Graphical Tools for the Analysis of Bi-objective Optimization Algorithms

[Workshop on Theoretical Aspects of Evolutionary Multiobjective Optimization]

Manuel López-Ibáñez Thomas Stützle
IRIDIA, CoDE, Université Libre de Bruxelles (ULB)
Brussels, Belgium
{manuel.lopez-ibanez,stuetzle}@ulb.ac.be

Luís Paquete
CISUC, Department of Informatics Engineering
University of Coimbra, Portugal
paquete@dei.uc.pt

Categories and Subject Descriptors

I.2.8 [**Computing Methodologies**]: Artificial Intelligence—*Problem Solving, Control Methods, and Search*

General Terms

Measurement

Keywords

Multi-Objective Optimization, Performance Assessment, Attainment Function, Graphical Exploratory Tools

1. INTRODUCTION

The performance assessment of stochastic local search (SLS) algorithms for multi-objective optimization problems is an active research topic [2, 4, 8, 11, 16]. The most challenging situation arises when comparing algorithms whose primary goal is to obtain the set of *Pareto optimal solutions*. A feasible solution is Pareto optimal if it is not dominated by any other feasible solution. A solution *dominates* another one if the former is not worse in any objective value than the latter and strictly better in at least one objective. Obtaining the Pareto optimal set is often impractical, and therefore multi-objective algorithms aim to produce a good approximation to it in the form of a nondominated set, that is, a set of mutually nondominated solutions. Evaluating the performance of multi-objective algorithms involves comparing the quality of the nondominated sets produced.

The dominance criterion straightforwardly extends to the comparison of nondominated sets [16]. Given two different nondominated sets, the former is better than the latter, if every solution in the latter is dominated by or equal to at least one solution in the former. Stricter and weaker dominance relations among nondominated sets may be defined [16]. Nonetheless, for high-performing algorithms, frequently neither set is better than the other, and hence, these sets are *incomparable* according to Pareto optimality.

GECCO'10, July 7–11, 2010, Portland, Oregon, USA.

Scalar quality measures were introduced to evaluate the quality of incomparable sets. Unary and binary quality measures [8, 16] assign a single scalar value to each nondominated set (or pair of nondominated sets). This value aims to measure a desirable property of nondominated sets. Properties widely acknowledged to be desirable are closeness to the Pareto optimal set, a wide spread of solutions in the objective space (as opposed to solutions clustered in a small region), and an even distribution of solutions in the objective space (as opposed to many small clusters of solutions).

The choice of quality measures introduces a strong bias because different measures may lead to different conclusions. Even restricting to a single measure, the interpretation of the results is difficult due to the inherent simplification introduced by summarising a multidimensional set into a single scalar value. A common approach is to assume that the quality measure is the evaluation criterion desired by the decision maker, and declare the best algorithm the one that obtains the best average value of a certain quality indicator, knowing in advance that this conclusion would change if different quality indicators were used. The use of multiple quality indicators may complicate rather than simplify the assessment, since the conclusions obtained by the quality indicators may disagree.

A fundamentally different approach to the quality assessment of multi-objective SLS algorithms derives from the concept of attainment function [6]. The attainment function extends the scalar concepts of mean and variance to random sets. The attainment function theory may completely characterize the statistical distribution of solutions in the objective space in terms of location, spread and mutual dependence [4]. Moreover, statistical testing and inference are possible [4, 5, 14, 15]. However, the use of attainment functions is still rather limited in practice. We present here two practical applications of the first-order attainment function [11] for analysing the output of SLS algorithms for bi-objective optimization problems. Programs implementing the techniques presented here are also available [12]. Later, we discuss what would be necessary to extend this work for more than two objectives and for other types of analysis.

