Peak-A-Boo! Generating Multi-Objective Multiple Peaks Benchmark Problems with Precise Pareto Sets

Lennart Schäpermeier[1 [0000-0003-3929-7465]],
Pascal Kerschke[1 [0000-0003-2862-1418]], Christian Grimme[2 [0000-0002-8608-8773]],
and Heike Trautmann[2 [0000-0002-9788-8282]]

[1] Big Data Analytics in Transportation, TU Dresden, Germany
{lennart.schaepermeier,pascal.kerschke}@tu-dresden.de
[2] Data Science: Statistics and Optimization, University of Münster, Germany
{christian.grimme,trautmann}@uni-muenster.de

Abstract. The design and choice of benchmark suites are ongoing topics of discussion in the multi-objective optimization community. Some suites provide a good understanding of their Pareto sets and fronts, such as the well-known DTLZ and ZDT problems. However, they lack diversity in their landscape properties and do not provide a mechanism for creating multiple distinct problem instances. Other suites, like bi-objective BBOB, possess diverse and challenging landscape properties, but their optima are not well understood and can only be approximated empirically without any guarantees.
This work proposes a methodology for creating complex continuous problem landscapes by concatenating single-objective functions from version 2 of the multiple peaks model (MPM2) generator. For the resulting problems, we can determine the distribution of optimal points with arbitrary precision w.r.t. a measure such as the dominated hypervolume. We show how the properties of the MPM2 generator influence the multi-objective problem landscapes and present an experimental proof-of-concept study demonstrating how our approach can drive well-founded benchmarking of MO algorithms.

Keywords: Multi-objective Optimization · Multimodal Optimization · Numeric Optimization · Benchmarking · Problem Generator.

1 Introduction

In order to adequately understand problem hardness and to specifically tailor algorithmic approaches with respect to the criteria characterizing different facets and levels of difficulty in multi-objective (MO) optimization, comprehensive and carefully designed benchmark sets are an essential prerequisite [1].

An MO benchmark suite ideally should be a) comprehensive with regard to the representativeness of relevant real-world problems, b) scalable regarding both decision and objective space dimensionality, c) capable of covering the most important characteristics of MO landscape properties in relevant combinations, d) extendable in size by providing a means to specifically generate a desired number of problem instances with certain landscape characteristics, and most importantly e) well understood by providing analytical expressions of both Pareto front (PF) and Pareto sets (PS), ideally including local structures.

However, meeting all these requirements is an extremely challenging MO problem (MOP) by itself, and we have to aim for optimal trade-off solutions. In single-objective (SO) optimization, BBOB [13] presents a benchmark suite that ticks off many boxes of the previously stated wish list such that, e.g., it is well understood in terms of problem difficulties, optima are known, an arbitrary number of instances per problem type can be generated, and scalability regarding decision space dimensionality is provided. Moreover, a standardized algorithm evaluation procedure exists, which is widely accepted within the community. The community, however, is still largely debating on a).

In MO optimization, there is, unfortunately, no straightforward counterpart. While real-world representativeness is also an issue in this domain, we are specifically concerned about items d) and e) as critical issues. In our view, existing benchmark suites turn out to be either not challenging enough, if PF and PS are known analytically (e.g., ZDT [29] and DTLZ [7]), or extremely challenging if requirement e) is omitted as, e.g., in the bi-objective BBOB [25]. Specifically, algorithm performance evaluation is very challenging as no ground truth exists for comparison. Also, item c) cannot be assessed properly as MO landscape characteristics can only be empirically and heuristically approximated.

This paper concentrates on the MPM2 generator [27,26] and on an MO benchmark set creation concept based thereon [16]. We will show that we can largely contribute to understanding MOPs by providing a method for deriving both PS and PF analytically and allowing for the approximation of optimal Hypervolume (HV) up to an arbitrary precision. Thereby, a ground truth is provided in combination with MPM2 being flexible regarding the generation of different types of landscape structures and problem characteristics. So far, we concentrate on continuous bi-objective MOPs as proof-of-concept while on top simultaneously illustrating generalization and scalability potential.

We start by giving some background on MO benchmark suites and algorithms in Section 2. Then, in Section 3, we introduce our methodology first for individual pairs of unimodal functions and subsequently for multiple peaks. This is followed by a proof-of-concept experimental study showing problem properties and algorithm performances in Section 4. Section 5 concludes this work and comments on future research perspectives building on our methodology.

2 Background

Before diving into our methodology, we will start by introducing some background on MO benchmark suites, in particular the MPM2 generator, and the MO optimizers utilized in the experimental section later in this work.

MO Benchmark Suites In a recent survey on continuous multimodal MO optimization [12], existing MO benchmark suites were discussed both in general as well as with a dedicated focus on multimodality. Therein, it has been concluded that the existing benchmark problems usually fall into one of two categories. The first group comprises hand-crafted MOPs with well-understood structural landscape properties, including analytically defined PSs. Typical representatives of this collection are historically well-established test suites such as DTLZ [7] or ZDT [29]. However, a severe limitation of these MOPs is, from a multimodal perspective, their lack of challenging landscape structures, as visually demonstrated in [22,19,20]. Likewise, the MOPs of the recently proposed MMF test suite [28] primarily exhibit extremely regular patterns; yet, structurally diverse problems are crucial for meaningful algorithm benchmarking [1]. Moreover, the scalability of the aforementioned MOPs is limited w.r.t. their number of optima and dimensionality of the decision space.

