Paulo Cortez

Modern Optimization with R

Second Edition

Springer

Use R!

Series Editors

Robert Gentleman, 23andMe Inc., South San Francisco, USA

Kurt Hornik, Department of Finance, Accounting and Statistics, WU Wirtschaftsuniversität Wien, Vienna, Austria

Giovanni Parmigiani, Dana-Farber Cancer Institute, Boston, USA

Use R!

This series of inexpensive and focused books on R will publish shorter books aimed at practitioners. Books can discuss the use of R in a particular subject area (e.g., epidemiology, econometrics, psychometrics) or as it relates to statistical topics (e.g., missing data, longitudinal data). In most cases, books will combine LaTeX and R so that the code for figures and tables can be put on a website. Authors should assume a background as supplied by Dalgaard's Introductory Statistics with R or other introductory books so that each book does not repeat basic material.

More information about this series at http://www.springer.com/series/6991

Paulo Cortez

Modern Optimization with R

Second Edition

Springer

Paulo Cortez
Department of Information Systems
University of Minho
Guimarães, Portugal

Preface

At present, we are in the age of data, where multiple individual and organizational activities and processes generate big data that can be processed using information technology (IT). We are also in a fast changing world. Due to several factors, such as globalization and technological improvements, both individuals and organizations are pressured for improving their efficiency, reducing costs, and making better informed decisions. This is where optimization methods, supported by computational tools, can play a key role.

Optimization is about minimizing or maximizing one or more goals and it is useful in several domains, including agriculture, banking, control, engineering, finance, management, marketing, production, and science. Examples of real-world applications include the optimization of construction works, financial portfolios, marketing campaigns, and water management in agriculture, just to name a few.

Modern optimization, also known as metaheuristics, is related with general-purpose solvers based on computational methods that use few domain knowledge from the addressed task, iteratively improving an initial solution (or population of solutions) to optimize a problem. Modern optimization is particularly useful for solving complex problems for which no specialized optimization algorithm has been developed, such as problems with discontinuities, dynamic changes, multiple objectives, and hard and soft restrictions, which cannot be handled easily by classical operational research methods.

Although modern optimization often incorporates random processes within their search engines, the overall optimization procedure tends to be much better than pure random (Monte Carlo) search. Several of these methods are naturally inspired. Examples of popular modern methods that are discussed in this book are simulated annealing, tabu search, genetic algorithms, genetic programming, grammatical evolution, NSGA II (multi-objective optimization), differential evolution, particle swarm, and ant colony optimization.

R is a free, open-source, and multiple platform tool (e.g., *Windows*, *macOS*, *Linux*) that was specifically developed for statistical analysis. Currently, there is an increasing interest in using R to perform an intelligent data analysis. In effect, the R community is very active and new packages are being continuously

created. For example, there are currently more than 16,170 packages available at the Comprehensive R Archive Network (CRAN) (http://www.r-project.org/), which enhance the tool capabilities. In particular, several of these packages implement modern optimization methods.

There are several books that discuss either modern optimization methods or the R tool. However, within the author's knowledge, there is no book that integrates both subjects and under a practical point of view, with several application R code examples that can be easily tested by the readers. Hence, the goal of this book is to gather in a single document (self-contained) the most relevant concepts related with modern optimization methods, showing how such concepts and methods can be implemented using the R tool. It should be noted that some of the explored modern optimization packages have a minimal documentation (e.g., with no vignettes or very short help pages). Thus, this book can be used as complementary source for better understanding how to use these packages in practice. Moreover, some metaheuristics (e.g., simulated annealing, genetic algorithms, differential evolution) are implemented in distinct packages. In such cases, the book addresses all these packages, showing their similarities and differences, thus providing a broader (and hopefully not biased) view for the readers.

This book addresses several target audience groups. Given that the R tool is free, this book can be easily adopted in several bachelor's or master's level courses in areas such as Artificial Intelligence, operations research, data science, decision support, business intelligence, business analytics soft computing, or evolutionary computation. Thus, this book should be appealing for bachelor's or master's students in computer science, information technology, or related areas (e.g., engineering, science, or management). The book should also be of interest for two types of practitioners: R users interested in applying modern optimization methods and non R expert data analysts or optimization practitioners that want to test the R capabilities for optimizing real-world tasks.

First Edition Feedback

The first edition of this book was published in 2014 and it included a total of 188 pages. The book edition covered a publication gap, since there were no written books about the usage of metaheuristics with the R tool. Indeed, after publication, the book received positive feedback (http://www3.dsi.uminho.pt/pcortez/mor/), including two book reviews:

- "The author provides valuable comments about the pros and cons of various optimization methods. . . . This book makes a good contribution in to the literature of modern optimization. It is well written and structured." – Diego Ruiz, Journal of Statistical Software, 2016 (https://www.jstatsoft.org/article/view/v070b03/v70b03.pdf); and

- "...the positive characteristics of Cortez's Modern Optimization with R make this book a must-have for graduate students and any researchers willing to delve into new algorithms to solve their optimisation tasks. I was pleasantly impressed by the book's organisation and clarity, by its comprehensive and coherent topic coverage, and by its many outstanding numerical illustrations using R." – Matthieu Vignes, Australian & New Zealand Journal of Statistics, 2017 (https://onlinelibrary.wiley.com/doi/epdf/10.1111/anzs.12186).

