Review of
“Experimental Methods for the Analysis of Optimization Algorithms”
Edited by Thomas Bartz-Beielstein, Marco Chiarandini, Luís Paquete and Mike Preuss
Springer, 2010.

Rubén Ruiz

Grupo de Sistemas de Optimización Aplicada.

Instituto Tecnológico de Informática (ITI).

Ciudad Politécnica de la Innovación, Edificio 8G. Acceso B.

Universitat Politècnica de València. Camino de Vera s/n, 46022 Valencia, Spain.

email: rruiz@eio.upv.es

May 11, 2011

1 Introduction

Assessing whether a given algorithm outperforms another goes way, way beyond simply reporting average values over small synthetic benchmarks. While this might seem a bold affirmation, it is actually an understatement. In the fields of operations research and management science, artificial intelligence and computer science it is still painfully common to find papers where authors compare approaches and put forward bold conclusions on the basis of the observed average performance of algorithms over common benchmarks. More than 15 years ago, authors started to wonder why other disciplines, for example, medical sciences or physics, carry out formalized and comprehensive experiments, with the aid of sound statistical techniques while in our fields, these tests are sadly most of the time the exception and not the norm. Some seminal works are those of Hooker (1994), Hooker (1995), Barr et al. (1995), and McGeoch (1996), to name just a few.

The book "Experimental Methods for the Analysis of Optimization Algorithms", edited by Thomas Bartz-Beielstein, Marco Chiarandini, Luís Paquete and Mike Preuss is a solid and comprehensive step forward in the right direction. The book not only covers adequate comparison of methodologies but also the tools aimed at helping in algorithm design and understanding, something that is being recently referred to as "algorithm engineering".

The book is 457 pages long, distributed into three parts, fifteen chapters and an appendix. It is excellently priced at 106.95 EUR by the publisher and available in most known online book stores. In the following, a brief outline of the chapters and contents are given, followed by a discussion and some conclusions.

2 Contents of the book

"Experimental Methods for the Analysis of Optimization Algorithms" is divided into three parts, preceded by an initial chapter 1 "Introduction" by the editors Thomas Bartz-Beielstein, Marco Chiarandini, Luís Paquete and Mike Preuss. It is a nice introduction to the need of experimental methods in algorithm design, calibration and understanding. In this chapter the aim of the book is clearly defined as a threefold objective: "First, to give importance to experiments and qualify them as complementary to theories, as ways to test and refine them, improve them, and make them more meaningful and useful in practice. Second, to intend to recognize the existence of a scientific process in the field of computing that consists of applying statistical testing and learning from error. To regard the learning that we achieve from experiments as valuable and trustworthy. Third, to contribute to making the analysis of experiments in the field of computing more rigorous, objective and reproducible, hence similar to what is seen in other natural sciences." Apart from stating the objectives of the book and highlighting the need for statistical tools, the chapter briefly outlines the parts and the contents of the remaining chapters.

After the introduction there is Part I of the book which contains five general chapters contributed by experienced authors in the field. Chapter 2 "The Future of Experimental Research" by Thomas Bartz-Beielstein and Mike Preuss concisely lays down the foundations of experimental research in algorithms and optimization. The problems of this type of research are identified and several tokens of the new experimentalism are introduced. Of particular interest is the presentation of the pitfalls of the experimentation with randomized algorithms. Also interesting is the discussion about the floor and ceiling effects, the confounded effects and, above all, the fairness in parameter settings. From this discussion one suddenly realizes that reporting the comparisons of new -and potentially calibrated- algorithms versus existing algorithms that were calibrated differently (or not calibrated at all by the original authors) is certainly misleading. The chapter closes with indications on how to conduct and report experiments. For many, the chapter opens important and frequently omitted steps in algorithms' and metaheuristics engineering, development and testing. Experienced readers in the field, specially those aware of the cited works of Hooker (1994), Hooker (1995), Barr et al. (1995), and McGeoch (1996), will find that the chapter beautifully

