of given vectors  $\xi_1, \dots, \xi_5 \in R^2$  which has minimal norm.

The first operation is essentially the same as a step in the gradient method of Section 7.1, and the same remarks as in Section 7.1.4 can be made about its implementation. The operation is comparatively uncomplicated only when  $m = 0$ , and also when  $G = E$  and  $m$  is small (of the order of a few units). The second operation is very much simpler. It consists in solving a quadratic optimization problem on a simplex. In actual fact, the situation is even simpler. Analysis of the proof of the theorem about convergence shows that there is no need to solve the corresponding problem exactly; it suffices to find a convex combination of the vectors  $P_i p_{ij} = \xi_i$ ,  $1 \leq j \leq M_i$ , with a norm less than  $16 L_i / 2^{M_3(Q)/2}$  (we are talking about the period  $i$  of the stage with input  $\bar{x}$ ). If there is no such combination, that fact is the signal that  $x_{i+1}$  already satisfies the condition  $V(x_{i+1}) \leq 3V(\bar{x})/4$ , and so the stage can in that case be broken off, with  $x_{i+1}$  printed out as the output.

On the other hand, it is also possible to find a convex combination of the vectors  $\xi_i$  which has a small norm in the following way. Let  $K$  be a square with centre at 0 and with sides of length  $32 L_i / 2^{M_3(Q)/2}$ . Finding a convex combination of the vectors  $\xi_i$  which lies in  $K$  is a linear-programming problem with 4 constraints on an  $(M_i - 1)$ -dimensional simplex, and so to solve it linear-programming methods can be applied (the dual methods are best of all, because the number of constraints is small, namely 5—4 inequalities and 1 equality. The constraints of non-negativity are disregarded, of course). Clearly, the solution of this problem is a convex combination of the vectors  $\xi_i$  which has the necessary bound for its norm. It is easy to see that such a modification of the method will ensure that it has the required accuracy if the right-hand side of the rule defining  $M_3(Q)$  in (4.4) is increased by 1.

We further remark that in describing the method we were concerned only with its ‘order of optimality’; we did not set ourselves the task of rationalizing to the maximum extent the construction described. This would have led to a

considerable complication which it would not have been too easy to set out. If desired, the reader will be able to think up more than one way of modifying the described construction so as to increase its efficiency in practice.

## 7.5 SOLUTION OF CONSTRAINED STRONGLY CONVEX AND SMOOTH PROBLEMS

### 7.5.1

We shall apply the method in Section 7.4 for solving minimax problems to solve problems in the classes  $H_{x_1}^t$ . We shall obtain a method with the bound  $O(\sqrt{Q} \ln^2 \bar{Q} \ln 1/v)$  for the laboriousness, i.e. a method improved as regards laboriousness (at least for large  $n$  and small  $v$ , see Theorem 7.2.6) only by  $O(\ln^2 \bar{Q})$  times. Unfortunately, the reduction, proposed below, of problems in the classes  $H_{x_1}^t$  to minimax problems does not go through for the classes  $H_{x_1}$  (it was precisely this fact which caused us to introduce the classes  $H_{x_1}^t$ ). The authors know of no way of extending the results of this section to the classes  $H_{x_1}$  when  $m > 0$ , although, in all probability, there is such a way.

### 7.5.2

We shall consider the class of problems

$$H_{x_1}^t \equiv H_{x_1}^t(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m).$$

Let  $Q = \max_{0 \leq i \leq m} Q_i$  be its modulus of strong convexity, and let  $\bar{Q} = \max\{Q, 2\}$ . We describe a method of solving problems of the class  $H_{x_1}^t$  with an accuracy  $v < 1$ . The method is constructed in the following way.

1°. A question is put about the problem  $f$  at the point  $x_1$ . Let

$$g = \left( g_0, \frac{3}{v} f_1, \dots, \frac{3}{v} f_m \right),$$

where  $g_0(x) = f_0(x) - f_0(x_1) + V$ . Clearly,

$$g \in H_{x_1}^t \left( G, m; l_0, \frac{3l_1}{v}, \dots, \frac{3l_m}{v}; Q_0, \dots, Q_m \right) \equiv H_{x_1}.$$

2°. To the minimax problem  $\bar{g}$  associated with the problem  $g \in H_{x_1}^t$ , the method of solving minimax problems given in Section 7.4, corresponding to the class  $H_{x_1}$  and constructed for an accuracy  $v^2/3$ , is applied. Let

$$M_Q(v) \leq c \ln^2 \bar{Q} \sqrt{Q} \left[ \ln \frac{1}{v} \right]$$

be the number of steps in this method ( $c$  is an absolute constant), and let  $\bar{x}$  be the result of its application to  $g$ . We remark that local information about  $g$  is

found in an obvious way from local information about  $f$ , and so stage 2° is 'informationally ensured' by the possibility of putting questions about  $f$ .

3°. A question about  $f$  is put at the point  $\bar{x}$ . If

$$\max_{1 \leq j \leq m} f_j(\bar{x}) \leq vV,$$

then the point  $\bar{x}$  is put out as the solution of the problem. Otherwise, \* is put out as the result.

**Proposition.** The method described solves all problems in  $H_{x_1}^V$  with accuracy  $v$  and with laboriousness

$$\leq \bar{c} \sqrt{Q} \ln^2 \bar{Q} \left[ \ln \frac{1}{v} \right],$$

where  $\bar{c}$  is an absolute constant.

*Proof.* The bound for the laboriousness is obvious. We estimate the error of the method on a problem  $f \in H_{x_1}^V$ . Let  $\tilde{x}$  be the result of applying the method to  $f$ . The problem  $f$  may be incompatible. In that case, rule 3° already guarantees the inequality

$$v(\tilde{x}, f) \leq v. \quad (5.1)$$

Now let  $f$  be compatible, and let  $x^*$  be its solution. Then

$$|f_0(x_1) - f_0(x^*)| \leq V, f_j(x_1) \leq V, 1 \leq j \leq m.$$

Therefore

$$0 \leq g_0(x^*) \leq 2V, g_j(x_1) \leq \frac{3}{v}V, 1 \leq j \leq m,$$

whereas  $g_j(x^*) \leq 0, 1 \leq j \leq m$ . Therefore

$$\bar{g}(x^*) \leq 2V \text{ and } \bar{g}(x_1) \leq 3V/v.$$

By the properties of stage 2° we have

$$\begin{aligned} \bar{g}(\bar{x}) &\leq \left(1 - \frac{v^2}{3}\right) \min_{y \in G} \bar{g}(y) + \frac{v^2}{3} \bar{g}(x_1) \\ &\leq \left(1 - \frac{v^2}{3}\right) \bar{g}(x^*) + \frac{v^2}{3} \bar{g}(x_1) \leq \left(1 - \frac{v^2}{3}\right) g_0(x^*) + vV. \end{aligned} \quad (5.2)$$

Hence

$$g_j(\bar{x}) \leq \left(1 - \frac{v^2}{3}\right) g_0(x^*) + vV \leq 2 \left(1 - \frac{v^2}{3}\right) V + vV \leq 3V, 1 \leq j \leq m,$$

i.e.

$$f_j(\bar{x}) \leq \frac{v}{3} \cdot 3V = vV, 1 \leq j \leq m.$$

Therefore  $\bar{x} = \tilde{x}$  and

$$f_j(\tilde{x}) \leq vV, 1 \leq j \leq m. \quad (5.3)$$

On the other hand, (5.2) with  $j = 0$ , and the fact that  $g_0(x^*) \geq 0$ , give

$$g_0(\tilde{x}) \leq \bar{g}(\tilde{x}) \leq g_0(x^*) + vV,$$

i.e.  $f_0(\tilde{x}) \leq f_* + vV$ , which, together with (5.3), proves (5.1). The proposition has been proved.

### 7.5.3

The constructed method for solving problems of the classes  $H_{x_1}^V$  is immediately applicable also for solving smooth convex problems, i.e. problems in the classes  $CS_{x_1}^V(G, m)$ . For, in Section 7.1.3 we learnt to associate with the method  $\mathcal{B}_v$  of solving problems of the class

$$H_{x_1}^{(1+\frac{v}{2})V} \left( G, m; \frac{vV}{\rho_{G, x_1}^2}, \dots, \frac{vV}{\rho_{G, x_1}^2}; \frac{2+v}{v}, \dots, \frac{2+v}{v} \right)$$

with accuracy  $v/3$  and laboriousness  $M$  a method  $\mathcal{B}'_v$  of solving problems of the class  $CS_{x_1}^V(G, m)$  with accuracy  $v$  and laboriousness  $M$ . The class  $H_{x_1}^{(1+v/2)V}$  has the modulus of strong convexity  $(2+v)/v$ . Therefore by choosing the method of the previous section as  $\mathcal{B}_v$ , we construct a method of solving problems  $CS_{x_1}^V(G, m)$  with accuracy  $v < 1$  and laboriousness

$$M(v) \leq \frac{c}{\sqrt{v}} \left[ \ln^3 \frac{1}{v} \right],$$

where  $c$  is an absolute constant.

### 7.5.4

Let us examine the possibilities of potential improvement of the methods in Sections 7.3–7.5. We shall consider, as the 'competitors', any deterministic methods of solving problems of the respective classes using deterministic local oracles. We shall see that in most of the natural situations the methods are not improvable. We shall take the accuracy  $v$  under consideration to be  $\leq 1/2$ ;  $c_i$  are absolute constants, and  $n = \dim E$ . We start with strictly convex classes.

1°.  $m = 0, G = E$ ; we are discussing solution of problems in the classes  $H_{x_1}^V(E, 0; l_0; Q_0)$  and  $H_{x_1}(H, 0; l_0; Q_0)$ . It is assumed that  $Q = Q_0 \geq 2$ . The method in Section 7.3 can be used to solve problems from either of these classes. The bound for the laboriousness will be

$$M_1(v) \leq c_1 \sqrt{Q} \ln Q \ln \frac{1}{v}.$$

By the results of Section 7.2 the laboriousness of the method when  $n \geq \sqrt{Q} \ln 1/\nu$  can, in principle, be reduced by not more than  $c_2 \ln Q$  times, and when  $n \geq \sqrt{Q}$ , by not more than  $c_2 \ln^2 Q$  times.

2°.  $m = 0$ ,  $G$  is any set admissible by the conventions adopted. For solving problems in the classes  $H_{x_1}(G, 0; l_0; Q_0)$  and  $H_{x_1}(H, 0; l_0; Q_0)$  the method in Section 7.4 for minimax problems can be used. The bound for its laboriousness will be (when  $Q \geq 2$ , which is henceforth assumed)

$$M_2(\nu) \leq c_3 \sqrt{Q} \ln^2 Q \ln \frac{1}{\nu}.$$

By the results of Section 7.2, under the conditions  $n \geq \sqrt{Q}$  and  $x_1 \in \text{int } G$ , its laboriousness on the class  $H_{x_1}$  can, in principle, be reduced by not more than  $c_4 \ln^2 Q$  times. A similar assertion is also true asymptotically as  $\nu \rightarrow 0$  for the class  $H_{x_1}^\nu$  (and even for all  $\nu \leq \frac{1}{2}$  if

$$\frac{2V}{l_0 \rho_{|\cdot|}^2(x_1, \partial G)} \leq 1.$$

3°.  $m > 0$ ,  $G$  is any set admissible by the conventions adopted. We know how to 'solve well' only problems from  $H_{x_1}^\nu(G, m; l_0, \dots, l_m; Q_0, \dots, Q_m)$ . The appropriate method, described in Section 7.5.2, has the bound for laboriousness (when  $Q = \max_{0 \leq i \leq m} Q_i \geq 2$ , which is henceforth assumed)

$$M_3(\nu) \leq c_5 \sqrt{Q} \ln^2 Q \ln \frac{1}{\nu}.$$

By the results of Section 7.2 under the conditions  $n \geq \sqrt{Q}$  and  $x_1 \in \text{int } G$ , the proposed method can, in principle, be improved as regards laboriousness, asymptotically as  $\nu \rightarrow 0$ , by not more than  $c_6 \ln^3 Q$  times. When  $\rho_{|\cdot|}(x_1, \partial G)$  is sufficiently large (for example, when  $G = E$ ) and  $Q_0 = Q$ , the result holds for all  $\nu \leq \frac{1}{2}$ .

*Remark.* The strongly convex programming methods in Sections 7.3–7.5 are efficient for classes with a modulus of strong convexity  $Q$  not too close to 1 (in 1°–3° we assumed that  $Q \geq 2$ ). This is the only case of interest, of course. A purely formal remark is that, when  $Q$  is close to 1 and under the appropriate hypotheses regarding  $\dim E$ , 'pride of place' passes to the gradient method in Section 7.1 (cf. the precise bounds for its laboriousness when  $Q$  is small with the precise bounds for  $I_0(\nu, \infty, Q)$  and  $J_0(\nu, \infty)$ ). Generally, when  $Q \leq Q_0$ , the gradient method 'loses' to the methods of Sections 7.3–7.5 by not more than  $\bar{c} \sqrt{Q}$  times, and so its deficiencies express themselves only in the asymptotic behaviour as  $Q \rightarrow \infty$ .

4°. It remains to elucidate how efficient the proposed method for solving problems in the classes  $CS_{x_1}^\nu(G, m)$  is. The bound for its laboriousness is

$$M_4(\nu) \leq c_7 \frac{1}{\sqrt{\nu}} \ln^3 \frac{1}{\nu}.$$

Comparing this bound with the bounds in Section 7.3, we see that under the conditions where  $G$  is a ball with centre at  $x_1$  and  $1/n^2 \leq \nu \leq \frac{1}{2}$ , the proposed method cannot, in principle, be improved by more than  $c_8 \ln^3(1/\nu)$  times.

*Remark.* In all the assertions about the sub-optimality of the methods in Sections 7.3–7.5 the requirement that  $\dim E$  be sufficiently large appeared. This is understandable: for small values of  $\dim E$ , methods like MCG for solving general convex problems begin to compete with the methods now under consideration; the bound for laboriousness of methods of the MCG type depends only on the dimension, and not on the smoothness of the convex problems being solved. Roughly speaking, the advantage of smoothness (or of strong convexity) shows up only on problems of a sufficiently high dimensionality.

### 7.5.5

In conclusion we consider the question of the possibility of reducing the requirements concerning the *a priori* information necessary for putting into operation the methods of strongly convex programming which we have constructed. We restrict ourselves to the case  $m = 0$ . The methods described can be applied if it is known *a priori* not only that the problem  $f$  under consideration is strongly convex, but also what the values of its parameters of strong convexity ( $l_f, Q_f$ ) are. Of course, it is not necessary to know very precisely the values of the parameters, namely, the quantities  $l_f$  and  $Q_f = L_f/l_f$ . It is sufficient to know estimates for them,  $l$  ( $\leq l_f$ ) and  $L$  ( $\geq L_f$ ). It is these which make up the *a priori* information for putting the method into operation (of course, it must also be known what accuracy of solution  $\nu$  is required). It must be borne in mind, however, that the estimation of the parameters  $l_f$  and  $L_f$  is, in practice, an extremely complicated problem; in addition, it is desirable that it be solved as accurately as possible. For, an 'incorrect' estimate (for some  $l > l_f$  or  $L < L_f$ ) is, in general, unacceptable: by starting from false *a priori* information of this kind, we shall be applying a method which, generally speaking, will not ensure the requisite accuracy. An estimate which is qualitatively correct (with  $l \leq l_f$  and  $L \geq L_f$ ) but rough (with  $Q \equiv L/f \gg L_f/l_f \equiv Q_f$ ), although it will indeed ensure the requisite accuracy, will lead to an uneconomical method (the laboriousness due to the roughness of the *a priori* information is increased by  $O(\sqrt{Q/Q_f})$  times).

Thus, the proposed methods are actually efficient only when the *a priori* estimates of the parameters of strong convexity of the function being minimized are sufficiently close to the true values. But in fact even a rough (but proper) estimation of these parameters is often very difficult in practice, and this substantially restricts the domain of applicability of the methods. In contrast to this, the standard methods of strongly convex programming (the gradient method, the conjugate-gradients method, etc.) do not suffer from this

drawback. No *a priori* information about the parameters of strong convexity of the function being minimized is needed before putting them into operation. In addition, the accuracy achieved after a given time will automatically be the greater, the smaller the modulus of strong convexity of the problem being solved (this statement must not be taken too literally—on particular problems one may ‘be lucky’, even though their moduli of strong convexity may be large—but in the sense of the estimates of the laboriousness on the corresponding classes).

Thus, no *a priori* information is needed in order to apply the standard methods. Of course, if a problem in ensuring the specified accuracy does arise, then there is no way of managing without an estimate of the parameters of strong convexity (or, at least, its modulus) since, as we know, the modulus of strong convexity influences substantially the potentially attainable solution-time. But the problem of ensuring the requisite accuracy is not generally posed in practice, presumably because of the difficulty with the *a priori* information. It is much more typical that ‘infinite-step’ methods are used; the solution of a problem is carried out for a certain length of time chosen from ‘other-worldly’ considerations, and whatever is obtained in the way of accuracy—that is the accuracy which is obtained. In practice this question is usually just not examined.

At the same time, the standard methods do not by any means always ensure (even as regards order of magnitude), for a problem with a given modulus of strong convexity, the potentially possible rate of growth of accuracy as a function of the solution-time (see the examples in Chapter 8). It becomes desirable to modify the method described so that it ceases to need *a priori* information, but basically preserves its efficiency. It turns out that this can easily be done.

We associate with the proposed method an ‘infinite-step’ strongly convex-programming method  $\mathcal{B}_G$  which has the following properties. The method is applicable to problems of the form

$$f(x) \rightarrow \min | x \in G \subset E, \quad (5.4)$$

in which  $f$  is any strongly convex function on  $E$ . The method  $\mathcal{B}_G$ , like the original one, uses an exact, first-order oracle, and questions can be put about  $f$  at any point of  $E$ . The method  $\mathcal{B}_G$  requires no *a priori* information. In solving a problem  $f$  with an arbitrary modulus of strong convexity  $Q_f = L_f/l_f$  the following relation holds for the result  $x_n(f)$  of the work of the method at the moment  $n$ :

$$x_n(f) \in G, \quad v(x_n(f), f) \leq v \quad (5.5)$$

for

$$n \geq c \sqrt{Q_f} \ln^3 \bar{Q}_f \ln \frac{1}{v} \left\{ \sqrt{\bar{Q}_f} \ln^3 \bar{Q}_f \ln \frac{1}{v} \right\},$$

provided only that  $v \leq \frac{1}{2}$ . Here  $c > 0$  is an absolute constant, and  $\bar{Q} = \max \{Q_f, 4\}$ . As always,

$$v(x, f) = \frac{f(x) - \inf_G f}{f(x_1) - \inf_G f},$$

where  $x_1$  is a point fixed once and for always in  $G$ . We remark that, even if we knew exactly  $(l_f, L_f)$  in advance, we would not be able to guarantee the accuracy  $v$  in the solution of  $f$  after a time less than  $c \sqrt{\bar{Q}} \ln(1/v)$ , where  $c$  is an absolute constant (at any rate, if  $Q_f \geq 2$ ,  $x_1 \in \text{int } G$ , and  $\dim E$  were sufficiently large). Thus, when  $\dim E$  is sufficiently large and when  $x_1 \in \text{int } G$ , the time, starting from which  $\mathcal{B}_G$  ensures an accuracy  $v \leq \frac{1}{2}$  in the solution of any strongly convex problem with a non-trivial ( $\geq 2$ ) modulus of strong convexity, is not essentially greater (up to a logarithmic factor) than the potentially attainable time of solving problems with a given modulus of strong convexity. We further remark that, if  $G = E$ , then the whole of  $\ln^3 \bar{Q}_f$  in (5.5) can be replaced by  $\ln \bar{Q}_f$ .

We pass on to the construction of  $\mathcal{B}_G$ . It recalls to mind the device of ‘patching together’ the finite-step methods which implement the complexity of the class for a specified accuracy—the device used in the optimal ‘infinite-step’ method in Section 1.5; but the present construction does not reduce to that device. For, here we are solving a more complicated problem: we shall effect a ‘patching together’ with regard to both the accuracy and the *a priori* information.

We construct the method  $\mathcal{B}_G$  in two stages. First we rid ourselves of the need to know both  $l_f$  and  $L_f$  and  $v$ ; it will suffice for us to know only  $Q_f$  and  $v$ . Later we shall eliminate even this *a priori* information.

1°. Let  $\mathcal{B}_G(l, L, v)$  denote the method of solving  $(l, L/l)$ -strongly convex problems of the form (5.4) with accuracy  $v$ , which was given in Section 7.3 (for  $E = G$ ), or in Section 7.4 (for  $E \neq G$ ); and let  $M(L/l, v)$  denote its laboriousness. We fix a pair of points  $x, y \in G$ ,  $x \neq y$ . Suppose it is given that  $Q \geq 2$  and  $v \leq \frac{1}{2}$ . We shall now construct a method  $\mathcal{B}_G(Q, v)$  of solving all problems of the form (5.4) and of modulus of strong convexity  $Q_f \leq Q/2$ , with accuracy  $v$  and with the bound for laboriousness  $M_1(Q, v) \leq c_1 \ln Q M(L/l, v)$ , where  $c_1$  is an absolute constant. When  $G = E$ , we can take  $M_1(Q, v) = c_1 M(L/l, v)$ .

When  $G \neq E$  the method  $\mathcal{B}_G(Q, v)$  is constructed as follows. A question is put about the problem under solution at the points  $x$  and  $y$ , and the number

$$\rho_f = \frac{\langle f'(x) - f'(y) | x - y \rangle}{|x - y|^2}$$

is formed. If  $\rho_f \leq 0$  is obtained, then solution is discontinued, since in this case  $f$  is automatically not strongly convex. If  $\rho_f > 0$  is obtained, we put  $\bar{l}_f = \rho_f/Q$ ,

$L_f = \rho_f Q$ , and apply to  $f$  in succession the methods

$$\mathcal{B}(l_f, Ql_f, v), \mathcal{B}(2l_f, 2Ql_f, v), \dots, \mathcal{B}(2^i l_f, 2^i Ql_f, v), 0 \leq i \leq \lfloor \log_2 Q \rfloor.$$

From the results of their work we choose the best as regards the value of  $f$ . This will then be the result of applying  $\mathcal{B}_G(Q, v)$  to  $f$ . It is clear that the laboriousness of the method described admits the necessary bound. We shall prove that this method solves any problem  $f$  with  $Q_f \leq Q/2$  with accuracy  $v$ . For, from the definition of  $l_f$  and  $L_f$  we have  $\rho_f \geq l_f \geq \bar{l}_f$ . Therefore, for some  $i \leq \lfloor \log_2 Q \rfloor$ , we shall have  $2^i \bar{l}_f \leq l_f$ ; but  $2^{i+1} \bar{l}_f \geq l_f$ . Hence

$$2^i Ql_f \geq \frac{1}{2} l_f Q \geq l_f Q_f = L_f,$$

and so  $f$  will be  $(2^i \bar{l}_f, Q)$ -strongly convex. Therefore the result of the  $i$ th of the methods which make up  $\mathcal{B}_G(Q, v)$  will have an error  $\leq v$ , and so the same is true of the method  $\mathcal{B}_G(Q, v)$  itself.

In the case  $E = G$  the matter is quite simple. For, in the method of Section 7.3, knowledge of the bounds for the parameters of strong convexity  $l_f$  and  $L_f$  themselves for the function  $f$  being minimized, and not just a knowledge of the upper bound  $Q$  of their ratio  $Q_f$ , is required in all at just two places, namely, in the choice of the radius of the disc in which the '2-dimensional' minimization of  $f$  takes place, and in the choice of the step in going from  $\hat{x}_i$  to  $x_i$  along the anti-gradient of  $f$  at the point  $\hat{x}_i$  (see Section 7.3.2). In the first case, it is actually sufficient to know  $\rho_f$  and  $Q$ . The radius of the disc can then be taken to be  $r = 2Q\rho_f^{-1}|f'(\hat{x}_i)|$ . In the second case, it is possible to replace a step in the direction of the anti-gradient by minimization of  $f$  on the interval  $[\hat{x}_i, \hat{x}_i - 2Q\rho_f^{-1}|f'(\hat{x}_i)|]$ , drawn by bisection and adjusted for a relative accuracy of order  $1/Q^4$ . It is left to the reader to give exact formulae for the relative accuracies of the 1-dimensional and 2-dimensional minimization in this method and to estimate the number of its steps. In doing so he will be able to satisfy himself that, as regards the order of dependence on  $Q$  and  $v$ , they are the same as for the original version of the method.

2°. We now 'patch together' the  $\mathcal{B}_G(Q, v)$  methods over  $Q$  and  $v$ , and construct the method  $\mathcal{B}_G$ . The method  $\mathcal{B}_G$  consists of successive stages; the  $n$ th stage ( $n = 1, 2, \dots$ ) consists of  $n$  procedures. The  $k$ th procedure  $\Gamma_{n,k}$ ,  $1 \leq k \leq n$ , of the  $n$ th stage consists in the application to the problem being solved of the method  $\mathcal{B}_G(Q_k, v_{n-k})$  with  $\ln(1/v_{n-k}) = 2^{n-k}$  and with  $Q_k$  determined from the equation

$$\sqrt{Q_k} \ln^3 Q_k = 2^k \text{ if } G \neq E, \text{ or } \sqrt{Q_k} \ln Q_k = 2^k \text{ if } G = E.$$

The result formed by the method  $\mathcal{B}_G$  at time  $t$  is the best (as regards the value of  $f$ ) of the results of all the procedures completed up to this time. On the initial segment of the trajectory (until  $\Gamma_{1,1}$  has been completed) the result is  $x_1$ .

Let us satisfy ourselves that the method described ensures any accuracy  $v \leq \frac{1}{2}$ , specified in advance, on any strongly convex function  $f$  in the time promised above. To do so, we choose the least integer  $k$  such that  $Q_k \geq 2Q_f$ , and the least integer  $m$  such that  $\ln 1/v \leq 2^m$ . We put  $n = m + k$ , and observe that the procedure  $\Gamma_{n,k}$  applied to  $f$  gives a result with error  $\leq v$  (the method  $\mathcal{B}_G(Q, v)$  was so defined). Therefore after the instant of finishing the  $n$ th stage, the error of  $\mathcal{B}_G$  on  $f$  will not exceed  $v$ . It remains to calculate this time  $t_n$ . For the sake of definiteness we consider the case  $G = E$  (the argument for the general case is exactly the same). From the estimates for the laboriousness of  $\mathcal{B}_G(Q, v)$  the procedure  $\Gamma_{q,p}$  takes a time

$$\leq c_2 \sqrt{Q_p} \left( \ln \frac{1}{v_{q-p}} \right) \ln Q_p,$$

where  $v_{q-p}$  is the accuracy to which  $\Gamma_{q,p}$  is adjusted, and  $c_2$  is an absolute constant. From the definitions of  $Q_p$  and  $v_{q-p}$ , this time is not greater than  $c_2 2^q$ . Thus the duration of the  $q$ th stage is not greater than  $c_2 q \cdot 2^q$ , and the time of finishing the  $q$ th stage does not exceed  $c_2 \sum_{j=1}^q j \cdot 2^j \leq c_3 q \cdot 2^q$ , where  $c_3$  is an absolute constant. Thus  $t_n \leq c_3 n \cdot 2^n$ . From the choice of  $n$ , we have  $n = m + k$ . Hence  $k$  is the least integer such that the root  $Q_k$  of the equation  $\sqrt{Q} \ln Q = 2^k$  is not less than  $2Q_f$ . It is easy to see that then

$$2^k \leq c_4 (\sqrt{Q_f} \ln \bar{Q}_f), \quad k \leq c_5 \ln (\sqrt{\bar{Q}_f} \ln \bar{Q}_f),$$

where  $c_4, c_5$  are absolute constants.

Similarly, from the definition of  $m$ ,

$$2^m \leq c_6 \ln \frac{1}{v}, \quad m \leq c_5 \ln \left( 2 \ln \frac{1}{v} \right).$$

Therefore

$$t_n \leq c_3 c_4 c_5 c_6 \left( \sqrt{Q_f} \ln \bar{Q}_f \ln \frac{1}{v} \right) \ln \left( \sqrt{\bar{Q}_f} \ln \bar{Q}_f \ln \frac{1}{v} \right),$$

as required.

In conclusion, we remark that we have been concerned only with 'order-of-magnitude' effects, and so, in the interests of clarity, we have not complicated matters by trying to rationalize the construction described. There are many ways in which such a rationalization could be effected, for example by an optimization of the 'scales' for measuring  $Q_k$  and  $v_{n-k}$ . It would be possible, let us say, to replace  $2^k$  and  $2^{n-k}$  in the equations defining  $Q_k$  and  $v_{n-k}$  by  $a^k, a^{n-k}$  with  $a > 1$ , and to choose  $a$  so as to improve the estimates. It would be possible to choose the best of the preceding results as the initial point for the next procedure, and to apply simple rules for testing whether the function under

consideration was  $(l, Q)$ -strongly convex or not (this would allow the time of ‘superfluous’ procedures in  $\mathcal{B}_G(Q, v)$  to be cut out); and so on. A reader who wishes to apply the described construction in practice will undoubtedly think up a set of cunning strategems of this sort.

### 7.6 PROOF OF THEOREM 7.4.3

In the following proof assertions from the appropriate parts of Section 7.4 are used without special explanation.

#### 7.6.1

Let  $\Delta = \left\{ \lambda \in (\lambda_0, \dots, \lambda_m) \mid \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\}$  and  
 $f_{\lambda, x}(y) = \sum \lambda_i f_{i, x}(y).$

By von Neumann’s lemma it is clear that

$$\min_{y \in G} f_x(y) = \min_{y \in G} \max_{\lambda \in \Delta} f_{\lambda, x}(y) = \max_{\lambda \in \Delta} \min_{y \in G} f_{\lambda, x}(y).$$

Therefore, for every  $x \in E$ , there is determined a non-empty set  $\Lambda(x) \subseteq \Delta$  such that, when  $\lambda \in \Lambda(x)$ , we have

$$f_x(Tx) = f_{\lambda, x}(Tx) = \min_{y \in G} f_{\lambda, x}(y)$$

(we have used the fact that  $f_{\lambda, x}(y) \leq f_x(y)$  for all  $y$ ). We put

$$s(x) = \min_{y \in G} f_x(y),$$

and

$$\rho_{x, \lambda} = \sum_{i=0}^m \lambda_i l_i Q.$$

We fix  $f \in H_{x_1}$ , and let  $x^*$  be the point where  $\bar{f}$  has minimum  $G$ , and let  $\bar{f}_* = \bar{f}(x^*)$ . Further,  $V_{\bar{f}}(\cdot)$  is written simply as  $V(\cdot)$ :  $V(x) \equiv \bar{f}(x) - \bar{f}_*$ .

#### 7.6.2

**Lemma.** Let  $x \in G$ ,  $y \in G$ . Then

(1)  $V(Tx) \leq V(x) - (f_{\lambda, x}(x) - f_{\lambda, x}(Tx)) \leq V(x)$  for all  $\lambda \in \Lambda(x)$ ;

(2) if  $V(Tx) \geq \frac{1}{2}V(x)$ , then for  $\lambda \in \Lambda(x)$

$$V(x) \leq 2Q(f_{\lambda, x}(x) - f_{\lambda, x}(Tx));$$

#### Proof of Theorem 7.4.3

(3) if  $V(Tx) \geq \frac{1}{2}V(x)$  and  $V(y) \leq 2V(x)$ , then

$$|x - y| \leq (\sqrt{5} + 1)Q|p(x)|.$$

*Proof.* The relation (1) is an obvious consequence of the chain of inequalities

$$\bar{f}(x) = f_x(x) \geq f_{\lambda, x}(x) \geq f_{\lambda, x}(Tx) = f_x(Tx) \geq \bar{f}(Tx).$$

(2) Let  $u \in G$  and let  $h = Q^{-1}(u - x) + x$ . By (1.1) we have

$$f_i(x) + \langle f'_i(x) | u - x \rangle + \frac{l_i}{2}(u - x)^2 \leq f_i(u),$$

or

$$Q \left[ \langle f'_i(x) | h - x \rangle + \frac{l_i}{2}Q(h - x)^2 \right] \leq f_i(u) - f_i(x).$$

Therefore

$$f_{i, x}(h) \leq \frac{1}{Q}f_i(u) + \left(1 - \frac{1}{Q}\right)f_{i, x}(x).$$

Then, when  $\lambda \in \Lambda(x)$ ,

$$f_{\lambda, x}(h) \leq \frac{1}{Q}\bar{f}(u) + \left(1 - \frac{1}{Q}\right)f_{\lambda, x}(x). \quad (6.1)$$

We write  $z = h - Tx$ . Then

$$f_{\lambda, x}(h) = f_{\lambda, x}(Tx + z) = f_{\lambda, x}(Tx) + \langle q_{x, \lambda} | z \rangle + \frac{1}{2}\rho_{x, \lambda} z^2,$$

where  $q_{x, \lambda} = (f_{\lambda, x})'(Tx)$  (the derivative of  $f_{\lambda, x}(y)$  with respect to  $y$  when  $y = Tx$ ). In exactly the same way

$$\begin{aligned} f_{\lambda, x}(x) &= f_{\lambda, x}(Tx) + \langle q_{x, \lambda} | x - Tx \rangle + \frac{1}{2}\rho_{x, \lambda}(x - Tx)^2 \\ &= f_{\lambda, x}(Tx) + \langle q_{x, \lambda} | p(x) \rangle + \frac{1}{2}\rho_{x, \lambda} p^2(x). \end{aligned}$$

Substituting these relations in (6.1), we obtain

$$\begin{aligned} \frac{1}{2}\rho_{x, \lambda} z^2 &\leq \frac{1}{Q}\bar{f}(u) + \left(1 - \frac{1}{Q}\right)f_{\lambda, x}(Tx) \\ &\quad + \left(1 - \frac{1}{Q}\right)\langle q_{x, \lambda} | x - Tx \rangle + \frac{1}{2}\rho_{x, \lambda} p^2(x) \left(1 - \frac{1}{Q}\right) \\ &\quad - \langle q_{x, \lambda} | z \rangle - f_{\lambda, x}(Tx) = \frac{1}{Q}(\bar{f}(u) - f_{\lambda, x}(Tx)) \\ &\quad + \langle q_{x, \lambda} | x - Tx - h + Tx \rangle - \frac{1}{Q}\langle q_{x, \lambda} | x - Tx \rangle \\ &\quad + \frac{1}{2}\rho_{x, \lambda} p^2(x) \left(1 - \frac{1}{Q}\right) = \frac{1}{Q}(\bar{f}(u) - f_{\lambda, x}(Tx)) \\ &\quad - \frac{1}{Q}\langle q_{x, \lambda} | u - Tx \rangle + \frac{1}{2}\rho_{x, \lambda} p^2(x) \left(1 - \frac{1}{Q}\right). \end{aligned} \quad (6.2)$$

Further,  $\langle q_{x,\lambda}|u-Tx\rangle \geq 0$  when  $u \in G$ , because  $Tx$  is the point where  $f_{\lambda,x}(y)$  has a minimum on  $G$ . Therefore

$$\frac{1}{2}\rho_{x,\lambda}z^2 \leq \frac{1}{Q}(\bar{f}(u)-f_{\lambda,x}(Tx))+\frac{1}{2}\rho_{x,\lambda}p^2(x)\left(1-\frac{1}{Q}\right). \quad (6.3)$$

We put  $u = x^*$  here. We then obtain from (6.3)

$$(f_{\lambda,x}(Tx)-\bar{f}(x^*)) \leq \frac{1}{2}Q\rho_{x,\lambda}\left(p^2(x)\left(1-\frac{1}{Q}\right)\right).$$

Therefore

$$\begin{aligned} \frac{1}{2}\mathbf{V}(x) &\leq \mathbf{V}(Tx) \leq f_x(Tx)-\bar{f}(x^*) = f_{\lambda,x}(Tx)-\bar{f}(x^*) \\ &\leq \frac{1}{2}Q\rho_{x,\lambda}p^2(x)\left(1-\frac{1}{Q}\right) \leq \frac{1}{2}Q\rho_{x,\lambda}p^2(x). \end{aligned} \quad (6.4)$$

On the other hand, by (1) we have

$$\mathbf{V}(x) \geq f_{\lambda,x}(x)-f_{\lambda,x}(Tx) = \frac{\rho_{x,\lambda}}{2}p^2(x)+\langle q_{x,\lambda}|p(x)\rangle \geq \frac{\rho_{x,\lambda}}{2}p^2(x),$$

and so (6.4) gives

$$\frac{1}{2}\mathbf{V}(x) \leq Q(f_{\lambda,x}(x)-f_{\lambda,x}(Tx)),$$

which is what is required in (2).

To prove (3) we consider (6.3) with  $u = y$ , i.e.

$$z = p(x)+\frac{1}{Q}(y-x).$$

We obtain

$$\begin{aligned} \frac{1}{2}\rho_{x,\lambda}z^2 &\leq \frac{1}{Q}(\bar{f}(y)-f_{\lambda,x}(Tx))+\frac{1}{2}\rho_{x,\lambda}p^2(x) \\ &\leq \frac{1}{Q}(\bar{f}(y)-\bar{f}_*)+\frac{1}{2}\rho_{x,\lambda}p^2(x) \leq \frac{2}{Q}\mathbf{V}(x)+\frac{1}{2}\rho_{x,\lambda}p^2(x) \\ &\leq 2\rho_{x,\lambda}p^2(x)+\frac{1}{2}\rho_{x,\lambda}p^2(x) \end{aligned}$$

(we have used (6.4)). Thus  $z^2 \leq 5p^2(x)$ , or

$$\left|p(x)+\frac{1}{Q}(y-x)\right| \leq \sqrt{5}|p(x)|.$$

Hence

$$\left|\frac{1}{Q}(y-x)\right| \leq (\sqrt{5}+1)|p(x)| \text{ and } |y-x| \leq Q(\sqrt{5}+1)|p(x)|,$$

as required. The lemma has been proved.