2. EMPIRICAL ATTAINMENT FUNCTION

The (first-order) attainment function [6] assigns a real value in $[0, 1]$ to each point in the objective space $\mathbb{R}^d$. This value describes the probability of a random set attaining (dominating or being equal to) that particular point in the objective space. The output of a multi-objective SLS algorithm for a particular problem instance may be characterized

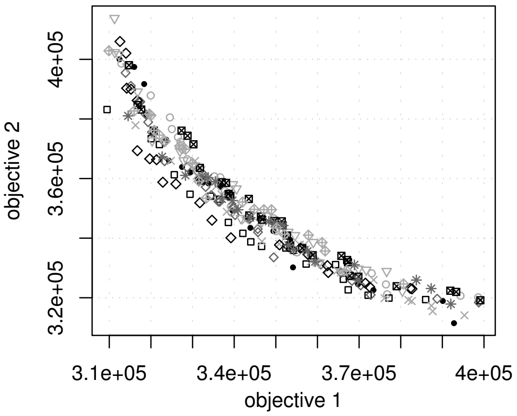

Figure 1: Ten independent outcomes obtained by an SLS algorithm applied to an instance of a bi-objective optimization problem.

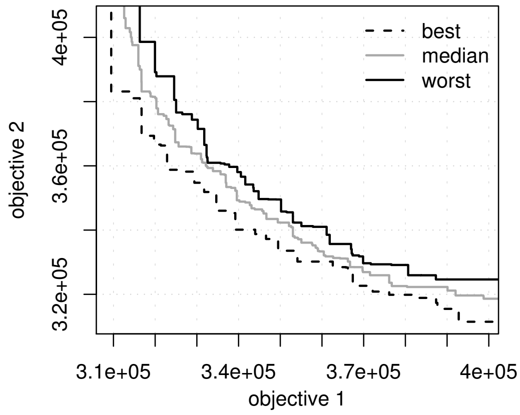

Figure 2: Best, median and worst attainment surfaces for the data described in Fig. 1.

as an attainment function. The attainment function corresponding to an algorithm is typically unknown, however, we can derive an empirical estimation of the attainment function from the outputs of several independent runs of an SLS algorithm. In the case of bi-objective optimization problems, the empirical attainment function (EAF) is fast to compute, and its graphical representation provides more intuitive information about the distribution of the output of an algorithm than unary (or binary) quality indicators.

3. PLOT OF ATTAINMENT SURFACES

The notion of attainment surface [2] predates the formal definition of the attainment function, but attainment surfaces are better defined in terms of the latter. Therefore, we define the k%-attainment surface as the lower boundary of the region in the objective space with a value of the attainment function of at least $k/100$. The empirical k%-attainment surface is the line delimiting the objective space attained by at least k percent of the runs of an SLS algorithm. This definition is an extension of the $k/100$ percentile of the empirical frequency distribution to the multi-objective case. We can define the *median* attainment surface as the boundary of the objective space attained by 50 percent of the runs. In a similar way, we define the *best* and *worst* attainment surfaces as the boundaries of the objective space attained by at least one run and by all runs, respectively.

Figure 1 displays the output of ten runs of an SLS algorithm on a bi-objective problem. Points belonging to the same set are represented by the same symbol and color. This kind of plots appears often in the literature to illustrate the output of a multi-objective SLS algorithm. However, it is difficult for the reader to assess the expected output of a single run, and it becomes more difficult with higher number of runs and points. Comparisons against a reference set or between multiple algorithms only complicate further the interpretation of an already crowded plot. For these reasons, researchers often decide to display only one "representative example" of the output sets, which is often arbitrarily chosen from all the runs.

Attainment surfaces, on the other hand, are a convenient way to summarise the outputs of several runs. Figure 2 shows the best, median and worst attainment surfaces calculated from the data in Fig. 1. The location of the median attainment surface gives an idea of the likely location of the output of a single run of the algorithm, whereas the distance between the worst and best attainment surfaces gives an indication of the variability of the results. As in the single-objective case, more robust indicators of variability are the first and third quartiles, which correspond to the 25%- and 75%-attainment surfaces. We argue that plots of attainment surfaces improve on clarity and conciseness over a direct plot of the output sets, while providing much more information than scalar quality measures.

4. PLOT OF EAF DIFFERENCES

When comparing two algorithms, it is possible to plot attainment surfaces side-by-side or within the same figure. However, this forces the reader to assess the differences between the algorithms by examining the intersections between the lines. A more convenient approach is to directly plot the differences between the EAFs. A large difference between the EAFs of two SLS algorithms at a certain point of the objective space indicates a larger probability of attaining this point with one algorithm than with the other.