The second group of benchmark problems comprises suites based on concatenations of SO benchmark functions, like the bi-objective BBOB [25]. These problem collections are usually scalable in dimensionality and they contain more diverse MOPs with potentially complex landscapes due to the flexible concatenation of functions. Yet, these benefits come at the cost of poorly understood structural properties. For instance, although the global optima of the SO BBOB functions are analytically known [13], the PS of the corresponding bi-objective BBOB instance must be empirically approximated.

After comparing the strengths and weaknesses of the existing MO benchmark suites, it became evident that our community is currently lacking a sophisticated problem generator, which is capable of constructing a scalable, comprehensive, and diverse set of multimodal MOPs with known landscape structures.

To fill this gap, we herein utilize Wessing's *Multiple Peaks Model* (MPM2) generator [27,26] for generating SO functions with configurable topologies and scalable dimensionality. Each of these SO functions is essentially the minimum of a configurable number of individual peak functions, i.e., unimodal functions with ellipsoidal structure, aligned in one of two topologies: funnel or random. The *funnel* type contains a large funnel in which all optima are grouped around a global optimum, such that the depth per peak decreases with increasing distance from the global optimum. In contrast, the *random* type distributes the depths and locations of local optima uniformly across the search space. The implementation of the MPM2 generator allows us to extract valuable information about each of the underlying peaks, such as the covariance matrix, radius, and height properties, which we will later use as a basis for identifying the PS of the generated MOPs. Finally, note that the decision space of a d-dimensional MPM2 problem is usually $[0, 1]^d$ and that objective values are restricted to $[0, 1]$.

Similar to [16], we create bi-objective benchmark problems by concatenating functions generated with MPM2. Consequently, the resulting MOPs tend to be part of the second category of benchmark problems. Still, due to its modular composition of multiple peaks, we can generate configurable and scalable MOPs with *known* PSs (see Sec. 3). Despite their relatively simple components, the landscapes of the MOPs generated with this approach do not necessarily exhibit simple, regular patterns, which are easy to exploit by an algorithm. Instead, our approach enables the creation of diverse test functions with irregular, non-separable landscapes, convex, concave, and/or split (globally or locally efficient) PFs. The respective decision spaces also offer a variety of structural challenges (see, e.g., Fig. 3). Our approach thus provides a valuable framework for creating various configurable and scalable MOPs that will be useful for meaningfully benchmarking MOEAs and studying their strengths and weaknesses.

Algorithms Next to classical MOEAs such as NSGA-II [6] and SMS-EMOA [10], which focus on the approximation of globally optimal solutions and convergence in objective space, specific MOEAs exist, which are explicitly suited for overcoming obstacles of multimodal MOPs and for exploiting local structures. A comprehensive overview of different MOEA categories, including multimodality aspects both from a multi-local and multi-global perspective, is given in [12,23].

Herein, we focus on specific MOEAs like Omni-Optimizer [8] and MOLE [21]. These approaches were shown to be more competitive than classical MOEAs w.r.t. convergence in objective space while simultaneously ensuring diversity in decision space [18,14]. Omni-Optimizer is conceptually similar to NSGA-II but additionally comprises a diversity preservation strategy in decision space. It is thus applicable to various types of MOPs. MOLE, however, is a gradient-based MO optimizer that actively explores and traverses locally efficient sets. Thereby, it exploits interactions between their respective basins of attraction for a directed descent towards dominating local (or even global) sets.

3 On the Pareto Set of Multiple Peaks Functions

As outlined in Sec. 2, MPM functions essentially take the minimum of multiple individual peak functions, whose shape and placement are up to the generator. Using the MPM2 generator, the individual peak functions obtain ellipsoidal level sets, which makes them approachable using analytical techniques. We will first illustrate how the PS of two unimodal peak functions can be derived and then discuss how this analysis scales with increasing numbers of peaks per function.

Bi-Objective Convex-Quadratic Problems A theoretical analysis of the PS between individual peak functions stemming from the MPM2 generator has been conducted before in [16]. Here, however, we base our discussion on the results of [24], which focus on bi-objective convex-quadratic problems. Consider the following bi-objective convex-quadratic problem $F(x)$ with search space $\mathcal{X} = \mathbb{R}^d$:

$$F(x) = (f_1(x), f_2(x)) \rightarrow min! \quad \text{with } f_i(x) = \frac{1}{2}(x - x_i^*)^T H_i (x - x_i^*), i = 1, 2,$$

where $x_1^*, x_2^* \in \mathbb{R}^d$ are the global optima of f_1 and f_2, respectively. Likewise, $H_1, H_2 \in \mathbb{R}^{d \times d}$ are positive definite symmetric Hessian matrices determining the shape and orientation of the quadratic functions.