### 7.6.3

**Lemma.** Let  $x \in G$  and  $y \in E$  be such that  $\langle p(y)|x-y\rangle \geq 0$ . Then, for all  $\lambda \in \Lambda(y)$ ,

$$\bar{f}(Ty)+\frac{1}{2}\rho_{y,\lambda}p^2(y) \leq \bar{f}(x).$$

*Proof.* Putting  $q_{y,\lambda}$  equal to the derivative of  $f_{\lambda,y}(z)$  with respect to  $z$  at the point  $z = Ty$ , we have

$$\langle q_{y,\lambda}|u-Ty\rangle \geq 0 \text{ for } u \in G$$

and

$$q_{y,\lambda} = \sum \lambda_i f'_i(y) - \rho_{y,\lambda}p(y).$$

Thus

$$\rho_{y,\lambda}p(y) = \sum \lambda_i f'_i(y) - q_{y,\lambda},$$

and so

$$\langle \sum \lambda_i f'_i(y) - q_{y,\lambda} | x - y \rangle \geq 0,$$

i.e.

$$\langle \sum \lambda_i f'_i(y) - q_{y,\lambda} | x - Ty \rangle + \langle \sum \lambda_i f'_i(y) - q_{y,\lambda} | Ty - y \rangle \geq 0.$$

But

$$\sum \lambda_i f'_i(y) - q_{y,\lambda} = \rho_{y,\lambda}p(y),$$

i.e.

$$\langle \sum \lambda_i f'_i(y) - q_{y,\lambda} | x - Ty \rangle \geq \rho_{y,\lambda}p^2(y),$$

or

$$\langle \sum \lambda_i f'_i(y) | x - Ty \rangle \geq \langle \sum \lambda_i f'_i(y) - q_{y,\lambda} | x - Ty \rangle \geq \rho_{y,\lambda}p^2(y).$$

Therefore

$$\langle \sum \lambda_i f'_i(y) | x - y \rangle + \langle \sum \lambda_i f'_i(y) | p(y) \rangle \geq \rho_{y,\lambda}p^2(y).$$

Hence, and because  $f$  is convex,

$$\sum \lambda_i f_i(x) - \sum \lambda_i f_i(y) + \langle \sum \lambda_i f'_i(y) | p(y) \rangle \geq \rho_{y,\lambda}p^2(y),$$

or

$$\sum \lambda_i f_i(x) \geq \sum \lambda_i f_i(y) + \langle \sum \lambda_i f'_i(y) | Ty - y \rangle + \frac{\rho_{y,\lambda}}{2}(Ty - y)^2 + \frac{\rho_{y,\lambda}}{2}(Ty - y)^2,$$

i.e.

$$\bar{f}(x) \geq f_{\lambda,y}(Ty) + \frac{1}{2}\rho_{y,\lambda}(Ty - y)^2 \geq \bar{f}(Ty) + \frac{1}{2}\rho_{y,\lambda}p^2(y),$$

as required. The lemma has been proved.

### 7.6.4.

**Lemma.** For all  $x, y \in E$  the following inequality holds:

$$|p(y) - p(x)| \leq 5|x - y| + \sqrt{24|x - y|^2 + 4|x - y||p(x)|}. \quad (6.5)$$

In particular,  $p(y)$  is continuous with respect to  $y$ .

*Proof.* First let  $s(y) \leq s(x)$ . We have

$$\begin{aligned} |f_{i,x}(z) - f_{i,y}(z)| &\leq |f_i(x) - f_i(y) - \langle f'_i(y) |x - y\rangle| \\ &+ \frac{l_i Q}{2} |(x-z)^2 - (y-z^2)| + |\langle f'_i(x) - f'_i(y) |z-x\rangle| \\ &\leq \frac{l_i Q}{2} (x-y)^2 + \frac{l_i Q}{2} (2|x-y||z-x| + |x-y|^2) \\ &+ l_i Q |x-y||z-x| = l_i Q |x-y|^2 + 2l_i Q |x-y||z-x|. \end{aligned}$$

Hence, when  $\lambda \in \Lambda(x)$ ,

$$\sum \lambda_i (f_{i,y}(z) - f_{i,x}(z) + l_i Q |x-y|^2 + 2l_i Q |x-y||z-x|) \geq 0. \quad (6.6)$$

We put  $z = Ty \equiv Tx + r$ . Then, from (6.6),

$$\begin{aligned} s(y) &\geq \sum \lambda_i f_{i,y}(Ty) \geq \sum \lambda_i f_{i,x}(Tx+r) \\ &- \sum \lambda_i l_i Q |x-y|^2 - 2 \sum \lambda_i l_i Q |x-y| |Tx+r-x|. \end{aligned}$$

But, when  $Tx+r \in G$ ,

$$\sum \lambda_i f_{i,x}(Tx+r) \geq s(x) + \sum \lambda_i \frac{l_i Q}{2} r^2,$$

and so

$$\begin{aligned} \sum \lambda_i f_{i,y}(Ty) &\geq \sum \lambda_i \left\{ \frac{l_i Q}{2} r^2 + s(x) \right. \\ &\left. - l_i Q |x-y|^2 - 2l_i Q |x-y| |Tx+r-x| \right\}. \end{aligned}$$

Hence it follows immediately that, for a certain  $i$ ,

$$s(y) \geq f_{i,y}(Ty) \geq \frac{l_i Q}{2} r^2 + s(x) - l_i Q |x-y|^2 - 2l_i Q |x-y| |Tx+r-x|.$$

Since, by hypothesis,  $s(y) \leq s(x)$ , we have

$$\frac{1}{2}r^2 - |x-y|^2 - 2|x-y|(|p(x)| + |r|) \leq 0,$$

i.e. for  $|r| = t$ ,

$$\frac{1}{2}t^2 - 2|x-y|t - 2|x-y|(|p(x)| + |x-y|) \leq 0,$$

and hence

$$t \leq 2|x-y| + \sqrt{4|x-y|^2 + 4|x-y|^2 + 4|x-y||p(x)|}.$$

But

$$|p(y) - p(x)| \leq |r| + |x-y|,$$

and so

$$|p(y) - p(x)| \leq 3|x-y| + \sqrt{8|x-y|^2 + 4|x-y||p(x)|}. \quad (6.7)$$

Thus (6.7) is true when  $s(y) \leq s(x)$ . Now let  $s(y) > s(x)$ . Then (6.7) gives

$$|p(y) - p(x)| \leq 3|x-y| + \sqrt{8|x-y|^2 + 4|x-y||p(y)|}.$$

It is possible that

$$|p(y) - p(x)| \equiv \Delta \leq 3|x-y|,$$

and then (6.5) is true *a fortiori*. Suppose it is not so. We put  $|x-y| = s$ . Then

$$0 \leq \Delta - 3s \leq \sqrt{8s^2 + 4s(|p(x)| + \Delta)},$$

whence

$$\Delta^2 - 6s\Delta + 9s^2 \leq 8s^2 + 4s|p(x)| + \Delta \cdot 4s$$

or

$$\Delta^2 - 10s\Delta + s^2 - 4s|p(x)| \leq 0.$$

Therefore

$$\Delta \leq 5s + \sqrt{25s^2 - s^2 + 4s|p(x)|},$$

$$\text{i.e. } |p(y) - p(x)| \leq 5|x-y| + \sqrt{24|x-y|^2 + 4|x-y||p(x)|},$$

as required.

### 7.6.5

We are now in a position to carry out the necessary analysis of the work of the method. We shall prove (and this, of course, is sufficient) that in any stage the residual is reduced by not less than  $\frac{1}{4}$ . More precisely, let the result of applying the stage with input  $\bar{x}$  to the problem  $f$  be  $\bar{y} \in G$ . Then

$$\mathbf{V}(\bar{y}) \leq \frac{3}{4} \mathbf{V}(\bar{x}). \quad (6.8)$$

To prove this, we shall assume the contrary, and obtain a contradiction. So suppose in the following that

$$\mathbf{V}(\bar{y}) > \frac{3}{4} \mathbf{V}(\bar{x}). \quad (6.9)$$

### 7.6.6

It is possible that the given stage was terminated according to the rule A (i.e. for a certain  $i$ ,  $L_i$  was  $= 0$ , or  $|p(x_i)|$  was  $= 0$ ). This means that  $\bar{y} = x_i$  (rule A). On the other hand,  $p(x_i) = 0$  implies, by virtue of the lemma in Section 7.6.3, in which we have to put  $y = x_i$  and take an arbitrary point of  $G$  as  $x$ , that  $Tx_i = x_i$  is the point where  $\bar{f}$  has a minimum of  $G$ . But then (6.9) is false. Termination according to the rule D.1 is analysed in the same way.

### 7.6.7

Thus,  $L_i > 0$  for all  $i$ , and  $p_{ij} \neq 0$  for all  $i, j$  for which  $p_{ij}$  is defined. It is possible that the  $i$ th period was not terminated in accordance with the rule C.2.1. Then

$x_{i+1}$  is the best (according to value of  $\bar{f}(\cdot)$ ) of the points  $TTx_i$ ,  $TTx_{ij}$ ,  $1 \leq j \leq M$ , and

$$\langle p(x_{ij})|x_i - x_{ij} \rangle \geq 0. \quad (6.10)$$

For, (6.10) holds because of C.2.3 since

$$\langle p(x_{ij})|x_i - x_{ij} \rangle = -t_{ij}\varphi_{ij}(t_{ij}).$$

But then, by the lemma in Section 7.6.3, we have

$$\bar{f}(Tx_{ij}) \leq \bar{f}(x_i),$$

and by the lemma in Section 7.6.2,

$$\bar{f}(T(Tx_{ij})) \leq \bar{f}(x_i).$$

Therefore

$$\bar{f}(x_{i+1}) \leq \bar{f}(\bar{x}_i).$$

### 7.6.8

Now suppose that the stage was terminated in accordance with the rule C.2.1, and that this took place at the  $i$ th period. Then  $\bar{y} = Tx_i$  and  $\bar{f}(x_i) \leq f(x_1) = f(\bar{x})$  (the latter in view of Section 7.6.7). Under our hypotheses, for a certain  $j$  we have  $\varphi_{ij}(-L_i)\varphi_{ij}(L_i) > 0$ , i.e. for a suitable choice of  $\bar{t} = \pm L_i$ ,

$$\langle p(x_i + \bar{t}e_i^{j-1})|e_i^{j-1}\bar{t} \rangle < 0.$$

We put  $y = x_i + \bar{t}e_i^{j-1}$ ; then  $\langle p(y)|x_i - y \rangle > 0$ , and so by the lemma in Section 7.6.3,  $\bar{f}(Ty) \leq \bar{f}(x_i)$ , i.e.  $V(Ty) \leq V(x_i)$ .

On the other hand,  $\bar{y} = Tx_i$  and (6.9) implies that

$$\frac{3}{4}V(\bar{x}) < V(\bar{y}) = V(Tx_i) \leq V(x_i) \leq V(\bar{x}),$$

i.e.

$$V(Tx_i) > \frac{1}{2}V(x_i).$$

Moreover,

$$V(Ty) \leq V(x_i).$$

By the assertion (3) in the lemma in Section 7.6.2 applied to the pair  $(x_i, T_y)$  we have

$$|x_i - Ty| \leq (\sqrt{5} + 1)Q|p(x_i)| \leq L_i.$$

On the other hand,

$$|x_i - Ty| = |(x_i - y) + p(x)| > |x_i - y| = L_i$$

(we have used the fact that  $\langle p(y)|x_i - y \rangle > 0$ ). Thus, on the one hand,  $|x_i - Ty| \leq L_i$ , and on the other hand,  $|x_i - Ty| > L_i$ . This contradiction

shows that the stage cannot have terminated in accordance with the rule C.2.1.

### 7.6.9

So, in all periods of the given stage,  $L_i$  was  $> 0$  and the rule C.2.1 was not applied. Also,  $p_{ij} \neq 0$  for all  $i$  and  $j$ . Let us analyse the  $i$ th period. We consider its  $j$ th procedure:  $\varphi_{ij}(t)$  changes sign on  $\Delta_i^j$ . Therefore, for a certain  $\bar{t} \in \Delta_i^j$ , we have  $\varphi_{ij}(\bar{t}) = 0$ . Let  $x'_{ij} = x_i + \bar{t}e_i^{j-1}$  and  $p'_{ij} = p(x'_{ij})$ . Then

$$\langle p(x'_{ij})|e_i^{j-1} \rangle = 0 \quad (6.11)$$

On the other hand,

$$|\bar{t} - t_{ij}| \leq \frac{L_i}{2^{M_3(Q)}}.$$

By the lemma in Section 7.6.4,

$$\begin{aligned} |p(x_{ij})| &\leq 5|x_{ij} - x_i| + \sqrt{24|x_{ij} - x_i|^2 + 4|x_{ij} - x_i||p(x_i)|} + |p(x_i)| \\ &\leq \frac{L_i}{(\sqrt{5} + 1)Q} + 5L_i + \sqrt{24L_i^2 + 4L_i} \frac{1}{(\sqrt{5} + 1)Q} L_i \leq 11L_i, \end{aligned} \quad (6.12)$$

i.e.

$$|p(x_{ij})| \leq 11L_i.$$

Again applying the lemma in Section 7.6.4, we obtain

$$\begin{aligned} |p(x_{ij}) - p(x'_{ij})| &\leq 5|x_{ij} - x'_{ij}| + \sqrt{24|x_{ij} - x'_{ij}|^2 + 4|x_{ij} - x'_{ij}||p(x_{ij})|} \\ &\leq \frac{5L_i}{2^{M_3(Q)}} + \sqrt{24 \cdot \frac{L_i^2}{2^{2M_3(Q)}} + \frac{44L_i^2}{2^{M_3(Q)}}} \leq \frac{8L_i}{2^{1/2M_3(Q)}}, \end{aligned}$$

since  $2^{M_3(Q)} > 120$ . Thus,

$$|p(x_{ij}) - p(x'_{ij})| \leq \frac{8L_i}{2^{1/2M_3(Q)}},$$

which, with (6.11), gives

$$|\langle p(x_{ij})|e_i^{j-1} \rangle| \leq \frac{8L_i}{2^{1/2M_3(Q)}}. \quad (6.13)$$

Hence it is clear that  $e_i^{j-1} \notin K_j$ . On the other hand, it is clear from the rule C.3 that  $K_j \subset K_{j-1}$ . Therefore the acute angle  $K_j$  is at least twice smaller than the angle  $K_{j-1}$ ,  $j \geq 2$ . Hence it follows that the angle  $K_j$  is not greater than  $2\pi/2^j$ . We shall show that  $K_{M_2(Q)}$  is void. For, suppose this is not so. In any case, the angle  $K_{M_2(Q)-1}$  is not greater than  $2\pi/2^{M_2(Q)-1}$ . The vector  $e_i^{M_2(Q)-1} = \bar{e}_i$  is the unit vector along its bisector, and every other unit vector  $e$  in  $K_{M_2(Q)-1}$  therefore satisfies the condition

$$|e - \bar{e}_i| \leq \frac{2\pi}{2^{M_2(Q)}}.$$

But then from (6.13) with  $j = M_2(Q)$  it follows that

$$|\langle p(x_{ij})|e \rangle| \leq \frac{8L_i}{2^{1/2M_3(Q)}} + \frac{2\pi}{2^{M_2(Q)}} |p(x_{ij})| \leq \frac{8L_i}{2^{1/2M_3(Q)}} + \frac{22\pi L_i}{2^{M_2(Q)}} \leq \frac{16L_i}{2^{1/2M_3(Q)}}$$

for every unit vector  $e \in K_{M_2(Q)-1}$ . But then  $K_{M_2(Q)-1} = \emptyset$ , contrary to hypothesis.

Thus, on termination of the series of procedures of the  $i$ th period, we always have  $K_{M_i} = \emptyset$  ( $M_i$  is the number of these procedures). Since  $K_{M_i} = \emptyset$ ,

$$\max_{e \in E_0^i, |e| \leq 1} \min_{\substack{M_i \\ \mu_j \geq 0, \sum_{j=1}^{M_i} \mu_j = 1}} \langle \sum \mu_j p_{ij} | e \rangle \leq \frac{16L_i}{2^{1/2M_3(Q)}}.$$

By von Neumann's lemma, the left-hand half of this inequality is equal to

$$\min_{\substack{\mu_j \geq 0, \sum_j \mu_j = 1}} |P_i \sum \mu_j p_{ij}| = |P_i \eta_i|.$$

Thus

$$|P_i \eta_i| \leq \frac{16L_i}{2^{1/2M_3(Q)}}. \quad (6.14)$$

### 7.6.10

We shall prove that

$$\frac{1}{132(\sqrt{5}+1)Q} \leq \frac{|p(x_{ij})|}{|p(x_i)|} \leq 11(\sqrt{5}+1)\bar{Q}. \quad (6.15)$$

For, the right-hand inequality is (6.12). We prove the left-hand inequality. Let  $y = Tx_{ij}$ . By the lemma in Section 7.6.3 and by (6.10),  $\bar{f}(y) \leq \bar{f}(x_i)$ . But, by the rule for choosing  $x_{i+1}$ , we have

$$\frac{3}{4}\mathbf{V}(\bar{x}) < \mathbf{V}(\bar{y}) \leq \mathbf{V}(Ty) \leq \mathbf{V}(y) \leq \mathbf{V}(x_i) \leq \mathbf{V}(\bar{x}),$$

and hence

$$\mathbf{V}(Ty) > \frac{1}{2}\mathbf{V}(y) \text{ and } 2\mathbf{V}(y) > \mathbf{V}(x_i).$$

Now applying assertion (3) of the lemma in Section 7.6.2 to the pair  $(y, x_i)$ , we have

$$|x_i - y| \leq (\sqrt{5}+1)Q|p(y)|. \quad (6.16)$$

But by the lemma in Section 7.6.4

$$\begin{aligned} |p(y) - p(x_{ij})| &\leq 5|y - x_{ij}| + \sqrt{24|y - x_{ij}|^2 + 4|y - x_{ij}| |p(x_{ij})|} \\ &= 5|p(x_{ij})| + \sqrt{24p^2(x_{ij}) + 4p^2(x_{ij})} \leq 11|p(x_{ij})|, \end{aligned}$$

which gives  $|p(y)| \leq 12|p(x_{ij})|$ .

Applying the lemma in Section 7.6.4 to the pair  $x_i, y$ , and using (6.16), we obtain

$$\begin{aligned} |p(x_i) - p(y)| &\leq 5(\sqrt{5}+1)Q|p(y)| \\ &+ (\sqrt{24(\sqrt{5}+1)^2 Q^2 p^2(y) + 4(\sqrt{5}+1)Qp^2(y)}) \leq 10(\sqrt{5}+1)Q|p(y)|; \end{aligned}$$

hence

$$|p(x_i)| \leq 11(\sqrt{5}+1)Q|p(y)|,$$

i.e.

$$|p(x_i)| \leq 132(\sqrt{5}+1)Q|p(x_{ij})|,$$

which proves the left-hand inequality in (6.15).

### 7.6.11

We fix  $\lambda^{ij} \in \Lambda(x_{ij})$ , and let  $\rho_{ij} = \rho_{x_{ij}, \lambda^{ij}}$ . We shall prove that

$$s_{ij} \geq \rho_{x_{ij}, \lambda^{ij}} \quad (6.17)$$

and

$$s_{ij} \leq \frac{\mathbf{V}(\bar{x})}{L_i^2} \cdot 1100(\sqrt{5}+1)^4 \bar{Q}^4. \quad (6.18)$$

For, by the lemma in Section 7.6.3,

$$\bar{f}(Tx_{ij}) \leq \bar{f}(x_i) - \frac{1}{2}\rho_{ij}p_{ij}^2,$$

whereas

$$s_{ij} = \frac{2(-\bar{f}(T(Tx_{ij})) + \bar{f}(x_i))}{p_{ij}^2} \geq \frac{2(-\bar{f}(Tx_{ij}) + \bar{f}(x_i))}{p_{ij}^2}.$$

A combination of the last two inequalities gives (6.17). Finally,

$$s_{ij}p_{ij}^2 = \frac{1}{2}(\bar{f}(x_i) - \bar{f}(T(Tx_{ij}))) \leq \frac{1}{2}(\bar{f}(x_i) - \bar{f}(x_{i+1})) \leq \frac{1}{8}\mathbf{V}(\bar{x})$$

and (6.18) follows from (6.15).

### 7.6.12

We have

$$f_q(x_{ij}) + \langle f'_q(x_{ij})|\bar{x} - x_{ij} \rangle + \frac{l_q}{2}(\bar{x} - x_{ij})^2 \leq f_q(\bar{x}),$$

$$f_q(x_{ij}) + \langle f'_q(x_{ij})|x^* - x_{ij} \rangle + \frac{l_q}{2}(x^* - x_{ij})^2 \leq f_q(x^*),$$

and so

$$\begin{aligned} 2f_q(x_{ij}) + \langle f'_q(x_{ij})|x^* - \bar{x} \rangle + 2\langle f'_q(x_{ij})|\bar{x} - x_{ij} \rangle \\ + \frac{l_q}{4}(x^* - \bar{x})^2 \leq f_q(\bar{x}) + f_q(x^*). \end{aligned}$$

From these inequalities

$$\begin{aligned} \frac{\rho_{ij}}{4Q}(x^* - \bar{x})^2 + \left\langle \sum_q \lambda_q^{ij} f'_q(\bar{x}_{ij}) | x^* - \bar{x} \right\rangle + 2 \left\langle \sum_q \lambda_q^{ij} f'_q(x_{ij}) | \bar{x} - x_{ij} \right\rangle \\ \leq \bar{f}(\bar{x}) + \bar{f}(x^*) - 2f_{\lambda^{ij}, x_{ij}}(x_{ij}). \end{aligned} \quad (6.19)$$

Let  $\eta_{ij}$  be the derivative of  $f_{\lambda^{ij}, x_{ij}}(y)$  with respect to  $y$  when  $y = Tx_{ij}$ . Then  $\langle n_{ij} | x - Tx_{ij} \rangle \geq 0$  when  $x \in G$  and

$$\sum_q \lambda_q^{ij} f'_q(x_{ij}) - n_{ij} = \rho_{ij} p_{ij}.$$

Moreover,

$$f_{\lambda^{ij}, x_{ij}}(x_{ij}) = f_{\lambda^{ij}, x_{ij}}(Tx_{ij}) + \frac{1}{2} \rho_{ij} p_{ij}^2 + \langle n_{ij} | p_{ij} \rangle.$$

Hence and from (6.19)

$$\begin{aligned} \frac{1}{4Q} \rho_{ij}(x^* - \bar{x})^2 + \langle \rho_{ij} p_{ij} + n_{ij} | x^* - \bar{x} \rangle + 2 \langle \rho_{ij} p_{ij} + n_{ij} | \bar{x} - x_{ij} \rangle \\ \leq \bar{f}(\bar{x}) + \bar{f}(x^*) - 2f_{\lambda^{ij}, x_{ij}}(Tx_{ij}) - \rho_{ij} p_{ij}^2 - 2 \langle n_{ij} | p_{ij} \rangle, \end{aligned}$$

i.e.

$$\begin{aligned} \rho_{ij} \left( \frac{(x^* - \bar{x})^2}{4Q} + \langle p_{ij} | x^* - \bar{x} \rangle + 2 \langle p_{ij} | \bar{x} - x_{ij} \rangle \right) \\ \leq \bar{f}(\bar{x}) + \bar{f}(x^*) - 2f_{\lambda^{ij}}(Tx_{ij}) + \langle n_{ij} | \bar{x} - x^* + 2x_{ij} - 2\bar{x} + 2Tx_{ij} - 2x_{ij} \rangle \\ \leq \bar{f}(\bar{x}) + \bar{f}(x^*) - 2\bar{f}(x_{i+1}) - \langle n_{ij} | \bar{x} - Tx_{ij} \rangle - \langle n_{ij} | x^* - Tx_{ij} \rangle \\ \leq \bar{f}(\bar{x}) + \bar{f}(x^*) - 2\bar{f}(x_{i+1}) \leq \mathbf{V}(\bar{x}) - 2\mathbf{V}(x_{i+1}) < -\frac{1}{2}\mathbf{V}(\bar{x}) \end{aligned}$$

(we have used the fact that

$$\langle n_{ij} | \bar{x} - Tx_{ij} \rangle \geq 0, \quad \langle n_{ij} | x^* - Tx_{ij} \rangle \geq 0$$

and

$$\mathbf{V}(x_{i+1}) > \frac{3}{4}\mathbf{V}(\bar{x}).$$

Thus

$$\langle p_{ij} | x^* - \bar{x} \rangle + 2 \langle p_{ij} | \bar{x} - x_{ij} \rangle + \frac{(x^* - \bar{x})^2}{4Q} \leq -\frac{\mathbf{V}(\bar{x})}{2\rho_{ij}} \leq -\frac{\mathbf{V}(\bar{x})}{2s_{ij}}$$

(we have used (6.17)). Moreover,  $\langle p_{ij} | x_i - x_{ij} \rangle \geq 0$ , and hence

$$\langle p_{ij} | x^* - \bar{x} \rangle + 2 \langle p_{ij} | \bar{x} - x_i \rangle + \frac{(x^* - \bar{x})^2}{4Q} \leq -\frac{1}{2} \frac{\mathbf{V}(\bar{x})}{s_{ij}}.$$

We take the weighted mean of these inequalities over  $j$  with the weights  $\mu_j$  to obtain

$$\langle \eta_i' | x^* - \bar{x} \rangle + 2 \langle \eta_i' | \bar{x} - x_i \rangle + \frac{(x^* - \bar{x})^2}{4Q} \leq -\frac{1}{2} \mathbf{V}(\bar{x}) \theta_i^2,$$

i.e.

$$\langle \eta_i' | x^* - \bar{x} \rangle + \frac{(x^* - \bar{x})^2}{4Q} \leq -\frac{1}{2} \mathbf{V}(\bar{x}) \theta_i^2 + |\Delta_i| |x^* - \bar{x}| + 2|\Delta_i| |\bar{x} - x_i|, \quad (6.20)$$

### Proof of Theorem 7.4.3

where  $\Delta_i = P_i \eta_i$ . We observe that, by (6.9),

$$\mathbf{V}(Tx_i) > \frac{1}{2} \mathbf{V}(x_i), \quad \mathbf{V}(x^*) \leq \mathbf{V}(x_i), \text{ and } \mathbf{V}(\bar{x}) \leq 2\mathbf{V}(x_i);$$

hence, by assertion (3) in the lemma in Section 7.6.2,

$$|x^* - x_i| \leq (\sqrt{5} + 1) Q |p(x_i)| = L_i,$$

and in exactly the same way,  $|\bar{x} - x_i| \leq L_i$ . Therefore

$$|x^* - \bar{x}| \leq 2L_i \text{ and } |\bar{x} - x_i| \leq L_i,$$

and so, because of (6.14), the right-hand side of (6.20) is not greater than

$$-\frac{1}{2} \mathbf{V}(\bar{x}) \theta_i^2 + \frac{64L_i^2}{2^{1/2} M_3(Q)}. \quad (6.21)$$

We now observe that, by (6.18),

$$\frac{1}{s_{ij}} \geq \frac{L_i^2}{\mathbf{V}(\bar{x}) \cdot 1100(\sqrt{5} + 1)^4 \bar{Q}^4},$$

and so,

$$\mathbf{V}(\bar{x}) \theta_i^2 \geq \frac{L_i^2}{1100(\sqrt{5} + 1)^4 \bar{Q}^4}. \quad (6.22)$$

Hence it is clear that, with the choice made for  $M_3(Q)$ , the quantity (6.21) is not greater than  $-\frac{1}{4} \mathbf{V}(\bar{x}) \theta_i^2$ .

Thus, (6.20) may be rewritten as

$$\langle \eta_i' | x^* - \bar{x} \rangle + \frac{(x^* - \bar{x})^2}{4Q} \leq -\frac{1}{4} \mathbf{V}(\bar{x}) \theta_i^2,$$

or

$$\langle \xi_i | x^* - \bar{x} \rangle + \frac{(x^* - \bar{x})^2}{4Q} \frac{1}{\theta_i} \leq -\frac{1}{4} \mathbf{V}(\bar{x}) \theta_i. \quad (6.23)$$

We now observe that

$$\sum_{i=1}^{M_1(Q)} \xi_i^2 < \frac{1}{4} \mathbf{V}(\bar{x}). \quad (6.24)$$

For, by the definition of  $s_{ij}$ ,

$$\begin{aligned} \eta_i^2 &= \left( \sum_j \mu_j p_{ij} \right)^2 = \left( \sum_j \sqrt{\frac{\mu_j}{s_{ij}}} \sqrt{\mu_j s_{ij} p_{ij}} \right)^2 \leq \sum_j \frac{\mu_j}{s_{ij}} \sum_j \mu_j s_{ij} p_{ij}^2 \\ &\leq \theta_i^2 \sum_j \mu_j (\bar{f}(x_i) - \bar{f}(x_{i+1})) = \theta_i^2 (\bar{f}(x_i) - \bar{f}(x_{i+1})), \end{aligned}$$

i.e.  $\xi_i^2 \leq \bar{f}(x_i) - \bar{f}(x_{i+1})$ , which proves (6.24).

By summing (6.23) and noting that  $\xi_i \perp \sum_{j=1}^{i-1} \xi_j$ , we find

$$-\sqrt{\frac{1}{4} \mathbf{V}(\bar{x})} |x^* - \bar{x}| + \frac{|x^* - \bar{x}|^2}{4Q} \sum_{i=1}^{M_1(Q)} \frac{1}{\theta_i} < -\frac{1}{4} \mathbf{V}(\bar{x}) \sum_{i=1}^{M_1(Q)} \theta_i,$$

and so the quadratic

$$\frac{1}{4Q} \left( \sum_{i=1}^{M_1(Q)} \frac{1}{\theta_i} \right) t^2 - \sqrt{\frac{1}{4} \mathbf{V}(\bar{x}) t + \frac{1}{4} \mathbf{V}(\bar{x}) \sum_{i=1}^{M_1(Q)} \theta_i}$$

has distinct real roots. Hence

$$\frac{1}{4} \mathbf{V}(\bar{x}) > \frac{\mathbf{V}(\bar{x})}{4Q} \sum_{i=1}^{M_1(Q)} \theta_i \sum_{i=1}^{M_1(Q)} \frac{1}{\theta_i} \geq \frac{\mathbf{V}(\bar{x}) M_1^2(Q)}{4Q},$$

because, for positive  $\theta_i$ ,

$$\sum_1^M \frac{1}{\theta_i} \sum_1^M \theta_i \geq M^2.$$

Thus  $M_1^2(Q) < Q$ , contrary to the definition of  $M_1(Q)$ . This contradiction proves the assertion.

# 8

## Efficiency of standard methods of strongly convex programming

In this chapter we analyze the laboriousness of several standard methods of convex programming. The purpose of the analysis is to elucidate whether these methods are in a position to realize the complexity of classes of strongly convex problems. It will appear that the answer to this question is negative in all the cases considered.

### 8.1 ON STANDARD METHODS OF STRONGLY CONVEX PROGRAMMING

Traditional methods of convex optimization can be divided, to a first approximation, into two classes, namely, methods for which pure ‘convergence theorems’ are proved without any estimates of the rate of convergence, and methods for which rates of convergence are established (as a rule, for their asymptotic behaviour with respect to the accuracy). The great majority of the methods in this second class are described as applied to strongly convex problems of unconstrained optimization, i.e. problems of the form

$$f(x) \rightarrow \min | x \in E \quad (f \text{ strongly convex}). \quad (I.1)$$

The purpose of this chapter is to estimate the laboriousness of some of the methods of this kind.

#### 8.1.1

We shall consider only first-order methods of solving the problem (I.1) (i.e. methods which use information about the values and gradients of  $f$  at the points surveyed). The traditional classification of these methods as regards rate of convergence is as follows:

- (i) a method converges linearly if the successive approximations  $x_i$  constructed by the method converge to the minimizing point  $x^*$  at the rate of a geometrical progression:  $|x_i - x^*| \leq Aq^i$ ,  $q < 1$ ;
- (ii) a method converges superlinearly if  $|x_i - x^*|$  decreases faster than any geometrical progression:  $|x_i - x^*| \leq A_q q^i$  for all  $q < 1$ ;
- (iii) a method converges quadratically if  $|x_i - x^*| \leq Aq^{(b)}$ ,  $0 < q < 1$ ,  $b > 1$ .

In the first case the bound for the laboriousness of the method for an accuracy  $v$  is of the form  $a \ln(1/v)$ , in the second case it is  $O(\ln(1/\mu))$ , and in the third case it is  $a \ln \ln(1/\mu)$ . The factor which we have denoted by  $a$  does not depend on the required accuracy, but does depend on other parameters of the class (the modulus of strong convexity of the problem  $f$  being solved, its dimension, the rate of variation  $f''$ , etc.).

It is customary to suppose that a difference in ‘rank’ between two methods according to the classification just given enables one to decide automatically which is the better method. Thus, a quadratically convergent method is regarded as more efficient than one which converges linearly. It seems to us that an evaluation of methods in this way—according to the nature of the asymptotic dependence of their laboriousness on the accuracy—by no means always corresponds to the truth of the matter. Apart from accuracy, the situation is characterized by other, no less important, parameters such as the modulus of strong convexity, the dimension of the problem under solution, etc., and it is no less important to examine their influence on the laboriousness than it is to examine the influence of the accuracy. Moreover, superlinear convergence, and *a fortiori* quadratic convergence, of first-order methods are themselves to a certain extent pathological phenomena.

For, firstly, methods of strongly convex programming can converge superlinearly only under certain additional requirements regarding the smoothness of the problems under solution, additional, that is, as compared with simple strong convexity. It is usual to demand a certain continuity (or even Lipschitz-continuity) of the Hessian of the function being minimized, and this is not a fortuitous circumstance. Indeed, as we saw in Section 7.2, the complexity  $N(v)$  of the class of  $(l, Q)$ -strongly convex problems (with  $Q > 2$ ) with regard to an *arbitrary*, local, deterministic satisfies the estimate

$$N(v) \geq c(l, Q, \dim E) \ln(1/v).$$

Therefore there can be no question of superlinear convergence of methods, not merely of the first order but even of any order however high, on the class of all  $(l, Q)$ -strongly convex problems; superlinear convergence can occur only if additional requirements are imposed regarding the smoothness of the problem. Moreover, the corresponding bounds for laboriousness will be very sensitive to the parameters which define this additional smoothness. From the practical

point of view this latter special feature of superlinear methods is not very palatable. As a rule, the strictly convex problems encountered in practice are as smooth as could be wished, but it is extremely laborious to estimate *a priori* the ‘smoothness parameters’, the Lipschitz constant for the second derivative, say. (Of course, estimating the parameters of strong convexity is also from time to time laborious, and so it may be worthwhile to apply methods of solving general convex problems even to strongly convex problems, accepting a reduction in speed, but, on the other hand, reducing the demand for *a priori* information about the problem under solution.) Therefore, from the practical point of view, it is not always convenient to use methods which are sensitive to difficultly observable characteristics of a problem.

However, high smoothness of the problem under solution is not by itself sufficient to give rise to superlinear methods of the first order. Superlinear convergence is always sensitive to the dimension of the problem. Indeed, we have seen that in solving the very smoothest convex problems, i.e. quadratic  $(l, Q)$ -strongly convex problems on  $R^n$ , an accuracy  $v$  is not attainable for first-order methods in a number of steps less than  $c\sqrt{Q} \ln(1/v)$  (we are assuming  $Q \geq 2$ ), where  $c$  is an absolute positive constant, if  $n > c\sqrt{Q} \ln(1/v)$ . Thus, a bound (uniform with respect to dimension) for the laboriousness of first-order methods even of quadratic programming cannot be better than the linear one  $O(\ln(1/v))$ . But in fact the dependence of the laboriousness of a method on the dimension is at least as important as its dependence on the accuracy. In this sense a quadratic estimate of the rate of convergence may even ‘discredit a method undeservedly’. Suppose, for example, there is a first-order method of solving  $(1, 2)$ -strictly convex quadratic problems, about which it is known only that the bound for its laboriousness is of the form  $a(\dim E) \ln \ln(1/v)$ . We have seen that, when  $\sqrt{2} \ln(1/v) \leq 2 \dim E$ , the complexity of this class relative to a first-order oracle is not less than  $c \ln(1/v)$ , where  $c > 0$  is an absolute constant. Now let  $v$  be determined by the relation  $\ln(1/v) = n = \dim E$ . Then we always have  $a(n) \ln \ln(1/v) \geq cn$ , i.e.  $a(n) \geq cn/\ln n$ . Thus the bound for the laboriousness of the method cannot be better than  $(cn/\ln n) \ln \ln(1/v)$ . This bound increases rapidly as  $n$  increases, and so for large  $n = \dim E$  the method (insofar as it is characterized by the original bound for its laboriousness) is worse, to an arbitrarily large extent, than the simplest gradient method (the bound for the laboriousness of which on the class considered is  $3 \ln(1/v)$ ).

Thus a ‘uniform (with respect to dimension) superlinear convergence’ can be enjoyed only by methods which operate with oracles of order higher than the first, and then only on classes of problems which are ‘smoother’ than strongly convex problems. In natural situations of this kind superlinearly convergent methods do actually exist; they are of the Newton-method type.

*Exercise 1.* Let  $G$  be a convex, closed body in a Hilbert space, let  $x_1 \in G$ , and

let  $F_m^{x_1}$  be the class of all problems of the form

$$f_0(x) \rightarrow \min | x \in G, f_j(x) \leq 0, 1 \leq j \leq m,$$

Such that the  $f_i(x)$  are twice continuously differentiable,  $(l_i, Q_i)$ -strongly convex functions with  $|f'_i(x_1)| \leq b_i$  and the  $f''_i$  satisfying Lipschitz conditions with Lipschitz constants  $L_i$ . Here the  $b_i, Q_i, l_i, L_i$  are given numbers. Prove that the complexity of solving problems in  $F_m^{x_1}$  with absolute error  $\varepsilon$  using a second-order oracle (which provides  $f(x), f'(x), f''(x)$ ) admits the bound

$$N(\varepsilon) \leq c(\rho_{1,1}(G), l, \bar{Q}, b, L) \ln \ln(1/\varepsilon), \quad \varepsilon < 10^{-1};$$

here  $(l, \bar{Q}, b, L)$  is the list of parameters specifying the class.

