Search Dynamics on Multimodal Multi-Objective Problems

P. Kerschke   kerschke@uni-muenster.de
Information Systems and Statistics, University of Münster, 48149 Münster, Germany

H. Wang   h.wang@liacs.leidenuniv.nl
LIACS, Leiden University, 2333 CA Leiden, The Netherlands

M. Preuss   mike.preuss@wi.uni-muenster.de
Information Systems and Statistics, University of Münster, 48149 Münster, Germany

C. Grimme   christian.grimme@wi.uni-muenster.de
Information Systems and Statistics, University of Münster, 48149 Münster, Germany

A. H. Deutz   a.h.deutz@liacs.leidenuniv.nl
LIACS, Leiden University, 2333 CA Leiden, The Netherlands

H. Trautmann   heike.trautmann@wi.uni-muenster.de
Information Systems and Statistics, University of Münster, 48149 Münster, Germany

M. T. M. Emmerich   m.t.m.emmerich@liacs.leidenuniv.nl
LIACS, Leiden University, 2333 CA Leiden, The Netherlands

Abstract
We continue recent work on the definition of multimodality in multi-objective optimization (MO) and the introduction of a test-bed for multimodal MO problems. This goes beyond well-known diversity maintenance approaches but instead focuses on the landscape topology induced by the objective functions. More general multimodal MO problems are considered by allowing ellipsoid contours for single-objective sub-problems. An experimental analysis compares two MO algorithms, one that explicitly relies on hypervolume gradient approximation, and one that is based on local search, both on a selection of generated example problems. We do not focus on performance but on the interaction induced by the problems and algorithms, which can be described by means of specific characteristics explicitly designed for the multimodal MO setting. Furthermore, we widen the scope of our analysis by additionally applying visualization techniques in the decision space. This strengthens and extends the foundations for Exploratory Landscape Analysis (ELA) in MO.

Keywords
Multi-Objective Optimization, Multimodality, Landscape Analysis, Hypervolume Gradient Ascent, Set Based Optimization.

1 Introduction

Multi-objective optimization is increasingly applied in domains where the single-objective functions are of complex, non-linear nature, and therefore most likely multimodal. A demonstrative example is given by the problem of antenna placement. If multiple antennas transmitting the same signal are employed, the strength of the signals is a multimodal function over space. We may also consider multiple types of signals, i.e., the mobile phone network and a signal from a local sensor network to be

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich
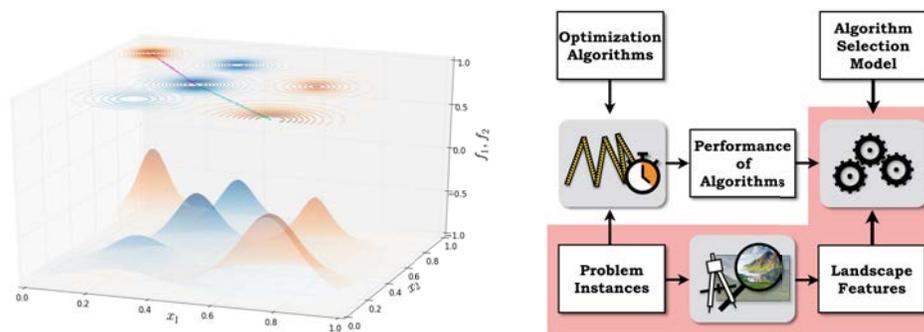

Figure 1 – Left: Example of a multimodal multi-objective landscape with single-objective functions plotted in orange and blue. Right: Schematic view of ELA (pink background) in the context of (continuous) algorithm selection.

maximized. In this case, the maximization of the signal strength is a multi-objective multimodal optimization problem. Due to the radial decay of signal strength around each antenna, its structure resembles that of the multi-objective multisphere problem visualized on the left of Fig. 1, which was recently introduced in Kerschke et al. (2016b). Similarly, *multi-objective, multimodal* problems occur also in high energy physics and quantum control (Laforge et al., 2011), drug design by docking considering energy and contact (Nicolaou and Brown, 2013) and in urban planning problems, when we want to choose a location near to different types of facilities (Maulana et al., 2015).

In either of the aforementioned scenarios, further information on the underlying problem is of high importance. In single-objective optimization, *Exploratory Landscape Analysis* (ELA, Mersmann et al., 2011) is known as a sophisticated technique for characterizing various properties of a continuous landscape by means of numerical features (e.g., the landscape's curvature, or the distribution of the local optima). These are initially computed on a small sample of evaluated points, and may be used to derive valuable information of a problem's landscape. Among others, Kerschke et al. (2016a) designed topological features that were used for detecting funnel structures. Using ELA features (in general) one can effectively enhance algorithm selection and/or configuration models (Liefooghe et al., 2015) as shown in the scheme on the right side of Fig. 1. However, the generalization of such techniques to the multi-objective domain remains an open research problem, and requires a thorough understanding of landscape features, first. In Kerschke et al. (2016b), formal definitions for multimodality in multi-objective optimization problems are introduced, which provide a first step towards generalizing the ELA framework to multi-objective optimization problems.

Being able to select the right algorithm on the basis of a small sample via applying ELA (as done for the single-objective case in Bischl et al., 2012) would be our vision. However, setting up the necessary features is not trivial because we have to explicitly target the interaction between the single-objective functions, for which not much is known, especially when the functions themselves are all multimodal. As we rely on an existing, highly configurable problem generator, we follow a *bottom-up* approach and first try to understand the combined effect the objective functions have on different types of algorithms, especially under specific variations of a similar problem composition. We treat the problem instances as white boxes, such that we can determine, e.g., how many local fronts an algorithm is able to find and how many solutions are distributed on either one. This enables a much more informed view onto the algorithm - problem interactions as would be possible otherwise. However, in a real-world set-

ting, this knowledge would not be available. Our general idea is thus to find out what measurable characteristics can describe the observed algorithm behavior well and then, later, use this knowledge in order to come up with ELA features that allow to choose the right algorithm for an unknown black-box problem on the basis of a small sample.

After summarizing the related work in Sect. 2, we extend the foundation laid in Kerschke et al. (2016b) in different ways:

- The necessary topological definitions for treating multimodality in the multi-objective context are developed further in Sect. 3.
- In Sect. 4, we look into the treated problems from an analytical perspective, and especially derive the Pareto fronts and efficient sets.
- As the shape of generated problems is generalized to ellipsoids, new visualization techniques explicitly take the decision space into account and help to understand the interactions between problem and algorithm characteristics (Sect. 5).
- Sect. 6 describes the two employed algorithms, especially the Hypervolume Indicator Gradient Ascent (HIGA-MO), in more detail.
- New problem and algorithm characteristics, so to say the white-box predecessors of new MO-related ELA features, are set up in Sect. 7.
- In Sect. 8, we experimentally analyze the behavior of the two algorithms on the new problems, and explain it with the newly introduced characteristics.

2   Related Work

In the past, the analysis of local properties of multi-objective optimization problems focused mainly on single point methods. The Fritz John and Karush Kuhn Tucker conditions form necessary and sufficient conditions for Pareto optimality in the continuous case, given the regularity conditions of differentiability and convexity of the objective functions (Miettinen, 1998). Such conditions can easily be restated to provide single point landscapes, e.g., by minimizing the residual of the angle between the objective function vectors in the unconstrained (bi-objective) case. Moreover, if full knowledge of the search landscape is available, the normalized dominance rank can be considered as a measure of closeness to the efficient set (Fonseca, 1995). More general conditions on local efficiency can be stated on level sets (Ehrgott, 2005). A set-oriented view of multimodality is however new, but it seems to better support the analysis of population-based algorithms for approximating the Pareto front.

In discrete optimization the problem of analyzing local properties of Pareto fronts has been further advanced. Single point analysis is classically done by stating non-dominance in some environment. Following this, Stadler and Flamm (2003) generalized Barrier trees from single-objective optimization to barrier forests of partially ordered landscapes, of which multi-objective optimization are a special case. The so-called Barrier forest allows to visualize the structure of basins of attraction for local search algorithms that accept only dominating points and how they are separated by barriers. Moves to non-dominated points are not allowed, which might limit the usability of these Barrier forests in the analysis of multi-objective optimization. A priori landscape analysis for discrete problems was also proposed by Tantar et al. (2008), with a focus on visualizing design space boundaries of combinatorial problems. Verel et al. (2011, 2013) instead, proposed set-oriented definitions of multimodality and local optimality, inspired by ideas of indicator-based multi-objective optimization and set-dominance expressed in earlier work. In recent work, discrete landscape features and neighborhood based search heuristics on binary search spaces are discussed, considering the ε-indicator as a measure of proximity to the Pareto front (Liefooghe et al., 2015;

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

Daolio et al., 2016). The generalization of these concepts to continuous domains is still new. Preuss et al. (2006) motivated the need for such studies by a detailed analysis of synthetical problems in low dimensions. However, in this work we do not explicitly focus on diversity issues of optimizers in multimodal settings (e.g., Ulrich et al., 2010; Zadorojniy et al., 2012) but rather investigate general search behavior of respective solvers. The previous work of Kerschke et al. (2016b), on which this article is based, presents a first step in the direction of understanding multimodal landscapes in multi-objective optimization in a more systematic way. Here the definition of local optimality of a point was generalized to the definition of a locally efficient set, which can be viewed as an attractor for population-based local search. Moreover, scalable test problems for multimodal single-objective optimization, originally introduced by Wessing (2015), were generalized to the multi-objective case. However, the rich structure of the objective space, as compared to single-objective optimization, allows to extend these basic definitions to a more comprehensive framework of reasoning about landscape features.

3 Multimodality

In this section we introduce the definition of multimodality for the multi-objective landscapes. The search and objective spaces of the multi-objective functions studied here are subsets of $\mathbb{R}^n$. Most of our definitions can also be generalized to other spaces, however, due to space limitations, this will not be part of this work.

Definition 1 (Connectedness and Connected Component). Let $A \subseteq \mathbb{R}^s$. The subset A is called *connected* if and only if there do not exist two open subsets U_1 and U_2 of $\mathbb{R}^s$ such that $A \subseteq (U_1 \cup U_2)$, $(U_1 \cap A) \neq \emptyset$, $(U_2 \cap A) \neq \emptyset$, and $(U_1 \cap U_2 \cap A) = \emptyset$; or equivalently there do not exist two non-empty subsets A_1 and A_2 of A which are open in the relative topology of A such that $(A_1 \cup A_2) = A$ and $(A_1 \cap A_2) = \emptyset$. Let B be a non-empty subset of $\mathbb{R}^s$. A subset C of B is a *connected component* of B iff C is non-empty, connected, and there exists no strict superset of C that is connected.

Now, let $\mathbf{f} : \mathcal{X} \to \mathbb{R}^m$ be a multi-objective function (which we want to 'minimize') with component functions $f_i : \mathcal{X} \to \mathbb{R}$, $i = 1, \ldots, m$ and $\mathcal{X} \subseteq \mathbb{R}^d$. Given a totally ordered set $(T, \leq)$, with total order $\leq$, the Pareto order $\prec$ on T^k for any $k \in \mathbb{N}$ is defined as follows: Let $\mathbf{t}^{(1)} = (t_1^{(1)}, \ldots, t_k^{(1)}), \mathbf{t}^{(2)} = (t_1^{(2)}, \ldots, t_k^{(2)}) \in T^k$. 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 its 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 set of all the (global) efficient points of $\mathcal{X}$ is denoted by $\mathcal{X}_E$ and is called the *efficient subset* of $\mathcal{X}$ (or *efficient* set of $\mathbf{f}$). The image of $\mathcal{X}_E$ under $\mathbf{f}$ is called the Pareto front of $\mathbf{f}$.