For convex quadratic functions, the position of the global optimum is determined by the singular point at which the gradient equals the zero vector. Further, the PF of two convex quadratic problems is convex [11], which enables us to express all globally optimal points by linear interpolation between the problems using a parameter $t \in [0, 1]$:

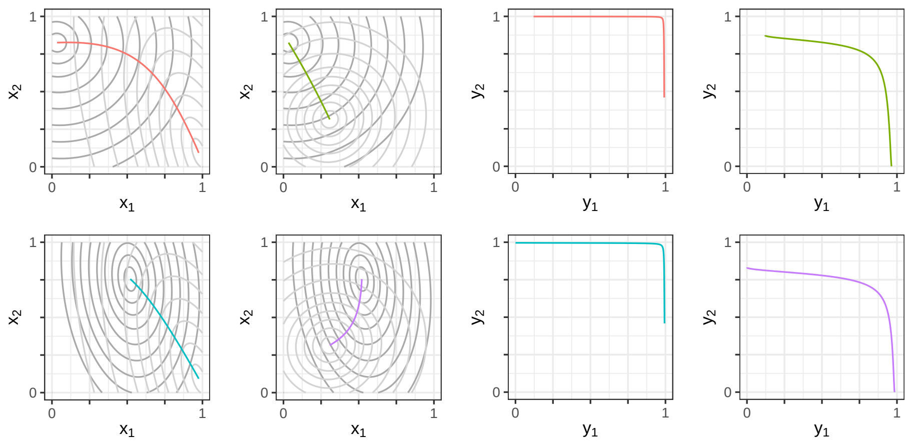
$$
\begin{aligned}
F_t(x) &= (1-t)f_1(x) + tf_2(x) \rightarrow min! \\
\Rightarrow &\nabla F_t(x) = (1-t)\nabla f_1(x) + t\nabla f_2(x) = 0 \\
&(1-t)H_1(x - x_1^*) + tH_2(x - x_2^*) = 0 \\
&[(1-t)H_1 + tH_2]^{-1}[(1-t)H_1 x_1^* + tH_2 x_2^*] = x
\end{aligned}
$$

In the work of [24], some additional results are shown, e.g., monotone objective transformations do not impact the placement of the PS, while PF properties, such as its shape, can be adjusted. As the peaks resulting from MPM2 constitute a convex quadratic function to which a monotone transformation is applied in the objective space only, the above analysis also describes the PS of a bi-objective unimodal peak function. Finally, note that this does not guarantee that the PS is contained within the usual $[0, 1]^d$ bounding box for MPM2 problems and the decision space boundaries may need slight adjustment when this unconstrained PS should still be reachable.

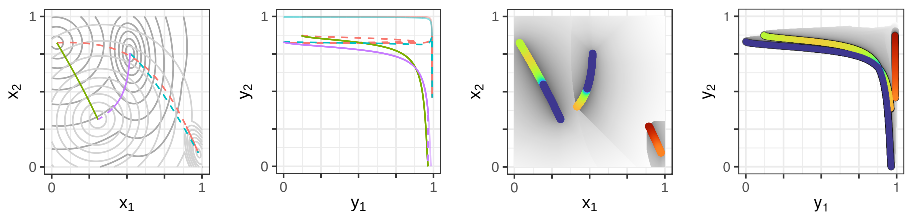
Multiple Peaks While we can only describe the PS analytically given a combination of two unimodal peak functions, we can leverage this knowledge for constructing the PS of bi-objective MPM2 problems. The first necessary insight is that this specific PS has to be a subset of the theoretical PSs which are generated by an exhaustive combination of each pair of peaks, cf. Figure 1.

This leaves us with considering all pairwise peak functions, i.e., one peak per objective, and their respective PSs, cf. Figure 1a. While new locally efficient solutions cannot be generated, some of the analytical PSs may become partially or completely inactive because of being dominated by objective values of another peak combination. Additionally, they may become globally dominated by the PSs of other peak pairs, thereby creating complex local and global interactions. Considering the example in Figure 1b, the blue set (corresponding to the bottom left image in Figure 1a) is completely inactive, while the red and violet sets are partially inactive, and only the green set stays fully active. These dynamics are mirrored in the PLOT visualization (see bottom right image in Figure 1b) [19], which only depicts the (local) PSs and PFs, respectively.

While these local dynamics can become very complex, it is not necessary to study them in detail, if one is only interested in deriving the PS and PF, i.e., the globally efficient points. We propose a simple, numerical procedure for deriving a set-based approximation building on the analytical description of the PSs of the peak pairs (see Figure 2). As a target, we choose to approximate the dominated hypervolume (HV) w.r.t. the reference point $(1, 1)$. We start by evaluating evenly spaced points w.r.t. the parameter t for each of the pairwise theoretical fronts. In our implementation, the minimum resolution is 4 points per set.



(a) Pairwise peak combinations and corresponding Pareto sets in dec. and obj. space.



(b) Left: Active (solid) and inactive (dashed) sets of the combined MPM2 problems. Sets are inactive if they are dominated by another pair of peaks. In the objective space, the inactive theoretical sets are printed semi-transparently. Right: PLOT [19] visualization, showing locally efficient sets (in color) and attraction basins (in grayscale).

