Natural Computing Series

More information about this series at http://www.springer.com/series/4190

Mike Preuss

Multimodal Optimization by Means of Evolutionary Algorithms

Springer

Mike Preuss
Lehrstuhl für Wirtschaftsinformatik und Statistik
Westfälische Wilhelms-Universität Münster
Münster, Germany

Series Editors
G. Rozenberg (Managing Editor)

Th. Bäck, J.N. Kok, H.P. Spaink
Leiden Center for Natural Computing
Leiden University
Leiden, The Netherlands

A.E. Eiben
VU University Amsterdam
The Netherlands

ISSN 1619-7127
Natural Computing Series
ISBN 978-3-319-07406-1 ISBN 978-3-319-07407-8 (eBook)
DOI 10.1007/978-3-319-07407-8

Library of Congress Control Number: 2015956174

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

to Carla

Foreword

Half a century ago, with more and more computers available at university and research labs, numerical optimization became en vogue. Direct search methods like those of Rosenbrock, Nelder & Mead, and M.J.D. Powell, to name just a few, helped to solve a lot of nonlinear, analytically intractable problems approximately. Unfortunately, such hill climbers found their way to only one local minimum or maximum in the vicinity of the given starting points in the search space. The case of multiple local optima seemed to be better treated by population-based methods like evolutionary algorithms, including genetic algorithms. Numerous experiments with properly tuned internal parameters of such black-box methods, as they were called, were published to demonstrate the suitability of those bio-inspired search procedures in multimodal landscapes. But there was no proof, no guarantee to find the optimum optimorum or global optimizer. Even worse, no satisfactory definition of all problems occurring in the case of multiple local optima existed. This situation remained the same for many years despite dozens of international conferences in the field of evolutionary and natural computation.

Mike Preuss' book is the first comprehensive treatment of many problems to handle multimodal optimization tasks by means of evolutionary algorithms in a structured manner. Indeed, there are a couple of different aspects to be obeyed when not only one local extremum exists. Sometimes only the global optimum is wanted, but it can be located at several locations (optimizers). Otherwise, all or only some of the optima are wanted. If not all of them, under which criteria should one select them? The author is probably the pioneer to create a taxonomy of the multitude of possible situations existing in multimodal optimization. One rather old idea in the field is niching: A population of seekers is split into subgroups each searching only in a subspace of the entire search space. There have been numerous such attempts, most of which are mentioned, characterized and evaluated, i.e., criticized. So far no satisfactory theory in the area of niching exists. This work is the first and so far only one to evaluate niching strategies rigorously to find out which ones are appropriate for which purpose. When one has found the promising basins of attraction of hopefully just one local extremum each, it is usual to start a traditional local optimum search, in this case one of the currently most successful evolution strategies (ES). Mike Preuss has combined his favorite basin detection method with such modern ES versions and sent them to benchmark-assisted international competitions—and won! Besides all previously mentioned leading-edge features of the book, this fact should attract interested readers particularly.

Hans-Paul Schwefel

Dortmund, Easter 2014

Foreword

Optimization problems arise in a wide variety of areas ranging from production, logistics, biology, and medicine, to engineering. The task in optimization is to find a solution, that is, an assignment of values to specific decision variables that gives the best possible value for a given objective function. In many cases, finding an optimal solution is a very difficult task due to nonlinearities in the objective functions and the possible occurrence of (many) locally optimal solutions that trap the search process. If finding a single optimal solution is already difficult, finding several or all optimal solutions is even more difficult; and it is this latter, so-called multi-modal optimization task that the author tackles in this book.

The focus in this book is apparently on evolutionary algorithms as solution methods and black-box continuous function optimization as the more specific problem class. This focus may, at first sight, limit the contributions of the work to these specific areas. However, this is not really true as the book contains a large number of more generic results and insights that make it relevant also beyond the field of evolutionary algorithms. In fact, any heuristic method for multimodal optimization should profit from techniques to identify basins of attraction of optima to make the search process more efficient, that is, from the techniques that are analyzed and designed here. One part of the contributions of this book develops formal models to analyze from a theoretical perspective the potential impact such techniques may have. This analysis is particularly interesting as it relates properties of the search space to the potential advantages of the considered techniques. While in the theoretical analysis specific techniques for basin identification and other tasks may be modeled, when it comes to actually solving multi-model optimization problems, effective algorithmic techniques need to be designed for implementing them. Another, generic contribution of the book is the development of the nearest-better clustering method for basin identification. This method is then used as a supporting tool for evolutionary algorithms for multi-modal optimization; however, it is directly applicable also to improve other heuristic search techniques. (Actually, the used evolutionary algorithms, in particular, the well-known co-variance matrix evolution strategy, could also be seen as efficient stochastic local search heuristics for black-box continuous optimization giving evidence for this claim.) The resulting multi-modal optimization method is particularly effective and shows excellent performance. This is confirmed by the fact that the resulting algorithm was the top-performer in a recent benchmark competition on multi-modal optimization.