presents these issues in a seamless approach.
Chapter 3 "Design and Analysis of Computational Experiments: Overview" by Jack P. C. Kleijnen is a much more technical one. It deals with the experimental design approach for computer experiments. The starting sections are more or less known to the DOE audience in the form of simple factorial designs or more specifically, fractional factorial designs with resolution III and above. The author then comments about the sequential bifurcation, an advanced technique for screening experiments with many studied factors. As the author states, this technique is not available yet in statistical software. Then the chapter moves on to much more advanced techniques like kriging and kriging designs, to be followed with the response surface methodology approach. The last techniques given are kriging with mathematical programming, a very advanced technique put forward by the author and the briefly mentioned Taguchi robust approach. While advanced, the chapter touches many interesting concepts, albeit briefly. The chapter is full of very interesting pointers for future research, something that is really interesting to researchers in the area.
Chapter 4 "The Generation of Experimental Data for Computational Testing in Optimization" by Nicholas G. Hall and Marc E. Posner proposes some simple properties that every synthetic benchmark for any optimization problem should have. The properties, or protocol as the authors call it, are simple to understand. The chapter then studies, with great detail, many of the synthetic benchmarks proposed for many different optimization problems, namely, generalized assignment, knapsack, supply chains, scheduling, graphs and networks, routing, data mining, stochastic programming and intractable problems. Chapter 5, "The Attainment-Function Approach to Stochastic Multiobjective Optimizer Assessment and Comparison" by Viviane Grunert da Fonseca and Carlos M. Fonseca deals about the assessment and comparison of multiobjective stochastic optimizers. Readers new to multiobjective comparisons will be amazed at how complex it is to compare two algorithms

when more than one objective is concerned. The chapter explains in detail the foundations and applications of the attainment function, which is enormously useful to characterize the performance of a multiobjective algorithm, both as regards spread as well as convergence to the optimal Pareto front. Of particular interest are the sections about hypothesis testing and the explanations of higher order attainment functions. This chapter will surely contribute to bolstering the usage of these functions since nowadays, as the authors state, unary and binary quality indicators are much more commonly employed. Future research directions encompass the study of several algorithms and several instances for the general application of attainment functions. Part I of the book, which closes with Chapter 6 "Algorithm Engineering: Concepts and Practice" by Markus Chimani and Karsten Klein, demonstrates how algorithmics is parting from a theoretical science to become more and more an engineering field. An excellent opening by the authors puts the more traditional algorithm development and worst case analysis in stark contrast with the real needs of large problem solving. The authors include the algorithm engineering cycle where not only real implementations of algorithms are sought but also actual learning and improvements can be obtained after these implementations. From this standpoint, the authors deal with several issues in algorithm engineering, like the special properties in the input data (often observed in real problems), large datasets, time critical settings and robustness. The authors provide plenty of pointers to all of these topics, which should satisfy the interest of those looking for additional materials. The chapter closes with two interesting applications and studies of algorithm engineering. The first one is the very well known shortest path algorithm in a graph. The second is another important topic, the problem of finding all occurrences of a pattern p in a long text T. The study depicts a clear situation in which algorithm engineering shines and demonstrates that it is here to stay.

After this interesting content, the book continues with Part II "Characterizing Algorithm Performance" comprised of three chapters. The first is chapter 7 "Algorithm Survival Analysis" by Matteo Gagliolo and Catherine Legrand. In this highly specialized chapter, the authors propose the application of the survival analysis to the runtime of an algorithm. The runtime of an algorithm can be modelized as a random variable. This has been already known for some time with the run time distributions (RTD) that were very well explained by Hoos and Stützle (2005). Here, the authors go much further beyond by employing survival analysis to iteratively select the best competing algorithm when solving a set of benchmark instances. The runtime distributions are therefore iteratively updated and refined for the algorithm portfolio. The whole process helps in selecting the best algorithms while saving valuable CPU time in the process.
Chapter 8 "On Applications of Extreme Value Theory in Optimization" by Jürg Hüsler contrasts with the previous chapter as in this case it studies the distributions of the objective values of the solutions obtained by simple random search methods or evolutionary strategies. More specifically, the chapter applies extreme value theory (EVT) to characterize the performance of the studied methods. The chapter contains a succinct, but fairly terse introduction to EVT and specific cases for stochastic optimizers, as it is the peaks over threshold method (POT). Once the basic foundations are laid out, the author studies a simple random search method, which is well suited to EVT as the objective values of each run are not self correlated. Some deep analytical results confirm the good performance of the EVT technique as a characterization tool of performance. Somewhat lighter analyses are carried out for simple evolutionary methods where the EVT techniques are shown to require a more arduous task due to the obvious self correlation of the results given. The studied techniques show great promise for the characterization of algorithm performance. Another take at multiobjective analysis is given in Chapter 9, "Exploratory Analysis of Stochastic Local Search Algorithms

in Biobjective Optimization" by Manuel López-Ibáñez, Luís Paquete, and Thomas Stützle. This chapter shows the application of the first order Empirical Attainment Functions (EAF) that were shown in detail in the previous chapter 5. While limited to one instance at a time at the moment, EAFs give plenty of information about the behavior of multiobjective stochastic optimizers in the objective space without the loss of information that occurs when all the Pareto front approximations are reduced to a single unary or binary performance indicator. The chapter provides some examples and explanations about publicly available codes put forward by the authors so as to foster the usage of EAFs in practice. The authors also study an extension referred to as the Differential Attainment Function, which allows the study of the differences that two different stochastic optimizers provide in the objective space.