Defining a local efficient point in $\mathcal{X}$ (or of $\mathbf{f}$) is as straightforward as defining local minimizers (maximizers) for single-objective functions. This is in contrast to defining local efficient *sets*, which are needed for the multi-criteria setting.

Definition 2 (Local Efficient Point). A point $\mathbf{x} \in \mathcal{X}$ is called *locally efficient point* of $\mathcal{X}$ (or of $\mathbf{f}$) if there is an open set $U \subseteq \mathbb{R}^d$ 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 set of all the local efficient points of $\mathcal{X}$ is denoted by $\mathcal{X}_{LE}$.

Definition 3 (Global Efficient Point). A point $\mathbf{x} \in \mathcal{X}$ is called *global efficient point* 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 set of all the global efficient points of $\mathcal{X}$ is termed *(global) efficient set* (or Pareto set) of $\mathbf{f}$ and denoted by $\mathcal{X}_E$.

Definition 4 (Local Efficient Set). A subset $A \subseteq \mathcal{X}$ is a *local efficient set* of $\mathbf{f}$ if A is a connected component of $\mathcal{X}_{LE}$ (= set of the local efficient points of $\mathcal{X}$).

Definition 5 (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 $P = \mathbf{f}(E)$.

Note that the local efficient set has been defined for the combinatorial search domain in Paquete et al. (2004). Furthermore the (global) Pareto front of $\mathbf{f}$ is obtained by taking the image under $\mathbf{f}$ of the union of connected components of $\mathcal{X}_E$. If $\mathcal{X}_E$ is connected and $\mathbf{f}$ is continuous on $\mathcal{X}_E$, the Pareto front is also connected. In this work we used the notion of connectedness to define the local efficient sets. There still remains the task of extending the notion of efficient set by looking at connectedness in the objective space. For instance it could happen that two different local efficient sets are mapped onto the same set in the objective space. This raises many questions, which need to be addressed in future work.

With a view towards algorithms which compute approximations to (local) efficient sets and/or (local) Pareto fronts you need to be able to tell whether a finite set is a subset of a connected component (i.e., a finite subset of $\mathcal{X}_{LE}$ is a subset of some local efficient set). A finite subset of a Euclidean space is never connected unless it consists of one point. Of course, if a set S is connected and it is a subset of the local efficient points of $\mathcal{X}$, then S is a subset of some local efficient set. In case we are dealing with neighborhood systems, finite sets could very well be connected (or even path connected).

Definition 6 (ε-connectedness). Let $\varepsilon \in \mathbb{R}_{>0}$ and $S \subseteq \mathbb{R}^q$ for some q. S is ε-connected if for any two points $s_i, s_k \in S$ there is a finite set of points $\{s_{i+1}, \ldots, s_{k-1}\} \subseteq S$ such that $d(s_i, s_{i+1}) \le \varepsilon, \ldots, d(s_{k-1}, s_k) \le \varepsilon$, where d is the Euclidean distance function on $\mathbb{R}^q$.

A finite subset S of $\mathcal{X}$ is a subset of $\mathcal{X}_{LE}$, if it consists of local efficient points of $\mathcal{X}$ and S is ε-connected – with ε being below a relatively small threshold $\varepsilon_0 > 0$.

Definition 7 (finite ε-Local Efficient Set). Let S be a *finite* subset of $\mathcal{X}_{LE}$. Then S is an *ε-local efficient set*, if $S \ne \emptyset$, and S is ε-connected.

4   Analysis on Simple Mixed-Peak Problems

In this section, the bi-objective problem that is used as our benchmark is introduced with detailed discussions on its properties. To facilitate the later analysis of the multi-objective landscapes, the analytical Pareto fronts and corresponding efficient sets are derived for this problem class.

4.1   Mixed-Peak Functions

In this paper, a sophisticated problem generator, called *Multiple Peaks Model 2* (MPM2, Wessing, 2015), is adopted to illustrate the proposed topological definitions and further analyze the behavior of explorative algorithms. Such a function class is a mixture of similar unimodal functions, i.e., the peaks, that have *convex* local level sets, which is typically combined with the well-known Karush-Kuhn-Tucker theorem to identify local efficient points. In addition, the complexity of the problem can be easily controlled by the number of peaks. The mixed-peak function is defined as an unconstrained function $f : \mathbb{R}^d \to \mathbb{R}$ that is subject to minimization:

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

$$g_i(\mathbf{x}) = h_i \left(1 + \frac{\left(\sqrt{(\mathbf{x} - \mathbf{c}_i)^\top \mathbf{\Sigma}_i (\mathbf{x} - \mathbf{c}_i)}\right)^{s_i}}{r_i} \right)^{-1}, \quad i = 1, \ldots, N. \tag{2}$$

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

The function g above defines a parameterized *quasi-concave* unimodal peak, whose negative leads to *quasi-convex* valleys on function f. According to the `optproblems` package (Wessing, 2016), it has the following parameters: (1) number of peaks $N \in \mathbb{Z}_{>0}$, (2) center $\mathbf{c}_i \in \mathbb{R}^d$, height $h_i \in [0,1]$ and radius $r_i \in [0.25\sqrt{d}, 0.5\sqrt{d}]$ per peak, with decision space dimension d, (3) "shape" $s_i \in [1.5, 2.5]$ per peak, controlling the landscape's steepness, (4) rotation of the elliptical level sets based on a *positive definite* matrix $\mathbf{\Sigma}_i$. In the following, we will use the norm notation $\|\mathbf{x} - \mathbf{c}_i\|_{\mathbf{\Sigma}_i} := \sqrt{(\mathbf{x} - \mathbf{c}_i)^\top \mathbf{\Sigma}_i (\mathbf{x} - \mathbf{c}_i)}$ as it can be considered as the Mahalanobis distance w.r.t. $\mathbf{\Sigma}_i$.

Ridges: As a result from the definition of f (Eq. 1), the landscape can contain ridges. The set of all ridges of f can be represented by:

$$\mathcal{R} = \left\{ \mathbf{x} \in \mathbb{R}^d \mid \exists i \in \{1, 2, \ldots, N\}, \text{ with } i \neq j \text{ and } g_i(\mathbf{x}) = g_j(\mathbf{x}) = \max_{1 \leq k \leq N} \{g_k(\mathbf{x})\} \right\},$$

i.e., the set of all points on which the value of f is simultaneously attained by at least two peak functions. According to Eq. 1, for any point that is not on the ridge, i.e., $\mathbf{x} \in (\mathbb{R}^d \setminus \mathcal{R})$, there is only one peak function that is effective or *active*. From now on, the active peak function at $\mathbf{x}$ is denoted as g_τ w.r.t. $\tau = \arg\max_{1 \leq i \leq N}\{g_i(\mathbf{x})\}$. In fact, ridges separate the decision space into many *active regions*, on each of which only a single peak function g is active:

$$\mathcal{A}_i = \left\{ \mathbf{x} \in \mathbb{R}^d \mid \forall k \in \{1, 2, \ldots, N\} \setminus \{i\}, \; g_i(\mathbf{x}) > g_k(\mathbf{x}) \right\}, \quad i = 1, 2, \ldots, N.$$

Note that the active regions $\mathcal{A}_i$'s are open and mutually disjoint and the union of all such active regions $\mathcal{A} = \cup_{1 \leq i \leq N} \mathcal{A}_i$ is equal to the set of non-ridge points.

Convex Local Level Sets: Given the quasi-concavity of each peak g_i, $1 - g_i$ has local convex level sets in $\mathbb{R}^d$. If the function $1 - g_i$ is restricted to an ε-Euclidean ball $B_\varepsilon(\mathbf{x}^*) = \left\{ \mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon \right\}$ for every $\mathbf{x}^* \in \mathbb{R}^d$ and every $\varepsilon > 0$, the resulting function $1 - g_i\big|_{B_\varepsilon(\mathbf{x})} : B_\varepsilon(\mathbf{x}) \to \mathbb{R}$ also has local convex level sets. Also, due the fact that the active regions $\mathcal{A}_i$'s are disjoint and open, for every non-ridge point $\mathbf{x}^*$, it is possible to find a $\delta > 0$ (depending on $\mathbf{x}^*$) such that $B_\delta(\mathbf{x}^*) \subset \mathcal{A}_\tau$ and $(B_\delta(\mathbf{x}^*) \cap \mathcal{A}_i) = \emptyset, \forall i \neq \tau$ (τ is the unique index of the active peak function at $\mathbf{x}^*$). Then the restricted f to $B_\delta(\mathbf{x}^*)$, $f\big|_{B_\delta(\mathbf{x}^*)}$ equals $1 - g_\tau\big|_{B_\delta(\mathbf{x}^*)}$ and it thus has local convex level sets. Therefore, we have the following conclusion:

$$\forall \mathbf{x}^* \in (\mathbb{R}^d \setminus \mathcal{R}) \; \exists \delta > 0 : \; f\big|_{B_\delta(\mathbf{x}^*)} \text{ has local convex level sets.} \tag{3}$$

For the points on the ridge, $\mathbf{x}^* \in \mathcal{R}$, the conclusion above does not hold because it is not possible to find a δ such that $B_\delta(\mathbf{x}^*)$ has no intersection with all $\mathcal{A}_i$'s except $\mathcal{A}_\tau$.

As the gradient of the mixed-peak function is required by both, the algorithms and analysis in the following, we derive it in the following:

$$\nabla f(\mathbf{x}) = \frac{h_\tau s_\tau}{r_\tau} \left(1 + \frac{\|\mathbf{x} - \mathbf{c}_\tau\|_{\mathbf{\Sigma}_\tau}^{s_\tau}}{r_\tau} \right)^{-2} \|\mathbf{x} - \mathbf{c}_\tau\|_{\mathbf{\Sigma}_\tau}^{s_\tau - 2} \mathbf{\Sigma}_\tau (\mathbf{x} - \mathbf{c}_\tau). \tag{4}$$

4.2 Mixed-Peak Bi-Objective Problem

By generating two different configurations for the parameters in Eq. 1, two different multimodal functions are constructed, defining a bi-objective optimization problem:

$$f_1(\mathbf{x}) = 1 - \max_{1 \leq i \leq N} \{g_i(\mathbf{x})\} \to \min, \qquad f_2(\mathbf{x}) = 1 - \max_{1 \leq i \leq N'} \{g_i'(\mathbf{x})\} \to \min.$$

Note that the peak functions g and g' (and its parameters N and N') are distinguished by the superscript. Next, the efficient set and Pareto front are derived analytically.

