Towards Analyzing Multimodality of Continuous Multiobjective Landscapes

Pascal Kerschke[1(✉)], Hao Wang[2], Mike Preuss[1], Christian Grimme[1],
André Deutz[2], Heike Trautmann[1], and Michael Emmerich[2]

[1] Information Systems and Statistics Group,
University of Münster, Münster, Germany
{pascal.kerschke,mike.preuss,christian.grimme,
heike.trautmann}@wi.uni-muenster.de
[2] LIACS, Leiden University, Leiden, The Netherlands
{h.wang,a.h.deutz,m.t.m.emmerich}@liacs.leidenuniv.nl

Abstract. This paper formally defines multimodality in multiobjective optimization (MO). We introduce a test-bed in which multimodal MO problems with known properties can be constructed as well as numerical characteristics of the resulting landscape. Gradient- and local search based strategies are compared on exemplary problems together with specific performance indicators in the multimodal MO setting. By this means the foundation for Exploratory Landscape Analysis in MO is provided.

Keywords: Multiobjective optimization · Multimodality · Landscape analysis · Hypervolume gradient ascent · Set based optimization

1 Introduction

In multiobjective optimization, respective algorithms are particularly challenged by multimodality of the underlying landscape caused by the interaction of objective functions. Thus, sophisticated Exploratory Landscape Analysis (ELA, [7]) features which are able to assess the level and type of multimodality based on an initial problem sample have huge potential for understanding algorithm behaviour, automated algorithm selection and algorithm design. Despite the success of ELA in single-objective continuous black-box optimization (e.g. [1,4]) multiobjective optimization has not been appropriately addressed apart from limited approaches in the combinatorial context (e.g. [6,9]) or expert-based characteristics such as the Pareto front shape, the dimensionality and some intuitions on multimodality. We here lay the groundwork for constructing such experimental features systematically by providing formal definitions of multimodality in terms of distinguishing between local and global efficient sets. A versatile problem generator is introduced for designing multimodal mixed sphere problems with predefined characteristics. Bringing together theoretical analysis and experiments, and contrasting gradient and local search based methods, highly increases understanding of the problem domain multimodality in multiobjective optimization as well as the explorative algorithm.

Section 2 introduces topological definitions, Sect. 3 details the problem generator and theoretically analyzes the resp. multimodal structures, Sect. 4 discusses the exploration algorithms, Sect. 5 presents algorithm and problem characteristics, and Sect. 6 provides experimental results. Conclusions are drawn in Sect. 7.

2   Multimodality

We present an approach of defining multimodality in that we distinguish between global and local efficient sets in $\mathbb{R}^d$ (the decision space). We are aware that most parts can be generalized to other spaces. We first recall some topological notions:

1. Let $A \subseteq \mathbb{R}^n$. The set A is called *connected* if and only if there do not exist two open, *disjoint* subsets U_1 and U_2 of $\mathbb{R}^n$ such that $A \subseteq U_1 \cup U_2$, $U_1 \cap A \neq \emptyset$, and $U_2 \cap A \neq \emptyset$.
2. Let $B \subseteq \mathbb{R}^n$. A subset $C \subseteq B$ is a *connected component* of B iff C is connected, and any subset of B which is a strict superset of C is not connected, and C is non-empty.