Fig. 1: Here, we demonstrate the effects of combining individual peak functions to one MPM2 problem on the Pareto set. The exemplary problem consists of two peaks per objective with a random topology and seeds 667172 and 540835.

For each set, we then individually compute the best possible intermediate points, i.e., the minimum of two consecutive points, that could still be contained within the set without dominating any other point in it. When interpolating on the same (theoretical) set, no more dominant points could be found, making the intermediates the most optimistic estimate between two adjacent points from a set. The (relative) HV gap is then given by the difference (percentage) of the HV dominated by i) the best possible intermediates, and ii) the actually evaluated points. This process is repeated with doubled resolution until the HV gap reaches a sufficiently small target value. To save on evaluations, sets whose best possible intermediates are fully dominated by the union of all evaluated points so far can be excluded in the respective iteration, as they cannot contribute to the PF.

Although we can pessimistically estimate the computational complexity proportional to the number of initial theoretical Pareto sets times the inverse of the

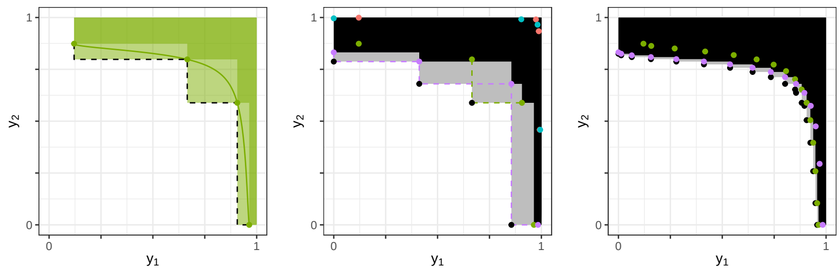


Fig. 2: Left: Illustration of the hypervolume gap (light shaded area) for the green front of the problem shown in Figure 1. Middle: The hypervolume gap (gray area) for a resolution of 4 points per set, where the black points denote the nondominated points from the best possible intermediates for each set. Right: The same gap for a resolution of 16 points per set. Note that the red and blue sets could be excluded as they cannot contribute to the Pareto front.

target gap, practical computations become surprisingly efficient. This is due to the observation that many of the theoretical local sets can be excluded early.

4 Experimental Study

In the following proof-of-concept study, we (1) demonstrate the properties of the proposed test suite subject to its main parameters and (2) conduct first performance analyses of several standard and multimodality-affine solvers.

Setup We created a total of 1,280 problems using the following configuration: We select the search space dimensionality $d \in \{2, 3, 5, 10\}$ to cover an increasing but still manageable decision space. The number of peaks $n_p \in \{1, 2, 4, 8, 16, 32, 64, 128\}$ per MPM2 function is exponentially scaled to enable analyzing the influence of SO multimodality on a log-scale. Note that n_p coincides for both constituent MPM2 functions to obtain a meaningful amount of unambiguous classes for further analyses. The same rationale applies to the topology, i.e., we select a specific topology parameter $t \in \{\texttt{funnel}, \texttt{random}\}$ for both problems simultaneously. Finally, we choose the random seeds $s \in \{1, \ldots, 20\}$ for the first objective, while the seed for the second objective is always set to $s + 1{,}000$. For all problems, we then approximate the HV of the PF to a relative gap of at most 10^{-4}, i.e., we have an uncertainty of less than 0.01% about the optimal HV of each problem. This approximation takes at most a few seconds per problem. We set the decision space boundaries to $[-0.2, 1.2]^d$, ensuring all Pareto sets are fully included.

To obtain the problem characteristics and to optimize the HV, we extract the parameters for all peaks of the MPM2 problems using an interface implemented in the R-package `smoof` [4]. Our experiments and analyses are also conducted in R and can be found at `https://github.com/schaepermeier/2023-emo-mpm2`.

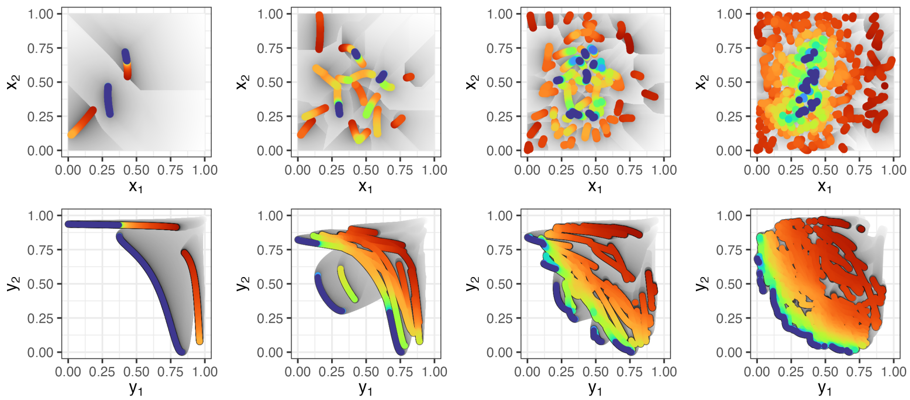
We run the algorithms introduced in Sec. 2 using an interface to the C implementations of NSGA-II and Omni-Optimizer provided by the `mco` [17] and `omnioptr` [5] packages, respectively. Further, we rely on the default implementation of SMS-EMOA provided by the `ecr` package [3]. Note that it uses random parent selection rather than the tournament scheme implemented in NSGA-II and Omni-Optimizer, which has to be taken into account in the experimental evaluation. We use MOLE with random uniform starting points as provided by the `moleopt` package in the default configuration, but setting its internal HV target parameter to 10^{-3}, to reduce time spent refining already found solutions. All MOEAs have their population set to 100, which is a common default.