One Peak Scenario   We first consider a simple case where each objective consists of one peak without any ridges in the domain. Then, the objectives degenerate to:

$$f_1(\mathbf{x}) = 1 - h\left(1 + \frac{\|\mathbf{x} - \mathbf{c}\|_{\boldsymbol{\Sigma}}^s}{r}\right)^{-1}, \quad f_2(\mathbf{x}) = 1 - h'\left(1 + \frac{\|\mathbf{x} - \mathbf{c}'\|_{\boldsymbol{\Sigma}'}^{s'}}{r'}\right)^{-1}.$$

According to the Karush-Kuhn-Tucker (KKT) condition (Ehrgott, 2005) for multi-objective optimization problems, a *necessary condition* for $\mathbf{x}^* \in \mathbb{R}^d$ being efficient is:

$$\exists \lambda_1 > 0, \lambda_2 > 0 : \ \lambda_1 \nabla f_1(\mathbf{x}^*) + \lambda_2 \nabla f_2(\mathbf{x}^*) = \mathbf{0}.$$

Substituting the condition above by the gradient expression (Eq. 4) leads to:

$$\lambda_1 C(\mathbf{x}^*) \cdot \boldsymbol{\Sigma}(\mathbf{x}^* - \mathbf{c}) + \lambda_2 C'(\mathbf{x}^*) \cdot \boldsymbol{\Sigma}'(\mathbf{x}^* - \mathbf{c}') = \mathbf{0}$$

$$\text{with } C(\mathbf{x}^*) := \frac{hs}{r}\left(1 + \frac{\|\mathbf{x}^* - \mathbf{c}\|_{\boldsymbol{\Sigma}}^s}{r}\right)^{-2} \|\mathbf{x}^* - \mathbf{c}\|_{\boldsymbol{\Sigma}}^{s-2}.$$

And C' is defined similarly to C by adding prime superscripts to all parameters. As a result, the condition above can further be simplified to:

$$\exists \lambda_1 > 0, \lambda_2 > 0 : \ \boldsymbol{\Sigma}(\mathbf{x}^* - \mathbf{c}) = -\frac{\lambda_2 C'(\mathbf{x}^*)}{\lambda_1 C(\mathbf{x}^*)} \boldsymbol{\Sigma}'(\mathbf{x}^* - \mathbf{c}'). \tag{5}$$

Let us denote $k := \lambda_2 C'(\mathbf{x}^*)/\lambda_1 C(\mathbf{x}^*)$. Thus, $\lambda_1, \lambda_2 > 0$ and $C, C' \geq 0$ result in $k \geq 0$. In addition, $C \to 0$ leads to $k \to \infty$, i.e., $\mathbf{x}^* \to \mathbf{c}$. Due to the fact that C and C' are continuous functions w.r.t. $\mathbf{x}^*$, k is also *continuous* in $\mathbb{R}^d$. Therefore, it must take any value between its minimum and maximum, resulting in $0 \leq k < \infty$. Taking the range of k into account, every point that satisfies Eq. 5 can be written as:

$$\forall k > 0 : \ \mathbf{x}^* = \mathbf{c} - \left(\frac{\boldsymbol{\Sigma}}{k} + \boldsymbol{\Sigma}'\right)^{-1} \boldsymbol{\Sigma}'(\mathbf{c} - \mathbf{c}'). \tag{6}$$

Note that the points above are not necessarily local efficient points (as defined in Sect. 3). However, their sufficiency can be shown as follows: for any point $\mathbf{x}^* \in \mathbb{R}^d$ satisfying Eq. 6 – remember, there is no ridge in this scenario – there exists an $\varepsilon > 0$ such that the restricted objective function $f_1\big|_{B_\varepsilon(\mathbf{x}^*)}$ has local convex level sets according to Eq. 3. Similarly, there exists an $\varepsilon' > 0$ such that $f_2\big|_{B_{\varepsilon'}(\mathbf{x}^*)}$ has local convex level sets. It is then possible to construct a Euclidean ball with radius $\varepsilon^* := \min\{\varepsilon, \varepsilon'\}$ such that: $f_1\big|_{B_{\varepsilon^*}(\mathbf{x}^*)}$ and $f_2\big|_{B_{\varepsilon^*}(\mathbf{x}^*)}$ both have local **convex** level sets. This implies that it is always possible to find a neighborhood around a point where the local level sets of both objectives are convex. Thus, it is sufficient to conclude that points satisfying Eq. 6 are *locally Pareto efficient* and the efficient set of the problem is expressed as:

$$\mathcal{X}_{LE} = \left\{\mathbf{c} - \left(\frac{\boldsymbol{\Sigma}}{k} + \boldsymbol{\Sigma}'\right)^{-1} \boldsymbol{\Sigma}'(\mathbf{c} - \mathbf{c}') \ \big| \ 0 \leq k < \infty\right\}. \tag{7}$$

Consequently, the Pareto front can implicitly be obtained by applying the objective functions to the efficient set from above. When the contour lines are spherical for both objective functions, the arguments here can be largely simplified. We omit such a special case, since it has already been discussed in detail in Kerschke et al. (2016b).

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich
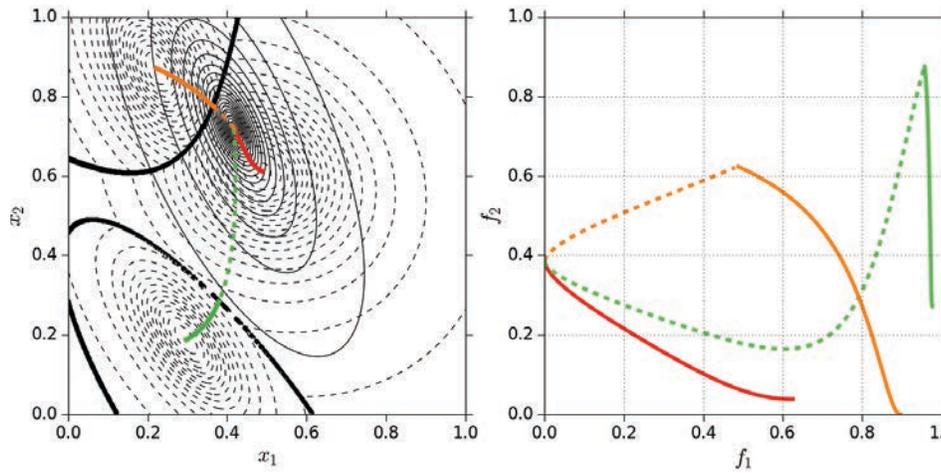

Figure 2 – Example of analytical Pareto fronts and efficient sets: the contour lines of f_1 (solid curves, 1 peak) and f_2 (dashed curves, 3 peaks) are drawn in the decision space (left) with ridges shown as thick solid curves. Three local efficient sets are drawn in different colors while the dashed extensions of them represent the pseudo-efficient sets. The corresponding (local) Pareto fronts are shown on the right.

Multiple Peaks If each of the objective functions consists of multiple peak functions, namely $N > 1$, the efficient set derived in Eq. 7 can be adapted in the following manner: suppose function f_1 and f_2 contain N and N' peaks, respectively. For each pair of peaks between two objective functions (e.g., g_i and g'_j), a *pseudo-efficient set* can be calculated according to Eq. 7 as if the rest of the peaks in both objective functions were not existing:

$$\mathcal{P}_{ij} = \left\{ \mathbf{c}_i - \left(\frac{\boldsymbol{\Sigma}_i}{k} + \boldsymbol{\Sigma}'_j \right)^{-1} \boldsymbol{\Sigma}'_j (\mathbf{c}_i - \mathbf{c}'_j) \mid 0 \leq k < \infty \right\},$$

where $\mathbf{c}_i$ and $\mathbf{c}'_j$ are the centers of the i-th and j-th peak of functions f_1 and f_2, respectively. Note that Eq. 7 requires that no ridge is present in the function domain and thus for the set defined above, it is not necessarily a local efficient set. Let us denote the active region of peak g_i and g'_j as $\mathcal{A}_i$ and $\mathcal{A}'_j$, respectively. Then the region on which g_i and g'_j are both active is $\mathcal{A}_i \cap \mathcal{A}'_j$. Consider the intersections of $\mathcal{P}_{ij}$ and the ridges $\mathcal{R}$ of f_1 for instance: at such points, any infinitesimal movement towards a different active region other than $\mathcal{A}_i \cap \mathcal{A}'_j$ will revert the direction of ∇f_1 and therefore this movement will improve both f_1 and f_2 values of the intersection points. This implies that the points in $\mathcal{P}_{ij}$ intersecting or crossing the ridges are not efficient for g_i and g'_j. In other words, the efficient set $\mathcal{X}^*_{ij} = (\mathcal{P}_{ij} \cap \mathcal{A}_i \cap \mathcal{A}'_j)$ associated with peak g_i and g'_j is the intersection of $\mathcal{P}_{ij}$ with the active regions of both peak functions. In addition, all local efficient sets can be enumerated by calculating the local efficient set associated with each pair of peaks between two objective functions: $\mathcal{X}^* = \bigcup_{i=1}^{N} \bigcup_{j=1}^{N'} \mathcal{X}^*_{ij}$. An example of this is illustrated in Fig. 2. Here, three pseudo-efficient sets are depicted in different colors (red, orange and green) and the orange and green sets are truncated by the ridges (thick black lines), where the valid local efficient sets are depicted as solid curves.

5   Visualizing the Decision Space of Multi-Objective Landscapes

Within recent work, a new approach for visualizing the decision space of multimodal multi-objective landscapes based on a scalar combination of its gradients was intro-

duced (Kerschke and Grimme, 2017). It depicts the interaction of overlapping multi-objective local optima and provides a first understanding of a problem's landmarks, such as ridges and valleys, in rather unexplored multi-objective settings. Figuratively speaking, the method visualizes the behavior of a "multi-objective" ball, which behaves like a gradient descent optimization algorithm, on multi-objective landscapes.

We thus compute the sum of the (per objective) normalized gradients – which always points into the dominating cone – for all points in an equidistant 1 000 by 1 000 grid across the decision space. The two gradients $\mathbf{v}_1$ and $\mathbf{v}_2$, i.e., the corresponding directions of the steepest descent, are approximated using their partial derivatives (although one could also use the analytical gradient from Eq. 4 for the MPM2 functions):

$$(\mathbf{v}_i)_k \approx \frac{f_i(\mathbf{x} + \delta \cdot \mathbf{e}_k) - f_i(\mathbf{x} - \delta \cdot \mathbf{e}_k)}{2\delta}, \qquad i = 1, 2 \text{ and } k = 1, 2, \ldots, d.$$

and normalized (to length one) afterwards. As both vectors are of length one, the length of the combined gradient vector reflects the angle between the two normalized vectors: a combined gradient of length two can only be achieved, if the two objective-wise gradients point in exactly the same direction (and thus enclose an angle of $0°$), while a combined gradient of length zero indicates objective-wise gradients that point towards opposite directions (i.e., $180°$). Furthermore, the combined gradient aims into the direction between the two (closest) objective-wise local optima and thereby indicates, which of the eight surrounding cells according to the Moore Neighborhood (Gray, 2003) is the next better option leading towards the attracting (at least local) optimum. Following a path of these combined gradients, one ultimately reaches one of the local efficient sets[1], which lie on connections between any pair of peaks (from the different objectives). Note that the connections are straight lines for the mixed-sphere problems or curved lines for the mixed-ellipse problems, respectively. As a path along the gradients leads to the (attracting) local efficient set, we use the cumulated path lengths as objective value (or "height") of our scalar representation of the multi-objective landscape. The scalarized problem can then be visualized within a two-dimensional heatmap or within a three-dimensional surface plot as shown in Fig. 3. We also enhanced our heatmap by adding the contour lines of the objective-wise mixed-sphere (or mixed-ellipse) problems, the combined gradient vectors (only every 50th value per dimension for better readability) and the true local efficient sets.