Pareto concepts are given next: Let $\mathbf{f} : \mathcal{X} \to \mathbb{R}^m$ be a multiobjective function where $\mathcal{X} \subseteq \mathbb{R}^d$ is the decision space. We will denote the component functions of $\mathbf{f}$ by $f_i : \mathcal{X} \to \mathbb{R}, i = 1, \ldots, m$. Given a totally ordered set $(T, \leq)$ where $\leq$ denotes the total order, we can define as usual the Pareto order, denoted by $\prec$, on T^k for any $k \in \mathbb{N}$ as follows. Let $\mathbf{t}^{(1)} = (t_1^{(1)}, \ldots, t_k^{(1)}), \mathbf{t}^{(2)} = (t_1^{(2)}, \ldots, t_k^{(2)})$ be elements of T^k. We say $\mathbf{t}^{(1)} \prec \mathbf{t}^{(2)}$ iff $t_i^{(1)} \leq t_i^{(2)}, i = 1, \ldots, k$ and $\mathbf{t}^{(1)} \neq \mathbf{t}^{(2)}$. Specializing this to the reals with their natural, total order we obtain the Pareto order on $\mathbb{R}^m$. A point $\mathbf{x} \in \mathcal{X}$ is called *Pareto efficient* or *global efficient* or for short *efficient* iff there does not exist $\tilde{\mathbf{x}} \in \mathcal{X}$ such that $\mathbf{f}(\tilde{\mathbf{x}}) \prec \mathbf{f}(\mathbf{x})$. The subset of $\mathcal{X}$ consisting of all the efficient points of $\mathcal{X}$ is denoted by $\mathcal{X}_E$ and is called the *efficient subset* of $\mathcal{X}$ (or the *efficient* set of $\mathbf{f}$). The image of $\mathcal{X}_E$ under $\mathbf{f}$ is called the Pareto front of $\mathbf{f}$. To define local efficient points in $\mathcal{X}$ and local efficient *sets* in the multiobjective case, we propose the following definitions:

Definition 1 (Efficiency of Points/Sets). *A point $\mathbf{x} \in \mathcal{X}$ is called a <u>locally efficient point</u> of $\mathcal{X}$ (or of $\mathbf{f}$) if there is an open set $U \subseteq \mathbb{R}^d$ with $\mathbf{x} \in U$ such that there is no point $\tilde{\mathbf{x}} \in U \cap \mathcal{X}$ such that $\mathbf{f}(\tilde{\mathbf{x}}) \prec \mathbf{f}(\mathbf{x})$. The subset of all the local efficient points of $\mathcal{X}$ is denoted by $\mathcal{X}_{LE}$.*

A point $\mathbf{x} \in \mathcal{X}$ is called a <u>global efficient point</u> of $\mathcal{X}$ (or of $\mathbf{f}$) if there is no point $\tilde{\mathbf{x}} \in \mathbb{R}^d \cap \mathcal{X}$ such that $\mathbf{f}(\tilde{\mathbf{x}}) \prec \mathbf{f}(\mathbf{x})$. The subset of all the global efficient points of $\mathcal{X}$ is termed <u>efficient set</u> of $\mathbf{f}$ and denoted by $\mathcal{X}_E$.

A subset $A \subseteq \mathcal{X}$ is a <u>local efficient set</u> of $\mathbf{f}$ if A is a connected component of $\mathcal{X}_{LE}$ (= the subset of $\mathcal{X}$ which consists of the local efficient points of $\mathcal{X}$).

Definition 2 (Local Pareto Front). *A subset P of the image of $\mathbf{f}$ is a local Pareto front of $\mathbf{f}$, if there exists a local efficient set E such that $P = \mathbf{f}(E)$.*

The (global) Pareto front (PF) of $\mathbf{f}$ is obtained by taking the image under $\mathbf{f}$ of the union of the connected components of the set of global efficient points of $\mathcal{X}$. If $\mathcal{X}_E$ is connected, then the (global) Pareto front of $\mathbf{f}$ is also connected, provided $\mathbf{f}$ is continuous on $\mathcal{X}_E$. One might also consider definitions related to connected components in the objective space [8]. However, we will omit this for brevity.

3   Analytics on Simple Mixed Sphere Problems

We use a sophisticated problem generator based on the Multiple Peaks Model 2 (MPM2, [10]) to illustrate the proposed topological definitions and further analyse the behavior of explorative algorithms.

$$f(\mathbf{x}) = 1 - \max_{1 \leq i \leq N} \{g_i(\mathbf{x})\}, \quad \mathbf{x} \in \mathbb{R}^d \tag{1}$$

$$g_i(\mathbf{x}) = H_i \left(1 + \left(\sqrt{(\mathbf{x}-\mathbf{c}_i)^T \mathbf{D}(\mathbf{x}-\mathbf{c}_i)}\right)^{s_i} / R_i\right)^{-1}, \quad i = 1, \ldots, N \tag{2}$$

Note that the g_i functions define peaks with center $\mathbf{c}_i$, depth H_i, radius R_i and shape s_i. $\mathbf{D}$ is the inverse of the covariance matrix while we concentrate on spherical peaks with isotropic level curves (*mixed spheres*), i.e. $\mathbf{D} = c\mathbf{I}, c \in \mathbb{R}_{\neq 0}$.