We use the following method to compute the differences between the EAFs of two algorithms. First, we compute the EAF of the union of the output sets obtained by all runs of both algorithms. Then, for each point in the objective space where the value of the EAF changes, we count how many runs of each algorithm attained that point. This allows us to compute the value of the EAF of the first algorithm at that point minus the value of the EAF of the second algorithm. Finally, we plot side-by-side positive and negative differences at the points we previously examined, encoding the magnitude of the differences using shades of grey: the darker a point, the larger is the difference [9, 10, 14].

Figure 3 illustrates the application of this method. The two side-by-side plots in the top part of Fig. 3 give the EAFs associated to two algorithms that were run 25 times with different random seeds on the same problem instance. Points in the EAFs are assigned a gray level according to their probability. In addition, we plot four different attainment surfaces. The lower line on both plots connects the best set of points attained over all runs of both algorithms (*grand* best attainment surface), and the upper one the set of points attained by any of the runs (*grand* worst attainment surface). Any differences between the algorithms are contained

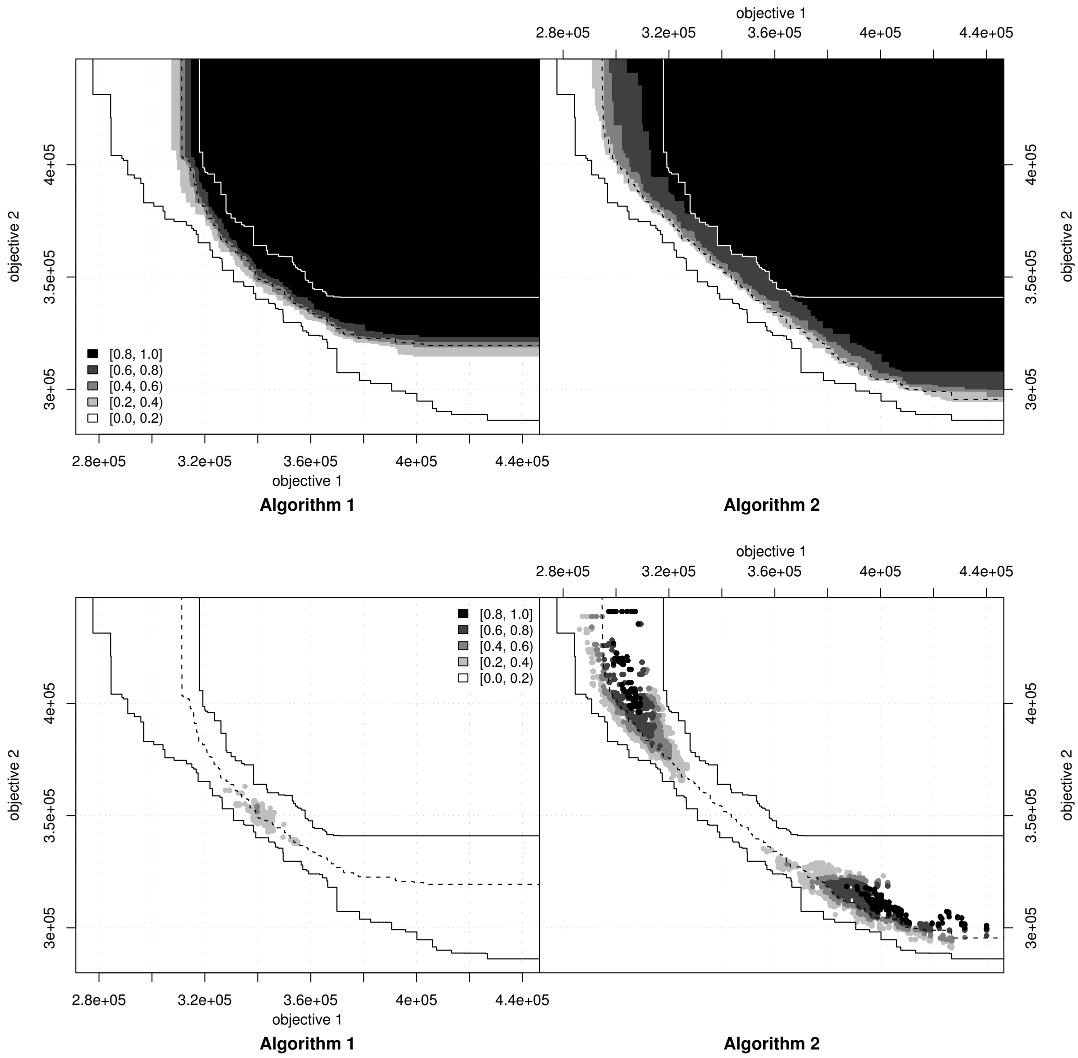

Figure 3: Visualization of the EAFs associated to the outcomes of two algorithms (*top*) and the corresponding differences between the EAFs (*bottom left*: differences in favour of Algorithm 1; *bottom right*: differences in favour of Algorithm 2). In the top, the gray level encodes the value of the EAF. In the bottom, the gray level encodes the magnitude of the observed difference.

within these two lines. The dashed line corresponds to the median attainment surface of each algorithm, that is, the lower boundary of the region where the EAF has value 0.5.

The bottom side-by-side plots of Fig. 3 show the location of the differences between the EAFs of the two algorithms. On the left, points denote positive differences between the EAF of Algorithm 1 over the one of Algorithm 2, and on the right the differences are in favor of Algorithm 2 over Algorithm 1. We only show points where the difference between the EAFs is larger than 20 percent. The magnitude of the differences is encoded in a grey scale given in the legend of the plot. We plot also the same attainment surfaces as for the top plots to facilitate comparison. From these plots, we can observe that Algorithm 2 performs better at the extremes of the nondominated sets, whereas there is a very small difference in favour of Algorithm 1 at the center of the nondominated sets. This indicates that Algorithm 2 is preferable to Algorithm 1, despite their outputs being mostly incomparable in terms of dominance relation. In addition, we know what, or more exactly where, is the difference between the two algorithms and how strong this difference is.