Apart from these contributions, I would like to highlight two main additional ones. The first one is that the author has a consistent personal view of the research on multi-modal optimization and clearly organizes the contributions described so far in the literature. This may seem a minor contribution at first sight, but in the context of multi-modal optimization it becomes an important one as (i) many contributions have been obtained within different fields and many researchers are apparently not aware of the existing links, and (ii) many notions such as niching are used in very different senses and therefore lead to confusion even inside the same research community. The second contribution concerns the experimental evaluation. Unfortunately, in the history of evolutionary computation and, more generally, heuristic search algorithms, a sound experimental methodology has not always received the attention that is actually required. This book is exemplary in the adoption of a sound experimental methodology (which actually the author has helped to develop) and it will hopefully help to convince fellow researchers to adopt such methodologies in their own research.

In conclusion, I think that this book contains a large number of in-depth research results, and if multi-modal optimization is your research subject, this book is clearly a milestone that has to be read. In addition, the book provides a wealth of additional contributions that will make it an enjoyable and beneficial read even beyond the particular research subject treated. I therefore wish the book all the deserved success and a large future audience.

Thomas Stützle

Brussels, June 2014

Preface

This book is the result of a *very* long journey into optimization, and, more specifically, into *evolutionary computation*. This journey would not have been possible without the support of my family. I am very grateful to my parents Herbert and Christa, my sister Jennifer and her family, and of course to my daughters Janinka and Merle. Of course there are many more people who acted as signposts and/or motivators, and they shall be mentioned as well. In order to avoid a boring long list, I will try to wrap their names into a short chronological report before giving an overview over the book itself.

The first event that connected me to (evolutionary) optimization was a radio feature I heard soon after starting my studies. It dealt with optimization by adaptation of concepts from nature carried out by Hans-Paul Schwefel and the people at his Chair of Systems Analysis at the TU Dortmund. Joachim Sprave then led me into the world of parallel evolutionary computing, and provided me with a very important insight: not everything that is written in a book is right just because it is printed. In this environment, I first met Beate Bollig, who later on consistently reminded me to finish my dissertation, up to when this was actually the case.

After going international (EU project: DREAM) due to Thomas Bäck, I had the chance to experience my first real scientific cooperation, for which I have to thank Márk Jelasity, Gusz Eiben, Ben Paechter, and Marc Schoenauer. Meanwhile, criticism of the experimental evaluation of optimization algorithms was on the rise. Thomas Bartz-Beielstein and I teamed up for years in order to provide techniques and guidelines for countering this criticism. Chapter 2 is my view on experimental work and I applied the described methodology to all experiments in this book.

With Günter Rudolph and Boris Naujoks, I explored the foundations of multi-objective optimization and learned how to deal with engineers in several real-world optimization projects. Ruxandra and Catalin Stoean are my long-term Romanian connection, together we have challenged many interesting problems and developed nice algorithmic techniques for multimodal optimization and evolutionary support vector machines. Of all the people enlisted here, Ofer Shir is probably the one I have collaborated with who was or is most devoted to niching in evolutionary algorithms. We tried to define niching and multimodal optimization at a time when the second term was not yet commonly in use.

Heike Trautmann and I first met in Singapore, only to find out that we worked in related fields at the same university. Besides steadily pushing me to finish my dissertation, she enriched my life in many ways, not the least of which is a much stronger inclination towards statistical techniques. With Jens Jägersküpper, I undertook an exciting excursion into theory, or rather, algorithm engineering. Catherine McGeoch taught me to pose the right questions, and from Thomas Stützle I learned that evolutionary algorithms are not always the answer, but often a useful start.

This work would have been finished much earlier if I had not been sidetracked by the fascinating world of game AI. While starting research in this area together with Simon Wessing and Jan Quadflieg, I had the opportunity to meet Julian Togelius and Georgios Yannakakis, and, a little bit later on, Paolo Burelli. I owe a lot to all five of you! That it could be finished at all is probably due to Hans-Paul Schwefel, who provided encouragement when it was needed, and Simon Wessing, Bernd Bischl, and Günter Rudolph, who helped me to resolve the last important questions.

Last but not least, I would like to thank Ronan Nugent for recognizing the scientific contribution of this book, and for guiding me through the publishing process.