We run all algorithms for 10,000 evaluations, i.e., 100 generations for NSGA-II and Omni-Optimizer. As performance measure, we consider the dominated HV w.r.t. the reference point $(1, 1)$ provided by the archive of evaluated points. For performance reasons, we compute the achieved HV every 100 function evaluations. We perform 15 repetitions per combination of problem and algorithm to ensure statistically reliable results.

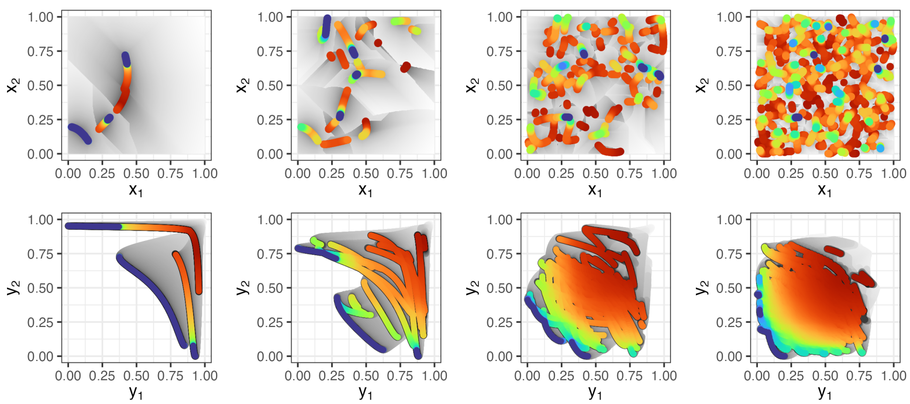
Analysis of Problem Properties To illustrate interesting properties of the generated problems, we perform two separate analyses: We start by visualizing two-dimensional problem landscapes to visually show the impact of the degrees of multimodality and the two problem topologies. We then focus on problem characteristics in objective space when scaling the search space dimensionality.

Figure 3 shows PLOT [19] visualizations of two-dimensional bi-objective problems generated with funnel and random topologies for 2, 8, 32, and 128 peaks, respectively. Here, it can be observed that the degree of multimodality greatly increases with the number of peaks used in the individual problems, as demonstrated by the number and location of visualized locally efficient points. Although the problems with lower multimodality still seem somewhat similar between topologies, varying distribution of the locally and globally efficient sets becomes apparent with an increasing number of peaks. The funnel problems tend to cluster the globally (and to an extent locally) efficient points, while the random problems show a much higher dispersion in the decision space, with many smaller areas contributing to the (global) PS. These representative problems also show a distinctive property in objective space: In random topologies the PF rapidly approaches the ideal point $(0, 0)$ with increasing number of peaks, while respective closeness is limited in the funnel topology.

Figure 4 visualizes the influence of generator parameters on PS properties. On the one hand, across dimensions and topologies, we can see that the number of locally efficient sets contributing to the Pareto set increases with the number of peaks. However, this effect decreases with increasing dimensionality and differences between the topologies in this regard almost vanish. Further, the approximated HV increases with the number of peaks. While it approaches the maximum possible HV of 1 for the random topology, funnel problems demonstrate a much wider range of HV values and a slower increase with the number of peaks. Again, the effect diminishes with increasing dimensionality.



(a) Problems with funnel topology in decision (top) and objective space (bottom).



(b) Problems with random topology in decision (top) and objective space (bottom).

Fig. 3: PLOT visualizations of problems with 2, 8, 32 and 128 peaks per objective (left-to-right) for funnel and random topologies. All problems use seed $s = 1$.

Finally, Figure 5 shows properties of the locally and globally efficient fronts, exemplarily for the random topology, with increasing dimension. Here, we can see that, using otherwise identical parameters, the PF is becoming more concave within our generator framework. This seems to be a property of the peak function, and should be investigated in detail in the future.