Such a fine-grained analysis would be impossible with most scalar quality indicators. The examination of the EAF differences reveals not only the magnitude of the differences between algorithms, but also where these differences are located in the objective space. Therefore, it is particularly helpful to point out problems on attaining certain regions of the objective space.

5. CONCLUSIONS AND OPEN QUESTIONS FOR FUTURE RESEARCH

Together with previous work by Knowles [7] and Fonseca et al. [4], this work describes earlier practical applications of the attainment function theory for the performance assessment of multi-objective SLS algorithms. We regularly use the methods described here for our own research as complementary tools to methods based on dominance relations among sets and quality measures [1, 10].

Nevertheless, the presented methods have some limitations. First, there are no publicly available algorithms for computation of the EAF for more than two objectives; however, there is ongoing work and such an algorithm will be available soon. Second, it is unclear what is the best practical way to make use of the information provided by the EAF for more than two objectives. Visualizations in three dimensions are possible [13], but cumbersome in non-interactive two-dimensional plots. For more than three dimensions, special methods would need to be developed to visualize the information provided by the EAF. The method of parallel coordinates has been used for plotting objective vectors in problems with many objectives [3], and it is the only existing proposal for plotting the EAF in high dimensions [15]. Third, the EAF is computed on the output of the algorithm for a single instance, and no method has been proposed so far to summarise the results across several instances. Finally, higher-order EAFs enable more complete performance analysis and statistical inference [4, 5]. However, to the best of our knowledge, these techniques have not been applied in any other practical study in the literature, besides its original publication [4]. In summary, there are both theoretical and practical challenges for the application of the EAF that call for more research and widely available tools.

Acknowledgments. This work was supported by the META-X project, an *Action de Recherche Concertée* funded by the Scientific Research Directorate of the French Community of Belgium. Thomas Stützle acknowledges support from the Belgian F.R.S.-FNRS, of which he is a Research Associate. The authors also acknowledge support from the FRFC project "*Méthodes de recherche hybrides pour la résolution de problèmes complexes*".