What is this book about? The field of multimodal optimization is just forming, but of course it has its roots in many older works, namely niching, parallel evolutionary algorithms, and global optimization. My aim is to bring all these together and thereby help to shape the field by collecting use cases, algorithms, and performance measures. In my view, it is very important to exactly define what the goals of such an optimization process are and also to obtain a good understanding of what the algorithms actually do during this process, especially with respect to the properties of the tackled optimization problems. More concretely, the main objectives of this work are listed in Sect. 1.4.

The algorithms I provide for basin identification and optimization are meant as a step forward, not as a definitive answer. I presume that there is still a lot of yet undiscovered potential in research on multimodal optimization, and I would like to encourage more research in this area.

Concerning the structure and usage of this book, the reader may find Sect. 1.5 useful, it contains a short description of the chapters and indicates which parts may be most interesting when addressing the different aspects of multimodal optimization.

Have a fun!

Mike Preuss

Bochum, July 2014

Contents

Nomenclature

γ	Euler-Mascheroni constant, approximately 0.577
λ	number of offspring individuals generated in one generation
μ	number of individuals who survive selection and are parents for the next generation
σ	step size/mutation strength
σ^0	initial step size/mutation strength
A	desired set of coupons
b	number of basins an (abstract) optimization problem possesses
c	number of basins covered by an algorithm at a certain time
D	number of search space dimensions of the treated problem
$d(\mathbf{x}_1, \mathbf{x}_2)$	distance between two search points
$f(\mathbf{x})$	objective (fitness) function (value) of search point $\mathbf{x}$
f^{*G}	global optimum, function value of (a) global optimizer
$G(r)$	distribution function for the nearest neighbor distance r
k	number of selected neighbors
P_t	population of search points at time t
$p_{BI}(\mathbf{x}_1, \mathbf{x}_2)$	probability of correctly identifying that two search points are located in the same basin
$p_{BR}(\mathbf{x}_1)$	probability of correctly detecting that the basin of a search point has already been found
R	redundancy factor
t_1	point in time during a heuristic optimization process when the first result can be delivered (usually very early)
t_2	first hitting time of the global optimum in a heuristic optimization process
t_3	latest first hitting time of all basins in a heuristic optimization process: all basins are discovered
t_c	cycle time, the number of repetitions before a random process arrives at the same state again
X	the whole search space
Z_n	expected waiting time for drawing n of a fixed set of coupons
$\mathbf{x}$	coordinate vector in the search space that determines a search point
$\mathbf{x}^{*L}$	local optimizer
$\mathbf{y}$	set of values for an abstract algorithm performance measure
$\mathscr{B}$	basin system, consisting of single basins B_i
$\mathscr{C}$	clustering, consisting of clusters (subsets C_i)

$\mathscr{D}$	set that contains all decided clusters of a clustering
AAHD	augmented averaged Hausdorff distance
AE	algorithm engineering
AHD	averaged Hausdorff distance
all-global	the target of the optimization is to detect all global optimizers
all-known	the optimization shall detect all existing optimizers, global and local
AOV	average objective value
APD	augmented peak distance
BA	basin accuracy
basin (of attraction)	the search space area from which a local search algorithm converges to a (local) optimizer
basin identification	detect the locations of the different basins of an optimization problem by identifying which search points belong to which basins
basin recognition	decide if the basin a search point belongs to is already known
BBOB	black box optimization benchmarking, an "*instance*" of COCO, held in the form of GECCO workshops (up to now 2009, 2010, 2012, 2013, 2015)[1]
BFGS	quasi-Newton method named after its inventors Broyden, Fletcher, Goldfarb, and Shannon
BFS	breadth-first search
BIPOP-CMA-ES	CMA-ES variant splitting the search effort between small and large, increasing population sizes
black box optimization	optimization without any knowledge about the system that generates objective function values, e.g., no analytical form or derivatives are given
BR	basin ratio
CCP	coupon collectors problem
CEC	annual (international) conference on evolutionary computation
ceteris paribus conditions	the experiment is repeated under exactly the same conditions, except for the starting time
CI	computational intelligence
CMA-ES	covariance matrix adaptation evolution strategy, introduced in Hansen and Ostermeier [103] and in details further developed since
COCO	comparing continuous optimizers, a platform for comparison of real-parameter global optimization algorithms, see http://coco.gforge.inria.fr/
COGA	cluster-oriented genetic algorithms
CSR	complete spatial randomness
DACE	design and analysis of computer experiments, deterministic precursor of SPO
dADE/nrand/*	DE/nrand/* with an additional dynamic archive
DBF	detected basin fraction
DE	differential evolution
DE/nrand/*	differential evolution variant that uses nearest neighbors as base vector for generating offspring
DECG/DELG/DELS	different differential evolution variants that emphasize parallel local searches
decided cluster	cluster of which the majority of constituents are located in the same basin (its main basin)
design	a set of design sites
design site	equivalent to experimental unit, here meaning the point in the algorithm parameter space that is tested