The mixed-sphere problem shown in Fig. 3 contains two peaks within the first (indicated by orange contour lines) and one peak within the second objective (white dashed lines). The coloring of the plots represents the ($\log_{10}$-transformed) cumulated length of the path of gradients to reach an at least locally efficient point and the corresponding path-length-to-color mapping is shown in the color bar on the right. In the following, we highlight some of the peculiarities that we detected within the plots.

Basins of attraction Within the shown example, two basins of attraction, each comprising one locally attracting connected set, are visible. The existence of these basins (along with their included connected sets) supports our thesis of the "multi-objective ball", which follows the combined gradients until it converges in a local efficient set. Interestingly, an area of attraction can comprise several disjoint local efficient sets – e.g., the light blue and orange segments in the left connected set within the heatmap of

---

[1] We used $\delta = 10^{-6}$ for the gradient approximation and considered points for which the length of the respective summed (normalized) gradient vectors was below 10^{-3} to be locally efficient. As we discretize the search space we might only end up in a point that is in the vicinity of the (true) efficient set.

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich
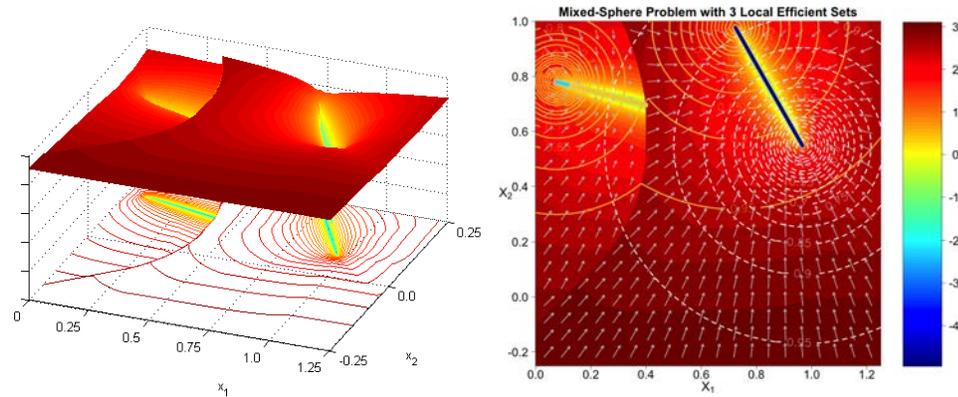
Figure 3 – Log-scaled gradient field, shown as a 3D surface plot (left) and a heatmap (right).

Fig. 3 – simply because (adjacent) parts of this connected set belong to different dominance layers. In this scenario, the right segment of the left connected set (i.e., the orange line) is dominated by the entire right connected set (dark blue line) as shown within the plot of the (theoretically) true local fronts in the objective space, Fig. 4.

Discontinuities   We furthermore discovered overlays of basins of attraction, resulting in "cliffs" within the landscape. These abrupt changes within the landscape are created by competing peaks within the single objectives. Even in the rather simple scenario from Fig. 3, which contains two peaks in the first and only one peak in the second objective, this behavior becomes visible. The two competing peaks of the first objective cause a shift in the gradient landscape - leading to a cliff along the line of equal height of its two peaks, i.e., the position where the spherical contour lines of the two peaks (from the first objective) intersect. If the two objectives would contain more peaks, the competition among the peaks would lead to an even more rugged landscape. As a result of this "discontinuity" in the landscape, minor changes in the (starting) position can cause the "multi-objective ball" to "roll" towards a completely different locally efficient set.

Expression of basins of attraction in the objective space   The coloring of the gradient landscape in Fig. 3 represents the ($\log_{10}$-transformed) cumulated lengths' of the gradient paths towards the nearest attracting local efficient set. As each observation exists in the decision *and* objective space, we are able to transfer the coloring scheme to the objective space by coloring the image in objective space according to each sample point's color within the decision space. The result is shown in the right plot of Fig. 4 and while both connected fronts (i.e., the images of the connected sets) become visible, one cannot identify the local fronts purely based on the coloring. However, by comparing the dominance relationship between the connected fronts, one could of course (manually) split the connected fronts into local fronts. Also, points within the vicinity of a local front are dominated, supporting our definition of a local front. Furthermore, one can see that points with an increased distance to the local front are colored in a darker shade of red, i.e., their cumulated path lengths towards the local optima are longer.

Interpretation of 3D visualization of multi-objective landscapes   Three-dimensional surface plots and the respective heatmaps are usually a good and helpful tool for detecting valleys, ridges or other characteristics of the analyzed problem landscape. However, in this case the objective function, which actually defines the "height" of the
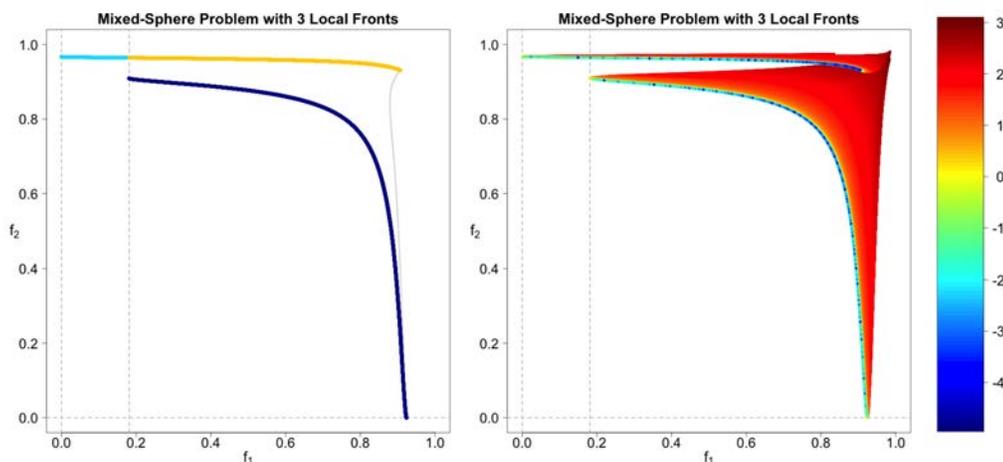
10

Figure 4 – Visualization of the theoretically true local fronts (left) and the mapping of the cumulated gradient field from the decision to the objective space (right) for the problem in Fig.3.

landscape, is the cumulated length of the path of the gradients, i.e., it is a mapping from the original two objectives into a scalar objective. Consequently, when interpreting plots, we have to keep in mind that they do not describe the landscape in the common single-objective sense but rather the interaction of the objectives.

Also, there are some technical limitations to this visualization approach: Ideally, all points belonging to an efficient set should have a combined gradient length of zero. Thus, all local efficient sets, as well as the corresponding local fronts, should have the same color. However, as the coloring within the plots indicate, none of the one million discrete points from the grid has a value below 10^{-5}. Due to numerical imprecision, none of these points has a gradient of exactly zero. By comparing heatmaps of multiple scenarios (e.g., the ones that are shown within Sect. 8), one can also detect that the sizes of the valleys (i.e., their lengths and widths) vary and thus, the probability of detecting these valleys, including their comprised local efficient sets, likely varies as well.

In general, we suggest to always consider multiple visualizations of the landscape, as each of the plots explains a different aspect of the problem and consequently, one can get a better understanding of the entire problem by looking at it from different angles.

6   Explorative Algorithm

In this section, two multi-objective optimizers are introduced that follow entirely opposing search dynamics: a gradient-based method that is able to converge to local and especially global efficient sets accurately, and a naïve stochastic hill-climbing approach in which each search point performs a simple (1+1)-selection. In general, neither of the two algorithms exploits the external archive technique and the population size was chosen by balancing the algorithm's running time and the reliability of the induced algorithmic features.

Hypervolume Indicator Gradient Ascent (HIGA-MO)   This algorithm computes the steepest ascent direction of the hypervolume indicator w.r.t. the decision vectors. Such a direction, called *Hypervolume Indicator Gradient*, is proposed in Emmerich et al. (2007) and Emmerich and Deutz (2012) and the practical gradient ascent algorithm based on it, called *Hypervolume Indicator Gradient Ascent Multi-Objective Optimization (HIGA-MO)* is

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

improved in Wang et al. (2017a,b). We first denote the approximation to the Pareto efficient set as a set of decision vectors $X = \left\{ \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(\mu)} \right\}$, $\mathbf{x}^{(i)} \in \mathbb{R}^d$, $i = 1, 2, \ldots, \mu$. In HIGA-MO, the set-based differentiation is considered, i.e., the decision vectors are concatenated into $\mathbf{X} = \left[\mathbf{x}^{(1)^\top}, \mathbf{x}^{(2)^\top}, \ldots, \mathbf{x}^{(\mu)^\top} \right]^\top \in \mathbb{R}^{\mu \cdot d}$, using the same ordering as in X. In this treatment, an approximation set X is considered as a single point in the product space $\mathbb{R}^{\mu \cdot d}$. Analogously, the corresponding objective values are encapsulated in $\mathbf{Y} = \left[\mathbf{y}^{(1)^\top}, \mathbf{y}^{(2)^\top}, \ldots, \mathbf{y}^{(\mu)^\top} \right]^\top \in \mathbb{R}^{\mu \cdot m}$, $\mathbf{y}^{(i)} = \mathbf{f}(\mathbf{x}^{(i)}) \in \mathbb{R}^m$, $i = 1, 2, \ldots, \mu$. Thus, one can explicitly define a vector-valued mapping $\mathbf{F} : \mathbb{R}^{\mu \cdot d} \to \mathbb{R}^{\mu \cdot m}$ as $\mathbf{Y} := \mathbf{F}(\mathbf{X})$. Furthermore, the hypervolume indicator H can be expressed as a continuous mapping from $\mathbb{R}^{\mu \cdot d}$ to $\mathbb{R}$, $\mathcal{H}_\mathbf{F}(\mathbf{X}) := H \circ \mathbf{F}(\mathbf{X}) = H(\mathbf{F}(\mathbf{X}))$, whose gradient

$$\nabla \mathcal{H}_\mathbf{F}(\mathbf{X}) = \left[\frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial x_1^{(1)}}, \ldots, \frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial x_d^{(1)}}, \ldots, \frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial x_1^{(\mu)}}, \ldots, \frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial x_d^{(\mu)}} \right]$$

$$= \left[\frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial \mathbf{x}^{(1)}}^\top, \ldots, \frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial \mathbf{x}^{(\mu)}}^\top \right]^\top . \tag{8}$$

is – under certain regularity conditions, e.g., if the decision vectors in X are non-dominated (Emmerich and Deutz, 2012) – defined. Each term in the right-hand-side of Eq. 8 is called *sub-gradient*, which is the local hypervolume change rate by moving each decision vector infinitesimally. Moreover, one can calculate the sub-gradients by applying the chain rule (Emmerich and Deutz, 2012; Wang et al., 2017b):

$$\frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial \mathbf{x}^{(i)}} = \frac{\partial \mathbf{Y}}{\partial \mathbf{x}^{(i)}} \frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial \mathbf{Y}} = \frac{\partial \mathbf{y}^{(i)}}{\partial \mathbf{x}^{(i)}} \frac{\partial \mathcal{H}_\mathbf{F}(\mathbf{X})}{\partial \mathbf{y}^{(i)}} = \sum_{k=1}^{m} \frac{\partial H(\mathbf{Y})}{\partial y_k^{(i)}} \nabla f_k(\mathbf{x}^{(i)}). \tag{9}$$

Note that the gradient of objective functions $\nabla f_k(\mathbf{x}^{(i)})$ can be approximated numerically if no analytical knowledge on the functions is available. In addition, the term $\partial H(\mathbf{Y})/\partial y_k^{(i)}$ are the partial derivatives of the hypervolume indicator H w.r.t. the $\mathbf{y}^{(i)}$'s, which are calculated as the length of the steps of the attainment curve (see Emmerich and Deutz (2012) for details). Consequently, the hypervolume indicator gradient is a *linear combination* of objective-wise gradients.