Algorithm Comparison Based on our knowledge of optimal HV values and regularly conducted HV evaluations during the optimization process, we can generate convergence plots as exemplarily provided in Figure 6 for 5D problems. It depicts the mean relative hypervolume gap, i.e., the mean percentage of hypervolume not yet covered, per problem w.r.t. function evaluations. Several insights can be gained: Firstly, problem hardness tends to increase with an increasing number of peaks per problem. This is particularly noticeable for the

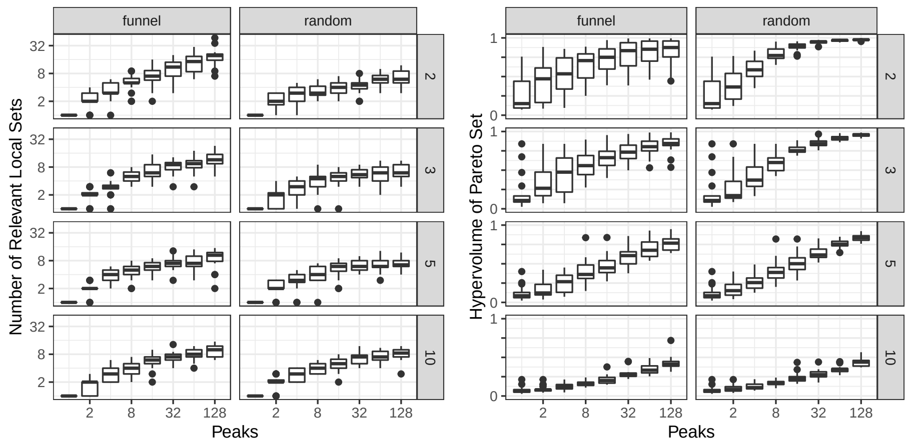


Fig. 4: Left: Number of local sets that contain globally efficient solutions. Right: Computed HV of PS. Rows indicate problem dimensionality, $d \in \{2, 3, 5, 10\}$.

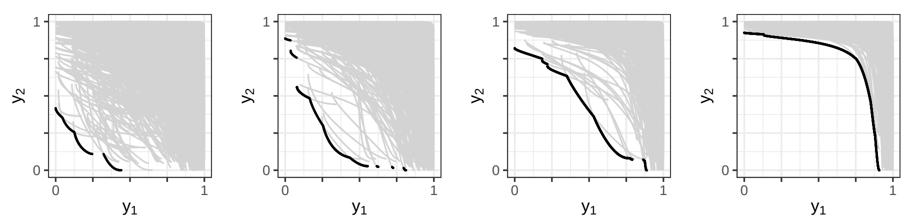


Fig. 5: Locally and globally efficient fronts for the problem with 32 peaks, $s = 1$ and random topology for dimensions 2, 3, 5, and 10. Higher-dimensional problems have fewer visible disconnects and overall a more concave front shape.

MOLE algorithm, which, as a purely local search approach, is slowed down by the amount of locally efficient points, while the performance of the evolutionary algorithms is less affected. Further, by comparing the achieved values at the end of the runs for the 32 and 128 peaks problems, we see that the random problems tend to be slightly harder to solve than the funnel problems for the EAs, while MOLE is less affected.

Figure 7 shows critical difference plots [9] for all problem dimensions and topologies. It validates that, in general, Omni-Optimizer and NSGA-II perform best, though only in 10D Omni-Optimizer is clearly superior. They are followed by SMS-EMOA and MOLE. For SMS-EMOA, the mentioned random parent selection scheme implemented in `ecr` might be the reason for its comparatively bad performance w.r.t. NSGA-II. Further, MOLE's performance relative to SMS-EMOA is improving with dimensionality, although they are always statistically tied. MOLE also tends to perform slightly better on random than on funnel topologies. Finally, while random search seems to have some merit in lower dimensions (2, 3), it is clearly the worst performer in higher dimensions (5, 10).

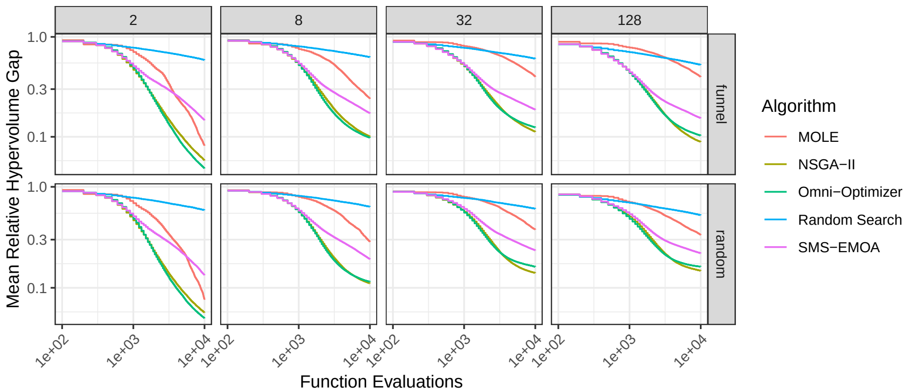


Fig. 6: Convergence plots of the aggregated algorithm performances for the 5D problems. Columns denote the number of peaks, while rows show the topology. Each algorithm was evaluated with 15 repetitions on 20 problems per group.