A bi-objective optimization problem $(f_1(\mathbf{x}|g_i), f_2(\mathbf{x}|g'_i)) \to \min$ results in choosing two different parameter sets – parameters of f_2 are labeled by the prime symbol. Exemplary problems are illustrated in Figs. 1, 2 and 3.

In order to evaluate and compare the Pareto fronts obtained by the optimization algorithms used in this paper, the analytical Pareto front (and efficient set) is derived in the following. First, we focus on the simplest case where each objective function consists of only one peak. In this case the Pareto efficient set PE is the line segment connecting $\mathbf{c}$ to $\mathbf{c}'$: $PE : \{\alpha\mathbf{c} + (1-\alpha)\mathbf{c}' \mid 0 \leq \alpha \leq 1\}$.

Then the parametric form of the Pareto front can be derived by mapping an arbitrary point in the efficient set $\hat{\mathbf{x}} = \alpha\mathbf{c} + (1-\alpha)\mathbf{c}'$ through the objective functions, which – using the Mahalanobis distance $d(\mathbf{c}, \mathbf{c}'; \mathbf{D}) = \sqrt{(\mathbf{c}'-\mathbf{c})^T \mathbf{D}(\mathbf{c}'-\mathbf{c})}$ – finally results after algebraic transformation in f_2 as a function of f_1, for $\mathbf{c} \neq \mathbf{c}'$:

$$f_2 = 1 - H' \left(1 + \left(d(\mathbf{c}, \mathbf{c}'; \mathbf{D}') \left(1 - \frac{R^{1/s}}{d(\mathbf{c}, \mathbf{c}'; \mathbf{D})} \left(\frac{H}{1-f_1} - 1\right)^{1/s}\right)\right)^{s'} / R'\right)^{-1}$$

The range of f_1 is $[\min\{f_1(\mathbf{c}), f_1(\mathbf{c}')\}, \max\{f_1(\mathbf{c}), f_1(\mathbf{c}')\}]$.

Using the expression above, we could calculate the red part of the global Pareto front in Fig. 1. For multiple peaks the (local) efficient sets still settle on line segments connecting each pair of peaks. It is difficult to derive the analytical expression because the *effective* peak might change when traversing along the line segment connecting peaks and multiple local efficient sets could exit on the same line segment (check Fig. 3 for example). However, it is possible to approximate the local efficient sets numerically by uniformly sampling on the line segments and taking the maximal non-dominated subset of the samples.
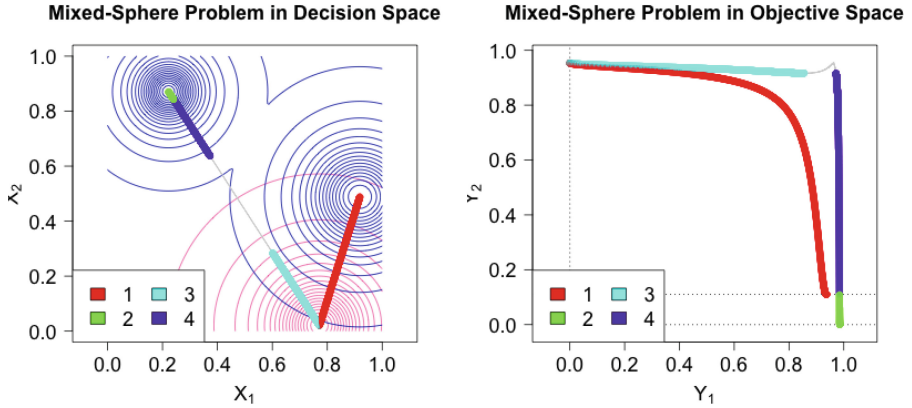
Fig. 1. Example of a simple mixed sphere problem in the decision (left figure) and objective space (right). Objectives are visualized in the decision space by pink (objective 1) and blue (objective 2) contour lines. The connections between peaks from the two objectives are shown as grey lines and the corresponding local efficient sets (or fronts) are colored. Here, the red and green parts form the disconnected global PF, whereas the cyan and purple parts show the remaining disconnected local PFs. The given scenario represents three disconnected local efficient sets (green/purple, cyan, red), two domination layers (red/green vs. cyan/purple) and four local Pareto fronts. (Color figure online)
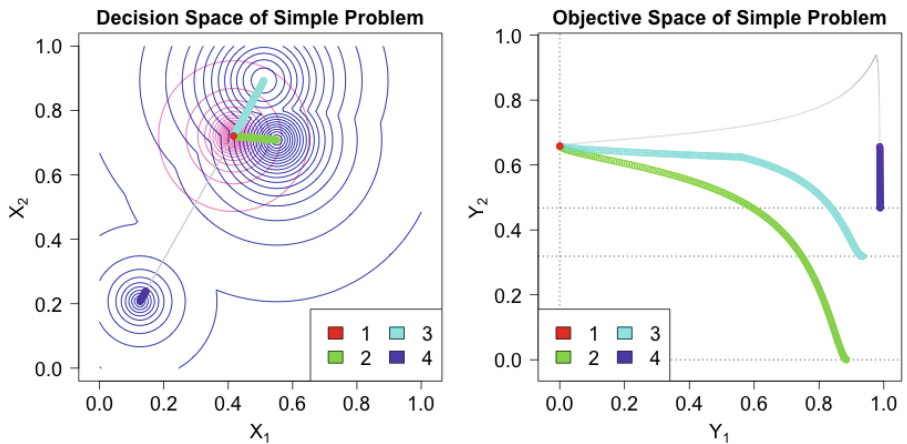