*Exercise 2.* Consider the class  $C_k$  of problems of the form

$$f(x) \rightarrow \min | x \in \mathbb{R}$$

such that  $|f'(0)| \leq 1, 1 \leq f''(t) \leq 2, |f^{(s)}(t)| \leq 1, 3 \leq s \leq k$ , and prove that the complexity of solving problems in this class with absolute error  $\varepsilon$  using any local, deterministic oracle admits the lower bound

$$N(\varepsilon) \geq c(k) \ln \ln(1/\varepsilon), \quad c(k) > 0.$$

Thus the ‘quadratic complexity’ is better than might be expected in natural situations from numerical methods.

### 8.1.2

In accordance with what we have said, we shall evaluate traditional methods of strongly convex programming on classes of  $(l, Q)$ -strongly convex problems of the (1.1) type, and we shall be concerned with the nature of the dependence of the laboriousness on the modulus of strong convexity  $Q$ . We shall not study particularly the effect of dimensionality on the laboriousness, restricting ourselves, when this is important, to a case of sufficiently high dimension. Under these conditions we shall be concerned with whether a given method realizes the ‘optimal’ bound for the laboriousness  $O(\sqrt{Q} \ln(1/\nu))$ . We recall that this bound is potentially unimprovable, and that it is ‘almost’ realized (up to a factor  $\ln Q$ ) by the method in Section 7.3. Running ahead, we shall prove that, for all those traditional methods for which an answer to the above question is known to us, that answer is negative.

### 8.1.3

The methods which we shall examine look like this. At the  $i$ th step a point  $x_{i-1}$  and the next direction of descent  $P_{i-1}$  have already been constructed. In the  $i$ th step ( $i \geq 2$ ), that point on the arc  $\{x_{i-1} + tp_{i-1} | t \geq 0\}$  at which the function

$f$  to be minimized has its minimum, is chosen as  $x_i$ ;  $x_1$  is some point specified in advance.

The methods differ from one another in the rules for constructing  $P_{i-1}$  as a function of the accumulated information (i.e. as a function of  $x_1, \dots, x_{i-1}, f'(x_1), \dots, f'(x_{i-1})$ ; the values of the  $f(x_j)$  are not generally made use of). Properly speaking, the scheme indicated gives not a method, but an idealization of a method: the 1-dimensional problem of minimizing  $f$  on a ray cannot be solved exactly in a finite time. Such an idealization is based on the fact that 1-dimensional problems can be solved quite quickly (even with quite high accuracies). We shall assume, as is usually done, that the exact solution of a 1-dimensional problem ‘occupies’ one step.

The simplest method of this kind is the gradient method:  $p_{i-1} = -f'(x_{i-1})$ . This method ‘has no memory at all’—the next successive point is sought on a ray chosen by means of information about the problem at the single preceding point. From the considerations in Section 7.1.5 it is easy to deduce that, on an  $(l, Q)$ -strongly convex problem  $f$ , the gradient method ensures an accuracy  $\nu$  (i.e. the initial residual  $f(x_1) - \inf_E f(x)$  is reduced  $1/\nu$  times) after a time

$$M(\nu) = \left\lceil \frac{\ln 1/\nu}{\ln(1 + 1/(Q-1))} \right\rceil.$$

After any substantially less time the method does not ensure accuracy  $\nu$  on the class  $H_{x_1}(E, 0, l, Q)$ ,  $Q > 1$ , if  $\dim E \geq 2$ . Thus the laboriousness of a gradient method ensuring an accuracy  $\nu < 1$  on the class  $H_{x_1}(E, 0, l, Q)$  when  $Q \geq 2$  and  $\dim E \geq 2$  is not less than  $cQ \ln(1/\nu)$ , where  $c > 0$  is an absolute constant. The gradient method ‘behaves badly as  $Q$  increases.’ The last of its drawbacks—excessive sensitivity to the degree of conditionality of the problem—is well-known.

*Exercise* Obtain a lower bound for the laboriousness of the gradient method by analysing its work on a 2-dimensional quadratic problem.

Let  $f(x) = \frac{1}{2} \langle Ax | x \rangle - \langle b | x \rangle$ , let  $x_1 = 0$ , and let  $x_i$  be the  $i$ th point of the work of the method on  $f$ . Then  $x_i = x_{i-1} - t_i(Ax_{i-1} - b)$ , where  $t_i$  is defined by the relation

$$\langle (Ax_{i-1} - b) - t_i A(Ax_{i-1} - b) | Ax_{i-1} - b \rangle = 0,$$

i.e.

$$t_i = \frac{(Ax_{i-1} - b)^2}{\langle A(Ax_{i-1} - b) | Ax_{i-1} - b \rangle},$$

and

$$x_i = x_{i-1} - \frac{(Ax_{i-1} - b)^2 (Ax_{i-1} - b)}{\langle A(Ax_{i-1} - b) | Ax_{i-1} - b \rangle}.$$

we put  $y_i = Ax_i - b$ . Then

$$y_i = Ax_{i-1} - b - \frac{(Ax_{i-1} - b)^2 A(Ax_{i-1} - b)}{\langle A(Ax_{i-1} - b) | Ax_{i-1} - b \rangle} = y_{i-1} - \frac{y_{i-1}^2}{\langle Ay_{i-1} | y_{i-1} \rangle} Ay_{i-1}.$$

Let

$$A = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, q = \frac{1-\alpha}{1+\alpha}.$$

It is easy to see that the vectors

$$y_i = \begin{pmatrix} -q^{i-1} \\ (-1)^i q^{i-1} \end{pmatrix}$$

satisfy the recursive relations given. Thus  $-Ax_{i+1} + b = q^i \begin{pmatrix} 1 \\ (-1)^i \end{pmatrix}$ . But

$$f(x) - \inf_y f(y) = \frac{1}{2} \langle A^{-1}(Ax - b) | Ax - b \rangle,$$

and so

$$\frac{f(x_{i+1}) - \inf_E f(x)}{f(x_1) - \inf_E f(x)} = q^{2i}.$$

Thus the error of the gradient method with  $M$  steps on a  $(1/Q, Q)$ -strongly convex problem (corresponding to  $\alpha = 1/Q$ ) is  $((Q-1)/(Q+1))^{2M-2}$ . Therefore the number of steps which will ensure an accuracy  $v$  on  $H_{x_1}(R^2, 0; l, Q)$  is not less than

$$\left\lceil \frac{\ln 1/v}{2 \ln(1 + 2/(Q-1))} \right\rceil, \quad \text{and this is the required lower bound.} \rangle$$

#### 8.1.4 Conjugate-gradient methods

These are a large group of methods of strongly convex programming which are obtained, so to say, by ‘analytic continuation’ of the CGM from quadratic problems to general problems. The conjugate-gradient method for quadratic minimization can be written in the form  $x_{i+1} = x_i + t_i p_i$ , where  $t_i$  is the minimizing point for  $f(x_i + tp_i)$  for all  $t \in \mathbb{R}$ , and  $p_i$  is the next direction of descent and is constructed, on the basis of the information accumulated, from suitable formulae. We did not give those formulae when we studied the CGM; only the geometry of the method, not its analytical form, were of importance to us. It is important to notice that the formulae defining the CGM can be written in many different ways which are equivalent (on quadratic problems). Each of these formal schemes for the CGM can later be regarded as defining an algorithm for general convex minimization.

It is clear that the various formal schemes will now generate, generally speaking, different algorithms (we recall that these schemes were equivalent, i.e. they formed the same sequence of directions of descent, only on quadratic problems.) Such a method (which is, to a certain degree, formal in character) of extending the CGM from quadratic problem to general problems is what we meant in speaking of ‘analytic continuation’ of the CGM. The idea of generalizations of this sort is easily understood. If the function to be

minimized is sufficiently smooth, then, in a small neighbourhood of its minimum, it will be ‘almost quadratic’. Therefore, if a method under consideration of the family converges at all, after a certain number of steps (when it begins to work close to the minimum), the work of the method will be almost the same as that of the usual CGM on a quadratic problem. If  $\dim E = n$ , the CGM with  $n$  steps solves any quadratic problem exactly; but the ‘analytic continuation’ of the method which we have been considering will not, of course, guarantee an exact solution after  $n$  steps even close to the minimum.

However, we can expect that after these  $n$  steps it will have greatly reduced the residual, and indeed, for most ‘analytic extensions’ of the CGM these expectations can be realized, and these extensions do give superlinear or quadratic estimates of the rate of convergence (it is assumed, of course, that the problem under solution is sufficiently smooth). (The reader can acquaint himself with the scheme of construction of many of the conjugate-gradient methods and with estimates of their rate of convergence by consulting [5].)

As already stated, we are interested in the behaviour of the methods under consideration on problems in the class  $H(E, 0; l, Q)$  when  $\dim E$  is quite large. More precisely, we want to know whether any of these methods realizes the ‘optimal’ bound for laboriousness,  $O(\sqrt{Q} \ln(1/v))$ . From this point of view the results mentioned about convergence are of little help to us, since essentially they relate only to the nature of the convergence in the neighbourhood of the optimum, and do not touch on (at least, not explicitly) the connection between the rate of convergence and the parameter  $Q$ . We shall therefore have to consider each method separately and to construct for it an example of a ‘slowly solvable’ problem. We shall carry out this awkward work for three conjugate-gradient methods. The choice of methods considered is to a certain extent arbitrary. In describing the methods we shall use the following notation:  $x_i$  is the  $i$ th point of the work of the method on the problem  $f$  considered (and we find it convenient to number them starting from 0 and not from 1),  $g_i = -f'(x_i)$ ;  $p_i$  is the direction of descent in the  $i$ th step, i.e.

$$x_{i+1} = x_i + t_i p_i, \quad (1.2)$$

where  $t_i$  is chosen as the minimizing point of  $f(x_i + tp_i)$  with respect to  $t$ ; and

$$\Delta x_i = x_{i+1} - x_i; \quad \Delta g_i = g_{i+1} - g_i.$$

#### 8.2. THE FLETCHER-REEVES METHOD

(See [27], Chapter II, for a description of this method.)

In this method  $p_i$  is chosen according to the formula

$$p_i = \frac{g_0}{|g_0|^2} + \dots + \frac{g_i}{|g_i|^2}. \quad (2.1)$$

We shall construct for every sufficiently large  $Q$  and for all  $l > 0$  an  $(l, Q)$ -strictly convex function on the plane  $R^2$ , such that the FR-method (Fletcher–Reeves method) of minimizing it with an accuracy  $c/Q$  requires  $\geq dQ^2$  steps (where  $c, d > 0$  are absolute constants). Thus, if we adopt the criteria which we have adopted for estimates, the FR-method is even worse than the gradient method.

### 8.2.1

We shall make one reservation straightaway. The FR-method, like some other conjugate-gradient methods, is often described as a process ‘with re-initialization’. The latter expression means that the process (1.2)–(2.1) is carried out not for the whole time, but for a certain number of steps  $N = k \dim E$  (where  $k$  is a natural number). After this the method starts afresh from the point  $x_N$ , as it did from  $x_1$ , and continues for a further  $N$  steps. Thus after each set of  $N$  steps (the ‘re-initialization period’) all the information accumulated is ‘forgotten’, and the solution is started from the ‘current point’ as it was from the original point. The ‘re-initialization’ is used so as to remove the unbounded consequences of effects arising from the fact that the problem being solved is not quadratic. Sometimes it is only because of this re-initialization that convergence of the method can be achieved with any initial approximation.

Incidentally, the method in Section 7.3 also contained re-initializations (its work during a period between two re-initializations was there called a stage). However, the re-initialization period of this method was not related to the ‘chance’ dimensionality of the problem, but to one of its intrinsic properties, *viz.* the modulus of strong convexity.

We shall not carry out the analysis separately for versions of methods which have re-initialization and for those which do not. The point is that we are only interested in the work of the methods of sufficiently large dimensionality; the latter can always be chosen so large that there are no re-initializations on the section of the trajectory considered. So we shall always restrict our attention to methods without re-initialization. Moreover, the examples themselves can have small dimensionality (in the present case, dimension 2). The point here is that an  $(l, Q)$ -strongly convex function  $f(x_1, \dots, x_k)$  can be put into correspondence with an  $(l, Q)$ -strongly convex function  $\bar{f}(x_1, \dots, x_n)$  of any number of variables  $n \geq k$  in the following way:

$$\bar{f}(x_1, \dots, x_n) = f(x_1, \dots, x_k) + \frac{l}{2} \sum_{j=k+1}^n x_j^2.$$

If the initial approximation lies in the plane of the first  $k$  variables (i.e. its coordinates, starting from the  $(k+1)$ th are equal to 0), then the work of any conjugate-gradient method on  $\bar{f}$  takes place all the time in the plane of the first  $k$  variables and coincides with the work of the method on  $f$  (this property can always be easily deduced from the description of the method). Thus the work

of a method without re-initialization on any problem of dimension  $k$  can be interpreted as the work of the same method, but now with re-initialization, on any ‘equally’ strictly convex problem of higher dimensionality. Therefore it suffices to study methods without re-initialization in any convenient number of dimensions. This remark also applies to the examples in the following sections.

### 8.2.2

We now return to the Fletcher–Reeves method. Suppose the work of the method does not stop in a finite time. In order to construct the example, we shall express all the parameters of the trajectory of the method by means of the angle between adjacent displacements. More precisely, let  $\varphi_{i+1}$  ( $0 < \varphi_{i+1} < \pi$ ) be the angle between  $\Delta x_i$  and  $\Delta x_{i+1}$ . It is easily verified that the following relations hold:

$$|g_{i+1}| = \frac{t_0 \cos \varphi_1 \dots \cos \varphi_{i+1}}{|\Delta x_0| \sin \varphi_{i+1}}, \quad (2.2)$$

$$|g_0| = \frac{t_0}{|\Delta x_0|}. \quad (2.3)$$

Let us suppose that a sequence of points  $x_0, x_1, \dots, x_i, \dots$  and a sequence of vectors  $g_1, g_2, \dots$  in the plane  $R^2$  are given such that

- (i)  $\langle \Delta x_i | \Delta x_{i+1} \rangle > 0$ ;
- (ii)  $\langle g_i | \Delta x_i \rangle > 0$ ;
- (iii)  $g_0 = \Delta x_0$ ,  $|g_0| = 1$ ;
- (iv)  $|g_i| = \frac{\cos \varphi_1 \dots \cos \varphi_i}{\sin \varphi_i}$ ,  $i \geq 1$ ;
- (v)  $\langle g_i | \Delta x_{i-1} \rangle = 0$ ,  $i \geq 1$ .

Here  $\varphi_i$  is the angle between  $\Delta x_i$  and  $\Delta x_{i-1}$ ,  $i \geq 1$ . We further assume that there is a convex function  $f(x)$  such that  $f'(x_i) = -g_i$ ,  $i \geq 0$ . We shall show that  $x_0, \dots, x_i, \dots$  is the trajectory of the FR-method on  $f$ . The proof is by induction over  $i$ . Suppose it is already known that  $x_0, \dots, x_i$  are the first  $i+1$  points of the trajectory of the FR-method on  $f$ . For  $i = 1$  this is true because of (iii) and (v). Now let  $i > 1$ . Let  $y_{i+1}$  be the  $(i+2)$ th point of the trajectory of the FR-method on  $f$ . In order to prove that  $y_{i+1} = x_{i+1}$ , we note that  $y_{i+1} - x_i$  must form an acute angle with  $\Delta x_{i-1}$  and an acute angle with  $g_i$ .

In view of (2.2) and the inductive hypothesis, this angle  $\psi$  is determined by the relation

$$\cot \psi = \frac{|g_i| |\Delta x_0|}{t_0 \cos \varphi_1 \dots \cos \varphi_{i-1}}.$$

By (iii) and the definition of  $t_0$ , we have  $|\Delta x_0| = t_0 = 1$ . Using (iv), we obtain  $\cot \psi = \cot \varphi_i$ , and so  $y_{i+1} - x_i$  is collinear with  $\Delta x_i$ . But then  $y_{i+1} = x_{i+1}$ , because, by the definition of the method,  $y_{i+1}$  must be the point on the ray  $\{x_i + t\Delta x_i\}$  at which  $f'(y_{i+1})$  is orthogonal to  $\Delta x_i$ . But by (v) this point is indeed  $x_{i+1}$ .

We now consider the following rule for constructing  $x_i$  (Figure 8.1):  $\{x_i\}$  is a polygonal 'spiral' winding round the point  $A$ ;  $\varphi_i$  is the angle between the directions from  $A$  to  $x_i$  and from  $A$  to  $x_{i+1}$ , and  $\Delta x_i$  is orthogonal to  $Ax_{i+1}$ . The vectors  $g_i$  are defined (for  $k \geq 1$ ) as  $\lambda_k(A - x_k)$ . Also,  $|x_1 - x_0| = 1$ , and  $g_0 = x_1 - x_0$ . Further, we take  $A = 0$ , and  $g_k = -\lambda_k x_k$ ,  $\lambda_k > 0$ ,  $k \geq 1$ . We shall choose the  $\varphi_k$  so that  $|g_k| = \alpha + 2\beta|x_k|$  and so that (iv) shall hold. Let  $|x_1| = \frac{1}{2}$ . It is required to choose the angles  $\varphi_1, \varphi_2, \dots$  so that the equations

$$\frac{\cos \varphi_1 \dots \cos \varphi_k}{\sin \varphi_k} = \alpha + \beta \cos \varphi_1 \dots \cos \varphi_{k-1}, \quad k \geq 1,$$

shall be satisfied, since  $|x_k| = |x_1| \cos \varphi_1 \dots \cos \varphi_{k-1}$  (for  $k = 1$ ,  $\cos \varphi_1 \dots \cos \varphi_{k-1} = 1$ ). In other words, the angles  $\varphi_k$  are found from the recursion relation

$$\cot \varphi_k = \frac{\alpha}{\cos \varphi_1 \dots \cos \varphi_{k-1}} + \beta, \quad \cot \varphi_1 = \alpha + \beta. \quad (2.4)$$

Putting  $t_i = \cos \varphi_i$ , we obtain, for  $k \geq 1$ ,

$$\frac{\sqrt{1-t_k^2}}{t_k} = \frac{t_1 \dots t_{k-1}}{\alpha + \beta t_1 \dots t_{k-1}}$$

(when  $k = 1$ , the right-hand side is equal to  $1/(\alpha + \beta)$ ). Hence

$$\frac{1}{t_k^2} = 1 + \frac{t_1^2 t_2^2 \dots t_{k-1}^2}{(\alpha + \beta t_1 \dots t_{k-1})^2},$$

i.e.

$$t_k^2 = \frac{(\alpha + \beta t_1 \dots t_{k-1})^2}{(\alpha + \beta t_1 \dots t_{k-1})^2 + t_1^2 \dots t_{k-1}^2}. \quad (2.5)$$

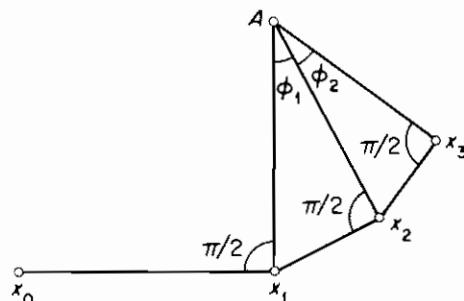


Figure 8.1

In addition to the sequence satisfying (2.5), we also consider the sequence  $\sigma_k = (t_1 \dots t_k)^{-1}$ . Then

$$\left( \frac{\sigma_k^2}{\sigma_{k-1}^2} \right)^{-1} = \frac{(\alpha + \beta/\sigma_{k-1})^2}{(\alpha + \beta/\sigma_{k-1})^2 + 1/\sigma_{k-1}^2},$$

or

$$\sigma_k^2 = \sigma_{k-1}^2 \left( 1 + \frac{1}{(\alpha \sigma_{k-1} + \beta)^2} \right) = \sigma_{k-1}^2 + \frac{\sigma_{k-1}^2}{(\alpha \sigma_{k-1} + \beta)^2} \leq \sigma_{k-1}^2 + \frac{1}{\alpha^2}.$$

Moreover,

$$\sigma_1^2 = \frac{1}{t_1^2} = (\alpha + \beta)^2 + 1.$$

Therefore

$$\sigma_k^2 \leq \frac{k-1}{\alpha^2} + (\alpha + \beta)^2 + 1$$

and

$$(t_1 \dots t_k)^2 \geq \frac{\alpha^2}{k-1 + \alpha^2(\alpha + \beta)^2 + \alpha^2}, \quad k \geq 1.$$

Thus the sequences  $\{x_i\}$  and  $\{g_i\}$  can be constructed so that

$$(1) \langle \Delta x_i | g_{i+1} \rangle = 0, \quad i = 0, 1, 2, \dots;$$

$$(2) |x_{i+1}| = |x_i| \cos \varphi_i, \quad i \geq 1;$$

$$(3) g_i = -\frac{\alpha + 2\beta|x_i|}{|x_i|} x_i, \quad i \geq 1;$$

$$(4) |g_i| = \frac{\cos \varphi_1 \dots \cos \varphi_i}{\sin \varphi_i}, \quad i \geq 1;$$

$$(5) (\cos \varphi_1 \dots \cos \varphi_i)^2 \geq \frac{\alpha^2}{(i-1) + \alpha^2(\alpha + \beta)^2 + \alpha^2}, \quad i \geq 1;$$

$$(6) g_0 = \Delta x_0, |g_0| = 1.$$

Moreover,  $\varphi_i$  is the angle between  $\Delta x_i$  and  $\Delta x_{i-1}$ . It is clear that the function  $f_0(x) = \alpha|x| + \beta x^2$  is such that  $f'_0(x_i) = -g_i, i \geq 1$ . We now choose a parameter  $\varepsilon$  with  $0 < \varepsilon < \frac{1}{2}$ , and we consider the function

$$f_\varepsilon(x) = \begin{cases} \alpha|x| + \beta x^2, & |x| \geq \varepsilon, \\ \frac{\alpha}{2\varepsilon} x^2 + \frac{\alpha\varepsilon}{2} + \beta x^2, & |x| \leq \varepsilon. \end{cases}$$

It is easy to see that  $f_\varepsilon(x)$  is  $(\beta, 4\alpha/\varepsilon\beta + 2)$ -strongly convex. Moreover,  $f_\varepsilon(x)$  coincides with  $f_0(x)$  when  $|x| \geq \varepsilon$ . Therefore, when  $i \geq 1$  is such that  $|x_i| \geq \varepsilon$ , we have

$$f'_\varepsilon(x_i) = f'_0(x_i) = -g_i. \quad (2.6)$$

Since  $|x_i| = |x_1| \cos \varphi_1 \dots \cos \varphi_{i-1} = \frac{1}{2} \cos \varphi_1 \dots \cos \varphi_{i-1}$ , we have

$$|x_i| \geq \frac{1}{2} \sqrt{\frac{\alpha^2}{(i-2) + \alpha^2(\alpha + \beta)^2 + \alpha^2}}, \quad i \geq 2,$$

and so (2.6) holds when

$$i < i(\varepsilon) = \left\lceil 2 + \left( \frac{\alpha^2}{4\varepsilon^2} - \alpha^2(\alpha + \beta)^2 - \alpha^2 \right) \right\rceil. \quad (2.7)$$

In order to convert  $f_\varepsilon$  into the required ‘bad problem’  $\bar{f}_\varepsilon$  for the FR-method, we have still to ensure that the condition  $\bar{f}'_\varepsilon(x_0) = -g_0$  is satisfied. We shall seek  $\bar{f}_\varepsilon$  in the form  $\bar{f}_\varepsilon(x) = \Delta(x) + f_\varepsilon(x)$ , where  $\Delta(x)$  is a convex function equal to 0 in the disc  $|x| \leq \frac{1}{2}$  and such that

$$\bar{f}'_\varepsilon(x_0) = -g_0,$$

i.e.

$$\Delta'_\varepsilon(x_0) = -g_0 - f'_\varepsilon(x_0) = -g_0 - \frac{\alpha x_0}{|x_0|} - 2\beta x_0 \equiv p.$$

Now let  $|x| \leq \frac{1}{2}$ . Then

$$\langle p|x - x_0 \rangle \leq \frac{1}{2}|p| + \langle x_0|g_0 \rangle + \alpha|x_0| + 2\beta|x_0|^2; |x_0|^2 = 1 + \frac{1}{4} = \frac{5}{4};$$

$$\begin{aligned} \langle x_0|g_0 \rangle &= -1; \quad |p|^2 = 1 + (\alpha + 2\beta|x_0|)^2 + \left( \frac{\alpha}{|x_0|} + 2\beta \right) \langle x_0|g_0 \rangle \\ &\leq 1 + (\alpha + 4\beta)^2. \end{aligned}$$

Thus

$$\begin{aligned} \langle p|x - x_0 \rangle &\leq \frac{1}{2}(1 + (\alpha + 4\beta)) + 2\alpha + 4\beta - 1 \\ &\leq -\frac{1}{2} + 3\alpha + 6\beta. \end{aligned}$$

We put  $\alpha = 1/24$ ,  $\beta = 1/48$ . Then  $\langle p|x - x_0 \rangle \leq -\frac{1}{4}$  and  $|p| \leq 2$ .

We consider a smooth function  $\theta$  on the real line which is equal to 0 for  $t \leq -\frac{1}{4}$ , and to  $t+1$  in a neighbourhood of  $t=0$ , and which is monotonic and convex. We can then take  $\Delta_\varepsilon(x) = \theta(\langle p|x - x_0 \rangle)$ . It is clear that  $\Delta_\varepsilon$  is  $c$ -smooth (where  $c$  is an absolute constant) and that  $\Delta'_\varepsilon(x_0) = p$ . Thus for every  $\varepsilon$  with  $0 < \varepsilon < \frac{1}{2}$ , a function  $\bar{f}_\varepsilon(x) = \Delta_\varepsilon(x) + f_\varepsilon(x)$  is defined, which is  $(1/48, c/\varepsilon)$ -strongly convex and is such that  $\bar{f}'_\varepsilon(x_i) = -g_i$ ,  $i = 0, 1, 2, \dots, i(\varepsilon)-1$ . Comparing (1)–(6) with (i)–(v), we see that  $x_0, \dots, x_{i(\varepsilon)-1}$  is the initial section of the trajectory of the FR-method on  $\bar{f}_\varepsilon$ . Further,  $\bar{f}_\varepsilon$  clearly has a modulus of strong convexity  $Q(\varepsilon) \leq \bar{c}/\varepsilon$ . Moreover,

$$\bar{f}_\varepsilon(x_0) - \min \bar{f}_\varepsilon \leq c_1,$$

where  $c_1$  does not depend on  $\varepsilon$ ; whereas, for  $|x| \geq \varepsilon$ , we have

$$\bar{f}_\varepsilon(x) - \min \bar{f}_\varepsilon \geq c_2,$$

where  $c_2 > 0$  does not depend on  $\varepsilon$ .

Thus, the FR-method on the problem  $\bar{f}_\varepsilon(x)$ , which has a modulus of strong convexity  $Q(\varepsilon) \leq \bar{c}/\varepsilon$ , ensures an accuracy  $c_3\varepsilon$  (where  $c_3 > 0$  is an absolute constant) not earlier than after  $i(\varepsilon)$  steps, where  $i(\varepsilon) = \lceil c_4/\varepsilon^2 \rceil$  and  $c_4 > 0$  does not depend on  $\varepsilon$ . If it is now given that  $Q \geq 2\bar{c}$ , we can take  $\varepsilon = \bar{c}/Q$ . With this  $\varepsilon$ , the function  $\bar{f}_\varepsilon$  provides an example of a function of two variables, with a modulus of strong convexity  $Q$ , on which the FR-method ensures an accuracy  $c/Q$  only after  $dQ^2$  steps (where  $c, d > 0$  are absolute constants).

It still remains to point out that the trajectory of the FR-method does not change when the function is multiplied by a constant. The constructed function can therefore be normalized so that it becomes  $(l, Q)$ -strongly convex, where  $l > 0$  is any given number. So the desired example has been constructed.

We see, incidentally, how perversely ‘after-effects’ can tell on the FR-method: for all  $i \geq 1$ ,  $\bar{f}'_\varepsilon(x_0)$  points exactly to the minimizing point of  $\bar{f}_\varepsilon$ , but, because of ‘memory’, the method constructs a trajectory which is all the longer, the ‘more orthogonal’ it is to the direction to the minimum (this is, of course, up to a certain moment, namely, up to the moment  $i/\varepsilon$ ).

### 8.3 THE POLAK–RIBI  RE METHOD

(The description of this method is taken from [27], Chapter II.) In this method  $p_0 = g_0$ ,  $p_{i+1} = g_{i+1} + \gamma_{i+1}p_i$ , where

$$\gamma_i = \frac{\langle g_{i+1} - g_i|g_{i+1} \rangle}{g_i^2}.$$

We shall show that, for any  $l > 0$  and for all sufficiently large  $Q$ , there is an  $(l, Q)$ -strongly convex function of three variables, on which the Polak–Ribi  re (PR) method ensures an accuracy  $v_0 > 0$  not earlier than after  $cQ$  steps ( $v_0, c > 0$  are absolute constants). Thus the PR-method does not realize the ‘optimal’ bound for the laboriousness, and, on the basis of the criteria adopted for estimates in this book, it is not better than the gradient method.

We pass on to the construction of the example.

#### 8.3.1

We shall prove that

$$\langle p_{i+1}|g_{i+1} - g_i \rangle = 0, \quad i \geq 0. \quad (3.1)$$

We observe that  $\langle p_i|g_{i+1} \rangle = 0$ ,  $i \geq 0$ , imply  $\langle p_{i+1}|g_{i+1} \rangle = g_{i+1}^2$ ,  $i \geq 0$ ; moreover,  $p_0 = g_0$ , and so  $\langle p_0|g_0 \rangle = g_0^2$ . Thus, for all  $i \geq 0$ ,

$$\langle p_i|g_i \rangle = g_i^2. \quad (3.2)$$

Therefore

$$\begin{aligned}\langle p_{i+1} | g_{i+1} - g_i \rangle &= \langle g_{i+1} + \gamma_{i+1} p_i | g_{i+1} - g_i \rangle \\&= g_{i+1}^2 - \langle g_{i+1} | g_i \rangle + \gamma_{i+1} \langle p_i | g_{i+1} \rangle - \gamma_{i+1} \langle p_i | g_i \rangle \\&= g_{i+1}^2 - \langle g_{i+1} | g_i \rangle - g_i^2 \frac{\langle g_{i+1} - g_i | g_{i+1} \rangle}{g_i^2} = 0,\end{aligned}$$

which proves (3.1). This means that the work of the method can be described in the following way. (It is assumed that the method does not stop after a finite number of steps with output at the minimizing point of the problem being solved.) The next—the  $(i+1)$ th—direction of descent lies in the plane going through  $\Delta x_i$  and  $g_{i+1}$ . In this plane the  $(i+1)$ th direction is picked out by the two conditions

- (1) it forms an acute angle with  $g_{i+1}$ , and
- (2) it is orthogonal to  $g_{i+1} - g_i$ .

These two conditions do determine the direction of descent uniquely. For, otherwise  $g_{i+1} - g_i$  would be orthogonal to  $\mathfrak{L}(\Delta x_i, g_{i+1})$ , and, since  $\Delta x_i$  is orthogonal to  $g_{i+1}$ , we would have  $\langle \Delta x_i | g_i \rangle = 0$ , i.e.  $g_i = 0$  (because, when  $g_i \neq 0$ ,  $\Delta x_i = t_i p_i$  with  $t_i > 0$ ), and this is not so, by hypothesis.

### 8.3.2

We can now reverse the argument. Namely, suppose a sequence of points  $x_0, x_1, \dots$  and a sequence of vectors  $g_0, g_1, \dots$  are given, and that

- (i)  $t_0 g_0 = \Delta x_0$ ,  $t_0 > 0$ ,  $g_0 \neq 0$ ;
- (ii)  $\langle g_{i+1} | \Delta x_i \rangle = 0$ ,  $i \geq 0$ ;
- (iii)  $\Delta x_{i+1} \in \mathfrak{L}(\Delta x_i, g_{i+1})$ ,  $\langle \Delta x_{i+1} | g_{i+1} \rangle > 0$ , and  $\langle \Delta x_{i+1} | g_{i+1} - g_i \rangle = 0$ .

Further suppose that there is a convex function  $f$  such that  $f'(x_0) = -g_0$ ,  $i \geq 0$ . Then  $\{x_i\}$  is the trajectory of the PR-method on  $f$ .

This assertion is proved by simple induction.

### 8.3.3

We now consider the following construction. Let  $(e_1, e_2, e_3)$  be an orthonormal basis in the 3-dimensional space  $R^3$ . We consider a cylindrical surface generated by a regular pentagon in the plane  $\mathfrak{L}(e_1, e_2)$ , with the generators collinear with  $e_3$ . We consider a ‘winding’  $x_0 x_1 x_2 x_3 x_4 x_5 \dots$  round this prism, which intersects the generators at a constant angle  $\psi$  (Figure 8.2). We then choose the vectors  $g_i$  so that the sequences  $\{x_i\}$  and  $\{g_i\}$  satisfy all the conditions (ii)–(iii). It turns out that  $\psi$  can be chosen so that  $g_{i+1} = t_i |g_i|$ , with  $i \geq 1$ , and  $t < 1$  sufficiently close to 1. We then learn how to construct a

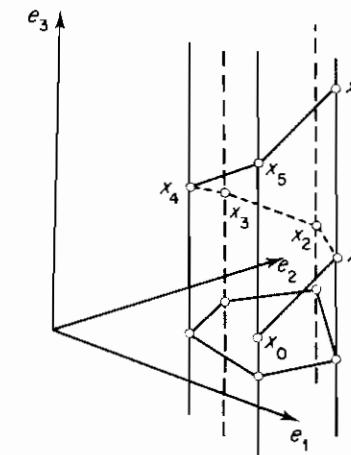


Figure 8.2

function which has  $g_i$  as its anti-gradient at  $x_i$ , and we correct it so that the condition (i) holds. With a suitable  $t$  this function gives the required example.

We pass on to the description of the construction. Let  $\theta = 2\pi/5 = 72^\circ$ , and let  $u_\theta$  denote a rotation through the angle  $\theta$  in the plane  $\mathfrak{L}(e_1, e_2)$ . Let  $q_0 \in \mathfrak{L}(e_1, e_2)$ ,  $|q_0| = 1$ , and let  $q_i = (u_\theta)^i q_0$ ,  $-\infty \leq i < \infty$ . We introduce a parameter  $\beta > 0$ , and put

$$\begin{aligned}\Delta x_i &= q_i + \beta e_3, & -\infty < i < \infty, \\r_i &= (q_i - \cos \varphi q_{i-1}) + \beta(1 - \cos \varphi)e_3, & -\infty < i < \infty.\end{aligned}\quad \left. \right\} \quad (3.3)$$

Here  $\varphi$  is chosen so that

$$0 < \varphi < \frac{\pi}{2} \quad \text{and} \quad \cos \varphi = \frac{\beta^2 + \cos \theta}{\beta^2 + 1}.$$

We notice the following facts.

(A)  $r_i \in \mathfrak{L}(\Delta x_{i-1}, \Delta x_i)$  and  $\Delta x_i \in \mathfrak{L}(\Delta x_{i-1}, r_i)$  (because  $r_i = \Delta x_i - \cos \varphi \Delta x_{i-1}$ ).

(B)  $\varphi$  is the angle between  $\Delta x_{i-1}$  and  $\Delta x_i$  (because

$$(|\Delta x_i| |\Delta x_{i-1}|)^{-1} \langle \Delta x_i | \Delta x_{i-1} \rangle = \frac{\cos \theta + \beta^2}{1 + \beta^2} = \cos \varphi.$$

(C)  $\langle r_i | \Delta x_{i-1} \rangle = 0 < \langle r_i | \Delta x_i \rangle$  (for,

$$\langle \Delta x_{i-1} | \Delta x_i - \Delta x_{i-1} \cos \varphi \rangle = |\Delta x_i| |\Delta x_{i-1}| \cos \varphi - |\Delta x_{i-1}|^2 \cos \varphi = 0,$$

since  $|\Delta x_i|$  does not depend on  $i$ ; further,

$$\langle \Delta x_i | \Delta x_i - \Delta x_{i-1} \cos \varphi \rangle = (\Delta x_i^2)(1 - \cos^2 \varphi) = \Delta x_i^2 \sin^2 \varphi.$$

We now prove that, with a suitable choice of  $t > 0$  and with  $\bar{g}_i = t^i r_i$ , we have  $\langle \Delta x_{i+1} | \bar{g}_{i+1} - \bar{g}_i \rangle = 0$ . Indeed, we must choose  $t$  from the condition that, for all  $i$ ,

$$\begin{aligned} & \langle q_{i+1} + \beta e_3 | t^{i+1} \{(q_{i+1} - \cos \varphi q_i) + \beta(1 - \cos \varphi) e_3\} \\ & - t^i \{(q_i - \cos \varphi q_{i-1}) + \beta(1 - \cos \varphi) e_3\} \rangle = 0, \end{aligned}$$

or

$$\begin{aligned} & tq_{i+1}^2 - t \cos \varphi \langle q_{i+1} | q_i \rangle - \langle q_{i+1} | q_i \rangle \\ & + \cos \varphi \langle q_{i+1} | q_i \rangle + t\beta^2(1 - \cos \varphi) - \beta^2(1 - \cos \varphi) = 0, \end{aligned}$$

i.e.

$$t - t \cos \varphi \cos \theta - \cos \theta + \cos \varphi \cos 2\theta + t\beta^2(1 - \cos \varphi) - \beta^2(1 - \cos \varphi) = 0,$$

or

$$t = \frac{\cos \theta - \cos \varphi \cos 2\theta + \beta^2(1 - \cos \varphi)}{\beta^2(1 - \cos \varphi) + 1 - \cos \varphi \cos \theta}.$$

Noting that

$$\cos 2\theta = \cos \frac{4\pi}{5} = \cos \left( \pi - \frac{\pi}{5} \right) = -\cos \frac{\theta}{2},$$

we obtain

$$t = \frac{\cos \theta + \cos \varphi \cos \frac{1}{2}\theta + \beta^2(1 - \cos \varphi)}{\beta^2(1 - \cos \varphi) + 1 - \cos \varphi \cos \theta}. \quad (3.4)$$

We shall prove that there is a  $\Delta > 0$  such that, for all  $t$  with  $1 - \Delta \leq t \leq 1$ , there are  $\beta_t$  and  $\varphi_t$  such that

$$\left. \begin{aligned} \cos \varphi_t &= \frac{\beta_t^2 + \cos \theta}{\beta_t^2 + 1}, \\ t &= \frac{\cos \theta + \cos \frac{1}{2}\theta \cos \varphi_t + \beta_t^2(1 - \cos \varphi_t)}{\beta_t^2(1 - \cos \varphi_t) + 1 - \cos \varphi_t \cos \theta}, \end{aligned} \right\} \quad (3.5)$$

and  $\beta_t > 0$ ,  $0 < \varphi_t < \pi/2$ , and  $\beta_t$ ,  $\varphi_t$  depend continuously on  $t$ .