References

[1] J. Dubois-Lacoste, M. López-Ibáñez, and T. Stützle. Adaptive "anytime" two-phase local search. In *Proceedings of LION 4*, volume 6073 of *Lecture Notes in Computer Science*. Springer, Heidelberg, 2010. Accepted.

[2] C. M. Fonseca and P. J. Fleming. On the performance assessment and comparison of stochastic multiobjective optimizers. In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *Proceedings of PPSN-IV*, volume 1141 of *Lecture Notes in Computer Science*, pages 584–593. Springer, Heidelberg, 1996.

[3] C. M. Fonseca and P. J. Fleming. Multiobjective optimization and multiple constraint handling with evolutionary algorithms-Part II: Application example. *IEEE Transactions on Systems, Man, and Cybernetics – Part A*, 28:38–44, Jan. 1998. doi: 10.1109/3468.650320.

[4] C. M. Fonseca, V. Grunert da Fonseca, and L. Paquete. Exploring the performance of stochastic multiobjective optimisers with the second-order attainment function. In C. C. Coello, A. H. Aguirre, and E. Zitzler, editors, *Proceedings of EMO 2005*, volume 3410 of *Lecture Notes in Computer Science*, pages 250–264. Springer, Heidelberg, 2005.

[5] V. Grunert da Fonseca and C. M. Fonseca. The attainment-function approach to stochastic multiobjective optimizer assessment and comparison. In T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuß, editors, *Experimental Methods for the Analysis of Optimization Algorithms*. Springer, 2010. To appear.

[6] V. Grunert da Fonseca, C. M. Fonseca, and A. O. Hall. Inferential performance assessment of stochastic optimisers and the attainment function. In E. Zitzler, K. Deb, L. Thiele, C. A. Coello, and D. Corne, editors, *Proceedings of EMO 2001*, volume 1993 of *Lecture Notes in Computer Science*, pages 213–225. Springer, Heidelberg, 2001.

[7] J. D. Knowles. A summary-attainment-surface plotting method for visualizing the performance of stochastic multiobjective optimizers. In *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications*, pages 552–557, 2005. doi: 10.1109/ISDA.2005.15.

[8] J. D. Knowles, L. Thiele, and E. Zitzler. A tutorial on the performance assessment of stochastive multiobjective optimizers. TIK-Report 214, Computer Engineering and Networks Laboratory, ETH Zürich, Feb. 2006. Revised version.

[9] M. López-Ibáñez and T. Stützle. An analysis of algorithmic components for multiobjective ant colony optimization: A case study on the biobjective TSP. In *9th International Conference, Evolution Artificielle, EA 2009*, volume 5975 of *Lecture Notes in Computer Science*, pages 134–145. Springer, Heidelberg, 2009.

[10] M. López-Ibáñez, L. Paquete, and T. Stützle. Hybrid population-based algorithms for the bi-objective quadratic assignment problem. *Journal of Mathematical Modelling and Algorithms*, 5(1):111–137, 2006. doi: 10.1007/s10852-005-9034-x.

[11] M. López-Ibáñez, L. Paquete, and T. Stützle. Exploratory analysis of stochastic local search algorithms in biobjective optimization. In T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuß, editors, *Experimental Methods for the Analysis of Optimization Algorithms*. Springer, 2010. To appear.

[12] M. López-Ibáñez, L. Paquete, and T. Stützle. EAF graphical tools, 2010. URL `http://iridia.ulb.ac.be/~manuel/eaftools`.

[13] L. Paquete. Algoritmos evolutivos multiobjectivo para afectação de recursos e sua aplicação à geração de horários em universidades (Multiobjective evolutionary algorithms to resource allocation and their application to university timetabling). Master's thesis, University of Algarve, 2001. In Portuguese.

[14] L. Paquete and T. Stützle. A study of stochastic local search algorithms for the biobjective QAP with correlated flow matrices. *European Journal of Operational Research*, 169(3):943–959, 2006.

[15] L. Paquete and T. Stützle. Design and analysis of stochastic local search for the multiobjective traveling salesman problem. *Computers & Operations Research*, 36(9):2619–2631, 2009.

[16] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2): 117–132, 2003.