Fig. 2. Local Pareto fronts in the decision (left) and objective space (right) for a rather simple mixed sphere problem consisting of one peak in the first objective (pink contour lines) and three peaks in the second objective (blue contour lines). The red area is caused by the fact that it belongs to the same local efficient set as the cyan area, and at the same time to the global dominance layer. (Color figure online)
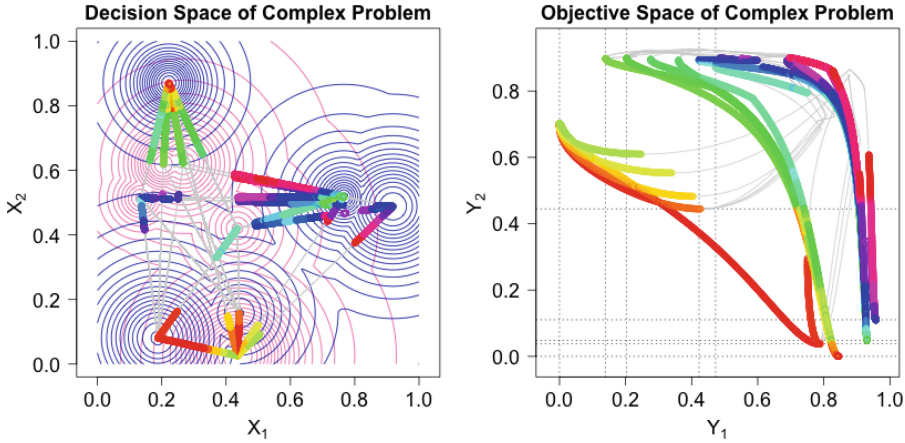
Fig. 3. Local Pareto fronts for a complex mixed sphere problem consisting of five peaks per objective, resulting in a total of 30 disconnected local efficient sets, 19 domination layers and 167 local Pareto fronts.