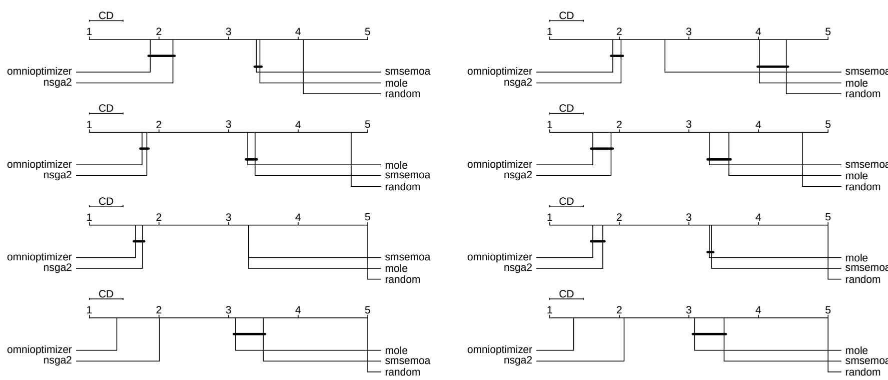


Fig. 7: Critical differences for the mean final HV gap per problem in the random (left) and funnel (right) topologies for dimensions 2, 3, 5, and 10 (top-to-bottom).

5 Conclusions

In this work, we introduced a new methodology for determining the globally optimal solutions of MOPs, which are created based on multiple peak problems, to an arbitrary precision in terms of dominated HV. We apply this methodology for developing tools that are able to generate a wide range of benchmark problems with ground truth regarding the PS and PF while simultaneously having complex landscape characteristics. The highly parametrizable generator facilitates the creation of problems with specific structural properties, which in turn is essential for conducting structured analyses of landscape properties of MOPs. Next to landscape analyses, the ground truth enables a systematic benchmarking of algorithms, which we demonstrated in a proof-of-concept study.

We consider our work fundamental to perspectively constructing diverse, yet well-understood, MO benchmark problems in order to enhance and meaningfully complement existing benchmark suites. Our presented framework offers promising perspectives for future research in various important areas ranging from benchmark design and understanding algorithm behavior to characterizing problem landscapes and measuring optimizer performances.

First, considering additional peak shape functions and other topologies enables and facilitates the straightforward construction of a broader scope and thus a more diverse set of benchmark problems. Integrating decision and objective space transformations provides an additional promising avenue for future extensions. Such transformations could be, for instance, the introduction of asymmetries (similar to those used in the single-objective BBOB test suite) into the previously constructed multiple peaks functions. We would also like to point out that the mathematical analysis of the PS can easily be extended to more objectives. However, in this case, it is not intuitively clear how to generalize the PS approximation.

In addition to constructing more comprehensive benchmark suites for *global* MO optimization, we are also interested in facilitating investigations of the *local* search dynamics. Therefore, analyzing an algorithm's convergence to locally efficient sets, e.g., using the Basin-Based Evaluation (BBE) method proposed in [14], represents another compelling and feasible extension of our framework. Aside from investigating the convergence of algorithms with BBE, considering performance metrics beyond HV and providing target values for an arbitrary precision represents another meaningful perspective for future work.

Another prospective research avenue could be the design of measurable landscape features to characterize (local and global) structural properties of purposefully constructed problems with different complexity and known ground truth w.r.t. efficient sets. This will be an essential intermediate step towards (i) characterizing MO problems in general (including high-dimensional problems that are not visualizable anymore), as well as (ii) developing feature-based approaches such as automated algorithm selection.

Finally, we emphasize that our experimental study is intended to illustrate first proof-of-concept takeaways. For future work, we envision our approach enabling a sound and reliable comparison of MO optimizers by evaluating them on a broader set of problems with known structural challenges and also configuring them via automated algorithm configuration methods [18,15,2]. This will ultimately lead to a better understanding of algorithmic components and pave the ground for better algorithm design.

Acknowledgment

The authors acknowledge support by the *European Research Center for Information Systems (ERCIS)*. Further, L. Schäpermeier and P. Kerschke acknowledge support by the *Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig*.