The third and last part of the book is titled "Algorithm Configuration and Tuning" and comprises the remaining six chapters. Chapter 10, "Mixed Models for the Analysis of Optimization Algorithms" by Marco Chiarandini and Yuri Goegebeur is an excellent work where the authors delve into advanced statistical designs that go way beyond the classical ones analyzed with ANOVA. More specifically, the authors study linear models where the instances that are tested in stochastic optimizers are considered as random factors. This results in mixed effects linear models or even in nested linear mixed models. The authors go to great lengths in explaining the theory behind all these models and present four different models, from one algorithm and one instance to many algorithms and many instances (actually, by instances the authors mean several factors affecting the characteristics of the instances). After the theoretical explanations, the authors use an optimization example from the 2-edge-connectivity graph augmentation problem. Well known heuristics with three different factors, as well as instance factors are studied using all four proposed statistical models. The explanations on the applications of the models are plentiful and detailed, with R-code listings and plots. These are well complemented by a web site with all codes and on-line materials with further explanations. Of particular interest are the results where the authors show that the significance observed by the mixed-effects linear models are wildly different than those observed with a more classical multifactor ANOVA where instance characteristics are considered as fixed factors. While one has to admit that the chapter is complex and deeply statistical, there is hardly any viable excuse for not using these designs, as not doing so would surely result in wrong conclusions when comparing or tuning stochastic optimizers. Conversely, chapter 11 "Tuning an Algorithm Using Design of Experiments" by Enda Ridge and Daniel Kudenko show, in a much simpler way, the application of more common and well known Design of Experiments (DOE) techniques. The chapter starts with basic explanations about full factorial and fractional factorial designs, to be followed by well known response surface designs. The authors apply these designs in a step by step approach to a case study about an ant colony algorithm with no less than 12 factors (two of them related to instance characteristics). Some more advanced topics as desirability functions are applied at the end for the final tuning of the case study. A very good read and more suitable as a DOE primer. A completely different approach is given in chapter 12 "Using Entropy for Parameter Analysis of Evolutionary Algorithms" by Selmar K. Smit and Agoston E. Eiben. This chapter is more centered around evolutionary algorithms, which subsume genetic algorithms, evolution strategies, evolutionary programming and genetic programming. The authors propose the usage and measurement of entropy to detect relevant parameters and suitable levels or good parameter values. The chapter contains basic theoretical explanations about Shannon, differential and joint entropy, entropy estimation and a detailed case study. Yet another different technique is explained and detailed in chapter 13, "F-Race and Iterated F-Race: An Overview" by Mauro Birattari, Zhi Yuan, Prasanna Balaprakash, and Thomas Stützle. F-Race algorithms have been known for some time already, as they date back to 2002. After presenting the foundations of the F-Race methods, the authors present an interesting extension, the Iterated F-Race. I-F-Race is shown in three different settings to outperform two other F-Race versions and the results show great promise. Quite amusingly, however, is the fact that I-F-Race also contains some parameters. So, in the end, the tuning algorithm needs also some tuning! Such are the complexities of proper and sound algorithm calibration. The chapter closes with an overview and outlook of F-race applications and pointers for future research venues. Chapter 14. “The Sequential Parameter Optimization Toolbox” by Thomas Bartz-Beielstein, Christian Lasarczyk, and Mike Preuss introduces the R Toolbox built around the sequential parameter optimization (SPO) introduced in chapter 2. The authors go step by step through the toolbox and employ some examples to illustrate how it works. The main interest lies in the automatic feature of the toolbox that requires little user intervention. Given the open-source implementation of the toolbox and the R package used for its development, there are few drawbacks, if any, for its final implementation and usage by the experimental community. The last chapter of the book is chapter 15 “Sequential Model-Based Parameter Optimization: an Experimental Investigation of Automated and Interactive Approaches” by Frank Hutter, Thomas Bartz-Beielstein, Holger H. Hoos, Kevin Leyton-Brown, and Kevin P. Murphy. It is actually a comprehensive study that compares tools for the tuning of algorithms for two methods proposed for different problem domains. The authors test the aforementioned SPO against the competing SKO (Sequential Kriging Optimization) approach. Both methodologies are tested on two different examples. In the process, the authors also propose refinements of the SPO, referred to as SPO 0.3, SPO 0.4 and SPO+. This latter method is shown to outperform the others. The authors also test the automatic versus interactive tuning approaches. Quite surprisingly, the automated approach performs rather well although the interactive process allows one to reach results much faster. The chapter is limited by the analysis of one single instance instead of a set of instances or benchmark. In any case, the authors comment to be working on extensions to observe performance across benchmarks, something really interesting and that should keep readers eagerly awaiting these extensions.