4 Explorative Algorithms

Hypervolume Indicator Gradient Ascent (HIGA-MO). Taking the advantage of the analytically computable gradient information of the mixed sphere problems, we choose the *Hypervolume Indicator Gradient Ascent* [2,3], because it is capable of generating well-distributed PF approximations and is almost free of control parameters. The corresponding pseudo-code is given in Algorithm 1. The basic idea is to maximize the Hypervolume indicator $\mathcal{H}$ of an approximation set of the true PF by using the gradient of $\mathcal{H}$. We denote the set of search points as $\{\mathbf{x}_1, \ldots, \mathbf{x}_\mu\}, \mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{X} = [\mathbf{x}_1^T, \ldots, \mathbf{x}_\mu^T]^T \in \mathbb{R}^{\mu d}$. As $\mathcal{H}$ can also be expressed as a function of the input vectors, one can calculate the Hypervolume indicator gradient $\partial \mathcal{H}(\mathbf{X})/\partial \mathbf{X}$. By applying the chain rule the so-called *subgradients* $\partial \mathcal{H}(\mathbf{X})/\partial \mathbf{x}_i = [\partial \mathcal{H}(\mathbf{X})/\partial \mathbf{y}_i] \cdot [\partial \mathbf{y}_i/\partial \mathbf{x}_i]$ [2] can be computed for $i = 1, \ldots, \mu$. In practice, a step-size control is used to adapt the step-size for each decision vector.[1]

HIGA-MO performs a fast local search and some individuals might get stuck in a local efficient set. However, in mixed sphere problems local efficient sets might be connected to the global one and the Hypervolume indicator gradient will steer the local efficient points towards the global one.

Stochastic Local Search (SLS). A simple local search strategy based on parallel perturbation and elitist selection is implemented. Essentially, each individual candidate solution of the current solution set is perturbed once per round. According to a simple (1+1)-selection scheme, for each pair of original and related perturbed solution the original solution is replaced when dominated by

---

[1] The HIGA-MO source code is available on moda.liacs.nl/index.php?page=code.

Algorithm 1. Hypervolume Indicator Gradient Ascent

```
 1  initialize the search points X uniformly in the search space
 2  while the termination criteria are not satisfied do
 3      Y ← evaluation of search points X
 4      layers {L_i}^q_{i=1} ← non-dominated-sorting of population R
 5      for i = 1 to q do
 6          for every element s_j in layer L_i do
 7              compute the subgradient ∂H(X)/∂x_j
 8              x_j ← x_j + σ · [∂H(X)/∂x_j]
 9          end for
10      end for
11  end while
12  return  {L_i}^q_{i=1}, X, Y with y_i = [f_1(x_i), f_2(x_i)]^T, i = 1, …, μ
```

the perturbed one. Initially, μ independently random solutions are generated using a Latin hypercube design. In every iteration, each solution is modified by an upper bounded normal distributed perturbation with maximum step size of σ. Here the step-size is fixed. After the elitist and parallel selection process based on domination, μ solutions are available for the next round until the maximum number of iterations is reached.

The rational of using this simple approach is to contrast the HIGA-MO search approach with a local search representative that is unable to traverse along local Pareto fronts. We expect this approach to get stuck in local efficient solutions.

5   Problem and Algorithm Characteristics

Problem Characteristics. In contrast to sophisticated ELA features [4,7], we know the underlying objective functions and solely intend to quantify some obvious differences in landscapes. The *count ratio* describes the problems by ratios related to the number of all local fronts or sets: `count_ratio.global` computes the percentage of fronts that are global PFs, `count_ratio.conn_ps` the percentage of sets connected to any of the global efficient sets, while `count_ratio.conn_pf` denotes the analogous percentage for PFs. The *length ratio* characteristics compute ratios of the lengths of the fronts and sets: `length_ratio.global_ps` computes the ratio of the lengths of all global Pareto sets and all local sets, whereas `length_ratio.global_pf` denotes the analogous ratio of global and local PFs. While `length_ratio.conn_ps` captures the ratio of the total length of all sets connected to any of the global efficient sets and the length of all local sets, `length_ratio.conn_pf` measures the analogous ratio in objective space.

Algorithm Characteristics. We propose characteristics in order to capture differences in local search behavior of the considered explorative algorithms.

The *population* characteristics describe the distribution of the final set of individuals of an algorithm run. They measure the percentage of individuals that are located in the ε-environment of any of the global PFs (`pop.global_front`), a front that is connected to any of the global PFs (`pop.conn_global_front`), and any local front in general (`pop.local_front`). The *coverage* characteristics

addresses the percentage of fronts that are reached by the final "population", i.e. at least one individual of that population is located in the ε-environment of the respective front.