References

1. Bartz-Beielstein, T., Doerr, C., Berg, D.v.d., Bossek, J., Chandrasekaran, S., Eftimov, T., Fischbach, A., Kerschke, P., La Cava, W., Lopez-Ibanez, M., Malan, K.M., Moore, J.H., Naujoks, B., Orzechowski, P., Volz, V., Wagner, M., Weise, T.: Benchmarking in Optimization: Best Practice and Open Issues (2020). https://doi.org/10.48550/arxiv.2007.03488
2. Blot, A., Hoos, H., Jourdan, L., Marmion, M., Trautmann, H.: MO-ParamILS: Multi-objective automatic algorithm configuration framework. In: et al. Joaquin, V. (ed.) Proceedings of the Learning and Intelligent Optimization, 10^{th} International Conference. pp. 32–47. Springer International Publishing, Ischia (2016)
3. Bossek, J.: ecr 2.0: a modular framework for evolutionary computation in R. In: Proceedings of the genetic and evolutionary computation conference companion. pp. 1187–1193 (2017)
4. Bossek, J.: smoof: Single- and Multi-Objective Optimization Test Functions. The R Journal **9**(1) (2017)
5. Bossek, J., Deb, K.: omnioptr: Omni-optimizer algorithm. `https://github.com/jakobbossek/omnioptr` (2021)
6. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation (TEVC) **6**(2), 182–197 (2002)
7. Deb, K., Thiele, L., Laumanns, M., Zitzler, E.: Scalable Test Problems for Evolutionary Multiobjective Optimization. In: Evolutionary Multiobjective Optimization, pp. 105 – 145. Springer (2005)
8. Deb, K., Tiwari, S.: Omni-Optimizer: A Generic Evolutionary Algorithm for Single and Multi-Objective Optimization. European Journal of Operational Research (EJOR) **185**, 1062–1087 (2008)
9. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. The Journal of Machine learning research **7**, 1–30 (2006)
10. Emmerich, M., Beume, N., Naujoks, B.: An EMO algorithm using the hypervolume measure as selection criterion. In: Proceedings of the International Conference on Evolutionary Multi-Criterion Optimization (EMO). pp. 62–76. Springer (2005)
11. Glasmachers, T.: Challenges of convex quadratic bi-objective benchmark problems. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 559–567 (2019)
12. Grimme, C., Kerschke, P., Aspar, P., Trautmann, H., Preuss, M., Deutz, A., Wang, H., Emmerich, M.: Peeking Beyond Peaks: Challenges and Research Potentials of Continuous Multimodal Multi-objective Optimization. Computers & Operations Research **136**, 105489 (2021)
13. Hansen, N., Finck, S., Ros, R., Auger, A.: Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions. Tech. Rep. RR-6829, INRIA (2009)
14. Heins, J., Rook, J., Schäpermeier, L., Kerschke, P., Bossek, J., Trautmann, H.: Bbe: Basin-based evaluation of multimodal multi-objective optimization problems. In: International Conference on Parallel Problem Solving from Nature. pp. 192–206. Springer (2022)
15. Hoos, H.H.: Automated algorithm configuration and parameter tuning. In: Hamadi, Y., Monfroy, E., Saubion, F. (eds.) Autonomous Search, pp. 37–71. Springer (2012)

16. Kerschke, P., Wang, H., Preuss, M., Grimme, C., Deutz, A.H., Trautmann, H., Emmerich, M.T.M.: Search Dynamics on Multimodal Multiobjective Problems. Evolutionary Computation **27**(4), 577–609 (Dec 2019)
17. Mersmann, O., Trautmann, H., Steuer, D., Bischl, B., Deb, K.: mco: Multiple criteria optimization algorithms and related functions, R package, version 1.0-15.6. `https://github.com/olafmersmann/mco` (2020)
18. Rook, J., Trautmann, H., Bossek, J., Grimme, C.: On the Potential of Automated Algorithm Configuration on Multi-Modal Multi-Objective Optimization Problems. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 356–359. ACM, New York, NY, USA (2022)
19. Schäpermeier, L., Grimme, C., Kerschke, P.: One PLOT to Show Them All: Visualization of Efficient Sets in Multi-objective Landscapes. In: International Conference on Parallel Problem Solving from Nature (PPSN). pp. 154 – 167. Springer (2020)
20. Schäpermeier, L., Grimme, C., Kerschke, P.: To Boldly Show What No One Has Seen Before: A Dashboard for Visualizing Multi-objective Landscapes. In: Evolutionary Multi-Criterion Optimization. pp. 632–644. Springer (2021)
21. Schäpermeier, L., Grimme, C., Kerschke, P.: MOLE: Digging Tunnels Through Multimodal Multi-Objective Landscapes. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO). pp. 592–600. ACM (2022)
22. Schäpermeier, L., Grimme, C., Kerschke, P.: Plotting Impossible? Surveying Visualization Methods for Continuous Multi-Objective Benchmark Problems. IEEE Transactions on Evolutionary Computation pp. 1–1 (2022)
23. Tanabe, R., Ishibuchi, H.: A Review of Evolutionary Multi-Modal Multi-Objective Optimization. IEEE Transactions on Evolutionary Computation (TEVC) **24**(1), 193 – 200 (2020)
24. Toure, C., Auger, A., Brockhoff, D., Hansen, N.: On bi-objective convex-quadratic problems. In: International Conference on Evolutionary Multi-Criterion Optimization. pp. 3–14. Springer (2019)
25. Tušar, T., Brockhoff, D., Hansen, N., Auger, A.: COCO: The Bi-Objective Black Box Optimization Benchmarking (bbob-biobj) Test Suite. arXiv preprint **abs/1604.00359** (2016)
26. Wessing, S.: The Multiple Peaks Model 2. Tech. Rep. TR15-2-001, TU Dortmund University, Germany (2015)
27. Wessing, S.: Two-stage methods for multimodal optimization. PhD thesis, University of Dortmund (01 2015). https://doi.org/10.17877/DE290R-7804
28. Yue, C., Qu, B., Yu, K., Liang, J., Li, X.: A Novel Scalable Test Problem Suite for Multimodal Multiobjective Optimization. Swarm and Evolutionary Computation **48**, 62 – 71 (2019)
29. Zitzler, E., Deb, K., Thiele, L.: Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. Evolutionary Computation (ECJ) **8**(2), 173 – 195 (2000)