STOCHASTIC GLOBAL OPTIMIZATION

Springer Optimization and Its Applications

VOLUME 1

Aims and Scope

Optimization has been expanding in all directions at an astonishing rate during the last few decades. New algorithmic and theoretical techniques have been developed, the diffusion into other disciplines has proceeded at a rapid pace, and our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in all areas of applied mathematics, engineering, medicine, economics and other sciences.

The series *Springer Optimization and Its Applications* publishes undergraduate and graduate textbooks, monographs and state-of-the-art expository works that focus on algorithms for solving optimization problems and also study applications involving such problems. Some of the topics covered include nonlinear optimization (convex and nonconvex), network flow problems, stochastic optimization, optimal control, discrete optimization, multi-objective programming, description of software packages, approximation techniques and heuristic approaches.

STOCHASTIC GLOBAL OPTIMIZATION

By

Anatoly Zhigljavsky
School of Mathematics
Cardiff University, UK

Antanas Žilinskas
Institute of Mathematics and Informatics
Vilnius, Lithuania

Springer

Anatoly Zhigljavsky
School of Mathematics
Cardiff University, UK

Antanas Žilinskas
Institute of Mathematics and Informatics
Vilnius, Lithuania

Library of Congress Control Number: 2007933821

ISBN-13: 978-0-387-74022-5 e-ISBN: 978-0-387-74740-8

Printed in the United States of America.

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Preface

This book aims to cover major methodological and theoretical developments in the field of stochastic global optimization. This field includes global random search and methods based on probabilistic assumptions about the objective function.

We discuss the basic ideas lying behind the main algorithmic schemes, formulate the most essential algorithms and outline the ways of their theoretical investigation. We try to be mathematically precise and sound but at the same time we do not often delve deep into the mathematical detail, referring instead to the corresponding literature. We often do not consider the most general assumptions, preferring instead simplicity of arguments. For example, we only consider continuous finite dimensional optimization despite the fact that some of the methods can easily be modified for discrete or infinite-dimensional optimization problems.

The authors' interests and the availability of good surveys on particular topics have influenced the choice of material in the book. For example, there are excellent surveys on simulated annealing (both on theoretical and implementation aspects of this method) and evolutionary algorithms (including genetic algorithms). We thus devote much less attention to these topics than they merit, concentrating instead on the issues which are not that well documented in literature. We also spend more time discussing the most recent ideas which have been proposed in the last few years.

We hope that the text of the book is accessible to a wide circle of readers and will be appreciated by those interested in theoretical aspects of global optimization as well as practitioners interested mostly in the methodology. The target audience includes graduate students and researchers in operations research, probability, statistics, engineering (especially mechanical, chemical and financial engineering). All those interested in applications of global optimization can also benefit from the book.

The structure of the book is as follows. In Chapter 1, we discuss general concepts and ideas of global optimization in general stochastic global optimization in particular. In Chapter 2, we describe basic global random search algorithms, study them from different view-points and discuss various probabilistic and statistical aspects associated with these algorithms. In Chapter 3, we discuss and study several more sophisticated global optimization techniques including random and semi-random coverings, random multistart, stratified sampling schemes, Markovian algorithms and finally the methods of generations. In Chapter 4, techniques based on the use of statistical models about the objective function are studied. The Introduction and Chapter 1 are written by both co-authors. Chapters 2 and 3 are written by A.Zhigljavsky, Chapter 4 is written by A.Žilinskas.

A.Zhigljavsky is grateful to his colleagues at Cardiff University (V.Savani, V.Reynish, E.Hamilton) who helped with typing and editing the manuscript and patiently tolerated his monologues on different aspects of global optimization. He is also grateful to his long-term friends and collaborators Luc Pronzato and Henry Wynn for stimulating discussions and to his former colleagues from St.Petersburg University – M.Chekmasov, V.Nevzorov, S.Ermakov, and especially to M.Kondratovich, V.Nekrutkin and A.Tikhomirov. Significant parts of Sects. 2.4, 2.5 and 3.3 are based on the joint work of A.Zhiglajvsky and M.Kondratovich; Sect. 3.4 is fully based on the results of V.Nekrutkin and A.Tikhomirov who very much helped with writing a summary of their results.