The hypervolume indicator gradient is well-defined for non-dominated subsets of the approximation set X due to the fact that the image of each decision vector contributes to the hypervolume. For any *strictly dominated* point, the sub-gradient associated with it is zero because such a point has no contribution to the hypervolume. In order to move such points towards the (global) Pareto efficient set, the well-known *non-dominated sorting* technique (Srinivas and Deb, 1994) is adopted to solve this issue. In principle, the Pareto set approximation is partitioned into multiple locally non-dominated *layers*. Then the hypervolume indicator sub-gradient at a point is computed w.r.t. the layer to which the point belongs.

The non-dominated sorting based approach is of particular interest to our landscape exploration task. To explore a multimodal multi-objective landscape, it is important to search for local efficient sets. In the non-dominated-sorting based approach, each layer has its own local hypervolume indicator and thus is *locally* optimized, which will not necessarily converge to the global efficient set. In this sense, the layers can be treated as candidate approximation sets to local efficient sets.

Stochastic Local Search (SLS) For comparison of local search behavior, we implement a simple local search strategy based on parallel perturbations. Essentially each
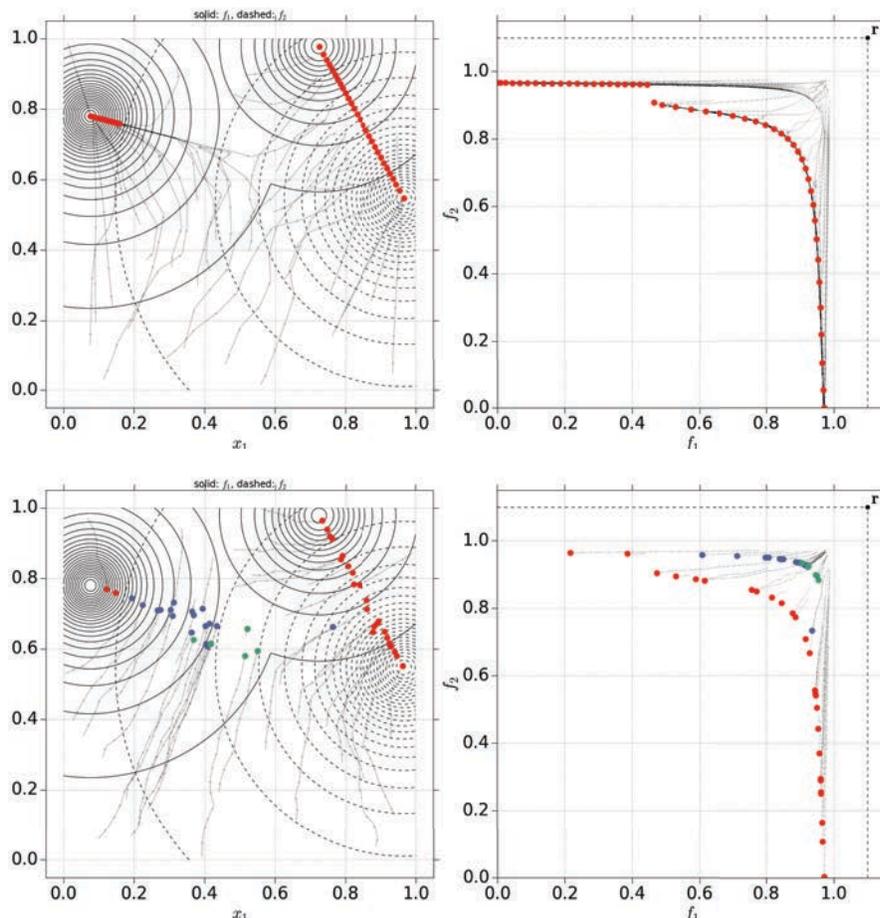
Figure 5 – Development of the two algorithms' populations (top: HIGA-MO; bottom: SLS) across their generations. For each of the algorithms, the results are shown within the decision (left) and objective space (right) based on a mixed-sphere problem with two peaks within the first and one peak in the second objective. The location of the true local efficient sets and fronts are already shown within Figures 3 and 4.

decision point of the current approximation set is perturbed once per round. According to a simple (1+1)-selection scheme, within each iteration, the original decision point is replaced when dominated by the perturbed one. Initially, μ decision points are generated using Latin hypercube sampling (LHS). In every iteration, each decision vector is mutated by a standard normal distribution that is truncated to $[-\sigma, \sigma]$. After the elitist and parallel selection process based on domination, μ decision points are available for the next iteration. The loop is repeated until a termination criterion (here: maximum number of rounds) is reached. The rationale of using this simple approach is to contrast HIGA-MO 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.

7   Problem and Algorithm Characteristics

In the preceding sections, the selected optimization problems (Sect. 4) and the applied algorithms (Sect. 6), along with their expected differences regarding the problem complexity or algorithm behavior, were introduced. Such (detailed) knowledge of the

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

problem landscapes and algorithm performances is crucial when trying to solve an *Algorithm Selection Problem* (Rice, 1976). That is, based on information of the problem landscapes and the algorithm performances, one can train a so-called *algorithm selection model*, which tries to select the best suited solver (out of a given portfolio of optimization algorithms) for an unseen optimization problem.

Unfortunately, the aforementioned 'information' are often derived manually – based on the authors observations and/or knowledge. However, in the future, that process should be automatized, e.g., by using exploratory landscape features as it is already common practice within single-objective optimization. As a first step towards the development of such features, we propose some characteristics, which can be seen as a numerical representation of the problem landscapes and algorithms. However, while sophisticated ELA features (Bischl et al., 2012; Kerschke et al., 2015; Mersmann et al., 2011) are computed based on a small sample of observations from the entire landscape, we consider the problems within this paper as white-box problems and thus can use information from the entire landscape to compute some representative numbers[2].

7.1 Problem Characteristics

The first group of characteristics aims at quantifying the problem landscapes. While the (connected) count characteristics should also apply to combinatorial landscapes, the length characteristics require a specific metric for computing the respective lengths.

Count Characteristics   These characteristics describe the landscapes by the number of local efficient sets (`count.les`), connected sets (`count.sets`), domination layer (`count.layer`) or peaks per objective (`count.peaks1`, `count.peaks2`). The `count.ps_rel` computes the ratio of local efficient sets that actually are global efficient sets (= Pareto sets). Here, a value of one means that all local efficient sets are non-dominated and thus form the Pareto set. Note that it is sufficient to compute the latter ratio for the local efficient sets as there exists a bijective function between the local efficient sets (in the decision space) and the local fronts (in the objective space).

Length Characteristics   As the pure number of fronts or sets might be misleading due to varying lengths – e.g., the light blue front within Fig. 4 is much shorter than the dark blue line – the next six characteristics focus on the actual lengths[3] of the local fronts and efficient sets. So, `length.les_total` represents the total length of all local efficient sets and `length.ps_rel` measures the relative length of Pareto sets, i.e., the total length of all global efficient sets divided by the total length of all local efficient sets (including the global ones). Thus, a value of one is equivalent to a landscape in which all points from the local efficient sets are globally non-dominated. The third characteristic of this group (`length.ps_ratio`) standardizes the former characteristic (`length.ps_rel`) by the analogon among the count characteristics (`count.ps_rel`)[4].

The remaining three characteristics of this group measure the analogous properties for the local fronts, i.e., the images of the local efficient sets: `length.lf_total` measures the total length of all local fronts, `length.pf_rel` is the ratio of the total length of the Pareto fronts compared to the total length of all local fronts and `length.pf_ratio` standardizes the latter by the count ratio of Pareto fronts and local fronts.

---

[2]One could for instance use the `R`-package `flacco` (Kerschke, 2017) for computing such features.

[3]Note that all the length features are approximated numerically by calculating the cumulative chordal distance of the samples on the curve. This distance converges asymptotically with rate $O(1/N^2)$ (N is the number of samples) for uniformly spaced samples (Kozera et al., 2003).

[4]From the points sampled along the theoretical Pareto front/efficient set for drawing the curves, we approximate the lengths by computing the sum of the Euclidean distances between respective neighbors.

Connected Front/Set Characteristics   As some algorithms (e.g., HIGA-MO) are able to travel along the local efficient sets (or local fronts), we also need information on sets or fronts that are connected to any of the Pareto sets or fronts. Thus, `conn_ps.count_abs` counts the number of local efficient sets that have a direct or indirect connection to at least one of the Pareto sets and `conn_ps.length` measures the total length of all of these sets. These two characteristics are also the foundation for the remaining ones: `conn_ps.count_rel` gives the proportion of the number of sets that are (somehow) connected to any of the Pareto sets and `conn_ps.length_rel` provides the same information based on the length of these sets. Analogously to the previous four characteristics, `conn_pf.count_abs` lists the number of local fronts that are connected to any of the Pareto fronts, `conn_pf.length` measures the total length of the aforementioned fronts and `conn_pf.count_rel` and `conn_pf.length_rel` provide the corresponding count and length ratios (compared to all local fronts).

7.2   Algorithm Characteristics

We also propose characteristics describing the distribution of an algorithm's final population across the problems' local fronts and efficient sets. Based on these, we want to get a better understanding of the algorithms' behavior across the different problems.

Population Characteristics   These characteristics measure the percentage of individuals from an algorithm's final population that are actually located in the vicinity of specific local efficient sets or local fronts. More precisely, `pop_glob.single_front` measures the ratio of individuals that are located in the proximity of the global Pareto front and `pop_glob.single_set` measures the analogon within the decision space. Similarly, the percentage of individuals from the final population that are located in the vicinity of any of the local fronts in general is measured by `pop_loc.single_front`, whereas `pop_loc.single_set` again measures the analogon in the decision space. The final characteristics of this category measure the ratio of individuals that are located close to a local front (`pop_glob.conn_front`) or set (`pop_glob.conn_set`) that is connected to any of the Pareto fronts or sets, respectively.

Coverage Characteristics   The remaining proposed characteristics describe the relation of fronts (or sets) and the final "population" from the opposite perspective. That is, they measure the percentage of fronts (or sets) that are covered[5] by at least one individual from the population. The first two characteristics, `cov_loc.single_set` and `cov_loc.single_front`, measure the ratio of local efficient sets or fronts that are covered by (at least) one individual. Analogously, `cov_glob.single_set` and `cov_glob.single_front`, measure the percentage of covered global efficient sets or Pareto fronts, respectively. The final characteristics (`cov_loc.conn_front`, `cov_glob.conn_front`, `cov_loc.conn_set` and `cov_glob.conn_set`) describe the connected fronts or sets, i.e., all fronts that are connected to each other are regarded as a single front, and then the previous four characteristics are computed for those aggregated fronts and sets, respectively. Note that, as the number of fronts (or sets) might be larger than the population size (i.e. the number of considered individuals), we standardize each characteristic by its corresponding highest achievable value (i.e., the minimum of population size and considered fronts or sets).