For, when  $t = 1$  the system (3.5) has a solution with the required properties. It can be expressed in the form

$$\cos \varphi_1 = \frac{1 - \cos \theta}{\cos \theta + \cos \frac{1}{2}\theta}, \quad (3.5')$$

$$\cos \varphi_1 = \frac{\beta_1^2 + \cos \theta}{\beta_1^2 + 1}, \quad (3.5'')$$

and direct calculation of the right-hand side of (3.5') shows that it lies between 0.5 and 0.9. Hence it is clear that (3.5') has a solution  $\varphi_1$  in the interval  $(0, \pi/2)$ , and that  $\cos \varphi_1 > \cos \theta$ . Because of this last relation, (3.5'') has a solution  $\beta_1 > 0$ . It remains to note that for the original system (depending parametrically on  $t$ ), the conditions of the implicit-function theorem hold at the point

$\varphi_1$ ,  $\beta_1$ , and so that system has a solution, which depends continuously on  $t$ , for all  $t$  sufficiently close to 1.

Now let  $1 - \Delta \leq t \leq 1$ . We put  $\varphi = \varphi_t$ ,  $\beta = \beta_t$ , and let  $\Delta x_i$ ,  $r_i$  be defined by (3.3) with the chosen  $\varphi$  and  $\beta$ .

### 8.3.4

We consider a trajectory  $\{x_i\}_{i \geq 0}$  in which

$$x_i = x_0 + \sum_{j=0}^{i-1} \Delta x_j,$$

and  $x_0$  is chosen in the plane  $\mathfrak{L}(e_1, e_2)$  so that the projections  $\bar{x}_i$  of the points  $x_i$  on to this plane are the vertices of a regular pentagon with centre at 0 and with side equal to 1 (it is clear that this can be done; Figure 8.3 shows the projection of the trajectory on to  $\mathfrak{L}(e_1, e_2)$ ). We put  $v_i = q_i - q_{i-1} \cos \varphi$ . (so that  $v_1$  is the projection of  $r_1$  on to  $\mathfrak{L}(e_1, e_2)$ ). We shall prove that  $\langle v_i | \bar{x}_i \rangle < 0$ .

For, having chosen an orthonormal basis  $(e, f)$  in the plane as shown in Figure 8.3 (where  $i = 5k$ ), we obtain

$$\begin{aligned} q_i &= -\cos \left( \frac{\pi - \theta}{2} \right) e + \sin \left( \frac{\pi - \theta}{2} \right) f = -\sin \frac{\theta}{2} e + \cos \frac{\theta}{2} f, \\ q_{i-1} &= \cos \left( \frac{\pi - \theta}{2} \right) e + \sin \left( \frac{\pi - \theta}{2} \right) f = \sin \frac{\theta}{2} e + \cos \frac{\theta}{2} f \end{aligned}$$

and

$$\langle v_i | e \rangle = -\sin \frac{\theta}{2} (1 + \cos \varphi) < 0,$$

and so also  $\langle v_i | \bar{x}_i \rangle < 0$ .

We also prove that, for  $t$  sufficiently close to 1, the line orthogonal to  $v_i$  and passing through  $\bar{x}_i$  is a support to the pentagon (i.e. if  $\bar{x}$  is a point in the

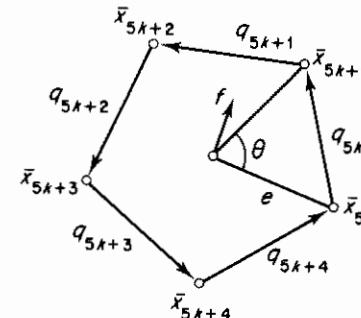


Figure 8.3

pentagon, then  $\langle \bar{x} - \bar{x}_i | v_i \rangle \geq 0$ . It suffices to prove that

$$\langle v_i | q_i \rangle > 0 \text{ and } \langle v_i | -q_{i-1} \rangle > 0.$$

We have

$$\langle v_i | q_i \rangle = q_i^2 - \langle q_i | q_{i-1} \rangle \cos \varphi_i = 1 - \cos \theta \cos \varphi_i > 0$$

and

$$\langle v_i | -q_{i-1} \rangle = -\cos \theta + \cos \varphi_i.$$

When  $t = 1$  we have

$$\cos \varphi_i = \frac{1 - \cos \theta}{\cos \theta + \cos \frac{1}{2}\theta} > \cos \theta$$

(this can be verified by direct calculation). Therefore  $\langle v_i | 1 - q_{i-1} \rangle > 0$  when  $t = 1$ , and then this inequality is also true for all values of  $t$  close enough to 1. Hence it follows that the function  $\psi_i(\bar{x}) = \langle v_i | -\bar{x} \rangle$  on the plane  $\mathfrak{L}(e_1, e_2)$  attains its maximum on the pentagon at the point  $\bar{x}_i$ .

We shall consider the convex function on  $\mathfrak{L}(e_1, e_2)$ :

$$\psi(\bar{x}) = \max_{0 \leq i \leq 4} \psi_i(\bar{x}).$$

It is equal to  $\langle -v_i | x \rangle$  in a neighbourhood of each point  $\bar{x}_i$ , and therefore its anti-gradient at  $\bar{x}_i$  is equal to  $v_i$ . It is clear that this function can be smoothed and converted into an  $(\bar{\ell}, \bar{Q})$ -strongly convex function  $\psi_t(\bar{x})$ , preserving the relation

$$\psi_t(\bar{x}_{i+1}) = \psi_t(\bar{x}_i), \quad \psi'_t(\bar{x}_i) = -v_i \quad (3.6)$$

and the condition  $\psi_t(0) = 0$ ,  $\psi'_t(0) = 0$ . Moreover, the numbers  $\bar{\ell}, \bar{Q}$  can be supposed independent of  $t$ , provided that  $1 \geq t \geq 1 - \Delta$ , where  $\Delta > 0$  is an absolute, sufficiently small, constant.

### 8.3.5

Now let  $t < 1$  and  $t \geq 1 - \Delta$ . We define  $\kappa > 0$  from the condition

$$\exp(-\kappa \beta_t) = t(\kappa = \kappa(t)).$$

We form the function  $F(x) = F(\bar{x}, y)$  on  $\mathbb{R}^3$  ( $\bar{x}$  is the projection of  $x$  on to  $\mathfrak{L}(e_1, e_2)$ , and  $y = \langle x | e_3 \rangle$ ) according to the formula

$$F(x) = \frac{\beta_t(1 - \cos \varphi_t)}{\kappa(t)} e^{-\kappa(t)y} + e^{-\kappa(t)y} (\psi_t(\bar{x}) - \psi_t),$$

where  $\psi_t = \psi_t(\bar{x}_i)$ ; the quantity on the right-hand side does not depend on  $i$ . We prove that  $F'(x_i) = -\bar{g}_i$ ,  $i \geq 0$ . Indeed, if  $x$  is such that  $\psi_t(\bar{x}) = \psi_t$ , then

$$F'_{\bar{x}}(x) = e^{-\kappa y} \psi'_t(\bar{x}),$$

$$F'_y(x) = -a_t e^{-\kappa y}, \text{ where } a_t = \beta_t(1 - \cos \varphi_t).$$

Therefore

$$F'_{\bar{x}}(x_i) = e^{-\kappa i \beta_t} \psi'_t(\bar{x}_i) = -t^i v_i,$$

$$-F'_y(x_i) = e^{-\kappa i \beta_t} a_t = t^i (1 - \cos \varphi_t) \beta_t,$$

and so

$$-F'(x_i) = t^i (v_i + (1 - \cos \varphi_t) \beta_t e_3) = t^i r_i = \bar{g}_i.$$

Further

$$F''(x) = e^{-\kappa y} \left( \frac{\psi''_t(\bar{x})}{-\kappa \psi'_t(\bar{x})} \right) \frac{-\kappa \psi'_t(\bar{x})}{a_t \kappa + \kappa^2 (\psi_t(\bar{x}) - \psi_t)}.$$

It is easy to see that, for all  $t$  sufficiently close to 1 (more precisely, for  $1 - \bar{\Delta} \leq t < 1$ , with  $\bar{\Delta} > 0$ ), in the domain  $G_t = \{ |y| \leq 1/\kappa(t) \text{ and } |\bar{x}| \leq 20 \}$  the function  $F$  is  $(c_1 \kappa(t), c_2/\kappa(t))$ -strongly convex, where  $c_1, c_2$  are absolute constants. It is also easy to see that  $F$  can be extended from this domain into a  $(c_1 \kappa(t), c_2/\kappa(t))$ -strongly convex function,  $F_t(x)$ , on the whole of  $\mathbb{R}^3$ . We note that all the points of the trajectory considered,  $x_0, \dots, x_{M(t)}$ , lie in the domain  $G_t$ ; here

$$M(t) = \left[ \frac{1}{\beta_t \kappa(t)} \right] = \left[ \ln \frac{1}{t} \right]. \quad (3.7)$$

The function  $F_t(x)$  is ‘almost’ the required example. The only thing remaining to be done is to correct  $F_t(x)$  in the neighbourhood of  $x_0$  so that for the corrected function  $\bar{F}_t(x)$  we have

$$\begin{aligned} \bar{F}'_t(x_0) &= \lambda \Delta x_0, \quad \lambda > 0 \\ \text{and} \quad \langle \bar{g}_1 - \lambda \Delta x_0 | \Delta x_1 \rangle &= 0. \end{aligned} \quad (3.8)$$

The second condition gives

$$\begin{aligned} \lambda = \lambda_t &= \frac{\langle \bar{g}_1 | \Delta x_1 \rangle}{\langle \Delta x_0 | \Delta x_1 \rangle} = \frac{t \langle r_1 | \Delta x_1 \rangle}{(1 + \beta_t^2) \cos \varphi_t} \\ &= \frac{t}{(\cos \varphi_t (1 + \beta_t^2))} \langle q_1 - q_0 \cos \varphi_t + \beta_t (1 - \cos \varphi_t) e_3 | q_1 + \beta_t e_3 \rangle \\ &= \frac{t}{(\cos \varphi_t (1 + \beta_t^2))} \{ 1 - \cos \varphi_t \cos \theta + \beta_t^2 (1 - \cos \varphi_t) \}. \end{aligned}$$

We shall try to find  $\bar{F}_t(x)$  in the form  $F_t(x) + \Delta(x)$ , seeking a convex function  $\Delta(x)$  such that (3.8) holds and such that  $\Delta(x)$  will be equal to 0 in a neighbourhood of  $x_1, x_2, \dots$ . If we succeed in doing this, then, by our construction (see (A)–(C)), we shall obtain that  $\bar{F}'_t(x_i) = -g_i$ ,  $M(t) \geq i \geq 0$ , where  $\bar{g}_i = g_i$ ,  $i \geq 1$ , and  $g_0 = \lambda_t \Delta x_0$ , and  $g_i$  and  $x_i$  satisfy (i)–(iii) for  $i \leq M(t)$ . Thus  $x_0, \dots, x_{M(t)}$  will be the initial section of the trajectory of the PR-method on  $\bar{F}_t$ .

In order to satisfy (3.8) we must have

$$-\Delta'(x_0) + \bar{g}_0 = \lambda_t \Delta x_0,$$

i.e.

$$\Delta'(x_0) = \bar{g}_0 - \lambda_t \Delta x_0 \equiv p(t). \quad (3.9)$$

In order to construct a convex function  $\Delta(x)$  which satisfies (3.9) and the condition  $\Delta'(x_i) = 0$ ,  $\Delta(x_i) = 0$ ,  $i \geq 1$ , it suffices to verify that, for some  $a > 0$  ( $a$  does not depend on  $t$ , at least for values of  $t$  sufficiently close to 1), we have  $\langle p(t)|x_i - x_0 \rangle \leq -a$ . It is sufficient to show that  $\langle p(t)|\Delta x_i \rangle \leq a_i$  (the  $a_i$  do not depend on  $t$ ), and

$$\begin{aligned} a_0 < 0, \quad a_0 + a_1 < 0, \quad a_0 + a_1 + a_2 < 0, \\ a_0 + a_1 + a_2 + a_3 < 0, \quad a_0 + a_1 + a_2 + a_3 + a_4 < 0 \end{aligned}$$

(since the  $\Delta x_i$  are periodic:  $\Delta x_{i+5} = \Delta x_i$ ). If we succeed in doing this, then, for  $t$  close to 1, we will be able to exhibit a  $\bar{c}$ -smooth, convex function which satisfies (3.9) and which is equal to 0 in a neighbourhood of  $x_i$ ,  $i \geq 1$  (it would be possible to take it in the form  $\varphi(\langle p(t)|x - x_0 \rangle)$  with a suitable choice of  $\varphi$ , see Section 8.2). Here  $\bar{c}$  is an absolute constant.

Direct calculation of  $\langle p(1)|\Delta x_i \rangle$  gives  $\langle p(1)|\Delta x_i \rangle \leq 0$  and  $\langle p(1)|\Delta x_0 \rangle < 0$ . Therefore, for values of  $t$  sufficiently close to 1, the required  $a_i$  can be found.

### 8.3.6

We summarize the results obtained. We have found a  $\Delta > 0$  and absolute constants  $c_0, c_1 > 0$  such that, for all  $t$  with  $1 > t > 1 - \Delta$ , there is a  $(c_0\kappa(t), c_1/\kappa(t))$ -strongly convex function  $\bar{F}_t$  on  $R^3$  which satisfies the following condition: along the first  $\lfloor \ln(1/t) \rfloor \equiv M(t)$  of the points  $x_i$ ,  $i \geq 1$ , of the trajectory of the PR-method on  $\bar{F}_t$  the relation

$$\bar{F}_t(x_i) = \frac{a_t}{\kappa(t)} \exp(-\kappa(t) \langle x_i | e_3 \rangle)$$

holds. Here  $\kappa(t)$  is as defined in Section 8.3.5,  $a_t = \beta_t(1 - \cos \varphi_t)$ , and  $\beta_t, \varphi_t$  are as defined in (3.5). We see that

$$v(x_{[M(t)/2]}, F_t) \geq \frac{\bar{F}_t(x_{[M(t)/2]}) - \bar{F}_t(x_{[M(t)]})}{\bar{F}_t(x_0) - \bar{F}_t(x_{[M(t)]})} \geq v_0,$$

and  $v_0 > 0$  does not depend on  $t$ . Moreover, it is clear that

$$\underline{c} \leq \frac{\kappa(t)}{(1-t)} \leq \bar{c},$$

$\underline{c}, \bar{c} > 0$  being absolute constants. Therefore the modulus of strong convexity of  $\bar{F}_t$  is of the order  $1/(1-t)$ , as also is  $M(t)$ ; and so, after a time which is of the same order as the modulus of strong convexity of  $\bar{F}_t$ , the PR-method is not in a position to ensure the accuracy  $v_0$  in the solution of  $\bar{F}_t$ .

At the same time, the modulus of strong convexity indicated can be made arbitrarily large by making  $1-t$  small. Thus, for all sufficiently large  $Q$ , there is an  $(l_Q, Q)$ -strongly convex function on which the PR-method with  $\bar{c}Q$  steps ( $\bar{c} > 0$  is an absolute constant) does not yield the accuracy  $v_0 > 0$ , where  $v_0$  does not depend on  $Q$ .

It remains to observe that the PR-method works on the function  $\lambda f(\lambda > 0)$  with the same accuracy as on the function  $f$ , and so the 'bad' function indicated can be taken to be an  $(l, Q)$ -strongly convex function, whatever value  $l > 0$  may have. The construction has been completed.

## 8.4 THE ZOUTENDIJK PROJECTION METHOD

(The description of Zoutendijk's method is taken from [12], Chapter 3.)

In this method  $p_0 = g_0$ , and  $p_i$  is the projection of  $g_i$  on to the orthogonal complement to  $\mathfrak{L}(\{\Delta g_s\}_{s=1}^{i-1})$ , and  $\Delta g_i = g_{i+1} - g_i$ . If we obtain  $p_i = 0$  but  $g_i \neq 0$ , then a re-initialization takes place at the moment  $i$ . We shall show that this method converges only owing to the re-initializations (ordinary re-initializations—after every  $(\dim F)$  steps—are meant). More precisely, we shall construct a  $\bar{Q}$ -strongly convex function  $f$  on the space  $I_2^{(\infty)}$ , on which the Zoutendijk method (the Z-method) does not converge at all. From our point of view, this means that the Z-method not only fails to realize the optimum bound for the laboriousness of solution of strongly convex problems, but also that it is in general of little use for solving problems of high dimensionality.

We pass on to the construction of an example. Let  $\{e_i\}_{i=1}^\infty$  be an orthonormal basis in  $I_2^{(\infty)}$ . We put

$$x_r = \sum_{j=r+1}^{\infty} 2^{-j/2} (j-r)e_j, \quad q_r = -g_r = \sum_{j=r+1}^{\infty} 2^{-j/2} e_j.$$

### 8.4.1

We shall prove that there is a  $\bar{Q}$ -strongly convex function  $\bar{f}$  on  $I_2^\infty$  ( $\bar{Q}$  is a certain constant) such that  $\bar{f}'(x_r) = -g_r$ ,  $0 \leq r < \infty$ . To do this, we shall try to find a convex function  $\xi(x)$  with a Lipschitz derivative such that  $\xi'(x_r) = t q_r - x_r$ , with  $t > 0$ . If we succeed in finding such a  $\xi$  for any  $t > 0$ , then we shall take

$$\bar{f}(x) = \frac{1}{t} \xi(x) + \frac{x^2}{2t}.$$

### 8.4.2

We pass on to the construction of  $\xi$ . We fix  $t > 3$  and put

$$\tilde{q}_i = t q_i - x_i = \sum_{j=i+1}^{\infty} 2^{-j/2} (t-j+i)e_j.$$

We calculate the numbers  $t_{ij} = \langle \bar{q}_i - \bar{q}_{i+1} | x_j \rangle$ ,  $0 \leq i, j < \infty$ .

(1) Let  $i+1 \leq j$ . We have

$$\begin{aligned}\bar{q} - \bar{q}_{i+1} &= t \cdot 2^{-(i+1)/2} e_{i+1} - \sum_{s=i+2}^{\infty} 2^{-s/2} e_s, \\ \langle \bar{q}_i - \bar{q}_{i+1} | x_j \rangle &= - \sum_{s=j+1}^{\infty} 2^{-s} (s-j) = -2^{-j+1}.\end{aligned}$$

So, when  $i+1 \leq j$ , we have

$$t_{ij} = -2^{-j+1}.$$

(2) Now let  $i+1 > j$ , i.e.  $i \geq j$ . Then

$$\begin{aligned}\langle \bar{q}_i - \bar{q}_{i+1} | x_j \rangle &= t \cdot 2^{-(i+1)} (i-j+1) - \sum_{s=i+2}^{\infty} 2^{-s} (s-j) \\ &= t(i-j+1) 2^{-(i+1)} - 2^{-(i+1)} (i-j+3) \\ &= 2^{-(i+1)} ((t-1)(i-j) + t-3).\end{aligned}$$

Thus, when  $i \geq j$ ,

$$t_{ij} = ((t-1)(i-j) + (t-3)) 2^{-(i+1)}.$$

Since  $t > 3$ ,  $t_{ij}$  decreases for a fixed  $i$  and  $j$  varying from 0 to  $i$ :

$$t_{ii} = (t-3) 2^{-(i+1)} = \min_{j \leq i} t_{ij}.$$

We put

$$t_i \equiv t_{ii} = (t-3) 2^{-(i+1)}.$$

#### 8.4.3

Now let  $\varphi_i(\tau)$  be a convex function, equal to 0 when  $\tau \leq 0$ , and coinciding with  $\tau$  when  $\tau \geq (t-3) 2^{-(i+1)} = t_i$ . We can assume  $\varphi_i(\tau)$  to be infinitely differentiable and such that

$$|\varphi'_i(\tau)| \leq c, \quad 0 \leq \varphi''_i(\tau) \leq L_i = c(t) 2^i$$

(indeed, we can take

$$\varphi_i(\tau) = \frac{t_i}{t_0} \varphi_0\left(\tau \frac{t_0}{t_i}\right).$$

We form the series

$$\sum_{i=0}^{\infty} \varphi_i(\langle \bar{q}_i - \bar{q}_{i+1} | x \rangle) \equiv \sum_{i=0}^{\infty} \psi_i(x).$$

The function  $\psi_i$  is convex and infinitely differentiable, and  $\psi_i(0) = 0$ ,  $\psi'_i(0) = 0$ . We have

$$\langle \psi''_i(x) h | h \rangle = \varphi''_i(\langle \bar{q}_i - \bar{q}_{i+1} | x \rangle) \sum_{k,l=1}^{\infty} (\bar{q}_i - \bar{q}_{i+1})_k (\bar{q}_i - \bar{q}_{i+1})_l h_k h_l.$$

Hence

$$\langle \psi''_i(x) h | h \rangle \leq L_i t^2 \sum_{k,l=i+1}^{\infty} 2^{-\frac{k+l}{2}} |h_k| |h_l|.$$

Therefore

$$\begin{aligned}\sum_{i=0}^N |\langle \psi''_i(x) h | h \rangle| &\leq c_1(t) \sum_{i=0}^N \sum_{k,l=i+1}^{\infty} 2^{i-\frac{k+l}{2}} |h_k| |h_l| \\ &\leq c_1(t) \sum_{k,l=1}^{\infty} |h_k| |h_l| \left( \sum_{i=0}^{\min(k,l)} 2^i \right) 2^{-\frac{k+l}{2}} \\ &\leq c_2(t) \sum_{k,l=1}^{\infty} |h_k| |h_l| \cdot 2^{-\frac{|k-l|}{2}} \leq c_3(t) |h|^2.\end{aligned}$$

Thus the series of second derivatives of the functions  $\psi_i$  converges absolutely along any direction, and has a bounded sum throughout the space. Hence and from the relations  $\psi_i(0) = 0$ ,  $\psi'_i(0) = 0$ , it follows that the series  $\sum_{i=0}^{\infty} \psi_i(x + th)$  and  $\sum_{i=0}^{\infty} \langle \psi'_i(x + th) | h \rangle$  converge uniformly on any interval of the form  $|t| \leq \Delta$ . Therefore the function

$$\psi(x) = \sum_{i=0}^{\infty} \psi_i(x)$$

is well-defined, convex, and continuous, and, for all  $x, h$

$$\langle \psi'(x) | h \rangle = \sum_{i=0}^{\infty} \langle \psi'_i(x) | h \rangle,$$

and  $\psi'(x)$  is Lipschitz with a constant  $c_3(t)$ .

At the same time,

$$\langle \psi'(x_j) | h \rangle = \sum_{i=0}^{\infty} \langle \psi'_i(x_j) | h \rangle = \sum_{i=0}^{\infty} \varphi'_i(t_{ij}) (\bar{q}_i - \bar{q}_{i+1}) = \sum_{i=j}^{\infty} (\bar{q}_i - \bar{q}_{i+1}) = \bar{q}_j.$$

Here we have used the fact that

$$t_{ij} \begin{cases} < 0 & \text{for } i < j \\ \geq t_i & \text{for } i \geq j \end{cases} \text{ and so } \varphi'_i(t_{ij}) = \begin{cases} 0, & i < j, \\ 1, & i \geq j. \end{cases}$$

Thus we may take  $t = 4$ , and adopt the corresponding  $\psi(x)$  as  $\xi(x)$ .

#### 8.4.4

Thus, the function  $\bar{f}(x)$  mentioned in Section 8.4.1 does actually exist. We put  $\mu_j = 2^{-j/2}$  and  $\bar{h} = \sum_{j=1}^{\infty} \mu_j e_j$ , and consider the function  $f(x) = -\langle \bar{h} | x \rangle + 2\bar{f}(x)$ . We shall prove that  $\{x_i\}$  is the trajectory of the Z-method on  $f$  with the initial point  $x_0$ .

What we have to prove is that  $x_1 - x_0$  is collinear with  $f'(x_0)$ , and that  $x_{i+1} - x_i$  is collinear with the projection of  $f'(x_i)$  on to the

orthogonal complement to  $L(f'(x_1) - f'(x_0), \dots, f'(x_i) - f'(x_{i-1}))$ , whereas  $\langle f'(x_{i+1})|x_{i+1} - x_i \rangle = 0$ .

We have

$$f'(x_i) = -\sum_{j=1}^{\infty} \mu_j e_j + 2 \sum_{j=i+1}^{\infty} \mu_j e_j = -\sum_{j=1}^i \mu_j e_j + \sum_{j=i+1}^{\infty} \mu_j e_j,$$

$$x_{i+1} - x_i = -\sum_{j=i+2}^{\infty} \mu_j e_j - \mu_{i+1} e_{i+1} = -\sum_{j=i+1}^{\infty} \mu_j e_j.$$

Putting  $i = 0$ , we find that  $f'(x_0)$  is collinear with  $\Delta x_0$ . Moreover,

$$\begin{aligned} & \langle f'(x_{i+1})|x_{i+1} - x_i \rangle \\ &= \left\langle -\mu_{i+1} e_{i+1} + \sum_{j=i+2}^{\infty} \mu_j e_j \right| - \sum_{j=i+2}^{\infty} \mu_j e_j - \mu_{i+1} e_{i+1} \rangle \\ &= \left\{ -\mu_{i+1}^2 + \sum_{j=i+2}^{\infty} \mu_j^2 \right\} = \left\{ -2^{-(i+1)} + \sum_{j=i+2}^{\infty} 2^{-j} \right\} = 0. \end{aligned}$$

Finally,

$$\Delta f_i \equiv f'(x_{i+1}) - f'(x_i) = 2 \left( \sum_{j=i+2}^{\infty} \mu_j e_j - \sum_{j=i+1}^{\infty} \mu_j e_j \right) = -2\mu_{i+1} e_{i+1}.$$

Therefore  $\mathfrak{L}(\Delta f_0, \dots, \Delta f_{i-1}) = \mathfrak{L}(e_1, \dots, e_i)$ , and the projection of  $f'(x_i)$  on to the orthogonal complement to  $\mathfrak{L}(\Delta f_0, \dots, \Delta f_{i-1})$  is  $\sum_{j=i+1}^{\infty} \mu_j e_j$ , which is collinear with  $x_{i+1} - x_i$ . Thus,  $\{x_i\}$  is the trajectory of the Z-method on  $f$ . But

$$f'(x_i) = -\sum_{j=1}^i \mu_j e_j + \sum_{j=i+1}^{\infty} \mu_j e_j \xrightarrow{i \rightarrow \infty} -h \neq 0,$$

and so the method does not converge, not even with respect to the functional. The required example has been constructed.

## 8.5 REMARKS ON THE DAVIDON-FLETCHER-POWELL METHOD

(The description of the method is taken from [27], Chapter II.)

### 8.5.1

We make a few remarks about one more conjugate-gradient method—the *Davidon-Fletcher-Powell method*; it is also called the *variable-metric method*. For brevity, we denote it by the abbreviation DFP. In this method the recurrent directions of descent  $p_i$  are constructed according to the formula

$p_i = H_i g_i$ , where  $H_i$  is the operator constructed from the recursive formulae

$$H_0 = I \quad (\text{the identity operator}),$$

$$H_{i+1} = H_i - \frac{H_i \Delta g_i \cdot H_i \Delta g_i}{\langle H_i \Delta g_i | \Delta g_i \rangle} + \frac{\Delta x_i \cdot \Delta x_i}{\langle \Delta x_i | \Delta g_i \rangle}. \quad (5.1)$$

Here  $\Delta g_i = g_{i+1} - g_i$ , and, for any vectors  $u, v$  the symbol  $u \cdot v$  denotes the operator defined by  $\langle u, v \rangle h = \langle v | h \rangle u$ .

The DFP is regarded as the most efficient conjugate-gradient method. We shall not be able to clear up completely the question whether it really does realize the ‘optimal’ bound for the laboriousness on the class of strongly convex problems.

The only thing that can be asserted here is that, if the DFP is indeed good, then it is so only for particular normalizations of the problems to be solved. More precisely, there is an absolute constant  $\bar{Q}$  such that there are strongly convex problems which have the modulus of convexity  $\bar{Q}$  and on which, if their dimensionality is sufficiently high, the DFP works ‘arbitrarily badly’, i.e. it does not attain a given accuracy  $v_0 > 0$  after any previously specified number of steps  $M$ . ( $v_0 > 0$  and  $\bar{Q}$  are arbitrary constants.)

The crux of the matter is this: in contrast to the methods previously considered, the DFP is ‘sensitive’ to changes in scale. The DFP trajectories on a problem  $f$  and on the problem  $\lambda f$  are, in general, distinct even when they have the same starting point. This circumstance is related to the presence of the third term on the right-hand side in (5.1). It is clear that the DFP trajectory on  $\lambda f$  coincides with the trajectory on  $f$  of a DFP $_{\lambda}$  which works according to the rules

$$p_i = H_i g_i, \quad H_0 = I,$$

$$H_{i+1} = H_i - \frac{H_i \Delta g_i \cdot H_i \Delta g_i}{\langle H_i \Delta g_i | \Delta g_i \rangle} + \frac{1}{\lambda} \frac{\Delta x_i \cdot \Delta x_i}{\langle \Delta x_i | \Delta g_i \rangle}$$

(the starting-point of all trajectories is a point fixed once and for all).

If we now fix  $f$  and consider the DFP trajectories on  $\lambda f$  for large  $\lambda$ , then they will be ‘almost the same’ as the trajectories on  $f$  of the method DFP $_{\infty}$  working according to the rules

$$p_i = H_i g_i, \quad H_0 = I,$$

$$H_{i+1} = H_i - \frac{H_i \Delta g_i \cdot H_i \Delta g_i}{\langle H_i \Delta g_i | \Delta g_i \rangle}.$$

(This pronouncement is not stated entirely correctly, but its meaning is clear. The following conclusions could easily be made rigorous, but we shall not bother to do this.) But it is easy to see that the ‘limiting case of DFP’, i.e. DFP $_{\infty}$ , is precisely Zoutendijk’s projection method (why?). For the latter we know how to construct an example of a function  $f$  with ‘arbitrarily bad’ convergence

and with a modulus of strong convexity  $\tilde{Q}$ . The function  $\lambda f$  for a sufficiently large  $\lambda$  will be the required example for the DFP.

This remark should not, of course, be interpreted as a criticism of the DFP. The only thing we have established is that efficient use of DFP requires a preliminary normalization of the problem being solved. The DFP 'senses' not only the ratio  $Q = L/l$  of the parameters of strong convexity of the  $(l, Q)$ -strongly convex problem being solved, but also the absolute values of these parameters.

It would be interesting to determine which normalization of the problem being solved is 'optimal' for the DFP, and whether it is true that, with such a normalization, the DFP implements the 'optimal' bound for the laboriousness of solution of problems with a given modulus of strong convexity. It would also be interesting to determine which of the standard methods of strongly convex programming implement the bound mentioned (or whether such methods exist at all). It is scarcely worth mentioning that the authors do not know the answers to these questions.

### 8.5.2

In concluding our brief and incomplete analysis of standard methods of strongly convex programming, we point out that the results obtained do not characterize the methods in all regards, but only from the particular point of view which we have chosen. It is true that this point of view seems to us to be the most natural one, in mathematical respects at least. It should, however, be borne in mind that we have not touched at all on such practically important aspects of methods as the simplicity of their computational organization and computational stability. Finally, the standard methods do not need, before they are put into operation, any *a priori* knowledge at all of the modulus of convexity of the function being minimized, at least, not explicitly (actually, certain estimates of this modulus are necessary all the same in accurate application of the methods: for example, it is necessary to choose properly the accuracy of solution of 1-dimensional problems, and so on). In this respect, of course, the practical advantage of the standard methods over the method described in Section 7.3 is quite appreciable (but it is to a certain extent cancelled by the concluding remarks in Section 7.5).

# 9

## Convex programming methods of zeroth order

In the previous chapters we have studied convex programming methods of the first order, i.e. methods using not only information about the values of the functionals in the problem under solution, but also information about their derivatives. In this chapter we shall be considering a situation where information about the derivatives is not accessible. From the practical point of view this situation would seem to be the more typical. At the same time it is objectively more complicated and it has been studied to a far less extent than the one considered earlier. For this reason our treatment will be more cursory, and the results less complete.

The plan of the exposition is as follows. The first two sections deal with methods using a deterministic oracle of zeroth order. In Section 9.1 the possibility is considered of constructing from this oracle a first-order oracle, with subsequent use of first-order methods. The main disadvantage of methods obtained in this way is their low stability relative to mistakes of the oracle. In Section 9.2 a theoretically more stable method of solving convex problems with a zeroth-order oracle is presented; this is based on the idea of the MMCG. In Sections 9.3–9.4 the case where a stochastic oracle of zeroth order is used is considered. In Section 9.3 a first-order stochastic oracle is constructed from the given zeroth-order stochastic oracle, and the possibility is considered of reducing the problem to the stochastic programming problems studied in Chapters 5–6. The methods obtained as a result have a laboriousness which is acceptable as regards its dependence on the dimensionality of the problem, but which increases too quickly with increasing demands on the accuracy of the solution. In Section 9.4, roughly speaking, a deterministic, zeroth-order oracle is constructed from the stochastic, zeroth-order oracle, and it is then used in application to the method of Section 9.2. The method so obtained has a laboriousness which is acceptable as regards its dependence on the required accuracy of solution, but which rapidly increases with increase in the dimensionality of the problem. The final part of Section 9.4 is devoted to a discussion of the results of the chapter.

We also touch on the question of the effects of bias in the oracle on the possibility of using the methods described.

### 9.1 METHODS OF ZEROTH ORDER: DETERMINISTIC ORACLE. I

Let  $\mathfrak{U}$  be a field of problems  $f = (f_0, \dots, f_m)$  of any of the classes considered previously of convex programming problems. We shall suppose that  $\mathfrak{U}$  is equipped with a suitable rule for varying the error of the approximate solutions (i.e. with a normalizing mapping  $f \rightarrow r(f)$ ). We consider the class of problems  $\bar{\mathfrak{U}}$  obtained by providing  $(\mathfrak{U}, r(\cdot))$  with an oracle  $\mathcal{O}$  with the observation function  $\psi(f, x) = f(x)$ ,  $x \in G$ , where  $G \subset E$  is the domain of definition of problems from  $\mathfrak{U}$ . It is assumed that  $E$  is a finite-dimensional space, and  $n = \dim E$ . Our purpose is to construct an efficient method of solving problems of the class  $\bar{\mathfrak{U}}$ .

#### 9.1.1

The first idea that comes into the head is to construct from the given zeroth-order, exact oracle a first-order oracle (it will no longer be an exact, but an approximate, one). The idea is simple: we already know how to solve efficiently problems of the class  $\bar{\mathfrak{U}}$  obtained from  $\mathfrak{U}$  by replacing the exact, zeroth-order oracle with an exact, first-order oracle. In other words, if we knew how to calculate not only the values, but also the derivatives, of the components  $f_j$  of the problem  $f \in \bar{\mathfrak{U}}$  being solved, then we would know how it should be solved. It is natural to try to construct approximate estimates of the derivatives of  $f_j$  from their values. The simplest thing to do is to replace these derivatives by difference quotients.

Thus, if it is required to construct an approximation  $\bar{f}'(x)$  to the derivative  $f_j$  at a point  $x \in G$ , then we may take it in the form

$$\bar{f}'(x) = \sum_{i=1}^n \frac{f_j(x + t_x e_i) - f_j(x)}{t_x} \varphi_j, \quad (1.1)$$

where  $e_1, \dots, e_n$  is any basis in  $E$ , and  $\varphi_j$  is a linear form on  $E$  such that

$$\langle \varphi_j | e_i \rangle = \begin{cases} 0, & i \neq j, \\ 1, & i = j, \end{cases}$$

and  $t_x \neq 0$ —the step of the difference quotient—is sufficiently small in absolute value. Other forms of difference approximations to the derivatives can, of course, be used.

When such approximations are used, the imitation of a single answer from a first-order oracle ‘costs’  $n+1$  questions to the zeroth-order oracle. It is thus

possible to hope that it will be feasible to use for solving problems in the class  $\bar{\mathfrak{U}}$  the methods for solving problems in the class  $\mathfrak{U}$ , i.e. first-order methods; in so doing, their laboriousness will be increased by  $n+1$  times.

#### 9.1.2

An approach of this sort encounters three problems. The first two are concerned with whether the approach itself is possible, and the third with its efficiency. These problems consist in the following.