A.Žilinskas thanks the Institute of Mathematics and Informatics at Vilnuis for facilitating his work on the book, and J.Mockus for introducing him to the field of global optimization many years ago. The work by A.Žilinskas has been partly supported by the Lithuanian State Science and Studies Foundation. The material on one-dimensional algorithms included into Chapter 4 is based mainly on joint publications by A.Žilinskas and J.Calvin. Before starting work on the book, the authors invited Jim Calvin to become a co-author. Although he rejected our invitation in view of his involvement in other projects, we consider him a virtual co-author of the mentioned part of the book.

Both authors thank Rebecca Haycroft and Julius Žilinskas as well as the two referees for their careful reading of the manuscript and constructive remarks. Especially, the authors are very grateful to the editor of the series Panos Pardalos for his encouragement with this project.

Cardiff, Vilnuis

Anatoly Zhigljavsky
Antanas Žilinskas

Contents

Introduction

Global optimization is a fast growing area. Its importance is primarily related to the increasing needs of applications in engineering, computational chemistry, finance and medicine amongst many other fields.

The area of global optimization is well documented in publications. There is the *Journal of Global Optimization* fully devoted to the developments in this area. Other journals on optimization, such as *Journal of Optimization Theory and Applications* and *Mathematical Programming* regularly publish research papers on global optimization. During the last few years, monographs on global optimization regularly appear predominantly in the Kluwer/Springer series *Nonconvex Optimization and Its Applications* edited by Panos Pardalos. Several conferences each year are either fully devoted to global optimization or have sessions on this subject. For the state of the art in theory and methodology of global optimization we refer to two volumes of *Handbook of Global Optimization* [122], [180] and to the paper by Floudas et al, *Global Optimization in the 21st Century* [79].

The problem of global optimization is difficult. Although classical optimization theory can not be directly applied in the problems of global optimization, the traditional tools such as convex analysis are extensively used in constructing global optimization methods. This approach constitutes an essential part of deterministic global optimization. For example, remarkable achievements have been made in constructing minimization algorithms for concave functions in convex regions and also in the minimization of differences of convex functions.

Deterministic global optimization is a well developed mathematical theory which has many important applications. Recently, several monographs by Floudas [78], Horst and Tuy [124], Strongin and Sergeyev [233], Tuy [256], and a text-book by Horst, Pardalos and Thoai [123] on deterministic global optimization have been published. The current state of affairs in deterministic global optimization is presented, e.g. in the book by Floudas and Pardalos [84], and a special volume of the *Mathematical Programming* (Ser.B, v. 103, 2005). However, in a situation close to 'black box' optimization the deterministic models often do not adequately represent the available information about the objective function.

In the cases where the objective function is given as a 'black box' computer code, the optimization problem is especially difficult. Stochastic approaches can often deal with problems of this kind much easier and more efficiently than the deterministic algorithms. Other big advantages of stochastic methods are related to their relative simplicity, their suitability for the problems where the evaluations of the objective function are corrupted by noise, and their robustness with respect to the growth of dimension.

Many algorithms where randomness and/or statistical arguments are involved have been proposed heuristically. Some algorithms are based on analogies with natural processes. Well-known examples of such algorithms are evolutionary optimization and simulated annealing. Heuristic global optimization algorithms are very popular in applications. Results of their application are frequently published in various engineering journals. Papers on evolutionary global optimization can be found in books and journals on evolutionary computing. Simulated annealing, as an important method, has been intensively studied by many authors. However, these studies often consider simulated annealing only, without the broader context of global optimization.

Although there are many publications on stochastic global optimization algorithms and their applications, they are scattered throughout various sources. It is not easy to grasp the state of the art in the field. Therefore, the authors believed that there was a serious need for a book presenting the main ideas and methods in stochastic global optimization from a unified view-point. The authors also believe that they have made an honest attempt to achieve this aim.

The stochastic global optimization techniques are not represented in literature nearly as well as the deterministic approaches. The recent monographs by Mockus [165] and Zabinsky [267] mainly cover the results related to the authors' research. The monographs by the authors of this book [248, 271, 273, 276] represent the stochastic approach to global optimization as it was fifteen-twenty years ago. The authors' current aim is to summarize the current state of affairs in stochastic global optimization and present the recent progress. For completeness of presentation, the key material of the above mentioned monographs is also included into this book. The book also contains a fair amount of new results.

The theory of stochastic global optimization is the main topic of the book. Although the applications of corresponding algorithms are very important, we have restricted the discussion of applications to a few short examples only. We nevertheless believe that the monograph will be useful to a wide circle of readers whose main interest lies in the applications of global optimization. The target audience also includes specialists in operations research, probability, statistics, engineering and other fields.