We use the coverage of global fronts (`cover.global_front`), fronts connected to any of the global fronts (`cover.conn_global_front`), connected local fronts (`cover.conn_local_front`) and local fronts in general (`cover.local_front`). As the number of fronts might be larger than the population size (i.e. the number of considered individuals), we standardize each characteristic by its maximum.

6 Experiments

Experimental Setup. Two exemplary instances with different levels of multimodality (low, very high) were generated using the MPM-2 generator [5,10], which is e.g. available in the `R`-package `smoof`. For our experiments, we used the settings shown in Table 1. Our explorative algorithms (cf. Sect. 4), were run with a population size $\mu = 50$ and an initial step size of 0.01 (SLS) and 0.001 (HIGA-MO). Note that the step-size is adaptive in HIGA-MO and will increase largely during the optimization while it remains unchanged in SLS.

Experimental Results. As stated in the previous sections, we applied different algorithms (HIGA-MO and SLS) on two opposing multimodal, multiobjective problems. The analyzed problems can in fact be divided into a simple (cf. Fig. 2) and a complex scenario (cf. Fig. 3) as the corresponding problem characteristics show. The red line (representing the simple scenario) within the parallel coordinate plot (cf. Fig. 4) is always above the blue line of the complex scenario, which means that a higher ratio of the local fronts (or sets) are part of the global non-dominated front (set). These findings are supported by some count characteristics, which are listed in Table 2. As each of the peaks of one objective is connected to each peak of the other objective (and each of those connections can contain multiple connected components), there exist 30 connected components within the complex scenario. Given the fact that the points of a connected component often belong to multiple domination layers, the components can be

Table 1. Parameter configuration for the setup of the MPM2-generator.

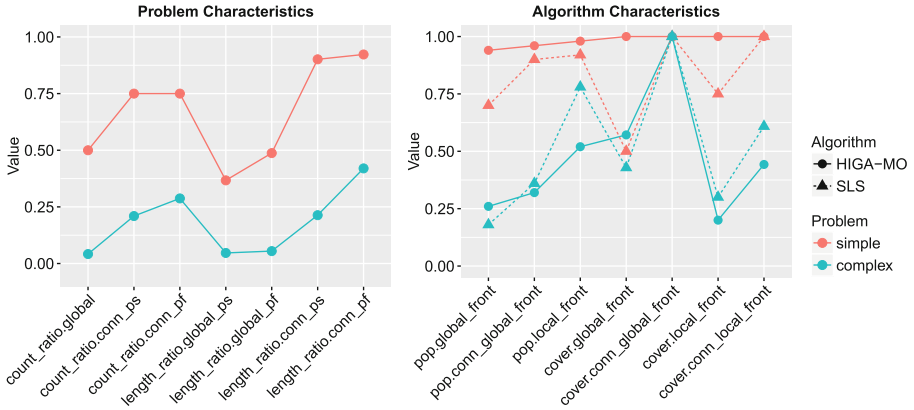| Name of parameter		Simple scenario		Complex scenario	
in general	in `R`	Obj. 1	Obj. 2	Obj. 1	Obj. 2
Number of peaks	`n.peaks`	1	3	5	5
Dimensions	`dimensions`	2	2	2	2
Topology	`topology`	`"random"`	`"random"`	`"random"`	`"random"`
Seed	`seed`	1	3	2	5
Rotated peaks	`rotated`	`FALSE`	`FALSE`	`FALSE`	`FALSE`
Shape of peaks	`peak.shape`	`"sphere"`	`"sphere"`	`"sphere"`	`"sphere"`

Fig. 4. The parallel coordinate plot on the left side distinguishes the two analyzed multiobjective, multimodal optimization problems from each other by means of rather simple problem characteristics, whereas the right figure visualizes the differences between HIGA-MO and SLS (Section 4) on these two problem instances. (Color figure online)

Table 2. Overview of the count chararacteristics of the problems.