Moreover, the book was adopted as a textbook by several bachelor's and master's level university courses, such as: "Discrete and Continuous Optimization" (University of California San Diego, USA); "Operations Research" (University of Loisville, USA); "Stochastic Optimization" (North Carolina State University, USA); "Advanced Optimization" (Warsaw School of Economics, Poland); "Machine Learning for Finance" (Barcelona Graduate School of Economics, Spain); and "Optimization Techniques" (National Institute of Technology Delhi, India). Furthermore, the book has received 65 Google scholar (https://scholar.google.com/) citations. In 2019, the book was among the top 25% most downloaded e-books of its respective collection. The book is also mentioned in the CRAN task view for Optimization and Mathematical Programming (https://cran.r-project.org/web/views/Optimization.html).

Updated and Revised Second Edition

This second edition consists of an updated version that includes 270 pages. While the core structure and material is essentially the same, the whole text was proofread and several portions were rewritten to reflect the teaching and research experience of the author. Moreover, several of the R code demonstrations were substantially updated. In effect, the first edition was written in 2013, when there were around 5,800 R packages. Since then, this number as almost tripled, with several of these newer packages implementing metaheuristics. Examples of packages that are now explored in this second edition include: **DEoptimR**, **GenSA**, **GA**, **MaOEA**, **NMOF**, **ecr**, **evoper**, **gramEvol**, **irace**, **mcga**, and **psotim**. New material was also added, such as: limitations of metaheuristics, interfacing with the Python language, creating interactive Web applications via the **shiny** package, a stronger emphasis on parallel computing (with several examples, including the island model for genetic algorithms), iterated racing, ant colony optimization, evolutionary breeding and selection operators, grammatical evolution, and several multi-objective algorithms (e.g., SMS-EMOA, AS-EMOA, NSGA-III).

The second edition code was written using the version 4.0.0 of the R tool, while the first edition was produced using the older version 3.0.0. Before updating the R demonstrations, all first edition R code examples (written in 2013) were re-executed using the more recent R version (4.0.0). In some cases, the executions used newer versions of the previously adopted packages (e.g., **genalg** latest version was

written in 2015). While 7 years have passed (since 2013), it was interesting to notice that the code examples executed well in almost all cases. This attests the robustness of the first edition code in the R tool technological evolution. The author expects that a similar phenomenon occurs with the second edition code. However, if needed and whenever possible, code updates will be provided at: https://www.springer.com/us/book/9783030728182.

How to Read This Book

This book is organized as follows:

Chapter 1 introduces the motivation for modern optimization methods and why the R tool should be used to explore such methods. Also, this chapter discusses key modern optimization topics, namely the representation of a solution, the evaluation function, and how to handle constraints. Then, an overall view of modern optimization methods is presented, followed by a discussion of their limitations and criticism. This chapter ends with the description of the optimization tasks that are used for tutorial purposes in the next chapters.

Chapter 2 presents basic concepts about the R tool. Then, more advanced features are discussed, including command line execution, parallel computing, interfacing with other computer languages, and interactive Web applications. This chapter is particularly addressed to non R experts, including the necessary knowledge that is required to understand and test the book examples. R experts may skip this chapter.

Chapter 3 is about how blind search can be implemented in R. This chapter details in particular two full-blind search implementations, two grid search approaches (standard and nested), and a Monte Carlo (random) search.

Chapter 4 discusses local search methods, namely hill climbing (pure and steepest ascent and stochastic variants), simulated annealing, and tabu search. Finally, it shows how modern optimization methods (including local search ones) can be tuned in terms of their internal parameters by using an iterated racing.

Chapter 5 presents population-based search methods, namely genetic and evolutionary algorithms, differential evolution, particle swarm optimization, ant colony optimization, estimation of distribution algorithm, genetic programming, and grammatical evolution. The chapter also presents examples of how to compare population based methods, how to handle constrains, and how to run population-based methods in parallel.

Chapter 6 is dedicated to multi-objective optimization. First, three demonstrative multi-objective tasks are presented. Then, three main multi-objective approaches are discussed and demonstrated: weighted-formula, lexicographic, and Pareto (e.g., NSGA-II and NSGA-III, SMS-EMOA, AS-EMOA).

Chapter 7 presents real-world applications of previously discussed modern optimization methods: traveling salesman problem, time series forecasting, and wine quality classification.

Each chapter starts with an introduction, followed by several chapter-topic related sections and ends with an R command summary and exercises sections. Throughout the book, several examples of R code are shown. The code was run using a 64 bit R (version 4.0.0) under on a *macOS* laptop. Nevertheless, these examples should be easily reproduced by the readers on other systems, possibly resulting in slight numerical (32 bit version) or graphical differences for the deterministic examples. Also, given that a portion of the discussed methods are stochastic, it is natural that different executions of the same code and under the same system will lead to (slight) differences in the results.

It is particularly recommended that students should execute the R code and try to solve the proposed exercises. Examples of solutions for proposed exercises are presented at the end of this book. All these code files and data examples are available at: https://www.springer.com/us/book/9783030728182.

Production

Several contents of this book were taught by the author in the last 12 years in distinct course units of master's and doctoral programs. At the master's level, it included the courses "adaptive business intelligence" (Master of Engineering and Management of Information Systems, University of Minho, Portugal) and "business intelligence" (Master of Information System Management, Lisbon University Institute, Portugal). The doctoral course was "adaptive business intelligence" (Doctoral Program in Computer Science, Universities of Minho, Aveiro and Porto, Portugal). Also, some of the older material was lectured at a tutorial given in the European Simulation and Modelling Conference (ESM 2011), held at Guimarães.

This book was written in LaTeX, using the **vim** editor (http://vim.org) and its US English spell checker. Most figures were made in R, while some of the figures were designed using **xfig** (http://www.xfig.org), an open source drawing tool.

Guimarães, Portugal   Paulo Cortez

Contents