Notation

$\mathbb{R}$	space of real numbers
$\mathbb{R}^d$	d-dimensional Eucledian space
A	feasible region (optimisation space); typically, A is a compact subset of $\mathbb{R}^d$ with non-zero volume
$f(\cdot)$	objective function given on A; this function is to be minimized
m	minimum of $f(\cdot)$ on A: $m = \min f = \min_{x \in A} f(x)$
x_*	any global minimizer of $f(\cdot)$; that is, x_* is any point such that $f(x_*) = m$
A_*	set of all global minimizers of $f(\cdot)$: $A_* = \{x_* \in A : f(x_*) = m\}$
$\mathcal{B}$	σ-algebra of Borel subsets of A
$\mathrm{vol}(Z)$	volume (d-dimensional Lebesgue measure) of $Z \in \mathcal{B}$
ρ	a metric on $\mathbb{R}^d$
ρ_2	Euclidean metric on $\mathbb{R}^d$
$\vert\vert \cdot \vert\vert$	Euclidean norm on $\mathbb{R}^d$
$\mathcal{F}$	set of all possible objective functions
$\mathrm{Lip}\,(A, L, \rho)$	class of functions satisfying the Lipschitz condition with known constant L in metric ρ: $\mathrm{Lip}\,(A, L, \rho) = \{f : \vert f(x) - f(z)\vert \le L\rho(x, z)\ \forall x, z \in A\}$
$B(x, \varepsilon)$	$= \{z \in A : \vert\vert z - x\vert\vert \le \varepsilon\}$, the ball (in Euclidean metric) in A of radius ε centred at x; more precisely, $B(x, \varepsilon)$ is the intersection of the set A with the ball in $\mathbb{R}^d$ of radius ε and centre at x
$B(\varepsilon)$	$= B(x_*, \varepsilon)$, the ball centered at the global minimizer x_*
$B(x, \varepsilon, \rho)$	$= \{z \in A : \rho(z, x) \le \varepsilon\}$, the ball (in metric ρ) in A of radius ε centred at x
$W(\delta)$	$= \{x \in A : f(x) \le m + \delta\}$
W_x	$= W(f(x) - m) = \{z \in A : f(z) \le f(x)\}$
$x_i (i = 1, ..., n)$	n points where the objective function $f(\cdot)$ has been evaluated
$y_i = f(x_i)$	result of the objective function evaluation at the point x_i
y_{on}	$= \min_{i=1\ldots n} y_i$, the smallest value of the objective function in n evaluations (*record value* or simply *record*)
x_{on}	the point x_i with smallest $i \le n$ such that $f(x_i) = y_{on}$ (*record point*)
$\mathrm{ess\,inf}\,\eta$	essential infimum of a random variable η: $\mathrm{ess\,inf}\,\eta = \inf\{a : \Pr\{\eta \ge a\} > 0\}$
c.d.f.	cumulative distribution function
$F^{-1}(s)$	$= \inf\{t : F(t) \ge s\}$, the inverse function of the c.d.f. $F(t)$
P	a probability distribution on A; more precisely, P is a probability distribution on the measurable space $(A, \mathcal{B})$
P_Z	the distribution on $Z \subseteq A$ defined by $P_Z(U) = P(U \cap Z)/P(Z)$ for all $U \in \mathcal{B}$, where P is a distribution on $(A, \mathcal{B})$ and $Z \in \mathcal{B}$
$\eta \overset{d}{=} \nu$	for random variables (vectors) η and ν means equality of their c.d.f.'s
$a_n \sim b_n,\ n \to \infty$	$\Longleftrightarrow$ the limit $\lim_{n\to\infty} a_n/b_n$ exists and equals 1; convergence in distribution is assumed if $\{a_n\}$ and $\{b_n\}$ are sequences of random variables
κ_n	$(1/n)$-quantile of the c.d.f. $F(\cdot)$: $\kappa_n = \inf\{u \vert F(u) \ge 1/n\}$
$\xrightarrow{\mathcal{D}}$	convergence in distribution; that is, $\xi_n \xrightarrow{\mathcal{D}} \xi\ (n \to \infty)$ for random variables ξ_n and ξ, if $\Pr(\xi_n \le x) \to \Pr(\xi \le x)$, $n \to \infty$, for all x such that $\Pr(\xi = x) = 0$
$\Phi(\cdot)$	the c.d.f. of the standard normal distribution: $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du$
PRS	pure random search
i.i.d.r.v.	independent identically distributed random variables (vectors)
l.h.s. / r.h.s.	left-hand side / right-hand side