Feature Computation   Obviously, some features (such as `conn_ps.count_rel`) require more computational resources than others. Nevertheless, so far the biggest part

---

[5]A front or set is "covered" if an individual is located in its ε-environment.

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

Table 1 – Parameters used by the MPM2-generator for creating the 40 mixed-sphere and mixed-ellipse problems. Each of the parameter combinations below was used to generate two problems: one with spherical and one with elliptical shape (`peak.shape`). If the peaks are aligned in an elliptical shape, the `rotated`-parameter was set to `TRUE`, otherwise, i.e., in case of spherical shaped peaks, it was set to `FALSE`. Across all 40 problems, the problem's dimension (`dimension`) was set to 2 and the `topology` parameter to `random`.

# **Peaks**																						
# **Peaks**	f_1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3
(`n.peaks`)	f_2	1	2	2	2	3	5	2	4	2	2	2	3	5	2	1	2	2	2	3	5
Seed	f_1	1	1	2	3	4	5	6	6	1	2	3	4	5	6	7	1	2	3	4	5
(`seed`)	f_2	3	5	6	7	8	9	8	7	5	6	7	8	9	4	4	5	6	7	8	9

of the resources is needed for matching the individuals to the correct set or front, respectively. These computations take even longer for line segments in which the points of different sets (or fronts) are so close to each other that individuals, which actually belong to different line segments, alternate. Here, the worst case scenario are alternations of irregular frequencies between line segments.

8 Experiments

The main goal of this paper is an improved understanding of multimodality in the context of multi-objective optimization. In order to perform experiments, we first created a benchmark consisting of easily configurable multi-objective and multimodal test problems. To be more precise, we manipulated the MPM2-generator (Wessing, 2015) in a way, that it produces bi-objective (instead of single-objective) multiple peak problems[6].

The rational behind using the mixed-sphere and mixed-ellipse problems rather than using other well-known multi-objective benchmarks, such as DTLZ (Deb et al., 2005) or ZDT (Zitzler et al., 2000), is the fact that the former allows to control the multimodality and thus hardness of the problems, whereas the latter ones are (at least for now) already too complicated for our purposes, which is a (hopefully) complete understanding of the interactions of the multimodal objectives. As it turns out, even for this rather simple setting, the problems quickly become (with an increasing number of peaks per objective) very difficult and highly multimodal.

8.1 Setup of Benchmark Problems

We created a benchmark with two groups of two-dimensional problems: 20 mixed-sphere problems and 20 mixed-ellipse problems. Each of the objectives of the 40 instances contains between one and five peaks. Furthermore, for a specific problem the contour lines of all peaks (of both objectives) are either spheres or rotated ellipses. Consequently, the local efficient sets can either be found somewhere on (a) a line segment (in case of sphere-shaped peaks) or (b) a curve (in case of ellipse-shaped peaks) between each pair of peaks from the different objectives.

Also, the generator was configured to place the peaks randomly within the decision space to account for multimodal landscapes. The alternative "default" option leads to nearby aligned peaks that would result in a funnel-like landscape – which is much more similar to a unimodal optimization problem rather than a multimodal one. The complete setup of this benchmark is given within Table 1.

---

[6]The MPM2-generator is for instance available in the python package `optproblems0.9` (Wessing, 2016) and within the R-package `smoof` (Bossek, 2017).

8.2   Setup of Multi-Objective Optimization Algorithms

We tested two conceptually different optimizers in order to investigate the challenges imposed by different degrees of multimodality on the optimization progress. Specifically, a stochastic local search (SLS) technique is contrasted to a gradient-based strategy (HIGA-MO). The population size is set to $\mu = 50$ for both algorithms while the initial step-size is set to 0.01 for SLS and 0.001 for HIGA-MO. While the step-size remains constant in case of the naïve SLS, HIGA-MO will actively perform a step-size adaptation[7] (Wang et al., 2017a). Thus, in the beginning it will rather explore the landscape by making larger steps and towards the end, it will exploit promising areas with rather small steps. The maximal number of iterations has been set to 500 iterations for HIGA-MO and – accounting for the rather naïve structural concept of the second solver – to 800 iterations for SLS. These budgets are chosen according to the ratio of expected running time (consumed to achieve the target convergence measure) between HIGA-MO and SLS w.r.t. the algorithm setting above. Note, that a systematic benchmark of solvers is not the focus of this work, but rather explaining algorithm behavior in general.

8.3   Experimental Results

In the following, we will first have a look at four scenarios, which we considered to be rather easy to grasp and among the most representative of the 40 benchmark problems, and visually study the behavior of HIGA-MO and SLS on those instances. Afterwards, we will analyze the problem and algorithm characteristics across the entire benchmark in more detail[8]. This should give us some first insights on whether our suggested problem characteristics can actually be used for distinguishing the problems from each other and also, which problem property might cause which algorithm behavior.

In addition to the results that we can show here, we provide more material, including various tables, figures, as well as videos for all 40 benchmark problems online[9].

8.3.1   Exemplary scenarios

As a first (explorative) step, we analyze four scenarios by visualizing their multi-objective gradient landscapes (as introduced in Sect. 5), as well as the trace of the population from the two algorithms.

The visualizations of the first scenario (ID 35 from the benchmark) are shown within the previous sections. More precisely, Figures 3 and 4 show the multi-objective gradient landscape of this problem within the decision and objective space, whereas Fig. 5 depicts the differences within the behavior of HIGA-MO and SLS. As Fig. 3 reveals, the analyzed scenario consists of two peaks in the first (visualized by orange contour lines) and one peak in the second objective (white dotted lines), resulting in two basins of attraction, comprising a set of connected local efficient sets each. Those connected sets are also visible by the yellow/green/blue valleys that lead towards the local efficient sets. Note that the coloring represents the depth of the valley. Due to the competition between the two peaks from the first objective, the right part of the left connected set (light orange line) is dominated by the entire connected set of the right basin (dark blue line). When looking at the traces of HIGA-MO (upper row of Fig. 5)

---

[7]The step-size adaption uses cumulative step-size control with parameters $\alpha = 0.5$ and $c = 0.2$.

[8]For the algorithm characteristics, an individual was considered to be in a set's (or front's) vicinity, if the difference between the respective individual and the closest point from the closest respective set (or front, respectively) was at most $5 \cdot 10^{-3}$ for each of the two dimensions (or objectives). In contrast to that, we were able to use a much more detailed grid for the computation of the landscape characteristics and hence, were able to use a much smaller threshold of 10^{-3}.

[9]https://www.wi.uni-muenster.de/department/statistik/additional-material

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

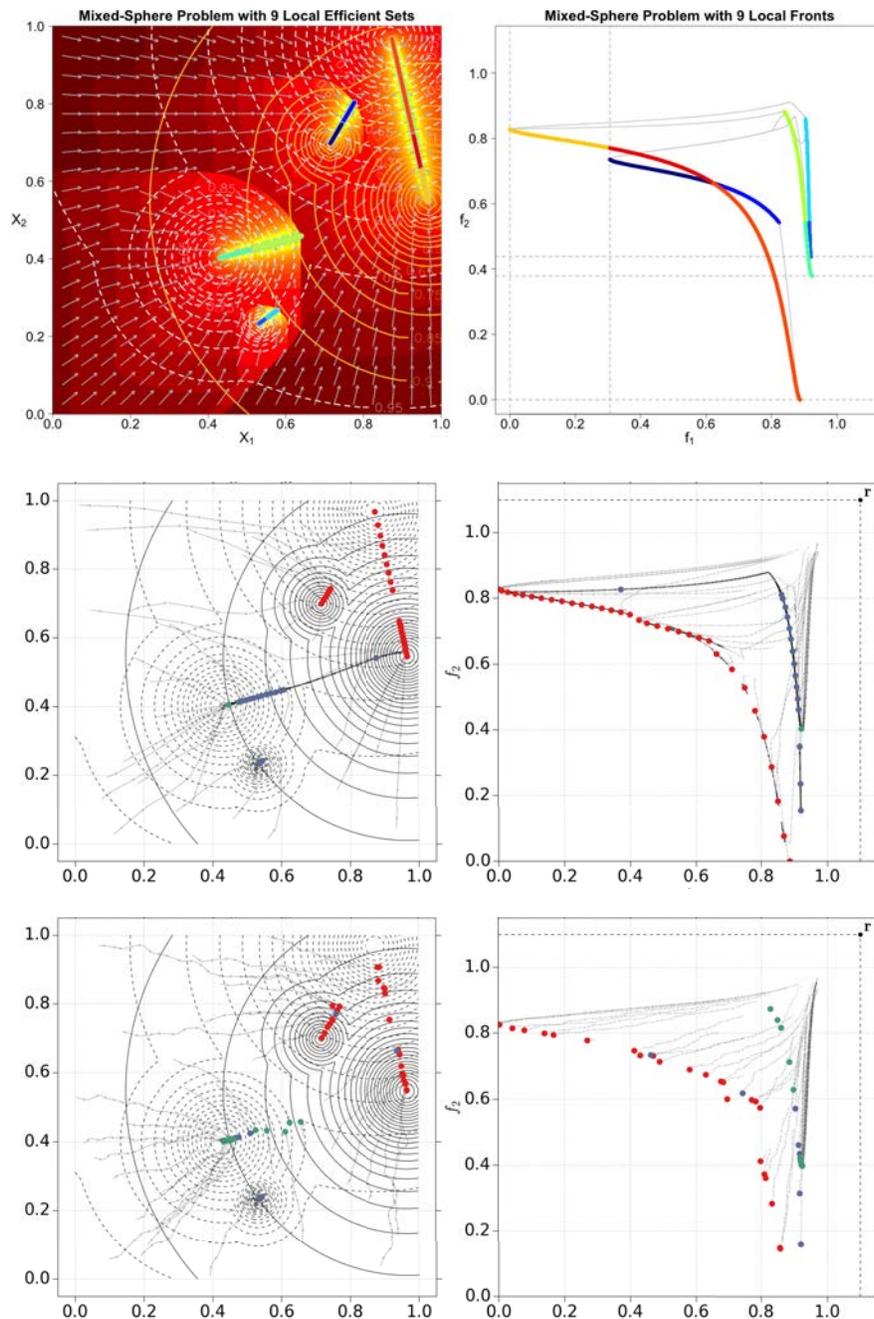**Figure 6** – A problem instance from the benchmark (ID 32) with sphere-shaped peaks (with `n.peaks = 2` and `seed = 4` for the first and `n.peaks = 3` and `seed = 8` for the second objective). The left column shows the heatmap based on the cumulated path lengths of the of multi-objective gradients (top), as well as a trace of HIGA-MO's (middle) and SLS' (bottom) population in the decision space of the landscape. The right column shows the theoretically existing four local efficient sets (top), as well as the behavior of the populations of HIGA-MO (middle) and SLS (bottom) – in the objective space.

and SLS (lower row), one can see that both algorithms succeed in finding the connected sets. But while SLS gets stuck in these local optima (as indicated by the blue points[10] within Fig. 5), HIGA-MO follows its goal, i.e., maximizing the dominated hypervolume, and consequently leaves the dominated part (light orange line) of the two sets in the left basin in order to travel towards the global efficient part (light blue line).