Count characteristic	Scenario	
	Simple	Complex
Optima (obj. 1 vs. obj. 2)	1 vs. 3	5 vs. 5
Domination layers	3	19
Connected components	3	30
Sets connected to global efficient set	2	66
Fronts connected to global efficient front	2	12
Local (global) efficient sets	4 (2)	167 (7)

split into numerous local efficient sets, resulting in 167 local sets – seven of them being non-dominated.

In addition, the mixed-sphere problems come with a nice property: all local efficient sets are located on connections of two peaks and thus many of them are connected. For instance, almost 40 % of all local efficient sets (66 out of 167) in the complex scenario are connected to (at least) one of the seven non-dominated sets. In consequence, smarter optimization algorithms only need to find one of those sets and can then "travel" along the connected sets until they converge in one of the non-dominated sets. As shown in Fig. 5, HIGA-MO is able to exploit that property. At the beginning, it performs similar to SLS and tries to find any of the aforementioned local efficient sets. Once it finds one of them, it travels along the connections – the so-called *channels* – and afterwards often converges in one of the non-dominated sets. The channels are visible (as strong black lines) in Fig. 5. In contrast to HIGA-MO, a "regular" algorithm (such as SLS) likely stops, once it hits one of the local efficient sets. These findings can also be detected by our measures as the right plot of Fig. 4 shows. While both

algorithms find the majority of the local efficient sets in either scenario (SLS finds more fronts in both scenarios), the individuals of HIGA-MO also often find the global Pareto sets, while the ones from SLS are often stuck in the local sets.

7 Discussion and Outlook

This paper provides a thorough definition of multimodality in the context of multiobjective optimization problems (MOPs). Moreover, analytical and
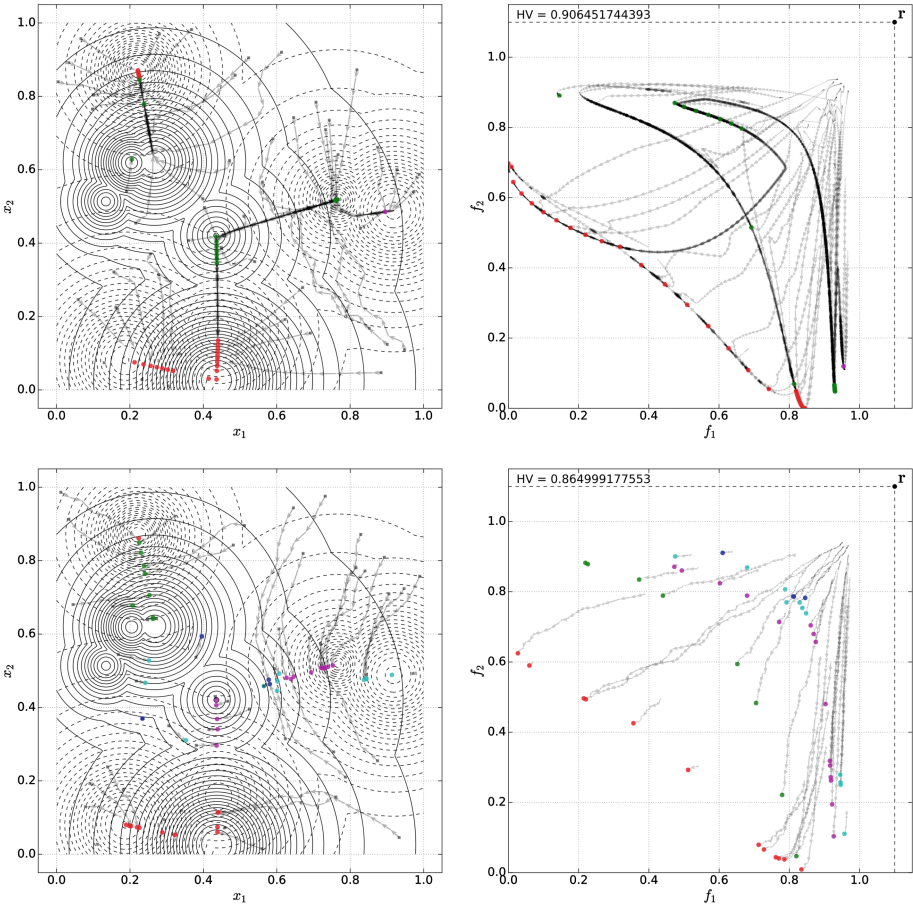