[1] web page of the 2015 issue: http://coco.gforge.inria.fr/doku.php?id=bbob-2015

DFS	depth-first search
DMM	detect-multimodal, short name for the hill-valley method
DOE	design of experiments, a set of techniques for setting up experiments, first introduced by Fisher [81]
DPI	dynamic peak identification
EC	evolutionary computation
ELA	exploratory landscape analysis
epistasis	related to separability, but defined over binary spaces: one phenotypical attribute is influenced by several genes or vice versa
ERT	expected running time
ES	evolution strategies
ETP	empirical tuning potential
F-Race	parameter tuning method
FMPM	funnel-based extension of the MPM generator
freestanding cluster	decided cluster whose main basin is different to all other clusters
GA	genetic algorithms
GECCO	annual (international) conference on evolutionary computation
GLOBAL	2-phase global optimization algorithm that employs single-linkage clustering in the global phase and BFGS for local seaches
global optimizer	location (possibly one of several) in the search space for which the objective function returns the global optimum
global optimum	best numerical value that is returned by an objective function
good-subset	the optimization shall detect a small subset of very good optimizers that are well spread over the search space
GP	genetic programming
hill-valley method	mechanism for detecting if two search points reside in the same basin by placing at least one point between them
ILS	iterated local search
IPOP-CMA-ES	CMA-ES variant with increasing population size
LHS or LHD	Latin hypercube sampling, Latin hypercube design, space-filling sampling method used within SPO as alternative to purely random (MC) sampling
local optimizer	location in the search space that corresponds to a local optimum
local optimum	objective function value of a point in search space that cannot be improved by making an infinitesimal step in any direction (note that this includes global optima)
locality principle	search points in the direct vicinity shall be more similar to each other than to more distant search points (in terms of objective values)
MC	Monte Carlo, meaning that a process (e.g., a sampling process) works completely at random
mixed cluster	cluster that does not have the majority of constituents in any single basin
MPM	multiple peaks model, test problem generator with randomly placed peaks
multimodal	objective function with at least 2 global optimizers
multimodal optimization	detecting several optimizers of a multimodal problem at once
multimodalCutProbs	NEA1/NEA2 parameter that determines how the DMM (hill-valley) method is used
NBC	nearest better clustering, a topological clustering method that makes use of objective values of a population next to the search space locations
NBC-CMA-ES	early version of the NEA1 algorithm

NEA1	niching evolutionary algorithm 1, first of two niching algorithms suggested by the author, basically a combination of initial random sample, NBC and CMA-ES, employing several populations concurrently
NEA2	niching evolutionary algorithm 2, suggested by the author, similar to NEA1 but doing local searches sequentially
niching (in optimization)	method to (implicitly or explicitly) recognize different basins of attraction and inject this information into an optimization algorithm
NND	nearest neighbor distance
one-global	the optimization shall find one global optimizer as fast as possible
optimizer	in the optimization context usually meant as *local optimizer*
optimum	contrary to common language, in the optimization context this is often understood as *local optimum*
PA	peak accuracy
ParamILS	iterated local search applied to the (algorithm) parameter space
PD	peak distance
PR	peak ratio
PSO	particle swarm optimization
QABR	quantity-adjusted basin ratio
QAPR	quantity-adjusted peak ratio
QMC	quasi-Monte Carlo, meaning that a deterministic process is used to emulate MC behavior
R5S	representative 5 selection
redundancy factor	ratio of actually performed local searches to necessary local searches (number of basins b)
REVAC	relevance estimation and value calibration, a tuning method
rule 2	extension of the NBC clustering method that takes the indegree of nodes in the nearest-better graph into account
SD	sum of distances
SDNN	sum of distances to nearest neighbor
search point	a location in the search space, in the real-valued case of zero volume, associated with at least one objective value
separability	separable functions can be solved by decomposing them into D 1-dimensional functions and aggregating the obtained optima, there is no interaction between the different variables
sigmaToDistance	NEA1/NEA2 parameter that controls how the step size is regulated according to the estimated basin size
SPD	Solow-Polasky diversity
SPO	sequential parameter optimization, model-based parameter tuning approach
sqr	semi-quartile range
surrounded cluster	decided cluster with the same main basin as at least one other cluster
TolFun	parameter of the CMA-ES stopping rules that refers to differences in objective function values
TSC/TSC2	topological species conservation algorithm
TSP	traveling salesperson problem, typical combinatorial optimization test problem
UCF	useful cluster fraction
unimodal	objective function with only one global optimizer
w.l.o.g.	without loss of generality
weak local optimum	the optimum does not correspond to a single search point but to a set of search points (e.g., a line or a plateau)