The scenario within Fig. 6 (ID 32) represents a slightly more difficult sphere-problem with two peaks in the first and three peaks in the second objective. In addition to this, the corresponding three-dimensional surface plot and the mapping from the decision into the objective space are shown in the top row of Fig. 9. The problem landscape also consists of two big basins of attraction. However, each one of them actually also contains a smaller basin. When looking at the objective space, one can also see that the corresponding local fronts (each of them colored in two shades of blue) are always closely located to local fronts from their respective surrounding bigger basins. Nevertheless, HIGA-MO again is able to find all Pareto fronts (including the dark blue front). It is also visible that HIGA-MO "pushes" a lot of individuals[11] from one basin towards the other one, or more precisely from the local efficient sets located near the peak at $\mu_1 \approx (0.4, 0.4)^T$, i.e., the turquoise and light green lines, towards the other peak at $\mu_2 \approx (1.0, 0.5)^T$ and from there along the adjacent global efficient set (yellow line). The few individuals from HIGA-MO's final population that are located along the light blue/middle blue and especially along the turquoise/light green sets might be caused by the limited number of generations. In case of the SLS, one would not expect any bigger changes with additional generations. Its points are located along all the local fronts (including the turquoise and light green fronts/sets), but it is not able to leave these (globally) dominated areas.

The two objectives from the next scenario (ID 7) are based on two and one ellipse-shaped peaks, respectively. As indicated by the multi-objective gradients within Fig. 7, this landscape also consists of two basins of attraction. Furthermore, one can see the ridge, i.e., the bended line starting approximately at $(0.0, 0.6)^T$ and ending at about $(0.9, 0.0)^T$, between the two basins. The problem basically contains two connected sets, but due to a partial overlap of their corresponding fronts (in the objective space), the intermediate section of the right connected set (red line) is globally dominated by the dark blue line. Analogously, the left connected set is cut in half (at least in the decision space), because its upper part (blue) is dominated by the light blue section. When looking at the traces of the algorithms' populations, it is peculiar that SLS barely finds any of the local fronts in general while HIGA-MO again finds all Pareto sets – and thus, obviously ignores the globally dominated sections of the connected sets.

Fig. 8 shows the final scenario (ID 5), which is based on two objectives with (slightly) ellipse-shaped peaks. The first objective consists of a single peak, the second contains three peaks. Again, the results of the algorithm runs are quite interesting: although the majority of individuals from HIGA-MO finds the global efficient set, not all of them succeed. In general, all of its individuals quickly converge to any of the local fronts, but while the individuals that reach the green local efficient sets, rather quickly travel towards the orange global efficient set – leading to the so-called *channeling effect*[12]

---

[10]The coloring of the points represents the dominance relation among the final population, i.e., red points represent the first, blue points the second and green points the third layer.

[11]The "push" is caused by the fact that HIGA-MO performs local searches along the front and once an individual crosses a ridge, it strives for the "better" front.

[12]By *channeling* we refer to the effect, in which multiple individuals walk along the same path – ultimately showing darker paths connecting the local fronts. Such "channels" result from local efficient sets that are connected to ridges.

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich
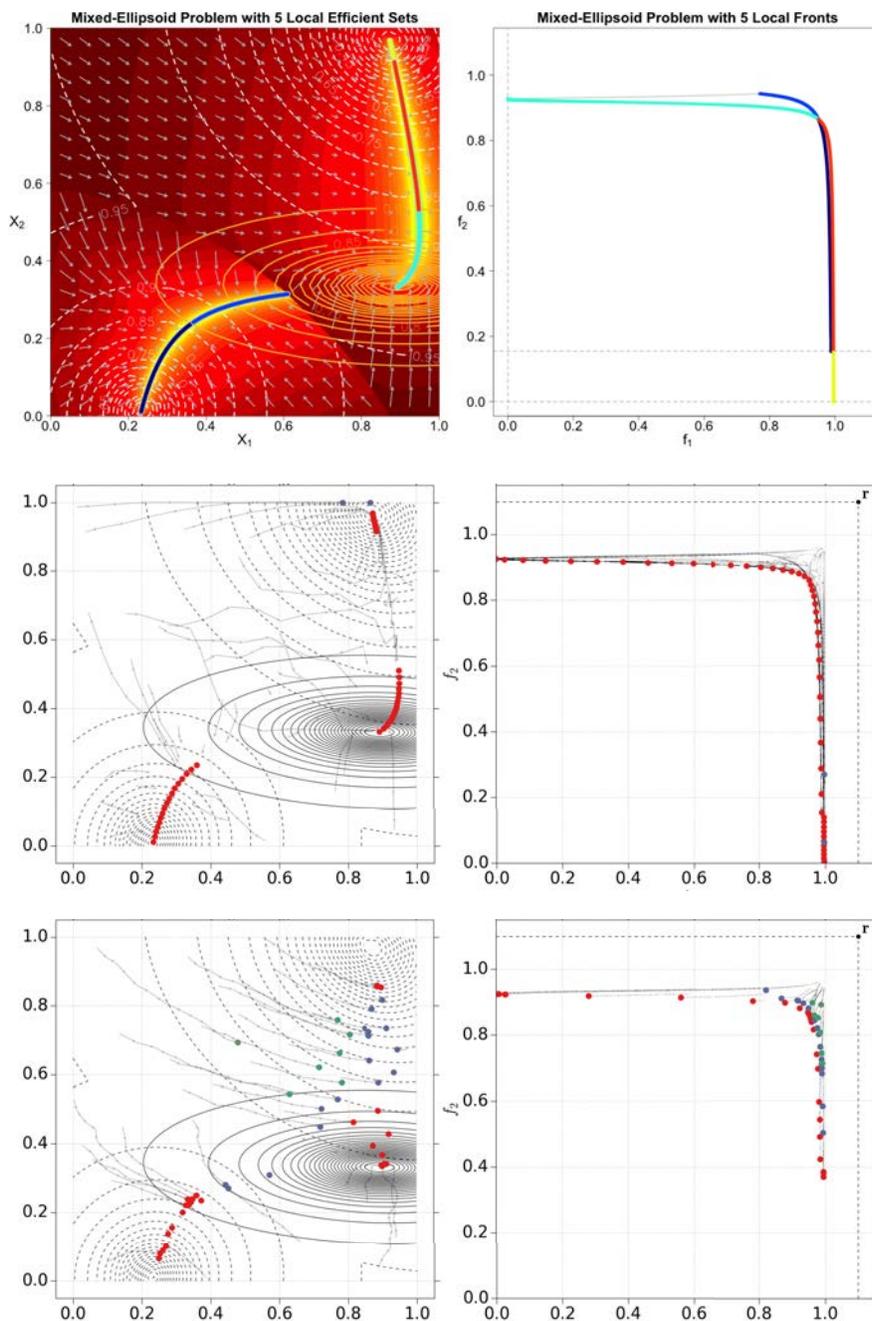

Figure 7 – A problem instance from the benchmark (ID 7) with (slightly) ellipse-shaped peaks (with `n.peaks = 1` and `seed = 6` for the first and `n.peaks = 2` and `seed = 8` for the second objective). The left column shows the heatmap based on the cumulated path lengths of the of multi-objective gradients (top), as well as a trace of HIGA-MO's (middle) and SLS' (bottom) population in the decision space of the landscape. The right column shows the theoretically existing four local efficient sets (top), as well as the behavior of the populations of HIGA-MO (middle) and SLS (bottom) – in the objective space.
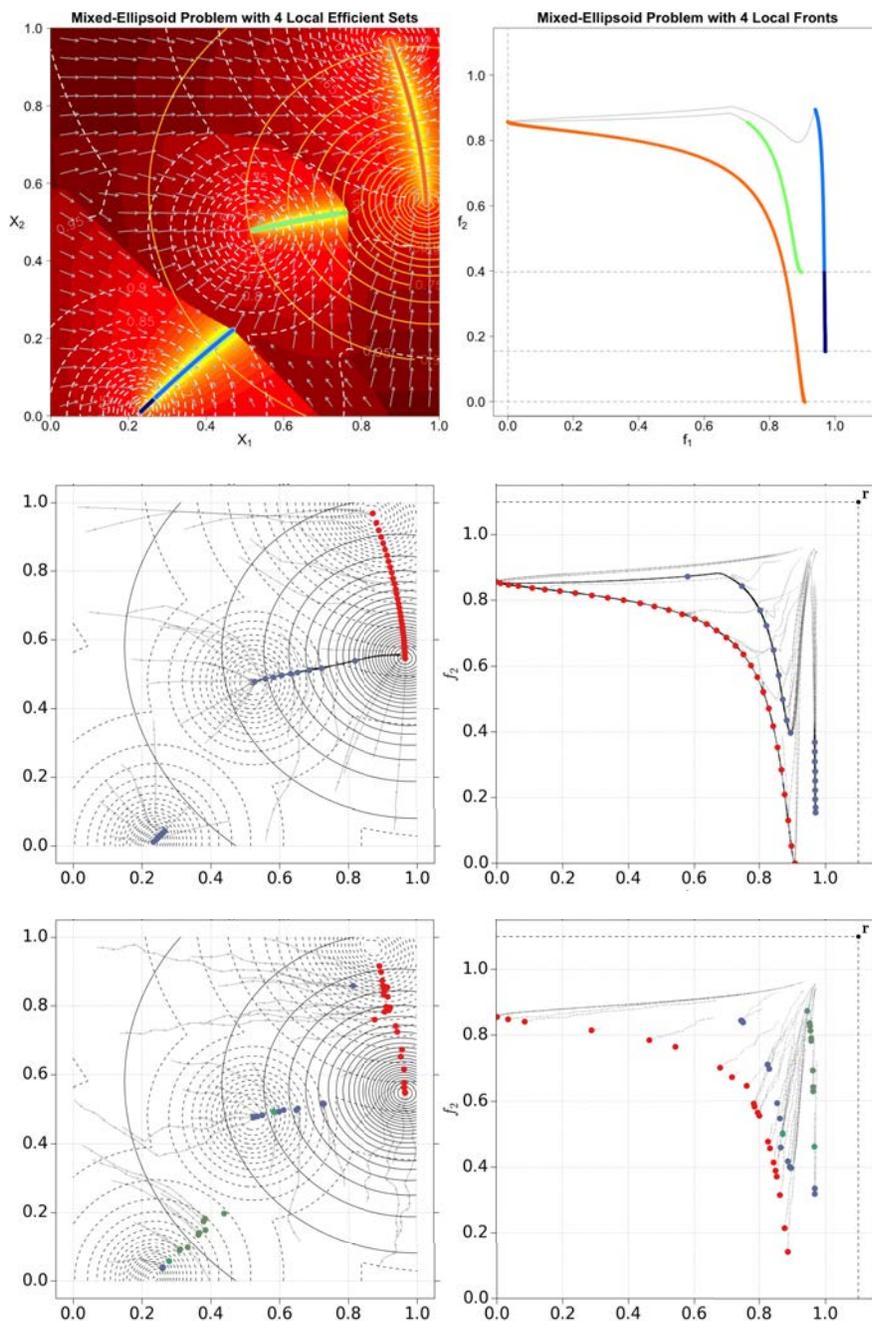
Figure 8 – A problem instance from the benchmark (ID 5) with (slightly) ellipse-shaped peaks (with `n.peaks = 1` and `seed = 4` for the first and `n.peaks = 3` and `seed = 8` for the second objective). The left column shows the heatmap based on the cumulated path lengths of the of multi-objective gradients (top), as well as a trace of HIGA-MO's (middle) and SLS' (bottom) population in the decision space of the landscape. The right column shows the theoretically existing four local efficient sets (top), as well as the behavior of the populations of HIGA-MO (middle) and SLS (bottom) – in the objective space.