Fig. 5. Population of HIGA-MO (top) and SLS (bottom), respectively. The grey arrows visualize the trace of each individual and the colored points represent the elements from the final population. The different colors indicate the different domination layers – red points belong to the global PF, green points to the second layer, etc. (Color figure online)

experimental approaches are presented which derive or approximate the global and local Pareto fronts of a MOP. Specifically, mixed sphere test problems of different dimensionality are designed and the behavior of a sophisticated Hypervolume gradient ascent approach and a stochastic local search variant are contrasted on problems consisting of different levels of multimodality. It is reflected that multimodality is a crucial factor determining the difficulty of a problem, especially in case the optimization algorithm relies on local search techniques.

Moreover, indicators are derived which allow to assess algorithm behavior w.r.t. the detection of global and local Pareto fronts which can further be used for performance assessment. In combination with specific indicators for problem characteristics, the basis for systematically constructing respective Exploratory Landscape Features is formed which has huge potential w.r.t. algorithm benchmarking, selection and design, also for higher dimensional problems.

References

1. Bischl, B., Mersmann, O., Trautmann, H., Preuss, M.: Algorithm selection based on exploratory landscape analysis and cost-sensitive learning. In: Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation, GECCO 2012, New York, NY, USA, pp. 313–320. ACM (2012)
2. Emmerich, M., Deutz, A.: Time complexity and zeros of the hypervolume indicator gradient field. In: Schuetze, O., Coello, C.A., Tantar, A.-A., Tantar, E., Bouvry, P., Moral, P.D., Legrand, P. (eds.) EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation III. SCI, vol. 500, pp. 169–193. Springer, Heidelberg (2014)
3. Emmerich, M.T.M., Deutz, A.H., Beume, N.: Gradient-based/evolutionary relay hybrid for computing pareto front approximations maximizing the s-metric. In: Bartz-Beielstein, T., Blesa Aguilera, M.J., Blum, C., Naujoks, B., Roli, A., Rudolph, G., Sampels, M. (eds.) HCI/ICCV 2007. LNCS, vol. 4771, pp. 140–156. Springer, Heidelberg (2007)
4. Kerschke, P., Preuss, M., Wessing, S., Trautmann, H.: Detecting funnel structures by means of exploratory landscape analysis. In: Proceedings of the 17th Annual Conference on Genetic and Evolutionary Computation, pp. 265–272. ACM (2015)
5. Kerschke, P., Preuss, M., Wessing, S., Trautmann, H.: Low-budget exploratory landscape analysis on multiple peaks models. In: Proceedings of the 18th Annual Conference on Genetic and Evolutionary Computation. ACM (2016, accepted)
6. Liefooghe, A., Verel, S., Daolio, F., Aguirre, H., Tanaka, K.: A feature-based performance analysis in evolutionary multiobjective optimization. In: Gaspar-Cunha, A., Henggeler Antunes, C., Coello, C.C. (eds.) EMO 2015. LNCS, vol. 9019, pp. 95–109. Springer, Heidelberg (2015)
7. Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., Rudolph, G.: Exploratory landscape analysis. In: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, GECCO 2011, New York, NY, USA, pp. 829–836. ACM (2011)

8. Rudolph, G., Naujoks, B., Preuss, M.: Capabilities of EMOA to detect and preserve equivalent pareto subsets. In: Obayashi, S., Deb, K., Poloni, C., Hiroyasu, T., Murata, T. (eds.) EMO 2007. LNCS, vol. 4403, pp. 36–50. Springer, Heidelberg (2007)

9. Verel, S., Liefooghe, A., Jourdan, L., Dhaenens, C.: On the structure of multiobjective combinatorial search space: MNK-landscapes with correlated objectives. Eur. J. Oper. Res. **227**(2), 331–342 (2013)

10. Wessing, S.: Two-stage methods for multimodal optimization. Ph.D. thesis, Technische Universität Dortmund (2015). http://hdl.handle.net/2003/34148