I. However small we may have chosen the step of the difference quotient to be,  $\bar{f}'(x)$  will always be just an approximation, either better or worse, to the derivative  $f'_j(x)$ . Therefore, a first-order method can be applied effectively only when it is stable in relation to errors in the information of the first order supplied to it; when the errors in this information are sufficiently small, it will, as before, solve the problem well.

II. Let us suppose that it is desired to use a stable, first-order method ('stable' in the sense just described), and we want to ensure an accuracy  $v$  in the solution of the problem. Suppose further that, by analysing the method, we are in a position to know what size errors in the estimates of the derivatives will be acceptable. Then it is required to choose the successive displacements  $t_{x_i}$  in such a way as to ensure that no errors greater than the acceptable ones occur in the estimates of the derivatives. Moreover, we must know how to choose such  $t_{x_i}$  'constructively', i.e. on the basis of the information available at the given step. If there is no such means of choosing the  $t_{x_i}$ , then it is not clear how the intended reduction of zeroth-order methods to first-order methods can be effected.

III. Suppose we have coped with problems I and II, and that the intended plan has been realized. Let us suppose that the plan has been applied to an efficient first-order method, i.e. a method which realizes the complexity of the class  $\bar{\mathfrak{U}}$ . Will the zeroth-order method so obtained realize (or almost realize) the complexity of the class  $\bar{\mathfrak{U}}$ ? Generally speaking, it is not clear why this should be the case.

We shall consider in general terms each of these three problems, starting with the first. Of course, generally speaking, a first-order method is not obliged to be stable relative to errors in the information. Methods can be envisaged which would work well only with exact information. However, with all the concrete methods known to us, and in particular, with all the first-order methods considered previously (exceptions are the classes  $H_{x_1}$  but not  $H_{x_1}^V$ ), matters turn out well in this respect, at least from the theoretical point of view. Not one of them requires absolutely accurate information. Moreover, for each

of them a simple analysis based on convergence enables us to indicate constructively that accuracy in the incoming information under which the use of the method employing the approximate information will lead to a result of the required accuracy in the same (order of) time as in the case of exact information. This level (admissible for the method and not zero) of errors in the information flow can be calculated explicitly as a function of the *a priori* information (i.e. as a function of the class of problems being solved) and of the required accuracy of solution.

Sometimes, it is true, a method has to be slightly modified to make it 'noise-stable' in the above sense. We shall not go into the questions of this stability for the first-order methods which we have constructed. As already mentioned, it is comparatively easy to do this—by analyzing the proofs of their convergence theorems. We point out that this has actually already been done for classes of general convex problems with either deterministic or stochastic oracles: from the very start the corresponding classes were equipped with an approximate, not an exact, oracle.

Thus, in the cases in which we are interested, the first of the problems mentioned above is solved positively.

As regards the second problem, the situation here is more complicated. We have to ascertain for what classes of problems it is possible to choose constructively the step for the difference quotient so as to ensure a specified accuracy of approximation in a support functional (as previously explained, this accuracy is found from an analysis of the stability of the corresponding first-order method).

Suppose that  $f$  is differentiable in a neighbourhood of a point  $x \in \text{int } G$  in which we are interested. Then the derivative  $f'_j(x)$  is

$$\lim_{t \rightarrow +0} \sum_{i=1}^n \frac{f_j(x + te_i) - f_j(x)}{t} \varphi_j.$$

Suppose that the modulus of continuity of the derivative of  $f_j$  is known *a priori*, i.e. a function  $\omega_j(s) \rightarrow 0$  such that

$$\|f'_j(x) - f'_j(y)\|_* \leq \omega_j(\|x - y\|)$$

for  $x, y \in \text{int } G$  (here  $\|\cdot\|$  denotes any norm in the corresponding space). It is clear that

$$\left| \frac{f_j(x + te_i) - f_j(x)}{t} - \frac{d}{dt} \Big|_{t=0} f_j(x + te_i) \right| \leq \omega_j(t \|e_i\|) \|e_i\|,$$

and so it is possible to indicate explicitly a step-size  $t_x$  for which the difference quotient will approximate the derivative with the required accuracy. Such, in particular, is the situation for classes of smooth convex (strongly convex) problems. In these classes the  $f'_j(x)$  are Lipschitz with constants known *a priori*

(the constants appear in the definition of the classes). Thus, for these two classes, the second of the problems mentioned above is also solved positively.

It is a different matter with classes of general convex problems. Here there is no *a priori* information at all about the modulus of continuity of the derivatives (and indeed there cannot be, since a class may contain problems with a discontinuous derivative). In this situation no choice of the steps for the difference quotients will ensure obtaining sufficiently exact estimates of the derivative. This is easily seen by considering the following example. Let  $E$  be a plane with an orthonormal basis  $(e_1, e_2)$ , and let  $f_0(x)$  be a convex function which has, as one of its level curves, the boundary of a square. In the calculation of the derivative of  $f_0$  at one of the corners of the square by means of (1.1) the estimate of the derivative will be zero for all positive values of the step  $t_x$ , although all the support functionals to  $f_0$  at 0 may be different from 0.

Thus the immediate answer to the question in which we are interested for the class of general convex problems is negative, and so it would seem at a first glance to be impossible to reduce the solution of these problems with a zeroth-order oracle to their solution by means of first-order methods. It turns out, however, that theoretically this difficulty can easily be overcome; the phenomena like those just described can occur only at certain 'exceptional' points  $x$ , and a simple randomization enables these points to be avoided 'almost certainly'. A detailed examination of this question will be given below.

We shall now say a few words about the last of the problems formulated. For a full investigation of the efficiency of the zeroth-order methods obtained in the way described above it would be necessary to compare the estimates for the laboriousness of these methods with the lower bounds (if possible exact) of the complexity of the class of problems  $\bar{\mathfrak{A}}$ . However, certain qualitative results can be obtained without a lengthy analysis of the complexity of  $\bar{\mathfrak{A}}$ .

We carry out the appropriate analysis for deterministic methods and the deterministic complexity. Similar considerations could also be carried out for stochastic methods and the stochastic complexity corresponding to them. Suppose, as above, that  $\bar{\mathfrak{A}}$  is the class of problems obtained from  $\mathfrak{A}$  by replacing the exact zeroth-order oracle by an exact first-order oracle. Let  $N_{\bar{\mathfrak{A}}}(v)$  and  $N_{\mathfrak{A}}(v)$  denote the deterministic complexities of functions in these classes. Let  $\{\mathcal{B}(v)\}_{v>0}$  be a family of methods of solving problems of the class  $\bar{\mathfrak{A}}$ ,  $\mathcal{B}(v)$  solving these problems with accuracy  $v$ . Suppose that  $l(v)$  is the laboriousness of  $\mathcal{B}(v)$ . We consider the *efficiency function* of the family in question, i.e. the ratio  $\theta(v) = N_{\bar{\mathfrak{A}}}(v)/l(v)$ ; the closer  $\theta$  is to 1, the 'better' the family  $\{\mathcal{B}(v)\}$ . We further assume that our scheme of imitating a first-order oracle by means of a zeroth-order oracle can be implemented and applied to the methods of the family  $\{\mathcal{B}(v)\}$ , and that it leads to a family  $\{\bar{\mathcal{B}}(v)\}$  of  $\bar{\mathfrak{A}}$ -methods,  $\bar{\mathcal{B}}(v)$  solving problems in  $\bar{\mathfrak{A}}$  with accuracy  $v$ . Because of the construction used, the laboriousness of  $\bar{\mathcal{B}}(v)$  is  $(n+1)l(v)$ , and the complexity  $N_{\bar{\mathfrak{A}}}(v)$  is clearly  $\geq N_{\bar{\mathfrak{A}}}(v)$ . Therefore the efficiency  $\bar{\theta}(v)$  of the family of  $\bar{\mathfrak{A}}$ -methods considered

satisfies the inequality

$$\theta(v) \geq \frac{\theta(v)}{n+1}.$$

Thus the proposed construction, provided it be applicable, reduces the efficiency by only  $n+1$  times. Thus, with accuracy up to a factor  $O(n)$ , the construction considered preserves the efficiency of the methods. In particular, if the family of methods  $\{\mathcal{B}(v)\}$  is asymptotically sub-optimal as  $v \rightarrow 0$  (i.e.  $\theta(v) > \theta_0 > 0$  for all  $v > 0$ ), then the same is also true of the family  $\{\bar{\mathcal{B}}(v)\}$ . In general, the 'degree of non-optimality' of  $\bar{\mathcal{B}}(v)$  can be only  $n+1$  times greater than that of  $\mathcal{B}(v)$ .

In the analysis of the behaviour of methods under conditions of large dimensionality (let us say, in investigating their efficiency asymptotically with respect to  $n$ ), the factor  $O(n)$  is already substantial, especially in situations where its occurrence leads to qualitative effects: the laboriousness of the methods, which previously were independent of the dimension, now begins to increase linearly with increase in the dimension (such situations arise when we consider classes of smooth convex (strongly convex) problems and general convex problems on  $l_p$ -bodies of high dimension). It is possible, however, to show that this phenomenon is unavoidable, since the complexity  $N_{\mathcal{B}}(v)$  itself increases in all natural situations with increase in dimensionality, this increase being at least linear. This is in contrast to the complexity  $N_{\bar{\mathcal{B}}}(v)$ , for which this circumstance does not always occur.

*Exercise 1.* Consider the class of problems of the form

$$f(x) = \langle \varphi | x \rangle \rightarrow \min | x \in V \subset E^n, V = \{x | \|x\| \leq 1\},$$

generated by vectors  $\varphi$  with  $\|\varphi\| = 1$ . Prove that the deterministic (and stochastic) complexities of solving problems of this class with an absolute error  $\epsilon$  is not less than  $cn$  if  $\epsilon < \epsilon_0$  (here  $c, \epsilon_0 > 0$  are absolute constants).

In view of the result formulated in Exercise 1, a linear growth in the complexity of convex programming as regards zeroth-order methods when the dimension increases is an unavoidable phenomenon, and the construction considered is not connected with specially unfavourable (from a roughly qualitative point of view) losses in efficiency. In view of what has been said, we shall not investigate specially the complexity of convex programming with zeroth-order oracles, contenting ourselves with estimates of method efficiency accurate up to a factor  $O(n)$ .

### 9.1.3

We now pass on to a presentation of general considerations regarding the implementation of the proposed plan as applied to classes of general (or

Lipschitz) convex problems (we have already pointed out that for classes of convex (or strongly convex) problems its implementation does not evoke difficulties). Clearly, it suffices to know how to imitate a first-order oracle of accuracy  $v_0 > 0$  for classes of general (or Lipschitz) convex problems by means of a zeroth-order oracle. It is sufficient to know how to do this in the case where  $G \subset R^n$  is a convex body, and it is required to imitate the answer of the oracle at interior points of  $G$ . Suppose that  $f(x)$  is a convex, continuous function on  $G$ ,  $v_0 > 0$ ,  $x_0 \in \text{int } G$ , and that  $V_f = \sup_G f - \inf_G f$ . We consider the following stochastic procedure for estimating the support functional to  $f$  at  $x_0$  (a procedure of this sort was first proposed by A. M. Gupal [10]).

1°. Let  $\Delta$  be a simplex, with centre at 0, which lies entirely within  $G$ . We calculate the values of  $f$  at the vertices  $y_0, \dots, y_n$  of this simplex and at the point  $x_0$ . We construct the number  $r = r_f = \max_i f(y_i) - f(x_0)$ . Clearly,  $r_f \geq 0$  and  $r \leq V_f$ . Moreover, if  $r_f = 0$ , then  $f$  is linear within the limits of  $\Delta$ , and from its values at the  $y_i$  its support functional at  $x_0$  can be calculated exactly (why?). We shall now suppose that  $r > 0$ .

2°. Let  $e_1, \dots, e_n$  be a fixed basis in  $R^n$ , and let  $x^1, \dots, x^n$  be the co-ordinates of  $x$  in this basis. For simplicity we shall suppose that  $x_0^i = 0$ ,  $1 \leq i \leq n$ . Let  $T_0 > 0$  be so small that the cube  $K_0 = \{x | |x^i| \leq 3T_0\}$  is contained in  $\Delta$ . We choose a  $T \leq T_0$ , and then  $\bar{x}$  is chosen in the cube  $K = \{x | |x^i| \leq T, 1 \leq i \leq n\}$  at random (according to the uniform distribution). We calculate the difference approximation to the support functional to  $f$  at  $\bar{x}$  in the form

$$\bar{\varphi} = \sum_1^n \frac{f(\bar{x} + \tau e_i) - f(\bar{x})}{\tau} \varphi_i,$$

where  $\varphi_i$  is a linear functional such that

$$\langle \varphi_i | e_j \rangle = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases}$$

We take  $\bar{\varphi}$  as the estimate provided by the described procedure for estimating the support functional to  $f$  at  $x_0$ . More precisely, we shall suppose that the answer of the imitated first-order oracle to a question about  $f$  at the point  $x_0$  is the affine function

$$g(y) = f(\bar{x}) + \langle \bar{\varphi} | y - \bar{x} \rangle - \rho,$$

where  $\rho > 0$  is the parameter of the procedure.

The procedure described has the parameters  $T, \tau, \rho > 0$  and it requires  $1 + (n+1) + (n+1) = 2n+3$  calculations of values of the function  $f$ . We emphasize that the procedure has a randomized character, and so the function  $g$ , the result of the procedure, is stochastic.

We shall prove that, with a suitable choice of the parameters  $T, \tau, \rho > 0$ , we can achieve the result that  $g$ , with probability as close to 1 as we please, will

satisfy the condition

$$A(v_0): \quad g(y) \leq f(y), \quad g(x_0) \geq f(x_0) - v_0 \mathbf{V}_f,$$

i.e. the modelled first-order oracle 'almost certainly' provides an estimate of the support functional with the required accuracy.

It is convenient to introduce a Euclidean structure into  $R^n$ , with  $e_1, \dots, e_n$  as an orthonormal basis. Let  $\|\cdot\|$  denote the corresponding Euclidean norm. Let  $d_G$  be the diameter of  $G$  in the Euclidean metric obtained.

**Proposition.** Let  $\alpha > 0$  and let the parameters of the described procedure (considered under the condition  $r > 0$ ) satisfy the relations

$$\begin{aligned} T &= \frac{v_0 r}{2} \left[ \frac{2n^2 r}{T_0} + \frac{2n\rho}{d_G} \right]^{-1}, \\ \tau &= \frac{\alpha T T_0 \rho}{3n^2 \sqrt{n d_G r}}, \quad \rho = \frac{v_0 r}{2}. \end{aligned} \quad (1.2)$$

Then the probability that the condition  $A(v_0)$  is satisfied by the result of applying this procedure to  $f$  does not exceed  $\alpha$ .

**Exercise 2.** Prove the proposition.

We point out that a procedure of the type described can be applied simultaneously to several convex functions  $f_0, \dots, f_m$  to obtain 'almost certain' estimates (to the required accuracy) of their support functionals. Moreover, the  $2n+3$  points at which it is necessary to calculate the values of these functions can be chosen to be the same for all  $m+1$  of these functions, and the parameters  $\rho, T, v$  can be chosen 'reckoning on the worst case'—on the least of the non-zero values of  $r_{f_0}, \dots, r_{f_m}$ . Further, the probability that  $A(v_0)$  is not satisfied for at least one of the functions  $f$  (that is, the uncertainty of the imitation) does not exceed  $(m+1)\alpha$ , where  $\alpha$  is the parameter according to which  $\rho, T$ , and  $\tau$  are chosen in (1.2). Thus, by a suitable choice of the parameters of the procedure, we can make this probability arbitrarily small.

#### 9.1.4

It is clear how to apply the above result to the construction of randomized zeroth-order methods of solving general (or Lipschitz) convex problems on  $G$ . Let  $v > 0$  be the required (relative) accuracy of solution of, let us say, general convex problems with  $m$  constraints on  $G$ . We put  $v_0 = v/4$ , and let  $\mathcal{B}_v$  denote the method constructed for an accuracy  $v/2$  in the solution of problems in classes of the type  $C^{v_0}(G, E, m)$  (see Section 2.2) (for simplicity, let us say a deterministic method); let  $l(v)$  be the upper bound for its laboriousness on this class. We choose an arbitrarily small  $\alpha > 0$ , and we shall imitate a first-order

oracle of relative accuracy  $v_0$  by means of a zeroth-order oracle, ensuring an uncertainty of the imitation  $\leq \alpha/l(v)$ . We consider a method  $\tilde{\mathcal{B}}$  obtained by applying  $\mathcal{B}(v)$  in combination with the imitation of the oracle for  $\mathcal{B}(v)$ . The laboriousness of  $\tilde{\mathcal{B}}$  will be  $\leq (2n+3)l(v)$ . It is, moreover, clear that, with probability  $\geq 1-\alpha$ , the result provided by  $\tilde{\mathcal{B}}$  will be a solution of the problem with a relative error  $\leq v/2$ . In particular, when  $\alpha = v/4$ , the mean relative error of  $\tilde{\mathcal{B}}$  on the class considered will not exceed  $v$  with laboriousness  $l(v)(2n+3)$ . Thus a simple randomization of first-order methods enables them also to be used in a case where only zeroth-order information is available. This randomization enables complications connected with a possible non-smoothness of the components in the problem to be avoided.

## 9.2 METHODS OF ZEROTH ORDER: DETERMINISTIC ORACLE. II

The simple ideas used in the previous section are of little use from the practical point of view. The point is that the zeroth-order oracle was there assumed to be exact (as also were all the calculations made using the information provided by it). Small errors in the calculations of values of the components in the problem can lead to colossal errors in the resulting estimates of the derivatives because of the smallness of the steps in the difference quotients. The latter have upper bounds of the order  $O(v_0^2 \alpha)$ . For the construction to be successful we must have  $v_0 \leq v$  and  $\alpha \leq v$ , i.e.  $\tau \leq O(v^3)$ . Also, the admissible errors in the calculation of the difference quotients is of the order  $v$ , i.e. the admissible error in the input information is of order  $\tau v = O(v^4)$ . Thus the construction described ensures successful solution only if the accuracy of the input information is of a considerably higher order than the accuracy required in the solution.

### 9.2.1

We shall describe another deterministic, zeroth-order method of solving general convex problems. The laboriousness of this method as regards the nature of its dependence on the accuracy  $v$  is the same (asymptotically as  $v \rightarrow 0$ ) as for the optimal MCG method. As regards the sensitivity of the laboriousness to dimensionality, it is considerably higher for the method about to be described than for the MCG (or for the randomized zeroth-order method obtained from the MCG by means of the construction described in the previous section). Thus the new method certainly does not realize the complexity asymptotically with regard to dimensionality. On the other hand, it is considerably more 'noise-stable' than the method of the previous section, at least in the formal sense of noise-stability: the errors, admissible for the method, in the current information are of the same order as the required accuracy of solution of the problem (a precise formulation will be given later).

We start by describing the class of problems of concern to us. Let  $G$  be a convex, closed, bounded body in  $R^n$ , and let  $C(G, R^n, m)$  be, as earlier in Section 2.2, the set of convex, continuous problems  $f = (f_0, \dots, f_m)$  on  $G$  with the same normalizing factors as in Section 2.2. We equip this set with a zeroth-order oracle

$$\psi(f, x) = f(x), \quad x \in G, \quad f \in C(G, R^n, m),$$

and denote the resulting class by  $\bar{C}(G, R^n, m)$ . We shall present a method of solving problems of this class. The rest of this exposition is based on the authors' paper [31].

### 9.2.2

The idea of the method is a development of that of MMCG (Section 2.5), and it consists in the following. For greater clarity, let  $m = 0$ . The answer of a first-order oracle to a question about  $f$  at a point  $x$  enables the solution of the problem to be localized in one of the parts into which  $G$  is divided by a certain hyperplane passing through  $x$ . In the other part  $f$  is not less than it is at  $x$ . By excluding this part from  $G$ , we have decreased the domain in which the solution lies. The answer of a zeroth-order oracle to a question about  $f$  at a point  $x \in G$  does not enable the domain of localization of the solution to be reduced. However, if we 'interrogate  $f$ ' at the vertices  $x_0, \dots, x_n$  of a certain simplex  $\Delta$ , then we do secure for ourselves a certain reduction in the domain of localization.

Let  $x_{i_0}$  be that vertex of the simplex at which  $f$  is maximal. On the side  $\Delta_0$  of  $\Delta$  opposite to  $x_{i_0}$  the value of  $f$  is not greater than at  $x_{i_0}$  (because  $f$  is convex, and so its maximum on  $\Delta$  coincides with its maximum on the vertices of  $\Delta$ ). Therefore  $f$  does not decrease in the passage from  $x \in \Delta_0$  to  $x_{i_0}$ ; but then it does not decrease along the ray  $[x, x_{i_0}]$  going from  $x$  towards  $x_{i_0}$ . Therefore in the polyhedral cone  $K$  with vertex  $x_{i_0}$  obtained by reflecting in  $x_{i_0}$  the cone of directions from  $x_{i_0}$  into  $\Delta$  we have

$$f(x) \geq f(x_{i_0}).$$

Therefore the cone  $K$  can be excluded from  $G$ , thus decreasing the domain of localization of the solution.

Unfortunately, the new domain of localization is only 'a little less' than the original one, and (what is worst of all) it is not convex; this creates substantial difficulties in implementing the idea 'in its pure form'.

Here, however, the basic idea of the MMCG comes to our aid. Suppose that  $G$  is a ball (or ellipsoid), and that  $\Delta$  is a simplex situated close to the centre of  $G$  (so that the cone  $K$  can be regarded, roughly speaking, as having a vertex at the centre of  $G$ ). Then  $G \setminus K$  is certainly contained in a convex body  $\bar{G}_1$  obtained by cutting off from  $G$  a certain 'cap'.  $\bar{G}_1$  could be taken as the new domain of localization of the solution. But we cannot iterate the process—because the

boundary of  $\bar{G}_1$  contains a plane part, and so the next 'cone to be rejected' might be supported exactly on this plane part, and it would not be possible to cut off a new 'cap'. It is therefore expedient to try to enclose  $\bar{G}_1$  in an ellipsoid  $G_1$  of minimal volume, and to regard this  $G_1$  as the new domain of localization of the solution. With such a course of action the geometrical situation at all steps will be of the same type. However, the actions described will be successful only if  $G_1$  proves to be less (in volume) than  $G$ , i.e. when the volumes of the successive domains of localization decrease in geometrical progression as the number of steps increases. It is easy to work out what the aperture of the cone  $K$  must be for the volume of  $G_1$  to be less than that of  $G$ . Unfortunately, it turns out that the polyhedral cone corresponding to a regular simplex is too small.

We are faced with a new problem. In order to implement the intended plan, we have to be certain that the cone  $K$  excluded is sufficiently large. It is impossible to choose the simplex  $\Delta$  so that an arbitrary one of the  $n+1$  cones corresponding to it would have the necessary size. Since we do not know in advance (before consulting the oracle) which of these  $n+1$  cones it will be possible to remove, it can very likely happen that the cone to be removed proves to be not big enough, and we shall not be able to decrease the domain of localization of the solution while preserving its form.

The simplest way out of this awkward situation is to calculate the values of  $f$ , not at the vertices of a small simplex situated close to the centre of  $G$ , but at the vertices of a polyhedron  $\Gamma$  having a large number of points on the boundary. As before, it will be possible to remove from  $G$  a cone obtained by reflecting in that vertex of  $\Gamma$  which gives the maximal value of  $f$  the cone of directions from that vertex into  $\Gamma$ . If we take this polyhedron to be one which 'well approximates a ball', then the cone described will be sufficiently big whichever of the vertices of  $\Gamma$  turns out to be maximal.

A similar device (in a rather different context) is used in a known method due to Kuzovkin and Tikhomirov [17] of solving convex extremal problems with a zeroth-order oracle. Its disadvantage is that the number of vertices of  $\Gamma$ , and so also the laboriousness of the method, has to increase exponentially with growth of  $n$ . To avoid this, we shall use a different approach, less obvious but more economical.

Let  $\Delta$  be a regular simplex with its centre of gravity at the centre  $x_0$  of the ball  $G$ , and let  $y^0$  be its maximal vertex. We consider a regular pyramid  $\Delta_0$  with vertex  $y^0$ , with the necessary aperture at the vertex, and with its altitude collinear with  $y^0 - x_0$ . Its base is chosen so that for the other vertices  $y_1, \dots, y_n$  the vectors  $y_i - y^0$  are orthogonal to  $y_i - x_0$  (i.e.  $y_i$  is the projection of  $x_0 - y^0$  on to the corresponding side of  $\Delta_0$ ). It is possible that  $y^0$  is the maximal vertex in the pyramid  $\Delta_0$ . We have then found a cone of the necessary size for removal. If  $y^0$  is not the maximal vertex, then we select the maximal vertex in  $\Delta_0$  and call it  $y^1$ , and construct a pyramid  $\Delta_1$  with vertex at  $y^1$  in the same way as the pyramid  $\Delta_0$  was constructed from  $y^0$ . If  $y^1$  is maximal in  $\Delta_1$ ,

everything is in order; if not, we select the maximal vertex in  $\Delta_1$  and construct  $\Delta_2$ , and so on.

The central fact is that this process cannot continue long. For,

$$f(y^0) < f(y^1) < f(y^2) < \dots, \text{ but } |y^{i+1} - x_0| = (\sin \psi)^{i+1} |y^0 - x_0|,$$

where  $\psi$  is the angle between  $y^0 - x_0$  and the lateral edge of  $\Delta_0$ . Therefore, for some  $i$  which can be estimated explicitly, we shall have  $|y^{i+1} - x_0| \leq 1/n |y^0 - x_0|$ , i.e.  $y^{i+1} \in \Delta$  (provided the process has not terminated earlier). But this is impossible, because by construction  $f(y^{i+1}) > f(y^0)$ , and  $y^0$  was the maximal vertex of  $\Delta$ . Thus, in the course of the process described, not later than after a certain number (estimated in advance) of steps, a cone, of the necessary size, for removal will have been constructed.

### 9.2.3

We now pass on to the description of the method. We shall consider the case where it is required to ensure a given absolute error  $\epsilon = (\epsilon_0, \dots, \epsilon_m) > 0$  in the solutions of problems in the class  $\mathcal{C}(G, R^n, m)$ . We preface the exposition of the method by a lemma which generalizes the lemma in Section 2.5.2.

**Lemma.** Let  $W_r$  be the Euclidean ball with centre at 0 and radius  $r$  in the  $n$ -dimensional Euclidean space  $R^n$  with an orthonormal basis  $e_1, \dots, e_n$ ,  $n > 1$ . Let  $Q(\varphi)$  be a cone with aperture  $2\varphi$ :

$$Q(\varphi) = \{x | \langle x | e_n \rangle \geq \cos \varphi |x|\}, \quad 0 \leq \varphi \leq \frac{\pi}{2}.$$

If  $\cos \varphi < 1/n$ , then the set  $W_r \setminus Q(\varphi)$  can be enclosed in an ellipsoid of volume  $\beta^n(\varphi) |W_r|_n$  having its centre at the point  $(-\gamma(\varphi)e_n)$ , where

$$\gamma(\varphi) = \frac{1 - n \cos \varphi}{1 + n},$$

$$\beta(\varphi) = 2 \left( \sin \frac{\varphi}{2} \right)^{\frac{n-1}{n}} \left( \cos \frac{\varphi}{2} \right)^{\frac{n+1}{n}} \frac{n(n-1)/(n+1)^{\frac{1}{n}}}{\sqrt{n^2 - 1}}.$$

We point out that the lemma in Section 2.5.2 is a particular case of this lemma, which is obtained when  $\varphi = \pi/2$ .

**Exercise 1.** Prove the lemma.

We shall be concerned with the particular case of the lemma corresponding to  $\varphi = \varphi_n = \cos^{-1}(1/(2n))$ .

It can be seen that

$$\beta(\varphi_n) = 1 - \frac{d_n}{n^2}, \quad d_n > 0,$$

$$\lim_{n \rightarrow \infty} d_n = \frac{1}{8}, \quad \gamma(\varphi_n) = \frac{1}{2(n+1)}.$$

It is clear that there is a  $\theta_n = \theta(n) > 0$  such that the assertion of the lemma for  $\varphi = \varphi_n$  remains true also when  $\beta(\varphi_n)$  in its formulation is replaced by  $\beta(\varphi_n) = \frac{1}{2}(1 + \beta(\varphi_n))$ , and the cone  $Q(\varphi_n)$  is allowed to be displaced arbitrarily subject to the condition that its vertex lies in a  $\theta(n)r$ -neighbourhood of the centre of  $W_r$ . When this is done,  $\theta(n)$  is  $\geq c/n$ , where  $c$  is an absolute constant.

Let  $\epsilon > 0$  be the specified absolute accuracy. We further assume that we have two similar ellipsoids  $\bar{W}_0$  and  $W_0$  such that  $\bar{W}_0 \subset G \subset W_0$ , and let  $|W_0| = \beta^n |\bar{W}_0|$ . We know (see Section 4.1.1) that  $\bar{W}_0$  and  $W_0$  can be chosen so that  $\beta \leq n$ . Let  $O_0$  be a co-ordinate system relative to which  $W_0$  is a ball of radius  $r_0$  with centre at 0. The method being described is made up of stages. The zeroth stage is a preparatory one. In the  $i$ th stage ( $i \geq 1$ ) the next domain of localization of the solution  $W_i \cap G$  is constructed; here  $W_i$  is an ellipsoid which, in the co-ordinate system  $O_i$ , is a ball of radius  $r_i$  with centre at 0 and which has the same volume element as in the original co-ordinates  $O_0$ . The oracle's answers about the values of the components in the problem being solved at points  $x$  will be denoted by  $\psi_j(x)$ ,  $0 \leq j \leq m$  (the reason for this strange notation is that we want to have the possibility of applying the method also in the case where we have only an approximate, and not an exact, oracle at our disposition).

*The zeroth step.* We consider a regular simplex  $\hat{\Delta}$  (regular, that is, in the  $O_0$  co-ordinates) inscribed in the ball  $\bar{W}_0$ . Let  $x^0, \dots, x^n$  be its vertices. At these points questions about the problem being solved are put to the oracle, and the numbers

$$\bar{r}_j(f) = \max \{0, \max_{0 \leq k \leq n} \psi_j(x^k)\}, \quad j = 1, 2, \dots, m,$$

are formed. We put

$$\bar{r}_0(f) = \max_{0 \leq k \leq n} \psi_0(x^k).$$

*Stage  $i$*  ( $1 \leq i \leq M_f$ ) ( $M_f$  is the number, formed during the course of the method's work, of stages in the solution of the problem). At the  $i$ th stage  $O_{i-1}$ ,  $r_{i-1}$ ,  $W_{i-1}$ , and also the centre  $x_i$  of the ball  $W_{i-1}$  are known. In the  $i$ th stage the following operations are carried out in turn. (All the considerations are carried out in the co-ordinates  $O_{i-1}$ , which are assumed to be Cartesian in the  $i$ th stage; thus the Euclidean structure of  $R^n$  will be proper to the particular stage.) We put  $\theta_n = \theta(n)$ .

1°. It is possible that the ball  $\bar{W}_{i-1}$  of radius  $\theta_n r_{i-1}$  with centre at  $x_i$  is not contained in  $G$ . In that case we find a point  $u_i \in \bar{W}_{i-1} \setminus G$  and consider the support functional  $\varphi_i$  to the function  $\rho_{\|\cdot\|}(\cdot, G)$  at the point  $u_i$  (here  $\|\cdot\|$  is any norm, fixed once and for all, on  $R^n$ ). Then  $\varphi_i \neq 0$  (because  $u_i \notin G$ ). We consider the cone  $K_i = \{ \langle \varphi_i | x - u_i \rangle \geq 0 \}$  with vertex at  $u_i$ . Clearly,  $K_1 \cap G = \emptyset$ . We choose as  $W_i$  the ellipsoid of least volume which contains  $\bar{W}_{i-1} \setminus K_i$ , and as  $O_i$

that co-ordinate system (with the same volume element as in  $O_{i-1}$ ) in which  $W_i$  is a ball. By Lemma 9.2.3 the radius  $r_i$  of the ball  $W_i$  in the  $O_i$  co-ordinates does not exceed  $\beta(n)r_{i-1}$ . We take the centre of  $W_i$  as  $x_{i+1}$ . We put  $j(i) = m + 1$ , and go to 3° (in stage 2.2 of the procedure given below).

2°. Now suppose that  $\tilde{W}_{i-1} \subset G$ . We inscribe in  $\tilde{W}_{i-1}$  a regular simplex  $\bar{\Delta}_i$ ; let  $\bar{z}_0, \dots, \bar{z}_n$  be its vertices. Questions about the problem  $f$  being solved are put to the oracle at the points  $x_i, \bar{z}_0, \dots, \bar{z}_n$ , and the numbers

$$r_i^j = \max_{0 \leq k \leq n} \psi_j(\bar{z}_k), \quad 0 \leq j \leq m$$

are formed (it is clear that  $r_i^j = \max_{x \in \bar{\Delta}_i} f_j(x)$  for an exact oracle). We put

$$j(i) = \begin{cases} 0, & \text{if } r_i^j \leq \varepsilon_j, \quad 1 \leq j \leq m, \\ j > 1 \text{ such that } r_i^j > \varepsilon_j \text{ if there is such a } j \in \overline{1, m}. \end{cases}$$

Let  $\Delta_i = \bar{\Delta}_i$  when  $j(i) > 0$ . Otherwise we inscribe in  $\bar{\Delta}_i$  a ball  $\hat{W}_{i-1}$  touching the faces of  $\bar{\Delta}_i$ , and we consider a regular simplex  $\Delta_i$  inscribed in  $\hat{W}_{i-1}$ . Let  $\tilde{V}_i$  denote a ball described round  $\Delta_i$ , and let  $\tilde{r}_i$  be its radius. It is clear that  $\tilde{r}_i \geq (\theta_n/n)r_{i-1}$  and that  $\tilde{V}_i \subset \hat{W}_{i-1}$ .

We now apply the following procedure.

2.1. For brevity we write  $g(y) = \psi_{j(i)}(y)$ . For any finite number of points of  $G$  we define the maximal point as that point of the set at which  $g$  is maximal. Further, in speaking of the choice of the maximal point among the points of a given set, we shall bear in mind that the oracle has to be interrogated at each of these points.

Let  $y^0$  be the maximal of the vertices of  $\Delta_i$ . We construct a regular  $(n+1)$ -faced pyramid  $\Delta^0$  with vertex  $y^0$ , with its altitude collinear with  $y^0 - x_i$ , and with the angle  $\varphi_n$  between the altitude and the lateral faces, and with the lateral edges orthogonal to the radii-vectors of their ends lying on the base (we recall that the exposition is being carried out in the Euclidean structure in  $R^n$  of the  $i$ th stage; the Cartesian co-ordinates in it are  $O_{i-1}$ ). We note that  $\Delta^0 \subset \tilde{V}_i$ . If  $y^0$  is the maximal one of the vertices of  $\Delta^0$ , then we shall say that  $\Delta^0$  is *normal*, and we shall put  $z_i = y^0$  and pass on to stage 2.2 of the procedure. Otherwise, we shall find the maximal vertex  $y^1$  of  $\Delta^0$ , and put  $\delta^0 = g(y^1) - g(y^0)$ . We construct for  $y^1$  a pyramid  $\Delta^1$  according to the same rules as in constructing  $\Delta^0$  for  $y^0$ . If  $\Delta^1$  is normal (i.e. if  $y^1$  is its maximal vertex), then we put  $y^1 = z_i$  and pass on to stage 2.2; otherwise we find the maximal vertex  $y^2$  of the pyramid  $\Delta^1$ , put  $\delta^1 = g(y^2) - g(y^1)$ , construct  $\Delta^2$ , and so on. We break off this process at the moment when we first obtain  $y^k \in \Delta_i$  (this will necessarily occur unless, of course, the process has been terminated earlier by a passage to stage 2.2). If in the course of constructing the successive points a normal pyramid were not obtained, then we would select from the constructed points  $y^0, \dots, y^{k-1}$  that to which the least of the numbers  $\delta^0, \dots, \delta^{k-1}$  corresponds. We then denote this point by  $z_i$  and pass on to stage 2.2.

2.2. By construction,  $z_i$  is a certain point  $y^s$ . We consider the cone

$$K_i = \{y \mid \exists \lambda > 0: \lambda(y^s - y) + y^s \in \Delta^s\}.$$

By the construction of  $\Delta^i$ , the cone  $K_i$  contains a circular cone with axis  $y^s - x_i$ , with the angle  $2\varphi_n$  as its aperture, and with its vertex at  $y^s$ . By construction,  $y^s \in \tilde{V}_i \subset \hat{W}_{i-1}$ .

By the lemma in Section 9.2.3, the domain  $W_{i-1} \setminus K_i$  is contained in an ellipsoid  $W_i$  of volume  $(\beta(\varphi_n))^n |W_{i-1}|_n$ . Let  $O_i$  be that co-ordinate system in which  $W_i$  is a ball with centre at 0. Also let the volume element in  $O_i$  be the same as that in  $O_{i-1}$ . Then the radius  $r_i$  of the ball  $W_i$  in the  $O_i$  co-ordinates is  $r_i = \beta(\varphi_n)r_{i-1}$ .

3°. We put  $a_i = \min \{\psi_0(z_s) \mid s \leq i, j(s) = 0\}$  (when  $j(s) > 0, 1 \leq s \leq i$ , we put  $a_i = +\infty$ ). Let  $\rho_i = \max \{0, \tilde{r}_0(f) - a_i\}$ , and let

$$v_i = \begin{cases} 1, & j(i) = m + 1, \\ \varepsilon_{j(i)}/\tilde{r}_{j(i)}(f), & j(i) \in \overline{1, m}, \\ \frac{\varepsilon_0}{\rho_i + \varepsilon_0}, & j(i) = 0. \end{cases}$$

We further put  $v^i = \min_{0 \leq s \leq i} v_s$ , where  $v_0 = 1$ . If  $r_i < (v^i/(4n\beta))r_0$ , we pass on to the concluding stage; otherwise, to the stage  $i + 1$ .