Finally, the book contains an appendix "A Brief Introduction to Inferential Statistics" by Dario Basso. Readers well versed on statistics need not to read this last part of the book. However, for novice statisticians, this appendix is invaluable. Without the typical deep statistical explanations and proofs, the author quickly attracts the reader to the core of simple inferential statistics, including a brief summary on statistical models, point estimation, hypothesis testing, regression and model fitting. In the very least, this appendix is an interesting, albeit not fully self-contained, complementary text to the rest of the book.

3 Discussion

The book is of interest to two distinct audiences. First and foremost, it is targeted at the whole operations research and management science, artificial intelligence and computer science communities with a loud and clear cry for attention. Strong, sound and reliable tools should be employed for the comparison and assessment of algorithms and also for more structured algorithm engineering.

Given the level of detail of some other chapters however, a second potential audience could be made up of those researchers interested in the core topic of algorithm assessment. The long list of contributors to this book includes top notch and experienced researchers that, together, set the trend in the field. As a result, those interested in this specific area of analysis of optimization algorithms should not miss this book under any circumstance.

A direct question arising while pondering over the purchase and ultimately the reading of this book is "Would I find what I need to test my method?". Well, sadly, there is no easy answer to this. While there are some chapters that go hands-on with even R code listings about how to do stuff, chances are that one is not going to easily encounter a point and click explanation. Algorithm testing and assessment is a complicated matter, as it is possibly like concluding that a given drug is effective or not after a clinical test in medical sciences. We have to finally accept, as a community, that there is little sense in devoting countless pages explaining methods and algorithms and then just a few paragraphs and a table for reporting average results. The careful, sound, detailed and comprehensive assessment of optimization algorithms is a necessity that requires attention and care. As a result, my opinion is that this book should be followed and that it should be at the top of every experimenter's table.

A final discussion is the fact that many of the techniques presented in this book still need further development and extensions before they can be mass applied by experimenters and before they are accepted as the de facto tools. Some of the presented techniques are limited to the analysis of a single instance or to a couple of algorithms only. In any case, it is not reasonable to expect these tools to be fully developed before we start using them. Additionally, since many of these tools are still being actively researched, there is no clear-cut indication about which one should be used over the others. For example, Should we use nested fractional designs, SPO, F-Race or any other method for calibrating an algorithm? The final recommendation is that all tools should be tinkered with, experimented with and worked with. In the worst case, one tool will validate the results of the others and in the best case, new interesting conclusions can be obtained.

4 Conclusions

The book "Experimental Methods for the Analysis of Optimization Algorithms", edited by Thomas Bartz-Beielstein, Marco Chiarandini, Luís Paquete and Mike Preuss is a comprehensive work into the field of the analysis of optimization methods. Well structured into three parts plus an appendix, its 15 chapters contain from primers of design and analysis of computer experiments to thorough and detailed treatments of advanced tools and technologies for algorithm engineering.

For readers interested in experimenting with algorithms in a sound and correct way, this book is a must read, as it is for those researchers working in the field of optimization algorithms analysis. Together with the comprehensive list of references at the end of each chapter, this book is an excellent starting point, as well as a solid reference guide for such audiences.

References

Barr, R. S., Golden, B. L., Kelly, J. P., Resende, M. G. C., and Stewart, J. W. R. (1995). Designing and reporting on computational experiments with heuristic methods. *Journal of Heuristics*, 1(1):9–32.

Hooker, J. N. (1994). Needed: An empirical science of algorithms. *Operations Research*, 42(2):201–212.

Hooker, J. N. (1995). Testing heuristics: We have it all wrong. *Journal of Heuristics*, 1(1):33–42.

Hoos, H. H. and Stützle, T. (2005). *Stochastic Local Search Foundations and Applications*. Morgan Kaufmann Publishers.

McGeoch, C. (1996). Toward an experimental method for algorithm simulation. *INFORMS Journal on Computing*, 8(1):1–15.