*Concluding stage.* The work of the method stops with output of the result

$$\bar{x} = \begin{cases} * & \text{if } j(i) > 0 \text{ when } i \leq M_f, \\ z_{i_0} & \text{where } i_0 \text{ is such that } j(i_0) = 0 \text{ and } \psi_0(z_{i_0}) = a_{M_f} \text{ (i.e. } \psi_0(z_{i_0}) \leq \psi_0(z_s) \text{ for all } s \text{ with } j(s) = 0 \text{) otherwise.} \end{cases}$$

Here  $M_f$  is the iteration number of the last of the working stages.

## 9.2.4

**Theorem.** (i) The method described, constructed for an accuracy  $\epsilon > 0$ , solves with this accuracy all problems in  $\bar{C}(G, R^n, m)$  with a laboriousness not exceeding

$$\Phi_{n, \infty}^0 \left( \frac{v(f, \epsilon)}{1 + v(f, \epsilon)} \right)$$

(here

$$\Phi_{n, \infty}^0(v) = \left[ cn^7 \ln n \ln \frac{4n\beta}{v} \right],$$

$c$  is an absolute constant, and

$$v(f, \epsilon) = \min_{0 \leq j \leq m} \frac{\varepsilon_j}{r_j(f)}$$

is the maximal relative error which ensures the given absolute error).

(ii) Let  $\beta \leq n(\bar{W}_0)$  and  $W_0$  such that this inequality holds can always be found). There is a polynomial  $P(n)$  which has the following property. Let the method described be applied to solving problems in the class  $\bar{\mathcal{C}}^\delta(G, R^n, m)$  obtained from  $\bar{\mathcal{C}}(G, R^n, m)$  by replacing the exact zeroth-order oracle by an approximate oracle

$$(\psi_0(f, x), \dots, \psi_m(f, x)) : \{f\} \times G \rightarrow R^{m+1}$$

with an absolute error  $\delta$  (the last phrase means

$$|\psi_j(f, x) - f_j(x)| \leq \delta_j \text{ for all } f \in \bar{\mathcal{C}}(G, R^n, m), x \in G \text{ and } j \in \{1, \dots, m\}$$

Suppose, further, that the required absolute error  $\epsilon$  of solutions of problems from  $\bar{\mathcal{C}}^\delta$  satisfies the condition  $\epsilon_j \geq P(n)\delta_j$ ,  $0 \leq j \leq m$ . Then the method described, constructed for an accuracy  $\epsilon$ , solves all problems in  $\bar{\mathcal{C}}^\delta$  with an error  $\leq 2\epsilon$  and with laboriousness

$$\leq \Phi_{n, \infty}^0 \left( \frac{v(f, \epsilon)}{1 + 2v(f, \epsilon)} \right).$$

Let us discuss the assertions of the theorem. We have constructed a deterministic method of solving general convex problems using a zeroth-order oracle and having the following property as regards efficiency: asymptotically (as  $v(f, \epsilon) \rightarrow 0$ ) its laboriousness coincides with the complexity of the class of problems considered (up to a factor which is bounded above by a polynomial in the dimension of the problem). The method is noise-stable in the following sense: it is in a position to solve a problem with an absolute error  $\epsilon > 0$  if the errors in the current information used by it do not exceed values proportional to  $\epsilon$ . It is true that the coefficient of proportionality—which is the index of the method's noise-stability—tends to 0 as the dimensionality of the problem increases. In Section 9.4 we shall see that, qualitatively, the latter effect is unavoidable.

Of course, the laboriousness of the method described depends strongly on  $n$ , and so, really, the method cannot be recommended for any appreciable values of  $n$ .

*Exercise 2.* Prove the theorem.

*Exercise 3.* Construct an analogue of the method described (and prove for it an appropriate version of the theorem) in the case where unconstrained problems ( $m = 0$ ) are to be solved with a specified relative error  $v > 0$ . Try to extend this result to the case of constrained problems (in the case of an approximate oracle, the authors did not succeed in obtaining the required extension, at least, not in the time they were prepared to spend on this question).

### 9.3 METHODS OF ZEROTH ORDER: STOCHASTIC ORACLE. I

In the preceding sections the case has been examined where exact information (or information distorted only by small errors) about the values of the functionals in the problem being solved was available to the method. We now consider a situation where this information is perturbed by random (and, in general, not small) noise, although the mean noise is equal to 0, and the noises in different cycles are independent.

#### 9.3.1

We begin by describing the class of problems considered. Let  $G \subset R^n$  be a convex, bounded, closed body, let  $G_I \supset G$  containing  $G$  be a body of the same type, and let  $\mathcal{F}_m(G_I)$  be the set of all  $(m+1)$ -dimensional, component-wise continuous, convex vector-functions on  $G_I$ . With each function  $f = (f_0, \dots, f_m)$  of this kind we associate the *problem*  $f$ :

$$f_0(x) \rightarrow \min |x \in G, f_j(x) \leq 0, 1 \leq j \leq m.$$

We shall suppose that problems from  $\mathcal{F}_m(G_I)$  are observed by means of a stochastic oracle  $\ell = ((\Omega, \mathbf{F}_\omega); \psi(x, f, \omega))$  with an observation function  $\psi(x, f, \omega) : \mathcal{F}_m(G_I) \times \Omega \times G_I \rightarrow R^{m+1}$ . The components  $\psi_0, \dots, \psi_m$  of the vector-function  $\psi$  are interpreted as observations of the values of the corresponding components of  $f$ . As always,  $(\Omega, \mathbf{F}_\omega)$  is assumed to be a Polish space with a regular, Borel, Lebesgue-complete, probability measure  $\mathbf{F}_\omega$ , and  $\psi(x, f, \omega)$  is assumed to be a Borel function with respect to  $\omega$  and  $x$ .

In order to be able to work with the oracle  $\ell$ , we must, as in the case of a first-order oracle, adopt definite hypotheses about the intensity of the noise and possible limits to the bias of the oracle  $\ell$ . We shall do this in the following way. Let  $L_0, \dots, L_m > 0$  and  $v_0 \geq 0$ ; also let  $r > 1$ . Let  $\bar{\mathcal{C}}_{r, v_0}^c(G, L_0, \dots, L_m, G_I)$  denote the set of all problems  $f$  from  $\mathcal{F}_m(G_I)$  for which, for all  $x \in G_I$ ,

$$(1) \quad \mathbf{M}_{\mathbf{F}_\omega} \max_{0 \leq j \leq m} \left( \frac{|\psi_j(x, f, \omega)|}{L_j} \right)^r \leq 1;$$

$$(2) \quad \max_{0 \leq j \leq m} \frac{|\mathbf{M}_{\mathbf{F}_\omega} \psi_j(x, f, \omega) - f_j(x)|}{L_j} \leq v_0.$$

Roughly speaking, as regards meaning, the  $L_j$  are *a priori* estimates of the (absolute)  $r$ th moments of the oracle's answers about the values of  $f_j$ , and  $v_0$  is an *a priori* estimate of the (relative) bias of the oracle  $\ell$ . As the normalizing factors for the class introduced, we select the parameters  $L_j$  themselves.

In examining methods of solving problems of this class, we shall assume the parameters  $r, v_0, L_j$  to have been specified *a priori*. It is clear that, by a simple normalization of the components in the problems being

solved, each class  $\bar{C}_{r,v_0}^\ell(G, L_0, \dots, L_m, G_l)$  can be reduced to a certain class  $\bar{C}_{r,v_0}^\ell(G, 1, \dots, 1, G_l) \equiv \bar{C}_{r,v_0}^\ell(G, G_l)$ . It is this normalized class of problems that we shall consider below.

### 9.3.2

We shall describe somewhat cursorily two approaches for solving problems of the class  $\bar{C}_{r,v_0}^\ell(G, G_l)$ . The first is based on the idea of imitating, for the problems considered, a stochastic oracle of the first order by the available stochastic oracle of the zeroth order, followed by use of methods of the stochastic approximation type (Chapters 5–6). In the second approach (discussed in the next section), roughly speaking, from the zeroth-order stochastic oracle a zeroth-order deterministic oracle is constructed, and then zeroth-order methods designed for an (approximate) deterministic oracle are used.

We start with the first approach, in which, from a zeroth-order stochastic oracle, a first-order oracle—still stochastic—is constructed. The idea of this construction has been used more than once in different versions of random descent methods. We shall present its general features without going into particular details. We fix a Euclidean structure in  $R^n$ , and let  $d$  be the diameter, and  $\alpha$  the asphericity, of  $G$  in the corresponding metric. We shall assume that  $f$  is a continuous, convex function on  $G_l$ , and  $\hat{\theta}$  is an *unbiased*, zeroth-order oracle. Let  $\rho > 0$  be a parameter. We consider the function  $f^\rho$  obtained from  $f$  by averaging:

$$f^\rho(x) = \frac{1}{|V_\rho|} \int_{\|\xi\| \leq \rho} f(x + \xi) d\xi, \quad (3.1)$$

where  $V_\rho = \{\xi | \|\xi\| \leq \rho\}_n$ , and  $|\cdot|_n$  is an  $n$ -dimensional Lebesgue volume. The function  $f^\rho(x)$  is defined, not over the whole domain  $G_l$  of the function  $f$ , but over a slightly smaller domain  $G_l^\rho$ , consisting of all points  $x \in G_l$  whose distance up to the boundary of  $G$  is not less than  $\rho$ . In particular, if  $\rho/d$  is sufficiently small, then  $f^\rho$  is defined in ‘a large part’ of  $G$ ; but if  $G_l$  contains a  $\rho$ -neighbourhood of  $G$ , then  $f^\rho$  is defined on the whole of  $G$ .

*Exercise 1.* Let  $G$  be contained in a ball of radius  $R$  with centre at 0 and let  $G$  contain a concentric ball of radius  $R/(2\alpha)$ . Prove that  $f^\rho$  is defined in the set  $\theta G$ , where  $\theta = 1 - 2\rho\alpha/R$ .

Since  $f$  is continuous,  $f^\rho$  can be made uniformly (over the whole of  $G_l^\rho$ ), arbitrarily close to  $f$  by choosing  $\rho > 0$  sufficiently small. Finally,  $G_l^\rho$  is convex, and  $f^\rho$  is convex in  $G_l^\rho$ . Moreover, it is easy to see that  $f^\rho$  is continuously differentiable in  $G_l^\rho$ , and its gradient is found from the formula

$$\nabla f^\rho(x) = \frac{1}{|V_\rho|} \int_{\|\xi\| = \rho} f(x + \xi) \frac{\xi}{\|\xi\|} ds_\rho(\xi) \quad (3.2)$$

(here  $ds_\rho(\xi)$  is an element of a spherical surface of radius  $\rho$  in  $R^n$ ). Passing from integration over the area of a sphere to integration with respect to the normalized area (the area of the whole sphere is taken to be 1), i.e. integration with respect to a uniformly distributed probability  $d\sigma(\xi)$  on  $s_1$ , we obtain

$$\nabla f^\rho(x) = \frac{n}{\rho} \int_{\|\xi\|=1} \xi f(x + \rho\xi) d\sigma(\xi). \quad (3.3)$$

*Exercise 2.* Prove (3.2)–(3.3).

Thus the gradient of  $f^\rho$  in  $G_l^\rho$  can be expressed in the form of an integral of values of  $f$  with a certain (vector) weight.

We are now in a position to describe the idea of the approach mentioned. It consists in the following.

1°. We choose, on the basis of the *a priori* information, the parameter  $\rho$  so small that the domain  $G_l^\rho$  of the problem  $(f_0^\rho, \dots, f_m^\rho)$  approximates  $G$  sufficiently well and the functions  $f_j^\rho$  themselves approximate the  $f_j$  sufficiently well in this domain  $G_l^\rho$ . More precisely, if  $v$  is the required accuracy of solution of the problem, then we choose  $\rho > 0$  and the numbers  $\alpha_j$ ,  $1 \leq j \leq m$ , such that, for any point  $x \in G_l^\rho \cap G$ , the relation

$$v(x, f) \leq v(x, \tilde{f}^\rho) + \frac{v}{2}$$

is satisfied, where

$$\tilde{f}^\rho: f_0^\rho \rightarrow \min |x \in G_l^\rho \cap G, f_j^\rho(x) \leq \alpha_j, 1 \leq j \leq m,$$

and  $v$  is the required accuracy of solution of the problem  $f$  (we suppose that the normalizing factors for  $\tilde{f}^\rho$ , as also for  $f$ , are unities).

We shall discuss later how to choose a suitable  $\rho$ .

2°. Having chosen  $\rho$  and  $\alpha_j$  as indicated, we shall reduce the problem of solving  $f$  with a (mean) error  $v$  to the problem of solving  $\tilde{f}^\rho$  with a mean error  $v/2$ .

We shall indicate a way of solving the latter problem. We shall model a first-order oracle for  $\tilde{f}^\rho$  in the following way. We select at random (relative to a uniform measure in  $V_\rho$ ) a point  $\xi \in V_\rho$ , and put to  $\hat{\theta}$  a question about  $f$  at the point  $\xi + x$ . The answer obtained is adopted as an observation of  $f^\rho(x)$ ; we denote it by  $a(\xi, \omega) = (a_0(\xi, \omega), \dots, a_m(\xi, \omega))$ , where  $\omega$  is the oracle noise in the corresponding cycle. Then we select at random (relative to the measure  $d\sigma$ ) a point  $\bar{\xi}$  on the spherical surface  $s_1 = \{\xi | \|\xi\| = 1\}$  and put a question to the oracle  $\hat{\theta}$  about  $f$  at the point  $x + \rho\bar{\xi}$ . Let  $b_0, \dots, b_m$  be the numbers communicated by the oracle. We form the vectors  $(n/\rho)b_j\bar{\xi}$  and regard them as observations of the gradients of the  $f_j^\rho$  at the point  $x$ . We denote them by  $b_j(\bar{\xi}, \omega')$ , where  $\omega'$  is the noise of the oracle  $\hat{\theta}$  in the corresponding cycle. We

shall prove that the oracle imitated in this way is indeed an unbiased, first-order oracle for the problem  $\tilde{f}^\rho$ . For,

$$\mathbf{M}a = \frac{1}{|V_\rho|_n} \int \mathbf{M}_{F_\omega} a(\xi, \omega) d\xi = \frac{1}{|V_\rho|_n} \int f(x + \xi) d\xi = f^\rho(x)$$

(here  $f^\rho$  denotes the vector-function  $(f_0^\rho, \dots, f_m^\rho)$ ). In the same way,

$$\mathbf{M}b_j = \frac{n}{\rho} \int_{\|\xi\|=1} d\sigma(\xi) \mathbf{M}_{F_\omega} a_j(\rho \xi, \omega') \xi = \frac{n}{\rho} \int_{\|\xi\|=1} d\sigma(\xi) f_j(x + \rho \xi) \xi = \nabla f_j^\rho(x).$$

Let us estimate the noise intensity of the oracle obtained in this way. Because of the properties of  $\hat{\ell}$ , we have

$$\mathbf{M}\|a\|_\infty = \frac{1}{|V_\rho|_n} \int \mathbf{M}_{F_\omega} \|a(\xi, \omega)\|_\infty d\xi \leq 1 \quad (3.4)$$

and

$$\mathbf{M}\|b_j\|^r = \left(\frac{n}{\rho}\right)^r \int_{\|\xi\|=1} d\sigma(\xi) \mathbf{M}_{F_\omega} |a_j(\rho \xi, \omega')|^r \leq \left(\frac{n}{\rho}\right)^r. \quad (3.5)$$

Thus, for a given  $\rho$ , the noise level of the modelled first-order oracle can be estimated explicitly (a consultation of this oracle ‘costs’ two consultations of the oracle  $\hat{\ell}$ ).

We are now in a position to solve the problem  $\tilde{f}^\rho$ , using the constructed first-order oracle, by first-order methods (designed for a stochastic oracle).

*Remark.* It is clear that the way just described of imitating a first-order stochastic oracle by means of a zeroth-order oracle is applicable not only to the ordinary problems of mathematical programming, but also to extremal problems with operator constraints (see Section 5.5). We leave it to the reader to formulate and prove the corresponding propositions.

### 9.3.3

We shall go into certain difficulties connected with the approach described. We see that for its implementation we must know how to choose constructively both  $\rho$  and the  $\alpha_j$  so that the problem  $\tilde{f}^\rho$  will approximate the problem  $f$  with the specified accuracy. Moreover, the smaller the  $\rho$  which we choose, the worse (as regards noise-intensity) will the constructed first-order oracle be (see (3.5)), and therefore the greater the laboriousness of a first-order method using this oracle will be for specified requirements regarding the accuracy of solution of the problem (the *relative accuracy*  $v$  of solution of the original problem is converted, roughly speaking, into an *absolute accuracy* required of the first-order method).

On the other hand, it is impossible for the parameter  $\rho$  to be other than small. For, even if we divert our attention from the effects connected with the

fact that  $G_i^\rho$  must approximate well to  $G$ , it is necessary that  $f_i^\rho$  shall well approximate (uniformly) to  $f_i$  in  $G_i^\rho$ , namely, with an accuracy of the order of  $v$ . Suppose, let us say, that  $G_i$  contains a sufficiently large neighbourhood of  $G$  so that we may suppose  $G_i^\rho \supset G$ . It is clear that if  $\rho$  is chosen so that  $|f_i^\rho - f_i| \leq v/4$  in  $G$  and if  $\alpha_j = v/4$ , then  $\tilde{f}^\rho$  provides the required approximation to  $f$ .

Of course, the choice of  $\rho$  from these conditions is sufficient for our purposes, but it is by no means necessary. For a given  $f$ , we might ‘by chance’ obtain an  $\tilde{f}^\rho$  of interest to us even with much greater values of the parameters. However, this really could happen only fortuitously. If it is required to choose  $\rho$  so that for every problem undergoing solution we could guarantee an acceptable approximation of  $f$  by means of  $\tilde{f}^\rho$ , then it is impossible, it would seem, to operate better than by choosing  $\rho$  from the considerations that the  $f_i^\rho$  are obliged to approximate the  $f_i$  uniformly in  $G_i^\rho$  with an accuracy of the order of  $v$ .

We now consider how  $\rho$  must be chosen so as to ensure the relations

$$|f_i^\rho - f_i| \leq av \text{ in } G_i^\rho, \quad 0 \leq j \leq m. \quad (3.6)$$

The corresponding choice of  $\rho$  depends on the *a priori* information about the smoothness of the  $f_i$ . Suppose it is known, let us say, that  $G_i$  contains a  $\bar{\rho}$ -neighbourhood of  $G$  and that the  $f_i$  are Lipschitz functions with constant  $L$  in a  $\rho_0$ -neighbourhood of  $G$  (we assume  $\rho_0 \leq \bar{\rho}$ ). We have no other information about the smoothness of the  $f_i$ . Then (3.6) is ensured by choosing  $\rho$  in the form

$$\rho_v = \frac{av}{L} \quad (3.7)$$

(we assume  $v$  to be small enough to have  $av/L \leq \rho_0$ ). Generally speaking, a choice of  $\rho$  greater than  $\rho_v$  will no longer ensure (3.6).

Suppose, further, that  $r = 2$  (we make this stipulation for the sake of greater clarity in the following argument). Combining the bound for laboriousness of the first-order method corresponding to this  $r$  (and in a Hilbert space  $E$ ), constructed for an absolute error of order  $v$ , with the estimate (3.4)–(3.5) of the noise-intensity level in the first-order oracle constructed earlier, we obtain that the laboriousness of the method will be

$$\geq d_1 \left( \frac{nd}{\rho_v v} \right)^2 \times d_2 \left( \frac{nLd}{v^2} \right)^2$$

(here and later the  $d_i$  depend on  $a$  and  $m$ , but not on  $n, L, d, v$ ). Thus in this case the asymptotic formula for the laboriousness as  $v \rightarrow 0$  is  $O(1/v^4)$ , and not  $O(1/v^2)$  as it was for first-order methods with a stochastic oracle. Apart from this unwelcome phenomenon, we see that, for the approach to be successful, we still need to know how to estimate  $L$ . If  $G_i$ , as above, contains a  $\bar{\rho}$ -neighbourhood of  $G$ , then, in a  $\bar{\rho}/2$ -neighbourhood of  $G$ , the  $f_i$  are Lipschitz

with a constant  $\leq c/\bar{\rho}$  (where  $c$  is an absolute constant), because  $|f_j|$  does not exceed 1 in  $G_1$  by virtue of the properties of the oracle  $\hat{\mathcal{O}}$ . Thus the laboriousness of a method obtained as a result of the implementation of the approach considered is not less than a quantity of the form  $d_3(n^2/v^4)(d/\bar{\rho})^2$ . In the most natural case,  $d/\bar{\rho} = O(1)$  and the estimate takes the form  $d_4 n^2/v^4$ .

Of course, if something more is known about the smoothness of the  $f_j$ , then the estimate can be improved. Suppose, let us say,  $G_1$  contains an  $O(d)$ -neighbourhood of  $G$ , and it is known that the  $f'_j(x)$  exist and are Lipschitz in  $G_1$  with constants  $O(1)$ . Then (3.6) is ensured by a choice of  $\rho$  of the form  $O(\sqrt{v})$ , and the laboriousness of the resulting method will be bounded below by a quantity of the form  $d_5 n^2/v^3$ . This last bound does not admit any further qualitative improvement, at least, not without extremely severe and hardly natural restrictions on the  $f_j$ .

In the cases considered,  $G_1$  contained a neighbourhood of  $G$ . If this is not the case (if, say,  $G_1 = G$ ), then the situation deteriorates still further (boundary effects begin to come into play). No problems in principle arise, but the laboriousness of the method obtained increases appreciably. When  $G = G_1$  for the whole class  $C_{2,0}^\ell(G, G_1)$  it is not possible in this way to make the laboriousness less than  $d_6 n^2/(v^6 \alpha^2)$ .

Thus even in the most favourable cases, the approach examined is not in a position to ensure a laboriousness which, as regards the order of its dependence on the accuracy, is better than  $O(1/v^3)$ . The dependence of the laboriousness on the dimension turns out, it is true, to be comparatively acceptable (with increase in  $n$  the laboriousness increases like  $(n \ln n)^2$ ).

We have examined the case where the original class was equipped with an unbiased oracle. Of course, in solving problems with accuracy  $v > 0$ , it would be possible to allow a certain non-zero bias  $v_0$ ; however, a simple analysis shows that the admissible  $v_0$  must be of a lower order than  $v$ —in the most favourable cases,  $v_0$  must be  $\leq O(v^{3/2})$  as  $v \rightarrow 0$ . Thus, the methods obtained by the described approach are inordinately demanding on the biases in the oracle (cf. this situation with the one analysed in the previous section when we examined the method obtained by imitating a first-order oracle by means of a zeroth-order oracle).

## 9.4 METHODS OF ZEROTH ORDER: STOCHASTIC ORACLE. II

### 9.4.1

The idea of reducing zeroth-order methods to first-order methods, on which the approach described in Section 9.3 was based, is extremely natural and popular. As we saw, it leads to methods whose laboriousness depends in an acceptable way on the dimension of the problem but increases quickly as the

required accuracy increases (the laboriousness increases substantially more quickly than it does in first-order methods using a stochastic oracle). The naturalness of the idea might suggest that this rapid growth of laboriousness with increased accuracy is in the natural order of things, and that it is not a defect of the approach. It turns out, however, that this supposition is incorrect: another way can be offered for constructing zeroth-order methods working with a stochastic oracle which ensure an accuracy  $v$  in the solution of problems in the classes  $C_{r,0}^\ell(G, G_1)$  with laboriousness not exceeding the quantities

$$\left\lceil \ln^2 \frac{I}{v} \left[ \ln(m+2) R_r(n)/v^{\bar{r}-1}, \right. \right.$$

where  $\bar{r} = \min(2, r)$ , and  $R_r(n)$  is some polynomial in  $n$ . We emphasize that this bound, for values of  $v$  less than a certain absolute constant  $v_* > 0$ , coincides, up to a factor of order  $O(\lceil \ln^2 1/v \rceil R_r(n))$ , with the *lower* bound for the complexity of classes of problems of convex stochastic approximation with a first-order oracle which were discussed in Section 5.3; and so, for a fixed  $n$  and for all  $v \leq v_*$  and for all  $m$ , the methods obtained cannot be substantially improved as regards laboriousness (namely, by not more than  $O(\ln^2 1/v)$  times).

Thus, methods of the type mentioned in Section 7.3 are indeed objectively bad asymptotically as  $v \rightarrow 0$ . The method which will be constructed below substantially excels them in the character of its asymptotic behaviour (moreover, the asymptotic behaviour of the laboriousness of this method as  $v \rightarrow +0$  cannot, in principle, be improved by more than a logarithmic factor). On the other hand, the polynomial  $R_r(n)$  is of quite a high degree (certainly more than 2), and so the new method is much worse than the former one as regards the nature of the dependence of the laboriousness on the dimensionality of the problem.

### 9.4.2

The idea of the method which we are about to construct is extremely simple. We have at our disposal a stochastic zeroth-order oracle (suppose, to begin with, that it is exact:  $v_0 = 0$ ). At the same time we know how to solve convex problems with an approximate, zeroth-order oracle, provided that its (absolute) error is sufficiently small compared with the required accuracy of solution. Namely, in considering problems in the class  $C_{r,0}^\ell(G, G)$  and with  $v$  as the required accuracy of solution of these problems it would be sufficient for us to know how to calculate the values of the functionals in the problem with absolute errors not exceeding  $v/(4P(n))$ , where  $P(n)$  is the polynomial in Theorem 9.2.4. If we succeeded in doing this, we would be able to solve the problem by the method in Section 9.2, constructed for the absolute errors

$(v/4, \dots, v/4)$ . The laboriousness of this method is

$$M(n, v) = S(n) \ln \frac{1}{v} [,$$

where  $S(n)$  is a certain polynomial in  $n$  (we have used the fact that the variation of the components  $f_j$  of a problem  $f \in \bar{C}_{r,0}^{\ell}(G, G)$  on  $G$  does not exceed 2).

It remains to learn how to calculate the values of the functionals in the problem with errors not exceeding  $v/(4P(n))$ . To do this, we proceed as follows: we observe the vector  $f(x)$  at the point  $x \in G$  in which we are interested a sufficiently large number of times  $Q$ , and we average in the proper way the obtained (independent) observations of  $f(x)$ . Of course, whatever number  $Q$  we may choose, there will, in general, be a certain positive probability that the results of the averaging will not give the required approximation for  $f(x)$ . By choosing  $Q$  sufficiently large, we shall be able to make this probability so small that similar 'nuisances' will not exert an appreciable influence on the mean error of the method.

More precisely, having at our disposition the oracle  $\mathcal{O}$  enables us, by making  $Q$  observations of the values of the components  $f_j$  of the problem  $f \in \bar{C}_{r,0}^{\ell}(G, G)$  being solved at an arbitrary point  $x \in G$ , to obtain  $Q$  independent realizations  $\xi_1, \dots, \xi_Q$  of a random variable  $\xi$  such that  $\mathbf{M}\xi = f(x)$  and  $\mathbf{M}\|\xi - f(x)\|_{\infty} \leq 2$ . Using the method described in Section 6.6.4 for processing the observations, we shall be able to construct an estimate  $\xi^Q$  for the vector  $f(x)$  such that

$$\Pr \left\{ \|\xi^Q - f(x)\|_{\infty} > \frac{v}{4P(n)} \right\} \leq \alpha,$$

if

$$Q \geq Q_r(n, v, \alpha) = \left[ \frac{W_r(n)}{v^{r/(r-1)}} \ln \frac{1}{\alpha} \ln(m+2) \right],$$

where  $W_r(n)$  is a certain polynomial in  $n$ . We then choose

$$\alpha = \alpha_v = v/8M(n, v) \text{ and } Q = Q(n, v, \alpha_v),$$

and then proceed in the way described above (i.e. we apply, for solving the problem  $f$ , the method  $\mathcal{B}$  in Section 9.2, constructed for the absolute errors  $(v/4, \dots, v/4)$ ). We shall obtain the information needed for this method by making  $Q$  observations of the vector  $f(x)$  at each of the 'interrogated' points and by supplying to the method the vector  $\xi^Q$  obtained as a result of the processing just described of the observations obtained).

It is clear that during the course of the work of the method, there will, with a probability

$$\geq 1 - M(n, v)\alpha_v^4 = 1 - \frac{v}{8},$$

be obtained an approximate solution of the problem  $f$  with an error  $\leq 3v/2$  (this will certainly occur if, in all the  $M(n, v)$  steps of the work of the method  $\mathcal{B}$ , the information supplied to the method does not differ component-wise from the true values by more than  $v/(4P(n))$ ). With additional probability the absolute error of the result will certainly not exceed 2, because  $|f_j| \leq 2$  on  $G$ . The mean error of the method is thus not greater than

$$\frac{v}{2} + 2 \cdot \frac{v}{4} \leq v,$$

whereas its laboriousness does not exceed

$$Q_r(n, v, \alpha_v)M(n, v) \leq R_r(n) \left[ \ln^2 \frac{1}{v} \right] / v^{r/(r-1)},$$

where  $R_r(n)$  is a polynomial in  $n$ . This is the result promised at the beginning of the section. It is also clear from the argument that the requirement  $v_0 = 0$  was not too important. It would be sufficient if an inequality of the type  $v_0 \leq \bar{P}_r^{-1}(n)v$  were satisfied,  $\bar{P}_r(n)$  being a suitable polynomial in  $n$ .

Thus the method described is (from the theoretical point of view) less sensitive to bias in the oracle (asymptotically as  $v \rightarrow 0$ ) than the method in Section 9.3.

#### 9.4.3

Let us draw some conclusions. The results presented above raise far more questions than they answer. Indeed, not one of the methods described earlier can lay claim to practical applicability except in the simplest situations (very rough accuracy or very small dimensionality). To be acceptable from a practical point of view, a method must, firstly, have a bound for laboriousness which does not increase too rapidly with growth in the dimensionality of the problem and increase in the accuracy demanded of the solution and, secondly, must not be too sensitive to errors in the information about the problem.

In speaking of errors in the information, we do not include under that heading 'regularly structured' errors (i.e. random errors with zero mean, which are independent from one step to the next; we know how to cope with errors of that sort, our methods being specially orientated towards them). We are speaking of errors regarding whose nature we cannot rightly make any *a priori* supposition whatsoever, except the natural hypothesis that they are small compared with the accuracy demanded of the solution. In Section 9.2 only such errors were considered, and in Sections 9.3–9.4 the biases in the oracle play the rôle of such errors. Definite hypotheses about the smallness of errors of this kind are obviously necessary: if the errors exceeded the required accuracy of solution of the problem, then it is readily understood that it would, in general, be impossible to solve the problem to that accuracy.

It is natural to measure the noise stability of any method by the ratio of the error-level, admissible in the limit for it, to the required accuracy of solution ('admissible' here is to be understood in the sense that this error-level does not lead to an essential loss of accuracy in the solution). This verbal definition of noise-stability is rather vague, but we shall not encumber this exposition with precise definitions. It is obvious that methods with low noise-stability cannot, as a rule, be used in practice.

How do things stand with the laboriousness and noise-stability of the zeroth-order methods described above? Unfortunately, rather badly. The laboriousness of the methods in Sections 9.2 and 9.4 increases very rapidly with increase in dimensionality of the problem, whereas the laboriousness of the method in Section 9.3 increases too rapidly with increase in the required accuracy. The methods constructed in accordance with the scheme of Section 9.1 do not, it would seem, have this drawback; but their noise-stability is extremely low (the level of information errors for them must be of the order of one-fourth power of the required accuracy). In general, the noise-stability of the methods considered leaves much to be desired: even the theoretically stable (asymptotically as  $v \rightarrow 0$ ) methods in Sections 9.2 and 9.4 have a noise-stability which falls rapidly with growth in the dimensionality of the problem.

It could be supposed that these difficulties were not defects in the methods studied, but rather were objectively inherent in all zeroth-order methods. However, we have no facts to support such an assertion; rather, there are reasons for asserting the contrary. For, in the case of a deterministic oracle we have the unstable methods of Section 9.1 with an acceptable (asymptotically as  $n \rightarrow \infty$  and as  $v \rightarrow 0$ ) laboriousness (in every case, rough estimates of the complexity of the corresponding class of problems with a zeroth-order oracle show that it is impossible in principle to reduce particularly strongly the laboriousness of these methods). On the other hand, we have the qualitatively more stable method of Section 9.2, the laboriousness of which increases unacceptably quickly with growth of  $n$ . It is natural to suppose that there must be methods which unite the laboriousness estimates of the methods in Section 9.1 with stability properties at least of the same kind as those of the method in Section 9.2. It would even be desirable to construct methods even more noise-stable than the one in Section 9.2. The latter is stable on the theoretical plane rather than on the practical plane; its noise-stability falls rapidly with growth in dimensionality. In this respect the ideal situation occurs with the first-order methods of solving general convex or Lipschitz convex problems in Chapters 2, 3, and 6. The noise-stability of these methods does not depend on the dimensionality or on any other of the parameters of the classes of the type indicated.

Unfortunately, such an ideal situation is impossible for zeroth-order methods: their noise-stability necessarily falls with increase in dimensionality (for all sufficiently wide classes of problems), provided that the laboriousness

of the methods remains within acceptable bounds. The proof of this last assertion is contained in the exercise at the end of this section. We notice, it is true, that the rate of decrease of the noise-stability with increase in dimensionality given by it is much lower than that which the method in Section 9.2 can realize.

In the stochastic situation of Sections 9.3–9.4 matters are roughly the same. The methods in Section 9.3 have a laboriousness which is acceptable as regards the nature of its dependence on  $n$  and not acceptable as regards the nature of its dependence on  $v$ ; and for the methods of Section 9.4 the opposite is true. It is natural to suppose that there is a method whose laboriousness in the nature of its dependence on  $n$  is at least the same as that of the method in Section 9.3, and in the nature of its dependence on  $v$  is the same as that of the method in Section 9.4.

Thus the results obtained do not so much solve the problem of the choice of convex optimization methods with a zeroth-order oracle as show what should be expected of a satisfactory solution of that problem; and this, properly speaking, is their only significance. Each of the methods suggested is in some respect unimprovable, but bad in other respects. We have not succeeded in combining their good qualities and eliminating the bad. Whether it is possible to do this, and—if so—how, we do not know. Thus, both in the theoretical, and even more in the practical, aspect the situation with a zeroth-order oracle is far from clear.

For comparison we point out that, in the case of a first-order oracle, matters stand much better (at least from the theoretical point of view). In this case we can, as a rule, point to methods having an upper bound for laboriousness which is of the same order of magnitude as the lower bound for complexity of the corresponding classes. The situation is also comparatively favourable as regards the stability of these methods relative to errors in the information about the problem. The classes of general convex and Lipschitz convex problems in Chapters 2, 3, 5, and 6 were, from the very beginning, considered for approximate deterministic (or biased stochastic) oracles. Moreover, a relative level of errors (measured there by a quantity  $v_0$ ) proportional to the required relative accuracy  $v$  (say, twice† smaller than the latter) did not lead to a loss of stability: a method intended for an exact oracle and accuracy  $v/2$  still ensured an accuracy  $v$  with the approximate oracle mentioned. The noise-stability (according to a natural definition of it) of these methods is therefore an absolute constant.

In the discussion of classes of strictly convex problems (Chapter 7) the situation regarding noise-stability was rather worse. We there assumed the oracle to be exact: but analysis of the methods constructed shows that a certain

† In Chapter 6 things are rather different: 'twice' is no longer enough, the constant has to be increased.

non-zero level of errors in the information can be tolerated. For a method (derived from those constructed in Chapter 7) for solving unconstrained strongly convex problems in the whole space, this level (with a natural normalization) turns out to be of the order  $v/D(Q)$ , where  $v$  is the required relative accuracy of solution,  $Q$  is the modulus of strong convexity of the problem, and  $D(Q)$  is a suitable function of  $Q$ . The noise-stability proves to be smaller, the worse the conditionality of the problem (i.e. the greater  $Q$  is). This is quite natural: the entire specific character of the class considered is connected with the regular behaviour of the derivative of the function being minimized. It is clear that errors in the observations of the derivative must be sufficiently small, if we want something of this regularity to remain. Thus there is nothing strange in the dependence of the noise-stability on the conditionality of the problem.

#### 9.4.4

*Exercise 1.* Let  $A_n$  be the set of ‘linear’ problems

$$f(x) \equiv \langle \varphi | x \rangle \rightarrow \min \|x\| \leq 1, x \in E^n,$$

generated by vectors  $\varphi$  with  $\|\varphi\| \leq 1$ . We equip  $A_n$  with an oracle  $\mathcal{O}_\delta$  which serves only the points  $V_n = \{x | \|x\| \leq 1\}$  and is such that

$$\mathcal{O}_\delta(f, x) = \begin{cases} f(x), & |f(x)| > \delta, \\ 0, & |f(x)| \leq \delta. \end{cases}$$

Then  $\mathcal{O}_\delta$  is a zeroth-order oracle with the absolute error  $\delta$ . Suppose it is necessary to solve problems of the class  $A_n^\delta$  obtained by equipping  $A_n$  with the oracle  $\mathcal{O}_\delta$  and that it is required to ensure an accuracy  $\varepsilon$  in the solution of these problems. Prove that the laboriousness  $M$  of any (deterministic) method  $\mathcal{B}$ , ensuring the accuracy  $\varepsilon$  in the solution of problems of the class  $A_n^\delta$ , satisfies the estimate

$$M \geq c \exp \left\{ dn \left( \frac{\delta}{\varepsilon} \right)^2 \right\},$$

where  $c, d > 0$  are absolute constants. Hence deduce that for a power rate of growth of the laboriousness with increase in  $n$ , the relation  $\delta \leq O(\sqrt{\ln n}/n)\varepsilon$  must hold. Thus the noise-stability of zeroth-order methods having a laboriousness which does not increase too rapidly with increase in  $n$  must necessarily decrease as  $n$  increases.

Let  $x_1, \dots, x_M$  be the trajectory of  $\mathcal{B}$  on the problem  $f \equiv 0$ . Consider a measure  $\bar{\mu}$  concentrated on the points  $x_1, \dots, x_M$ . It is known that, for every probability measure  $\mu$  on the ball  $V_n$  and for every  $\kappa > 0$ , there is a unit vector  $\varphi_{\mu, \kappa}$  such that, when

$$\Pi_\kappa(\varphi) = \{x \in V_n | |\langle x | \varphi \rangle| > \kappa\}$$

we have

$$\mu(\Pi_\kappa(\varphi_{\mu, \kappa})) \leq \bar{c} \exp(dn\kappa^2),$$

$\bar{c}, d > 0$  being absolute constants. We put  $\kappa = \delta/(2\varepsilon)$  and  $\bar{\varphi} = \varphi_{\bar{\mu}, \kappa}$ . Then

$$\bar{\mu}(\Pi_\kappa(\bar{\varphi})) \leq \bar{c} \exp(-dn\kappa^2),$$

and when

$$\bar{c} \exp(-dn\kappa^2) < 1/M, \quad (4.1)$$

$\Pi_\kappa(\bar{\varphi})$  does not contain the points  $x_1, \dots, x_M$ . Suppose (4.1) holds. We consider the pair of problems

$$f_+(x) = 2\varepsilon \langle x | \bar{\varphi} \rangle \text{ and } f_-(x) = -2\varepsilon \langle x | \bar{\varphi} \rangle \in A_n.$$

Along the trajectory of  $\mathcal{B}$  on the problem  $f \equiv 0$  the information provided by the oracle  $\mathcal{O}_\delta$  about  $f_+$  and  $f_-$  is zero, for along this trajectory  $x_1, \dots, x_M$  we have, by construction,

$$|2\varepsilon \langle \bar{\varphi} | x_i \rangle| \leq 2\varepsilon\kappa = \delta.$$

Therefore  $x_M$  is the result of applying  $\mathcal{B}$  both to  $f_+$  and to  $f_-$ . But then, clearly, for one of these problems  $x_M$  is not an approximate solution with error  $\varepsilon$ , and this contradicts the definition of  $\mathcal{B}$ . Thus (4.1) is false, i.e.

$$M \geq c \exp \left( \frac{dn}{4} \left( \frac{\delta}{\varepsilon} \right)^2 \right),$$

as required.  $\square$

# Appendix: Mathematical supplement

In this auxiliary section we give a résumé of definitions and properties of some relatively special mathematical concepts which are used in the book, and we also explain some of the frequently used notation.

## A.1 POLISH SPACES, BOREL FUNCTIONS, MEASURES

A knowledge is assumed of the concepts of metric spaces (in particular, of compacts), open and closed sets, convergence, etc., as given in, say, Chapters I and II of [16].

### A.1.0 Notation relating to a metric.

If  $X$  is a metric space,  $G$  a subset to  $X$ , then  $\text{int } G$  denotes the *interior* of  $G$ , i.e. the set of points  $\{x \mid \exists \rho > 0: \{y \mid \rho(x, y) \leq \rho\} \subset G\}$ .  $\bar{G}$  denotes the *closure* of  $G$ ,  $\partial G$  the boundary of  $G$ , i.e. the set  $G \setminus \text{int } G$ .  $\rho(x, G)$  denotes the distance from  $x$  to  $G$ , i.e. the number

$$\rho(x, G) = \inf \{\rho(x, y) \mid y \in G\}.$$

$\rho(G)$ , the *radius* of  $G$ , is the lower bound of the radii of balls whose centres are in  $G$  and which are contained in  $G$ :

$$\rho(G) = \inf \{r \mid \exists x \in G: \forall y \in G \rho(x, y) \leq r\}.$$

The *diameter* of  $G$  is the number

$$d(G) = \sup \{\rho(x, y) \mid x, y \in G\}.$$

The *asphericity* of  $G$  is, roughly speaking, the ratio of the radii of the minimum ball containing  $G$  and the maximum ball contained in  $G$ :

$$\alpha(G) = \inf \{\alpha \mid \exists x, y, r: \{z \mid \rho(z, x) \leq r\} \subset G \subset \{z \mid \rho(z, y) \leq \alpha r\}\}.$$

### A.1.1 Polish spaces.

A metric space is called a *Polish space* if it is separable and complete. The class of all Polish spaces is denoted by  $\Pi$  and the clause ‘ $X$  is a Polish space’ is written as  $X \in \Pi$ . The metric on  $X$  is written  $\rho_X(\cdot, \cdot)$  (sometimes simply as  $\rho(\cdot, \cdot)$  if it is clear from the context which space  $X$  is involved). Examples of Polish spaces are: metric compacts, the spaces  $R^n$  with the usual metric, and separable Banach spaces (see Section A.2 below).

Other examples can be obtained by applying to a Polish space the operation of taking a (closed) subspace (a closed subspace of a Polish space is itself a Polish space) and the operation of *direct multiplication*. Let  $X_i \in \Pi$ ,  $i = 1, 2, \dots$ . Consider the set-theoretic product  $X^\infty = \prod_{i=1}^\infty X_i$  of the spaces  $X_i$ ; a point of  $X^\infty$  is a sequence  $x^\infty = (x_1, x_2, \dots)$ ,  $x_i \in X_i$ ). We can define a metric on  $X^\infty$  by

$$\rho_{X^\infty}(x^\infty, y^\infty) = \sum_1^\infty 2^{-i} \frac{\rho_{X_i}(x_i, y_i)}{1 + \rho_{X_i}(x_i, y_i)}.$$

Convergence in this metric is precisely ‘co-ordinate-wise’ convergence. When speaking of direct products later, we shall always have in view the description of a construction (of course, the idea is also applicable to a finite number of factors).

### A.1.2 Borel functions.

See [7, 11] in this connection.

#### A.1.2.1

Let  $X$  be a set,  $\Sigma$  a family of its subsets.  $\Sigma$  is called a  *$\sigma$ -algebra* if

- (1)  $X \in \Sigma$ ;
- (2)  $A \in \Sigma$  implies  $X \setminus A \in \Sigma$ ;
- (3)  $A_1, A_2, \dots \in \Sigma$  implies  $\cup_{i=1}^\infty A_i \in \Sigma$ .

Let  $\{\Sigma_\alpha\}_{\alpha \in J}$  be a family of  $\sigma$ -algebras of subsets of  $X$ . We consider the intersection of all the  $\Sigma_\alpha$  over  $\alpha \in J$ . This also is a certain family of subsets of  $X$ , and it too is obviously a  $\sigma$ -algebra. In particular, for any family  $\Sigma$  of subsets of  $X$  there is a least  $\sigma$ -algebra  $\bar{\Sigma}$  which contains  $\Sigma$  (namely, the intersection of all the  $\sigma$ -algebras which contain  $\Sigma$ ).  $\bar{\Sigma}$  is called the  *$\sigma$ -algebra spanned by  $\Sigma$* .

#### A.1.2.2

Now let  $X \in \Pi$ . A *Borel  $\sigma$ -algebra*  $B(X)$  of subsets of  $X$  is the  $\sigma$ -algebra spanned by all the closed subsets (or, equally, all the open subsets) of  $X$ . The elements of  $B(X)$  are called *Borel subsets* of  $X$ .

Let  $X, Y \in \Pi$  and let  $f: X \rightarrow Y$  be a mapping. The mapping  $f$  is a Borel mapping if the total pre-image  $f^{-1}(Y)$  of every open subset of  $Y$  (and then, of an arbitrary Borel subset of  $Y$ ) is a Borel subset of  $X$ . In particular, a continuous mapping of  $X$  into  $Y$  is a Borel mapping.

The following properties of Borel mapping easily follow from the definition.

**Propositions.** Let  $Y_i, X_i \in \Pi$ ,  $i = 0, 1, 2, \dots$

- (1) If  $f_0: X_0 \rightarrow X_1$  and  $f_1: X_1 \rightarrow X_2$  are Borel mappings, then so is the composition  $f_1 \circ f_0: X_0 \rightarrow X_2$  ( $f_1 \circ f_0(x) = f_1(f_0(x))$ ).
- (2) If  $f_i: X_0 \rightarrow X_1$ , and if, for all  $x \in X_0$ , the limit  $\lim_{i \rightarrow \infty} f_i(x) \equiv f(x)$  exists, then  $f: X_0 \rightarrow X_1$  is a Borel mapping.
- (3) If  $f_i: X_i \rightarrow Y_i$ ,  $1 \leq i < \infty$ , are Borel mappings, then  $f^\infty: \prod_{i=1}^{\infty} X_i \rightarrow \prod_{i=1}^{\infty} Y_i$ , defined as  $f^\infty(x_1, x_2, \dots) = (f_1(x_1), f_2(x_2), \dots)$ , is a Borel mapping.
- (4)  $f: X_0 \rightarrow \prod_{i=1}^{\infty} X_i$  is a Borel mapping if and only if all the mappings  $\pi_N \circ f: X_0 \rightarrow \prod_{i=1}^N X_i$ , where the  $\pi_N(x^\infty) = (x_1, \dots, x_N)$  are the natural projections of  $\prod_{i=1}^{\infty} X_i$  on to  $\prod_{i=1}^N X_i$ , are Borel mappings.
- (5) Let  $A_i \in B(X_0)$ ,  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , and  $\cup_{i=1}^{\infty} A_i = X_0$ . Suppose also that  $f_i: X_0 \rightarrow X_1$  is Borel. Consider the mapping  $f: X_0 \rightarrow X_1$  defined thus: the restriction of  $f$  to  $A_i$  coincides with the restriction of  $f_i$  to  $A_i$ ,  $1 \leq i < \infty$ . Then  $f$  is Borel.

The facts enumerated in these propositions show that the ‘natural’ operations on Borel functions do not take us out of the class of Borel functions. By means of them it is easy to show that all the rules constructed in the proofs of the theorems in Chapter 1 and the rules relating to the methods of Chapters 5 and 6 are Borel. We give a model of such an argument. Let  $X, Y, Z \in \Pi$  and let  $f(x, y): X \times Y \rightarrow Z$  and  $g: X \rightarrow Y$  be Borel mappings. We want to show that the function  $\bar{f}(x) = f(x, g(x))$  is also Borel. So, consider the sequence of mappings

$$\begin{aligned}\varphi_0: X &\rightarrow X \times X (\varphi_0(x) = (x, x)), \\ \varphi_1: X \times X &\rightarrow X \times Y (\varphi_1(x, x') = (x, g(x'))), f: X \times Y &\rightarrow Z.\end{aligned}$$

It is clear that they are all Borel ( $\varphi_0$  is continuous,  $\varphi_1$  by Proposition (3), and  $f$  by hypothesis). Therefore, by Proposition (1),  $f \circ \varphi_1 \circ \varphi_0 \equiv \bar{f}$  is also a Borel function.

#### A.1.2.3

Let us consider, in particular, scalar Borel functions. Let  $X \in \Pi$ , and let  $B(X, R)$  be the set of Borel functions on  $X$  with values on the real line. It is sometimes useful to characterize  $B(X, R)$  as the minimal function space  $\bar{B}$  which contains all continuous bounded functions (we denote this space by  $C_b(X)$ ) and which is closed relative to sequential pointwise convergence (i.e.

$B(X, R)$  is such that

$$f_i \in A, i = 1, 2, \dots \text{ and } \lim_{i \rightarrow \infty} f_i(x) \equiv f(x) \text{ for all } x$$

together imply that  $f(x) \in A$ . On the other hand,  $A \subset X$  is Borel if and only if the characteristic function of the set  $A$

$$\chi_A(x) = \begin{cases} 0, & x \notin A, \\ 1, & x \in A, \end{cases}$$

lies in  $B(X, R)$ . It is thus possible to give a new definition of Borel sets by defining them by means of the concept of a scalar Borel function (the latter, as we have just seen, can be defined independently of the concept of Borel sets).

#### A.1.2.4

We introduce another useful concept. Let  $T(X) \subset B(X, R)$  be a family of bounded Borel functions. Let  $\bar{T}(X)$  be the minimal set of functions which contains  $T(X)$  and is such that, if  $f_i \in \bar{T}(X)$ , if  $|f_i| \leq c$  for some  $c < \infty$ , and if  $f_i(x) \rightarrow f(x)$  as  $i \rightarrow \infty$  for all  $x \in X$ , then  $f \in \bar{T}(X)$ . We shall say that  $T(X)$  generates  $C_B(X)$  if  $\bar{T}(X) \supset C_B(X)$ . For example,  $C_B(X)$  generates itself. It can be proved that, if  $X_i \in \Pi$ ,  $i = 1, 2, \dots$ , then the set of functions on  $X^\infty$  of the form  $\prod_{i=1}^k f_i(x_i)$ , where  $k$  is an arbitrary natural number, and  $f_i \in C_B(X_i)$ , generates  $C_B(X^\infty)$ ,  $X^\infty = \prod_{i=1}^{\infty} X_i$ .

#### A.1.3 Measures.

##### A.1.3.1

Let  $X$  be a set,  $\Sigma$  a  $\sigma$ -algebra of subsets of  $X$ . The pair  $(X, \Sigma)$  is called a *measure space*. A *measure* (more precisely, a  *$\sigma$ -additive measure*) on  $(X, \Sigma)$  is a non-negative function which is defined on subsets of  $X$  that belong to  $\Sigma$ , which may take the value  $+\infty$ , and which has the following property of additivity: if  $A_i$ ,  $i = 1, 2, \dots$ , belong to  $\Sigma$  and if  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , then  $\mu(\cup A_i) = \sum_{i=1}^{\infty} \mu(A_i)$  (if the latter series diverges, then its sum is regarded as equal to  $+\infty$ ). A measure  $\mu$  is said to be  *$\sigma$ -finite* if the whole space  $X$  is the union of a finite number of sets of finite measure. A measure  $\mu$  is said to be *Lebesgue-complete* if it follows from  $A \in \Sigma$ ,  $\mu(A) = 0$ , and  $B \subset A$  that  $B \in \Sigma$  (then  $\mu(B) = 0$  automatically). A measure  $\mu$  is said to be a *probability measure* if  $\mu(X) = 1$  (then for all  $A \in \Sigma$  the inequality  $0 \leq \mu(A) \leq 1$  holds).

##### A.1.3.2

Let  $\bar{R}$  denote the extended real line, i.e. the set of real numbers supplemented by  $+\infty$  and  $-\infty$ . Let  $(X, \Sigma)$  be a measurable space and let  $f: X \rightarrow \bar{R}$  be

a function  $f$  is said to be  $\sigma$ -measurable if the pre-image of any set of the form  $\{t > a\}$  on  $\bar{R}$  belongs to  $\Sigma$  (then the pre-image of any Borel set of  $\bar{R}$  belongs to  $\Sigma$ ). One can try to integrate any such function with respect to the measure  $\mu$  in the following way. First let  $f \geq 0$ . A non-negative  $(X, \Sigma)$ -measurable function  $\varphi$  which takes only a finite number of values is called a *simple function*. Let these values be  $t_1, \dots, t_r$ , and let  $A_i = \{x \in X | \varphi(x) = t_i\}$ . We put

$$\int \varphi(x) d\mu(x) = \sum_{i=1}^r t_i \mu(A_i)$$

(here  $0.(+\infty) = 0$ ). An  $(X, \Sigma)$ -measurable function  $f \geq 0$  is said to be *summable* if it is the limit (for all points of  $X$  except possibly for the points of a set of measure zero) of a non-decreasing sequence of simple functions  $\varphi_i$  which has a finite limit

$$\lim_{i \rightarrow \infty} \int \varphi_i(x) d\mu(x).$$

This limit is called the *integral of  $f$  with respect to the measure  $\mu$*  and it is denoted by  $\int f(x) d\mu(x)$  (this definition is well formulated: it can be shown that the limit does not depend on the set  $\{\varphi_i\}$  of functions which have the stated properties). An  $(X, \Sigma)$ -measurable function  $f: X \rightarrow \bar{R}$  is said to be *summable* if each of the two functions  $f_+ = \max\{f, 0\}$  and  $f_- = -\min\{f, 0\}$  is summable. In that case the integral  $\int f(x) d\mu(x)$  is defined as

$$\int f_+(x) d\mu(x) - \int f_-(x) d\mu(x).$$

Let  $L(X, \mu)$  be the space of  $\mu$ -summable functions. This turns out to be a linear space (relative to the natural pointwise operations), which, if it contains  $f$ , also contains the modulus  $|f|$ ; and

$$I(f) = \int f(x) d\mu(x)$$

is a linear functional on this space. It is non-negative ( $I(f) \geq 0$  when  $f \geq 0$ ) and it is continuous in the following sense: let  $f_i \in L(X, \mu)$  and let  $|f_i| \leq f$ ,  $f \in L(X, \mu)$ . Suppose, further, that  $f_i(x) \rightarrow_{i \rightarrow \infty} f(x)$   $\mu$ -almost everywhere. (Note: a property which depends on a point of the space  $(X, \Sigma)$  with a measure  $\mu$  is said to hold  $\mu$ -almost everywhere if it holds for all  $x \in X \setminus A$ , where  $A \in \Sigma$  and  $\mu(A) = 0$ .) Then  $f(x) \in L(X, \mu)$  and  $I(f) = \lim_{i \rightarrow \infty} I(f_i)$  (Lebesgue's theorem). Moreover, if  $|f| \leq g \in L(X, \mu)$  and if  $f$  is  $(X, \Sigma)$ -measurable, then  $f \in L(X, \mu)$ .

#### A.I.3.3 Measures on Polish spaces.

Let  $(X, \Sigma, \mu)$  be a measurable space with a  $\sigma$ -additive probability measure  $\mu$ , and let  $X \in \Pi$ . The measure  $\mu$  (more precisely, the triple  $X, \Sigma, \mu$ ) is said to be

*Borel* if  $B(X) \subset \Sigma$  (i.e. the measure is defined for all Borel sets). The measure  $\mu$  is said to be *regular* if  $A \in \Sigma$  implies that there are closed sets  $A_i \subset A$  such that  $\mu(A) = \lim_{i \rightarrow \infty} \mu(A_i)$  (there are then also *compacts*, i.e. compact sets  $A_i$  which have this property). It is known that every  $\sigma$ -additive (probability) measure on  $B(X)$ ,  $X \in \Pi$ , is regular. Moreover, the term 'measure' applied to a Polish space  $X$  means a 'Lebesgue-complete regular Borel probability measure on  $X$ '.

Let  $X$  be a Polish space and let  $C_B(X)$  denote, as before, the space of continuous bounded functions on  $X$  with values in  $R$ . If  $\mu$  is a measure on  $X$ , then the functional

$$I(f) = \int f(x) d\mu(x)$$

on  $C_B(X)$  has the following properties:

- (1)  $I(f)$  is linear;
- (2)  $I(f) \geq 0$  when  $f \geq 0$ ;
- (3)  $I(1) = 1$  (here 1 on the left-hand side denotes the function which is identically equal to 1);
- (4) if  $f_i \in C_B(X)$  and if  $f_1 \geq f_2 \geq \dots$ , and if  $\lim_{i \rightarrow \infty} f_i(x) = 0$ , then

$$\lim_{i \rightarrow \infty} I(f_i) = 0 \quad (\text{Lebesgue's theorem}).$$

The converse is also true: every functional on  $C_B(X)$  which satisfies the conditions (1)–(4) generates a measure  $\mu$  according to the formula

$$I(f) = \int f(x) d\mu(x).$$

Moreover, to different measures there correspond different functionals.

#### A.I.3.4 Compositions of measures.

Let  $X$  and  $Y$  be Polish spaces, and let  $\Phi_{y|x}$  be a family of measures on  $Y$  depending on a parameter  $x \in X$ . This family is said to be a *Borel family with respect to  $x$*  if, for every  $f \in C_B(Y)$ , the function

$$\tilde{f}(x) = \int f(y) d\Phi_{y|x}$$

is Borel with respect to  $x$ . If  $\Phi_{y|x}$  is Borel with respect to  $x$ , and if  $f$  is a bounded scalar Borel function on  $X \times Y$ , then

$$\tilde{f}(x) = \int f(x, y) d\Phi_{y|x}$$

is Borel with respect to  $x$ . Now let  $\Phi_x$  be a measure on  $X$  and let  $\Phi_{y|x}$  be a family of measures, Borel with respect to  $x$ , on  $Y$ . Then for every function  $f \in C_B(X \times Y)$  the functional

$$\int \left\{ \int f(x, y) d\Phi_{y|x} \right\} d\Phi_x$$

makes sense.

By using Lebesgue's theorem it is easy to check that this functional has the properties (1)–(4), and that it is therefore generated by some measure. This measure is denoted by  $\Phi_x \otimes \Phi_{y|x}$ . Suppose, in particular, that  $\Phi_y \equiv \Phi_{y|x}$  does not depend on  $x$ . Then  $\Phi_x \otimes \Phi_{y|x}$  is a measure composed by the ordered pair of measures  $\Phi_x$  on  $X$  and  $\Phi_y$  on  $Y$ . Thus, from a pair of Polish spaces with measures  $(X, \Phi_x)$  and  $(Y, \Phi_y)$  we can construct a third Polish space  $(X \times Y, \Phi_x \otimes \Phi_y)$ ; it is called the *product of the measure spaces*  $(X, \Phi_x)$  and  $(Y, \Phi_y)$ .

Conversely, let  $\Phi_{x,y}$  be a measure on  $X \times Y$ . We construct from it a measure on  $X$  by the formula  $\Phi_x(A) = \Phi_{x,y}(A \times Y)$  for all  $A$  such that the right-hand side is defined. It is easy to see that a measure on  $X$  has indeed been obtained. It is called the *measure on  $X$  induced by  $\Phi_{x,y}$* . It turns out that there is a family of measures  $\Phi_{x|y}$  on  $Y$ , which are Borel with respect to  $x \in X$ , such that  $\Phi_x \otimes \Phi_{y|x} = \Phi_{x,y}$ .  $\Phi_x$  and  $\Phi_{y|x}$  are defined by  $\Phi_{x,y}$  ‘almost uniquely’: if  $\Phi_x$ ,  $\Phi_{y|x}$ , and  $\tilde{\Phi}_x$ ,  $\tilde{\Phi}_{y|x}$  are such that  $\Phi_x \otimes \Phi_{y|x} = \tilde{\Phi}_x \otimes \tilde{\Phi}_{y|x}$ , then  $\Phi_x = \tilde{\Phi}_x$  and  $\Phi_{y|x} = \tilde{\Phi}_{y|x}$ ,  $\Phi_x$ -almost for all  $x$ . Now let  $X_i \subset \Pi$ , let  $X^i = \prod_{j=1}^i X_j$ , and let  $X^\infty = \prod_{j=1}^\infty X_j$ . We suppose that on the spaces  $X^i$  we are given measures which are compatible in the following sense:  $X^i$  is a natural direct factor in  $X^j$ ,  $j > i$  ( $X^i$  is identified with the set of the first  $i$  factors among the  $j$  factors whose product is  $X^j$ ), so that it is possible to consider measures  $\Phi_{x^i}$  induced by  $\Phi_{x^j}$  on  $X^i$ . *Consistency of the measures*  $\Phi_{x^i}$  means that  $\Phi_{x^i} = \Phi_x$  for all  $i$  and  $j > i$ .

As an example of such a family we can take the family of measures induced on  $X^i$  by a measure  $\Phi_{x^\infty}$  on  $X^\infty$ . In fact, it turns out that there are no other examples: by a classical theorem due to Kolmogorov, every consistent family of measures (consistent in the above sense) is generated, in the way described, by a certain (uniquely determined) family of measures on  $X^\infty$ . This fact was used in Section 1.3.3.2.

#### A.1.3.5 On the coincidence of measures.

Sometimes it is useful to have convenient sufficient conditions for two measures to be the same. Let  $X$  be a Polish space and let  $T(X)$  be any family of bounded Borel functions on  $X$ , generating  $C_B(X)$ . Let  $\Phi_x$  and  $\tilde{\Phi}_x$  be two measures on  $X$ . In order to show that  $\Phi_x = \tilde{\Phi}_x$ , it suffices to verify that

$$\int f(x)d\Phi_x = \int f(x)d\tilde{\Phi}_x$$

for all  $f \in T(X)$  (for then the family of functions  $f \in B(X, R)$  such that

$$\int f(x)d\Phi_x = \int f(x)d\tilde{\Phi}_x$$

contains  $\bar{T}(X)$ , i.e.  $C_B(X)$ , and so  $\Phi_x = \tilde{\Phi}_x$ ).

We give an example. Let  $X, Y \in \Pi$  and let  $\Phi_{x,y}, \Phi_x, \Phi_{y|x}$  be measures (family of measures) on the respective spaces. Suppose we want to prove that  $\Phi_{x,y} = \Phi_x \otimes \Phi_{y|x}$ . To do this, it suffices to check that, for any  $f \in C_B(X)$

and  $g \in C_B(Y)$ ,

$$\int f(x)g(y)d\Phi_{x,y} = \int f(x) \{ \int g(y)d\Phi_{y|x} \} d\Phi_x$$

(for, it was pointed out in Section A.1.2.4 that the set of functions of the form  $f(x)g(y)$ , where  $f \in C_B(X)$ ,  $g \in C_B(Y)$ , generates  $C_B(X \times Y)$ ).

## A.2 BANACH SPACES

The reader is assumed to be familiar with the basic ideas and facts of linear algebra and the main definitions relating to Hilbert space, as given in, say, [16], Chapters III and IV.

### A.2.1

Let  $E$  be a linear real space. A function  $x$  on  $E$  which takes real values is called a *norm* if

- (1)  $\|x\| \geq 0$  and  $\|x\| = 0 \Leftrightarrow x = 0$ ;
- (2)  $\|\lambda x\| = |\lambda| \|x\|$ ,  $\lambda \in \mathbb{R}$ ;
- (3)  $\|x + y\| \leq \|x\| + \|y\|$ .

The space  $E$  with a norm  $\|\cdot\|$  is called a *normed space*. The norm  $\|\cdot\|$  generates a metric on  $E$ :  $\rho(x, y) = \|x - y\|$ . With this metric, linear operations in  $E$  are continuous.

Every norm  $\|\cdot\|$  on  $E$  determines a certain set  $V = \{x \mid \|x\| \leq 1\}$ , the *unit ball* in  $E$ . The set  $V$  has the following properties:

- (1)  $V$  is convex, i.e.  $x, y \in V$  and  $\theta \in [0, 1]$  imply that  $\theta x + (1 - \theta)y \in E$ ;
- (2)  $V$  is balanced ( $-V \equiv \{x = -u \mid u \in V\} = V$ );
- (3)  $V$  is absorbent (i.e.  $\forall x \in E \exists \lambda > 0: \lambda x \in V$ );
- (4)  $V$  is bounded along rays (i.e.  $\forall x \neq 0 \exists \lambda > 0: \lambda x \notin V$ );
- (5)  $V$  is closed along rays (i.e.  $\forall x \in E$  the set  $\{\lambda \mid \lambda x \in V\}$  is closed along the axis).

Conversely, every set  $V \subset E$  having the properties (1)–(5) is the unit ball corresponding to some (uniquely determined) norm, namely, the norm

$$\|x\| = \inf \{\lambda > 0 \mid \lambda^{-1}x \in V\}.$$

### A.2.2

A norm on  $E$ , as already stated, determines a metric (and therefore it distinguishes a class of convergent sequences of elements of  $E$ ). Two norms on  $E$  are said to be *equivalent* if they generate the same topology on  $E$  (i.e. the store of convergent sequences in both norms is the same). It can be shown that two norms  $\|\cdot\|$  and  $\|\cdot'\|$  are equivalent if and only if the unit ball for either of them is

a bounded set in the other, or what comes to the same thing, if for some  $c > 0$ ,

$$\frac{1}{c} \|x\|' \geq \|x\| \geq c \|x\|' \text{ for all } x \in E.$$

### A.2.3

A space  $E$  with a norm  $\|\cdot\|$  is called a *Banach space* if  $E$  is complete as a metric space with the metric  $\|x - y\|$ . The property of being a Banach space is a topological property of a normed space and does not depend on the actual method of norming (i.e. if  $E$ ,  $\|\cdot\|$  is a Banach space and if  $\|\cdot\|'$  is equivalent to  $\|\cdot\|$ , then  $E$ ,  $\|\cdot\|'$  also is a Banach space).

### A.2.4 The Hahn–Banach theorem.

Let  $E$  be a linear space, and let  $p(h)$  be a real-valued function on  $E$ . The function  $p(h)$  is said to be a *sub-linear function* if

- (1)  $p(\lambda h) = \lambda p(h)$ ,  $\lambda \geq 0$ ;
- (2)  $p(h_1 + h_2) \leq p(h_1) + p(h_2)$ ,

in other words,  $p(h)$  is convex.

For linear (and convex) analysis the following theorem is of fundamental importance.

**The Hahn–Banach theorem.** Let  $E$  be a linear space,  $E_1$  be one of its subspaces, and let  $p(h)$  be a sub-linear function on  $E$ . Suppose, further, that  $f$  is a linear functional on  $E_1$  such that

$$\langle f|h \rangle \leq p(h) \text{ for all } h \in E_1$$

(here and elsewhere  $\langle f|x \rangle$  is the value of the linear functional  $f$  on the vector  $x$ ). Then there is a linear functional  $\bar{f}$  on the whole of  $E$  which coincides with  $f$  on  $E_1$  and is such that

$$\langle \bar{f}|h \rangle \leq p(h) \text{ for all } h \in E.$$

### A.2.5 The dual space.

Let  $E$ ,  $\|\cdot\|$ , be a normed space, and let  $f$  be a linear functional on  $E$ . The functional  $f$  is said to be *continuous* if  $\langle f|x \rangle$  is a continuous function of  $x \in E$  (it is sufficient that it be continuous at 0). An equivalent definition is that the function  $\langle f|x \rangle$  is bounded on the unit ball in  $E$ . The set of all linear continuous functionals on  $E$  with the natural linear operations itself forms a linear space. It is called the space *dual* to  $E$ ,  $\|\cdot\|$ , and is denoted by  $E^*$ . On  $E^*$  the norm

$$\|f\|_* = \sup \{ \langle f|x \rangle | x \in E, \|x\| \leq 1 \}$$

is introduced. Relative to this norm  $E^*$  turns out to be a Banach space. It is important to notice that the store of elements (and the topology) of the space  $E^*$  is determined only by the store of elements and the topology of  $E$  (i.e. when  $\|\cdot\|$  is replaced by an equivalent norm  $\|\cdot\|'$  the store of elements and the linear structure of  $E^*$  do not change, and the norm  $\|\cdot\|_*$  is replaced by an equivalent norm  $(\|\cdot\|')_* = \|\cdot\|'_*$ ).

We now consider the question of the store of elements in  $E^*$ . It turns out that there are adequately many of them: for every  $x \in E$  there is a  $\varphi \in E^*$ :  $\|\varphi\|_* = 1$  and  $\langle \varphi|x \rangle = \|x\|$ .

For, consider the 1-dimensional linear subspace  $E_1$  spanned by  $x$  (we can suppose that  $x \neq 0$ ). On  $E_1$  there is determined a linear functional  $\varphi(\lambda x) = \lambda \|x\|$ ,  $\lambda \in \mathbb{R}$ . The norm of this functional (on  $E_1$ ) is equal to 1 and  $\varphi(x) = \|x\|$ . It remains to notice that from the Hahn–Banach theorem follows this

**Corollary.** If  $E_1 \subset E$  is a subspace, and if  $\varphi$  is a continuous linear functional on  $E_1$ , then  $\varphi$  can be extended, without increasing the norm, on to the whole of  $E$ .

For, if  $a$  is the norm of  $\varphi$  on  $E_1$ , then  $\langle \varphi|h \rangle \leq a\|h\|$  when  $h \in E_1$ . The function  $a\|h\|$  is sub-linear on  $E$ , and by the Hahn–Banach theorem  $\varphi$  extends to a linear functional  $\bar{\varphi}$  on the whole of  $E$  such that  $\langle \bar{\varphi}|h \rangle \leq a\|h\|$ ,  $h \in E$ . In view of this,  $\|\bar{\varphi}\|_* \leq a$ , and this shows that  $\bar{\varphi}$  is the extension of  $\varphi$  required by the corollary.

Thus the following ‘symmetric’ relations hold:

$$\begin{aligned} \sup_{x \in E, \|x\| \leq 1} \langle \varphi|x \rangle &= \|\varphi\|_* \\ \sup_{\varphi \in E^*, \|\varphi\|_* \leq 1} \langle \varphi|x \rangle &= \|x\|. \end{aligned} \quad (2.1)$$

Moreover, from the definition of the linear operations in  $E^*$  it is clear that the form  $\langle \varphi|x \rangle : E^* \times E \rightarrow \mathbb{R}$  for a fixed  $x \in E$  is a linear functional on  $E^*$  which is continuous by virtue of (2.1) (similar to the fact that, for a fixed  $\varphi$ , this form is a linear continuous functional on  $E$ ).

Thus, a vector  $x \in E$  generates a linear continuous functional on  $E^*$ . The norm of this functional is  $\|x\|$ . In other words, we have constructed a canonical isometric (norm-preserving) embedding of  $E$  in the second dual space  $(E^*)^* \equiv E^{**}$ . A space  $E$  is said to be *reflexive* if the image of  $E$  under this embedding is the whole of  $E^{**}$  (i.e. every linear continuous functional on  $E$  is determined by a vector  $x \in E$  according to the formula  $\varphi \rightarrow \langle \varphi|x \rangle$ ). It is known that, if  $E$  is reflexive, then so is  $E^*$ , and conversely.

### A.2.6 The weak topology.

As well as the ordinary topology (determined by the norm) of a Banach space  $E$  we can also consider the weak topology —which is the weakest topology in  $E$  in

which all functions of the form  $x \rightarrow \langle \varphi | x \rangle$ ,  $\varphi \in E^*$ , are continuous. If  $E^*$  is separable, then this topology on all *bounded* (with respect to the norm) subsets of  $E$  can be specified by means of the metric

$$\rho(x, y) = \sum 2^{-i} |\langle \varphi_i | x - y \rangle|,$$

where  $\{\varphi_i\}$  is a countable, dense in the unit ball of  $E^*$ , family of functionals. The linear operations in  $E$  are continuous in the weak topology. Convex closed (with respect to the norm) subsets of  $E$  are also closed in the weak topology.

On the space  $E^*$  dual to  $E$  we can consider, as well as the ordinary and the weak topology, the  $*$ -weak topology, the weakest of all topologies for which all the functions  $\varphi \rightarrow \langle \varphi | x \rangle$ ,  $x \in E$ , are continuous. If  $E$  is separable, the  $*$ -weak topology on *bounded* subsets of  $E^*$  can be specified by the metric

$$\rho(\varphi, \psi) = \sum 2^{-i} |\langle \varphi - \psi | x_i \rangle|,$$

where  $\{x_i\}$  is a countable set of vectors dense in the unit ball of  $E$ .

It is a very important fact that the unit ball in  $E^*$  is always compact in the  $*$ -weak topology. In particular, if  $E$  is reflexive, then the unit ball in  $E$  is compact in the weak topology of  $E$ . (For, it is clear that the weak topology in  $E$  coincides with the  $*$ -weak topology in  $E^{**} = E$ .) Combining this circumstance with the fact that convex subsets of  $E$ , which are closed relative to the norm, are weakly closed, we obtain the following important result. A bounded, convex, closed subset of a reflexive Banach space is compact in the weak topology.

We remark further that in the finite-dimensional case all the topologies mentioned coincide. Every finite-dimensional space is reflexive.

### A.2.7 Direct products of Banach spaces.

Let  $E_i$ ,  $\|\cdot\|_i$ ,  $1 \leq i \leq k < \infty$ , be Banach spaces. Their set-theoretic direct product  $E = \prod_{i=1}^k E_i$  is provided with a natural linear structure (linear operations proceed co-ordinate-wise), and a natural topology for the direct product is given by a norm, say  $\|x^k\| = \max_i \|x_i\|_i$ , where  $x^k = (x_1, \dots, x_k)$ ,  $x_i \in E_i$ . The space dual to  $E$  is

$$\prod_{i=1}^k E_i^* \text{ with the norm } \sum_i \|\varphi_i\|_*, \quad \varphi = (\varphi_1, \dots, \varphi_k), \quad \varphi_i \in E_i^*.$$

If the  $E_i$ ,  $1 \leq i \leq k$ , are reflexive, then so is  $\prod_{i=1}^k E_i$

### A.2.8 Examples.

We give some examples of actual Banach spaces used in the book. They all belong to a scale of spaces  $L_p$ . This scale is defined as follows. Let  $(X, \Sigma, \mu)$  be a

measurable space with a  $\sigma$ -finite,  $\sigma$ -additive, Lebesgue-complete measure  $\mu$ . We shall consider two scalar  $\Sigma$ -measurable functions  $f_1, f_2$  to be equivalent if  $f_1 = f_2$   $\mu$ -almost everywhere. Let  $M(X, \mu)$  be the space whose elements are equivalence classes in  $X$  of  $\Sigma$ -measurable scalar functions with the natural (pointwise) linear operations.

Suppose, further, that  $1 \leq p < \infty$ . The space  $L_p(X, \mu)$  consists of all the elements of  $M(X, \mu)$  for which

$$\|\bar{f}\|_p = \left\{ \int |f(x)|^p d\mu(x) \right\}^{1/p} < \infty$$

(on the left  $\bar{f}$  denotes an element of  $M(X, \mu)$ ; on the right  $f$  is any representative of the class  $\bar{f}$ ; the right-hand side does not depend on which representative is chosen). For  $p = \infty$  the space  $L_\infty(X, \mu)$  is defined as the set of elements of  $M(X, \mu)$  for which the number

$$\|\bar{f}\|_\infty = \text{ess sup } |f(x)| = \inf_{A \in \Sigma, \mu(A) = 0} \sup_{x \in X \setminus A} |f(x)| \text{ is finite.}$$

The spaces  $L_p(X, \mu)$ ,  $1 \leq p \leq \infty$ , turn out to be linear subspaces of  $M(X, \mu)$  which are Banach spaces relative to the norm  $\|\cdot\|_p$ . When  $1 \leq p < \infty$ , the space dual to  $L_p(X, \mu)$  is canonically identified with  $L_q(X, \mu)$ , where  $q = p/(p-1)$ , i.e.

$$\frac{1}{p} + \frac{1}{q} = 1.$$

In this identification an element  $\bar{f} \in L_q$  acts, as a functional on a vector  $\bar{g} \in L_p$ , according to the formula

$$\langle \bar{f} | \bar{g} \rangle = \int f(x) g(x) d\mu(x) \quad (f \in \bar{f}, g \in \bar{g}).$$

The identification mentioned preserves the norm. In particular,

$$\int |f(x)g(x)| d\mu(x) \leq \|f\|_q \|g\|_p, \quad \frac{1}{p} + \frac{1}{q} = 1$$

(Holder's inequality). Hence it follows that if  $\mu$  is a probability measure and if  $1 \leq p \leq p' \leq \infty$ , then  $L_p(X, \mu) \supseteq L_{p'}(X, \mu)$  and  $\|f\|_p \leq \|f\|_{p'}$  when  $f \in L_{p'}$ .

The case  $p = 2$  is special. When  $p = 2$  the space  $L_p(X, \mu)$  is a Hilbert space, with the scalar product and norm

$$\begin{aligned} \langle \bar{f} | \bar{g} \rangle &= \int f(x) g(x) d\mu(x), \quad f \in \bar{f}, \quad g \in \bar{g} \\ \|\bar{f}\|_2 &= (\langle \bar{f} | \bar{f} \rangle)^{1/2}. \end{aligned}$$

Every Hilbert space  $H$  is canonically identified with its dual (a vector  $x \in H$  acts as a functional on a vector  $y \in H$  according to the formula  $\langle x | y \rangle = (x | y)$ ; the right-hand side is the scalar product in  $H$ ). This identification of  $H$  with  $H^*$  when  $H$  is taken to be the space  $L_2(X, \mu)$  turns into the identification mentioned above of  $(L_2(X, \mu))^*$  with  $L_2(X, \mu)$ .

Let us consider the case of finite-dimensional spaces  $L_p(X, \mu)$  again. Every such space of linear dimension  $n$ ,  $1 \leq n < \infty$ , is isometrically isomorphic to the co-ordinate space  $R^n$  with the norm

$$\|x\|_p = (\sum |x^i|^p)^{1/p}, x = (x^1, \dots, x^n) \in R^n.$$

When  $p = \infty$ ,  $(\sum |x^i|^p)^{1/p}$  is replaced by  $\max_i |x^i|$ .  $R^n$  normed in this way is denoted by  $l_p^{(n)}$ . The space dual to  $l_p^{(n)}$  is  $l_q^{(n)}$ ,  $1/p = 1/q = 1$ . An element  $\varphi \in l_q^{(n)}$  acts as a functional on an element  $x \in l_p^{(n)}$  according to the formula

$$\langle \varphi | x \rangle = \sum_{i=1}^n \varphi^i x^i.$$

When  $1 < p < \infty$ , the spaces  $L_p(X, \mu)$  are reflexive. This is not so when  $p = 1$  or  $p = \infty$  (in the finite-dimensional case).

### A.2.9 Integration of functions with values in a Banach space.

Let  $E$ ,  $\|\cdot\|$ , be a separable Banach space, and let  $(X, \mu)$  be a Polish space with a measure. (We recall that a measure on a Polish space means to us a Lebesgue-complete, Borel, regular, probability measure. Vector functions can be integrated in even more general situations, but we do not need to do this.)

Let  $M(X, E)$  denote the set of equivalence classes (relative to coincidence  $\mu$ -almost everywhere) of Borel functions on  $X$  with values in  $E$ . Let  $L_p(X, \mu, E)$  be the set of all elements  $\bar{f} \in M(X, E)$  such that the number

$$\|\bar{f}\|_p = \|\{\|f(x)\|\}\|_p, f \in \bar{f}$$

is finite. It can be shown that  $L_p(X, E)$  is a linear space (relative to the natural linear operations) and, with the norm  $\|\cdot\|_p$ , it is a Banach space, and  $L_p(X, \mu, E) \supset L_{p'}(X, \mu, E)$  for  $p' \geq p$ . Also  $\|\bar{f}\|_{p'} \geq \|\bar{f}\|_p$  for  $p' \geq p$ ,  $\bar{f} \in L_{p'}$ .

Functions  $f \in \bar{f} \in L_p(X, \mu, E)$  can be integrated with respect to the measure  $\mu$  in the following way. We shall call  $f$  a *step-function* if it is a Borel function which takes a finite number of values  $t_1, \dots, t_r, \dots$ . For a step-function  $f \in \bar{f} \in L_1(X, \mu, E)$  the integral  $\int f(x) d\mu(x)$  is defined as

$$\sum_{i=1}^{\infty} \mu\{x | f(x) = t_i\} t_i$$

(the series converges absolutely because  $f \in L_1$ ). Further, it can be proved that every function  $f \in \bar{f} \in L_1$  can be expressed as the sum of a series  $\sum_{i=1}^{\infty} f_i$  of step-functions with a finite sum  $\sum_{i=1}^{\infty} \|f_i\|_1$ . It can be proved that the series  $\sum_{i=1}^{\infty} \int f_i(x) d\mu(x)$  converges absolutely in  $E$ , and that its sum does not depend on the choice of the  $f_i$  with the indicated properties. It is taken by definition to be the integral  $\int \bar{f}(x) d\mu(x)$ . The integral does not depend on the choice of  $f \in \bar{f} \in L_1$ , and

$$\|\int \bar{f}(x) d\mu(x)\| \leq \|\bar{f}\|_1.$$

Thus the mapping  $\bar{f} \rightarrow \int f(x) d\mu(x)$  is a continuous linear operator  $I$  from  $L_1(X, \mu, E)$  into  $E$ , and  $\|If\| \leq \|\bar{f}\|_1$ , if  $f \in L_p(X, \mu, E)$ .

### A.2.10

In addition to the notation already given (which is standard throughout the book), the following notation is extensively used. Let  $G \subset E$ , and let  $\|\cdot\|$  be a norm on  $E$ . Then  $\rho_{\|\cdot\|}(x, G)$ ,  $\rho_{\|\cdot\|}(G)$ ,  $d_{\|\cdot\|}(G)$ ,  $\alpha_{\|\cdot\|}(G)$  denote respectively the distance from  $x$  to  $G$ , the radius of  $G$ , the diameter of  $G$ , and the asphericity of  $G$  (see Section A.1.0) in the metric on  $E$  given by the norm  $\|\cdot\|$ . Any point  $x \in G$  such that  $G$  is contained in a ball of radius  $\rho_{\|\cdot\|}(G)$  with centre at  $x$  is called a *centre* of  $G$ . It can be shown that, if  $E$  is reflexive and if  $G$  is convex, closed, bounded and not empty, then there is at least one centre of  $G$ .

# Principal notation

Arabic numerals indicate (from left to right) chapters, sections, and subsections: A indicates the appendix. Notation which is entirely standard, or which is used within a single chapter only, is not listed.

	Principal notation	377
$\mathbb{R}$	MCG	2.3 the method of centres of gravity
$\mathbb{R}^n$	MMCG	2.5 the modified method of centres of gravity
$E^n$	$\overline{\text{MD}}_v, \overline{\text{MD}}_p, \text{MD}_{1,n}$ $\widetilde{\text{MD}}_v, \widetilde{\text{MD}}_p, \text{MD}_{1,n}$ $\text{MD}_v, \text{MD}_v(v), \text{MD}^{p,r}$	3.3 versions of different kinds of the method of mirror descent for solving convex extremal problems
inf, sup	$\text{MD}_{V,V_I}, \text{MD}_{V,V_I}(v)$ $\text{MD}_{V,V_I}, \text{MD}_{V,V_I}(v)$	3.4 } 5.2 } 6.2 } 6.3 } versions of the method of mirror descent for solving convex-concave games
$[t]_+$	$\omega_v(t), \gamma_v(s)$	3.2 } smoothness characteristics of a regular function $V$
$]t[$	$\omega_{V,r}(t), \gamma_{V,r}(s)$	5.1 }
$\mathfrak{L}\{e_1, \dots, e_k\}$	$C_{\text{Lip}}(G, E, \ \cdot\ , m)$ $C(G, E, m)$	2.2 weighted field of linear (general) convex problems with $m$ constraints on a set $G$ in a space $E$ , $\ \cdot\ $
int $G$	$E^*, \ \cdot\ _*$	A.2 normed space dual to $E$ , $\ \cdot\ $
$\partial G$	$\langle \varphi   x \rangle$	A.2 value of a linear functional $\varphi \in E^*$ on a vector $x \in E$ (in a Hilbert space, the scalar product)
$\bar{G}$	$\rho_{\ \cdot\ }(G)$	A.2.10 radius of a set $G$ relative to the norm $\ \cdot\ $
$E, \ \cdot\ $	$d_{\ \cdot\ }(G)$	A.2.10 diameter of a set $G$
$l(\mathcal{B}, \mathcal{A}), \Gamma(\mathcal{B}, \mathcal{A})$	$\alpha_{\ \cdot\ }(G)$	A.2.10 $\ \cdot\ $ -asphericity of a subset $G$
$v(\mathcal{B}, \mathcal{A}), \bar{v}(\mathcal{B}, \mathcal{A})$	$\rho_{\ \cdot\ }(x, G)$	A.2.10 distance from $x$ to $G$
$N(v), \bar{N}(v)$	$\alpha_{p,k}(\ \cdot\ )$ $\alpha_{p,\ \cdot\ ,k}(G)$ $\alpha_{p,k}(G)$ $\alpha(G)$	4.1.1 various kinds of characteristics of asphericity
$\equiv$	$\nabla f$	— gradient of a scalar function on a Hilbert space
	$\partial_G f(x)$	2.1 set of support (on $G$ ) functionals to a convex function $f$ at the point $x$
	$\mathbf{V}_G(f)$	2.1.3 variation of a function $f$ on a set $G$
	$\mathbf{L}_{\ \cdot\ }(f)$	2.2.2 Lipschitz constant of $f$ relative to $\ \cdot\ $
	$L_p(T, \mu)$	A.2 Banach space of $p$ th-power integrable functions on a space $T$ with a $\sigma$ -additive $\sigma$ -finite measure $\mu$ , $1 \leq p \leq \infty$
	$\ \cdot\ _p$	A.2 the natural $L_p$ norm
	$I_p^{(n)}$	A.2 the standard $n$ -dimensional space $L_p$

	Principal notation
$X^i$	A.1 the product of a space $X$ by itself $i$ times, $0 \leq i \leq \infty$
$x^i = (x_1, \dots, x_i)$	— a point of the space $X^i$
$f = (f_0, \dots, f_m)$	1.2.1 a mathematical programming problem, generated by a vector-function $f$ (and also the function itself)
$f_*$	1.2.1 optimal value of the objective functional of the problem $f$
$\mathcal{O} = ((\Omega, \mathbf{F}_\omega), I, \psi(x, f, \omega))$	1.3.1 oracle with a noise space $(\Omega, \mathbf{F}_\omega)$ , observation functions $\psi(x, f, \omega)$ , and information space $I$
$r(f) = (r_0(f), \dots, r_m(f))$	1.2.3 normalizing factors for a class of problems under consideration
$\mathfrak{U}(\mathcal{F}, G, m, E)$	1.2.2 a field of problems with $m$ constraints on a set $G \subset E$ , the problems being generated by $(m+1)$ -dimensional functions of a family $\mathcal{F}$
$\varepsilon(x, f) = (\varepsilon_0(x, f), \dots, \varepsilon_m(x, f))$	1.2.3 vector of the absolute errors of a point $x$ regarded as an approximate solution of a problem $f$
$v(x, f)$	1.2.3 relative error of a point $x$ regarded as an approximate solution of a problem $f$
$\mathcal{B}(\tilde{\mathcal{B}})$	1.3.3 a deterministic (resp. stochastic) method of solving problems of the class in question
$\int_{\mathcal{B}} d\mathbf{F}_t$	1.3.4 a mixture of deterministic methods
$l(\tilde{\mathcal{B}}, f), \bar{l}(\tilde{\mathcal{B}}, f)$	1.4.1 mean (resp. strong) laboriousness of a method $B$ on a problem $f$
$v(\tilde{\mathcal{B}}, l), \bar{v}(\tilde{\mathcal{B}}, f)$	1.4.1 mean (resp. strong) error of a method $\mathcal{B}$ on a problem $f$
$C_{\text{Lip}}^\circ(G, E, \ \cdot\ , m) \quad C^\circ(G, E, m)$	2.2 classes of problems obtained by equipping the respective fields with an oracle $\mathcal{O}$
$C \in C_{\text{Lip}}^{\circ_0}(G, E, \ \cdot\ , m) \quad C \in C^{\circ_0}(G, E, m)$	2.2 Abbreviated formulation of the statement: 'C is a class of problems of the type $C_{\text{Lip}}^{\circ_0}(G, E, \ \cdot\ , m)$ resp. $C^{\circ_0}(G, E, m)$ '.

$D(G \times G_I, E, E_I, \ \cdot\ , \ \cdot\ _I)$	6.1.2 field of Lipschitz convex-concave games
$\tilde{C}_{r,L}^{\circ,v_0}(G, E, \ \cdot\ , m)$	5.1.3 class of stochastic convex problems
$D_{r,L}^{\circ,v_0}(G \times G_I, E, \ \cdot\ , E_I, \ \cdot\ _I)$	6.1.2 classes of problems of solving games
$H_{x_1}^V(G, m, l_0, \dots, l_m, Q_0, \dots, Q_m)$	7.1 classes of smooth (strongly) convex extremal problems
$H_{x_1}^V(G, m, l_0, \dots, l_m, Q_0, \dots, Q_m)$	
$CS_{x_1}^V(G, m)$	
$\tilde{x}, \tilde{\tilde{x}}, \sim, \lesssim$	4.1.4
$g_{j,x}^f(y)$	2.2.5
$G_*$	1.2.3
$\pi_G(x)$	3.3.1
$\mu_G(x)$	3.3.1

# References

## References

381

1. Akhiezer, N. I. (1965). *Lectures on the theory of approximation*, 'Nauka', Moscow. (English translation of an earlier edition: *Theory of approximation*, Ungar, New York, 1956.)
2. Arrow, K. J., Hurwicz, L., and Uzawa, H. (1958). *Studies in linear and non-linear programming*, Stanford University Press, Stanford, California.
3. Bakhalov, N. S. (1964). On optimal estimates of the rate of convergence of quadrature processes and integration methods of the Monte Carlo type. In the miscellany: *Numerical methods of solving differential and integral equations and quadrature formulae*, 'Nauka', Moseow. pp. 5–63. (In Russian.)
4. Bruck, R. (1977). On weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space. *J. Math. Anal. and Appl.*, **61**, No. 1.
5. Danilin, Yu. M. (1971). Methods of conjugate gradients for solving minimization problems. *Kibernetika*, No. 5. pp. 122–136. (In Russian.)
6. Day, M. (1973). *Normed linear spaces*, 3rd edition, Springer Verlag, Berlin, Heidelberg, New York.
7. Dunford, N., and Schwartz, D. (1963). *Linear operators*. Interscience, New York.
8. Dvoretzky, A. (1960). Some results on convex bodies and Banach spaces. *Proc. Int. Symp. on Linear Spaces, Jerusalem, 1960*. pp. 123–160.
9. Ermolev, Yu. M. (1976). *Methods of stochastic programming*, 'Nauka', Moscow. (In Russian.)
10. Gupal, A. M. (1977). A method of minimizing almost-differentiable functions. *Kibernetika*, No. 1. (In Russian.)
11. Halmos, P. R. (1950). *Measure theory*, Van Nostrand, Princeton, New Jersey.
12. Himmelblau, D. (1972). *Applied non-linear programming*. McGraw-Hill, New York.
13. Ivanov, V. V. (1972). On optimal algorithms for minimizing functions of a certain class. *Kibernetika*, No. 4. pp. 81–94. (In Russian.)
14. John, J. F. (1948). Extremum problems with inequalities as subsidiary conditions. In: *Studies and essays presented to R. Courant on his 60th birthday*, Interscience, New York.
15. Karp, R. M. (1974). On the computational complexity of combinatorial problems. *Proc. Symp. on Large-scale Networks, Northwestern University, Evanston, April 1974*. Special issue of *Network*, **5**, No. i, 45–68.
16. Kolmogorov, A. N., and Fomin, S. V. (1972). *Elements of the theory of functions and functional analysis*, 3rd edition, 'Nauka', Moscow. (In Russian.) English translation of earlier edition, Graylock Press, Rochester, N.Y., 1957.
17. Kuzovkin, A. I., and Tikhomirov, V. N. (1967). On the number of calculations for finding the minimum of a convex function. *Ekonomika i matem. metody*, **III**, No. 1, 95–103. (In Russian.)
18. Levin, A. Yu. (1965). On an algorithm for minimizing convex functions. *Doklady Akad. Nauk SSSR*, **160**, No. 6. English translation: *Soviet Maths*, **6**, No. 1, 286–290. American Mathematical Society.
19. Levin, L. A. (1973). Universal sorting problems. *Problemy peredachi informatsiya*, **IX**, No. 3, 115–116. (In Russian.)
20. Martin, D. W., and Tee, G. I. (1961). Iterative methods for linear equations with a positive-definite symmetric matrix. *Computer J.*, **4**, No. 3.
21. Mityagin, B. S. (1969). Two inequalities for the volume of convex bodies. *Matem. Zametki*, **5**, No. 1. English translation: *Mathematical notes of the Academy of Sciences USSR*, **5**, No. 1, 61–64. Plenum, New York.
22. Nemirovskiy, A. S., and Yudin, D. B. (1977). Methods of optimization adaptable to problems of 'substantial' dimensionality. *Avtomatika i telemekhanika*, No. 4, 75–87. (In Russian.)
23. Nemirovskiy, A. S., and Yudin, D. B. (1979). Efficient methods of solving convex-programming problems of high dimensionality. *Ekonomika i matem. metody*, **XV**, No. 1. (In Russian.)
24. Nemirovskiy, A. S., and Yudin, D. B. (1978). Cesàro convergence of gradient methods of approximating to saddle-points of convex-concave functions. *Doklady Akad. Nauk SSSR*, **239**, No. 5, 1056–1059. English translation: *Soviet Maths*, **19**, No. 2, 482–486.
25. Oblonskaya, L. Ya. (1969). Comparison of rapidity of convergence of conjugate-gradient methods and the gradient method for quadratic functions. In the miscellany: *Questions of accuracy and efficiency*, *Trudy simposiuma*, Volume 4. Kiev. (In Russian.)
26. Oblonskaya, L. Ya. (1969). Rate of convergence of the conjugate-gradient method for quadratic functionals. *Trudy 2y zimney shkolu po matem. program. i smezhnym voprosam*, No. 3, 550–568. (In Russian.)
27. Polak, E. (1971). *Computational methods in optimization; a unified approach*, Academic Press, New York.
28. Shor, N. Z. (1970). On the rate of convergence of the generalized method of gradient descent with extension of the space. *Kibernetika*, No. 2. (In Russian.)
29. von Neumann, J. (1961). On the theory of strategic games. In the miscellany: *Matrix games*, Fizmatgiz, Moscow. (In Russian.)
30. Yudin, D. B., and Nemirovskiy, A. S. (1976). Estimating the computational complexity of mathematical-programming problems. *Ekonomika i matem. metody*, **XII**, No. 1, 128–142. (In Russian.)
31. Yudin, D. E., and Nemirovskiy, A. S. (1976). Computational complexity and efficiency of methods of solving convex extremal problems. *Ekonomika i matem. metody*, **XII**, No. 2, 357–369. (In Russian.)
32. Yudin, D. B., and Nemirovskiy, A. S. (1977). Computational complexity of strictly convex programming. *Ekonomika i matem. metody*, **XIII**, No. 3, 550–569. (In Russian.)
33. Yudin, D. B., and Nemirovskiy, A. S. (1977). Efficiency of random search in control problems. *Izvestiya Akad. Nauk SSSR, tekhnicheskaya kibernetika*, No. 3, 3–17. (In Russian.)
34. Yudin, D. B., and Nemirovskiy, A. S. (1978). Efficiency of randomization of control. In the miscellany: *Problems of random search*, Zinatne, No. 7, Riga. (In Russian.)

35. Zvonkin, A. K., and Levin, L. A. (1970). Complexity of finite objects and proof of the concepts of information and randomness by means of the theory of algorithms. *Uspekhi matem. nauk*, **XXV**, No. 6, 85–127. English translation: *Russian math. surveys*, **25**, No. 6, 83–124. London Mathematical Society.

## Index

- $\alpha$ -strictly convex function, 48
- absolute error of a point as an approximate solution of a problem ( $\varepsilon(x, f)$ ), 13
- \*-absolute error ( $\varepsilon_*(x, f)$ ), 13
- accuracy: ordinary, strong, 25
- \*-accuracy, 25
- accuracy of a method  $\mathcal{R}$  on the class  $\mathfrak{U}$ , 25
- accuracy of an estimate, 244
- affine functional, 44
- asphericity, 125, 362
- Banach space, 370
- Banach space, regular, 89
- Banach spaces, direct product of, 372
- Banach spaces, examples of, 372
- body, 60
- Borel family of measures, 367
- Borel functions, 364
- Borel mappings, 364
- Borel measure, 367
- Borel properties of a function, 364
- Borel properties of a measure, 367
- Borel properties of subsets, 350
- Borel  $\sigma$ -algebra, 363
- Borel subsets, 363
- canonical forms of methods, 30
- Cauchy's inequality, 44
- centre of a set, 375
- centres-of-gravity method for general convex problems, 57
- Césaro convergence, 211
- characteristics of a given method, 23, 24
- class of general convex mathematical programming problems  $C^e(G, E, m)$ , 52
- class of Lipschitz-convex mathematical programming problems  $C_{\text{Lip}}^e(G, E, \|\cdot\|, m)$ , 52, 93
- class of mathematical programming problems, 25
- class of mathematical programming problems of solving games, 206 *et seq.*
- class of mathematical programming problems of the type  $C^{v_0}(G, E, m)$ , 55
- class of mathematical programming problems of the type  $C_{\text{Lip}}^{v_0}(G, E, \|\cdot\|, m)$ , 55
- class of mathematical programming problems, smooth  $S_c^R(G; L_0, \dots, L_m)$ , 35
- class of smooth convex mathematical programming problems  $CS_{x_i}(G, m)$ , 257
- class of smooth mathematical programming problems  $S_c^k(G; L_0, \dots, L_m)$ , 35
- class of stochastic, convex mathematical programming problems  $\tilde{C}_{r, L}^{v_0}(G, E, \|\cdot\|, m)$ , 185
- class of strongly convex mathematical programming problems  $H_x(G, m; l_0, \dots, l_m, Q_0, \dots, Q_m)$ , 255
- class withstanding rotation, 261
- classes of convex-concave games, 206 *et seq.*
- closed linear hull  $\mathcal{L}$ , 261
- closure of a set, 362
- coincidence of measures, 368
- compatible mathematical programming problem, 11

complexity of a class of problems, 26  
 complexity of a class of problems, strong,  
   deterministic  $N(v)$ , 26  
 complexity of a class of problems, strong,  
   stochastic  $\tilde{N}(v)$ , 26  
 complexity of a class of strictly  
   compatible convex problems, 150  
 complexity of classes of general convex  
   problems (exact 1st-order oracle),  
   125 *et seq.*  
 complexity of classes of Lipschitz-convex  
   problems, 137  
 complexity of classes of multi-extremal  
   problems, 35  
 complexity related to a local oracle, 32  
 composition of measures, 367  
 cone, non-negative, 225  
 conjugate-gradients method CGM, 261,  
   312  
 constrained strongly convex and smooth  
   problems, 285  
 constrained mathematical programming  
   problem, 51  
 constraints, 10, 51  
 convex combination, 43  
 convex extremal problem, 51  
 convex function, 44  
 convex Lipschitz mapping, 226  
 convex mathematical programming  
   problem, 49  
 convex operator inequalities, 225–233  
 convex programming, 42 *et seq.*  
 convex programming methods of zeroth  
   order, 333 *et seq.*  
 convex set, 43  
 convex-concave function, 206  
 convex-concave game, 206  
 convex-concave Lipschitz pay-off  
   function, 208  
 convexity of a function, 44  
 convexity of a set, 43  
 convexity, strict, of a function, 48  
 convexity, strong, of a function, 255  
 Davidon–Fletcher–Powell method, 330  
 deterministic method, 17, 19  
 deterministic oracle, 50

diameter of a set, 362  
 dimension of a problem, 51  
 direct multiplication of spaces, 363  
 direct product of Banach spaces, 372  
 distance to a set, 362  
 distribution, laboriousness-error, 23  
 distribution, noise, 16  
 distribution of trajectories, 19, 20  
 distribution of trajectories, with noise, 19,  
   20  
 domain of a problem, 10  
 dual cone, 226  
 dual space, 370  
 Dvoretzky's theorem, 127, 179  
 efficiency of standard methods, 307  
 embedding, 170  
 epigraph, 40, 44  
 equivalence classes, 373  
 equivalence of methods, 30  
 error, absolute, of a point as an  
   approximate solution of a problem  
   ( $\epsilon(x, f)$ ), 13  
 error, \*-absolute ( $\epsilon_*(x, f)$ ), 13  
 error, mean, of a method  $v(\mathcal{A}, f)$ ,  $v_*(\mathcal{A},$   
    $f)$ , 23  
 error, mean, of a method on a class  
    $v(\mathcal{A}, \bar{\mathcal{A}})$ ,  $v_*(\mathcal{A}, \bar{\mathcal{A}})$ , 25  
 error, mean, on a mixture of classes, 31  
 error of a trajectory, 18  
 \*-error of a trajectory, 18  
 error, relative, scalar  $v(x, f)$ , 15  
 error, relative, vector  $v(x, f)$ , 14  
 error, scalar measure of, 13  
 error, strong, of a method on a problem  
    $\tilde{v}(\mathcal{A}, f)$ ,  $\tilde{v}_*(\mathcal{A}, f)$ , 23  
 error, strong, on a class  $\tilde{v}(\mathcal{A}, \bar{\mathcal{A}})$ ,  
    $\tilde{v}_*(\mathcal{A}, \bar{\mathcal{A}})$ , 25  
 error, vector measure of, 13  
 estimate at observation time  $n$ , 244  
 extremal problems with operator  
   constraints, 233–241  
 feasible point, 10  
 field of general convex problems  
    $C(G, E, m)$ , 52

field of Lipschitz-convex problems  
    $C_{\text{Lip}}(G, E, \|\cdot\|, m)$ , 52  
 field of mathematical programming  
   problems, 11, 227  
 field, weighted, 15  
 finite-dimensional space, 372  
 first-order oracle, 50  
 Fletcher–Reeves method, 313  
 function,  $\alpha$ -strictly convex, 48  
 function, Borel, 364  
 function, convex, 44  
 function, Lipschitz, 52  
 function,  $(l, Q)$ -strongly convex, 254  
 function,  $L$ -smooth, convex, 254  
 function, observation, 16  
 function, regular, corresponding to  
    $E, \|\cdot\|$ , 89  
 function,  $\sigma$ -measurable, 366  
 function, simple, 366  
 function, sub-linear, 370  
 function, summable, 366  
 functional, affine, 44  
 functional, objective, 10  
 functional, support, 45  
 functional constraints, 10  
*G*-trajectory, 18  
 game, convex-concave, 206  
 game, solution of, 207  
 general convex problem with linear  
   convergence, 42 *et seq.*  
 germ of a function, 46  
 gradient method, 95, 106, 258  
 graph, 40  
 Hahn–Banach theorem, 43, 370  
 higher-order oracle, 50  
 Hilbert space, 373  
 Hölder's inequality, 373  
 hyperplane separating  $x$  and  $G$ , 43  
 incompatible problem, 11  
 information space  $I$ , 16  
 integral with respect to a measure, 366  
 integration of functions with values in a  
   Banach space, 374

integration with respect to a measure,  
   366, 373–374  
 intensity of an oracle, 183  
 interior of a set, 362  
 Jensen's inequality, 44  
 joint distribution of trajectories with  
   noise, 19, 20  
 $k$ -dimensional  $p$ -asphericity of a norm,  
   125  
 Kelly's method,  $x$   
 $L$ -smooth convex functions, 254  
 $l$ -step method, 28  
 $(l, Q)$ -strongly convex functions, 254  
 laboriousness-error distribution, 23  
 laboriousness of a method on a class,  
   mean  $l(\mathcal{A}, \bar{\mathcal{A}})$ , strong  $\tilde{l}(\mathcal{A}, \bar{\mathcal{A}})$ , 25  
 laboriousness of a method on a mixture,  
   31  
 laboriousness of a method on a problem,  
   mean  $l(\mathcal{A}, f)$ , strong  $\tilde{l}(\mathcal{A}, f)$ , 23  
 laboriousness of a trajectory, 18  
 Lebesgue-complete measure, 365  
 Lebesgue theorems, 366–367  
 lemma on indistinguishability, 33  
 Lipschitz constant  $L_{\|\cdot\|}(f)$ , 44  
 Lipschitz function, 52  
 Lipschitz mapping, 226  
 local oracle, 17  
 mathematical programming problem, 10  
 mathematical programming problem,  
   compatible, 11  
 mathematical programming problem,  
   convex, 49  
 mathematical programming problem,  
   constrained, 51  
 mathematical programming problem  
   solution of games, 207  
 mathematical programming problem,  
   solvable, 11  
 mathematical programming problem,  
   unconstrained, 51  
 MD-methods, 83 *et seq.*  
 MD<sub>V</sub>-methods, 93

$\overline{\text{MD}}_V$ -methods, 107  
 $\overline{\text{MD}}_{1,n}$ -methods, 112  
 $\overline{\text{MD}}_p$ -methods, 112  
MD-methods, Hilbert-space version, 95  
MD-methods for classes of general convex problems, 107  
MD-methods for solving games, deterministic oracle, 210–219  
MD-methods for solving games, stochastic oracle, 219–224  
mean error of a method on a class  $v(\tilde{\mathcal{B}}, \tilde{\mathcal{A}}), v_*(\tilde{\mathcal{B}}, \tilde{\mathcal{A}})$ , 25  
mean error on a mixture of classes, 31  
mean error of a method on a problem  $v(\tilde{\mathcal{B}}, f)$ , 23  
mean \*-error of a method on a problem  $v_*(\tilde{\mathcal{B}}, f)$ , 23  
mean laboriousness of a method on a problem  $l(\tilde{\mathcal{B}}, f)$ , 23  
measure, 365  
measure, Borel, 367  
measure, regular, 367  
measure on a Polish space, 366  
measure space, 365  
measure, coincidence of, 368  
measures, composition of, 367  
measures, consistency of, 368  
method, conjugate-gradients, CGM, 261, 312  
method, Davidon–Fletcher–Powell, 330  
method, deterministic, 17  
method, Fletcher–Reeves, 313  
method,  $l$ -step, 28  
method of a given order, 51  
method of centres of gravity, MCG, 59–62  
method of centres of gravity, adaptive form, 80  
method of centres of gravity MCG\*, 70  
method of centres of gravity, modified, MMCG, 73–82  
method of centres of gravity, variants, 65  
method of mirror descent, MD, 82 *et seq.*  
method of mirror descent for solving constrained stochastic problems, 219–224

method of mirror descent for solving games, 210–219  
method of mirror descent for solving general convex problems, 107–124  
method of mirror descent for solving Lipschitz-convex problems, 93–98, 100, 105  
method of mirror descent for solving unconstrained stochastic problems, 188–197  
method, Polak–Ribi  re, 319  
method, regular, deterministic, 30  
method, stochastic, 17  
method, zeroth order, 51  
method, Zoutendijk, 327  
methods, equivalence of, 30  
minimax problems, 277  
Minkowski function, 49  
mixture of methods, 21  
mixture of problems, 31  
modified method of centres of gravity, MMCG, 73–82  
modulus of continuity of a convex function, 44  
modulus of smoothness of a regular function, 90  
modulus of strong convexity of the class  $H$ , 256  
natural oracle, 183  
noise distribution, 16  
noise space of an oracle, 16  
non-negative cone, 225  
norm in a linear space, 369  
norm, uniformly convex, 91  
normalizing factors  $r_i(f)$ , 14  
normalizing mapping  $r(f)$ , 15  
normed space, 369  
objective functional, 10  
observation function, 16  
optimal value of objective function  $f_*$ , 11  
oracle, 16  
oracle, deterministic, 16, 50  
oracle, first-order, 50  
oracle, higher-order, 50

oracle, intensity of, 183  
oracle, local, 17  
oracle, natural, 183  
oracle of type  $(G, v_0)$ , 55  
oracle of type (int  $G$ ,  $v_0$ ), 55  
oracle, stochastic, 16, 50  
oracle, unbiased, 183  
oracle, zeroth-order, 50  
ordering, 225  

-asphericity, affine, of a body,  $\alpha(G)$ , 126, 362  

-asphericity of a body  $\alpha_{\parallel \cdot \parallel}(G)$ , 125  

-asphericity of a  $k$ -dimensional body  $\alpha_{p,\parallel \cdot \parallel,k}(G)$ , 126  

-asphericity of a norm,  $k$ -dimensional  $\alpha_{p,k}(\|\cdot\|)$ , 125–126  

-asphericity of an affine  $k$ -dimensional body  $\alpha_{p,k}(G)$ , 126  
Polak–Ribi  re method, 319  
Polish space, 363  
Polish space, measure on, 366  
pre-ordering, 225  
preparatory transformation, 242  
probability measure, 365  
problem, compatible, 11  
problem, constrained, 51  
problem, incompatible, 11  
problem, smooth convex, 255  
problem, solvable, 11  
problem, unconditional, 51  
problem, unconstrained, 51  
product of measure spaces, 368  
‘proper’ application of MD-methods, 113–114  
quadratic programming, 261  
radius of a set, 362  
reflexive space, 371  
regular  $\mathcal{A}$ -mixture, 30  
regular deterministic method, 30  
regular function corresponding to  $E$ ,  $\|\cdot\|$ , 89  
regular measure, 367  
regular spaces, 89

regularization of a problem, 257  
‘re-initialization’, 314  
relative error, 228  
relative error, scalar  $v(x, f), v_*(x, f)$ , 15  
relative error, vector  $v(x, f)$ , 14  
reliability of an estimate, 244  
result of a trajectory, 18  
resultative trajectory, 18  
σ-additive measure, 365  
σ-algebra, 363  
σ-measurable function, 366  
saddle-point, 200  
sagitta, 91  
set of approximate solutions of a problem  $(G_*)$ , 12  
Shmulyan’s theorem, 91  
simple function, 366  
simplex, 60  
Slater condition, 70  
smooth convex problem, 255  
smooth function, 255  
smoothness properties, 51  
solution of a game, 207  
solution of a problem, 10  
solvable problem, 11  
space, Banach, 370  
space  $C_B(X)$ , 365  
space, dual, 370  
space, Hilbert, 373  
space, information, 16  
space,  $L(X, \mu)$ , 366  
space  $L_p(X, \mu)$ , 373  
space, normed, 369  
space  $\Omega$  of oracle noise, 16  
space, Polish, 363  
space, reflexive, 371  
space, regular, Banach, 89  
step-function, 374  
stochastic complexity of a class, 26  
stochastic method, 20  
stochastic oracle, 16, 50  
strict convexity of a function, 48  
strictly compatible problem, 69  
strong characteristics of a method on a class, 25

- strong characteristics of a method on a problem, 23  
 strong convexity of a function, 255  
 strong deterministic complexity of a class of problems  $N(v)$ , 26  
 strong error of a method on a class  $v(\tilde{\mathcal{B}}, \tilde{\mathcal{A}}), v_*(\tilde{\mathcal{B}}, \tilde{\mathcal{A}})$ , 25  
 strong error of a method on a problem  $\tilde{v}(\tilde{\mathcal{B}}, f), \tilde{v}_*(\tilde{\mathcal{B}}, f)$ , 23  
 strong laboriousness of a method on a problem  $I(\tilde{\mathcal{B}}, f)$ , 23  
 strong stochastic complexity of a class of problems  $\tilde{N}(v)$ , 26  
 strongly convex function, 255  
 strongly convex problems, 255 *et seq.*  
 strongly convex programming;  
     unconstrained problems, 272  
 sub-gradient, 46  
 sub-linear function, 370  
 successful preparatory transformation, 242  
 summable function, 366  
 support functional, 45  
 symbol of inaction,  $\phi$ , 18  
*t*-regular function, 159  
 Theorems:  
     Dvoretzky, 127, 179  
     Hahn–Banach, 370  
     Lebesgue, 366–367  
 Shmulyan, 91  
 von Neumann, 207  
 topology, weak, 371  
 topology, \*-weak, 372  
 trajectories, distribution of, 19, 20  
 trajectory, 18, 19  
 trajectory, error of, 18  
 trajectory, laboriousness of, 18  
 trajectory, result of, 18  
 unbiased oracle, 183  
 unconditional problem, 51  
 unconstrained mathematical programming problem, 51  
 uniformly convex norm, 91  
 variation of a function on a set  $V_G(f)$ , 44, 52  
 von Neumann’s theorem, 207  
 weak topology, 371  
 \*-weak topology, 372  
 weighted field of problems, 15  
 zeroth-order methods with deterministic oracle, 334–348  
 zeroth-order methods with stochastic oracle, 349–357  
 zeroth-order oracle, 50  
 Zoutendijk method, 327