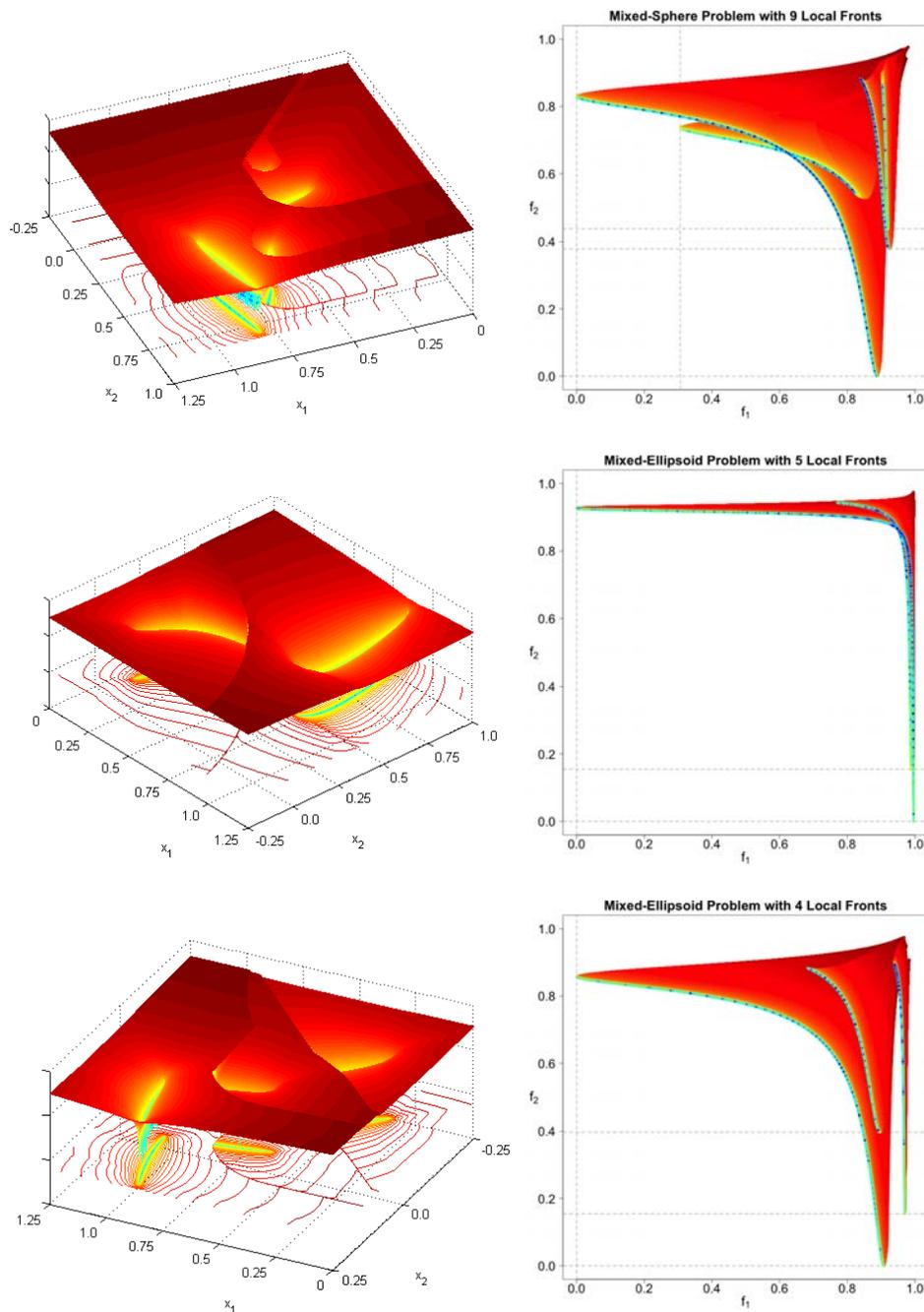P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

Figure 9 – Depiction of the gradient field landscape as three-dimensional surface plots (left) and the corresponding mapping into the objective space (right) for three of the four exemplary scenarios from the benchmark. The plots show the results for three problems from our benchmark (top: ID 32, middle: ID 7, bottom: ID 5).
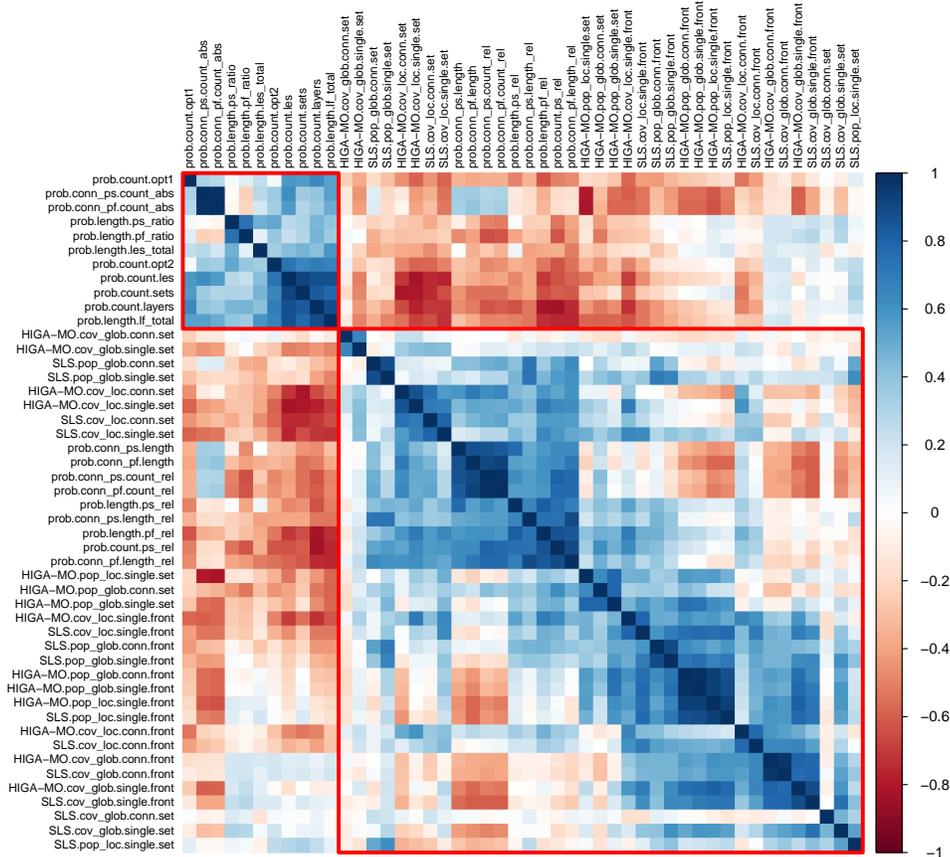
Figure 10 – Visualization of correlations (Friendly, 2002) between problem (prefix `prob`) and algorithm characteristics. The latter were computed for HIGA-MO (`HIGA-MO`) and the naïve SLS (`SLS`). The characteristics are ordered based on a complete-linkage clustering approach (Jobson, 1992) and the correlation matrices of its two most obvious clusters are framed by red squares.

– individuals which come across the light or dark blue local efficient sets do not manage to leave that area completely and instead focus on spreading along the "better" half of the efficient set (according to the *true* local fronts) – i.e., the dark blue segment – as depicted by the blue points within the HIGA-MO plots in the middle row of Fig. 8.

Summarizing the findings across all four analyzed scenarios, we were able to describe the behavior of the two algorithms based on our visualization approaches. Furthermore, we could show that HIGA-MO, as a multi-objective global optimizer, in most cases finds the global optima, whereas SLS, as a (multi-objective) local search algorithm, usually converges in a local optimum. Thus – although not surprising – multimodality has different impacts w.r.t. conceptually different search strategies.

8.3.2 Analyzing the problem and algorithm characteristics

As shown within the first part of this section, our benchmark allows to distinguish between different algorithms. So far, these differences were mainly based on visual inspections across a subset of the benchmark. In the following, we will analyze the landscape and algorithm characteristics, which were introduced in Section 7, to get a basic idea of possible causes for certain algorithm behavior and show the relevance of

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

Table 2 – Top 10 strongest (positive or negative) correlations among the problem characteristics.

	Characteristic 1	Characteristic 2	Correl.
1	prob.conn_ps.count_abs	prob.conn_pf.count_abs	0.990
2	prob.conn_ps.count_rel	prob.conn_pf.count_rel	0.989
3	prob.conn_pf.length	prob.conn_pf.count_rel	0.925
4	prob.count.les	prob.count.sets	0.911
5	prob.conn_ps.count_rel	prob.conn_pf.length	0.903
6	prob.count.ps_rel	prob.conn_pf.length_rel	0.896
7	prob.conn_ps.length_rel	prob.conn_pf.length_rel	0.893
8	prob.count.layers	prob.length.lf_total	0.884
9	prob.count.les	prob.count.layers	0.874
10	prob.conn_ps.length	prob.conn_pf.length	0.872

our benchmark in that it covers a broad range of problem instances within our chosen classes. Such findings would be a good indication for possible multi-objective landscape features. Note that the following investigations are based on sophisticated visualization techniques, which in most cases successfully reduce the underlying dimensionality. However, for completeness, we also provided the exact values for each of the 20 problem and 28 algorithm characteristics (14 per algorithm) across all 40 benchmark problems within the additional material[9].

In a first step, we visualized the correlation matrix, which is based on all pairwise (Pearson) correlations among all 48 characteristics, within Fig. 10. The colors of the boxes represent the correlations. That is, while blue boxes correspond to positive correlations, red boxes correspond to negative correlations and the intensity of the color indicates the magnitude or strength of the correlations, i.e., highly correlated characteristics yield to darker boxes (Friendly, 2002). Furthermore, the characteristics are aligned based on a (complete linkage) hierarchical clustering approach (Jobson, 1992). Given this clustering approach, two clusters were revealed: a smaller cluster, completely consisting of eleven problem characteristics (five out of six count characteristics, four out of six length characteristics and two out of eight connected front/set characteristics) and a bigger cluster, comprising the remaining 37 characteristics. The corresponding correlation matrices of these two clusters are highlighted by red squares within Fig. 10.

By analyzing the magnitude of the correlations in more detail, one can find indications for possible redundancy among the characteristics. For instance, the number of local efficient sets that have a direct or indirect connection to the Pareto set (`prob.conn_ps.count_abs`) and the counterpart in the objective set (`prob.conn_pf.count_abs`) are highly correlated (0.990) to each other. Also, the characteristics with the corresponding proportions (instead of the absolute numbers) of the corresponding sets/fronts (`prob.conn_ps.count_rel` and `prob.conn_pf.count_rel`) are highly correlated (0.989). An overview of the ten highest correlated problem characteristics can be found in Table 2.

Focussing on the algorithm characteristics, it is noticeable that HIGA-MO shows a strong positive correlation (0.982) between the percentage of individuals that are located in the neighborhood of a local front that has a connection to any of the Pareto fronts (`HIGA-MO.pop_glob.conn.front`) and the individuals that are located close to a Pareto front (`HIGA-MO.pop_glob.single.front`). This effect is plausible: as explained within the description of the four exemplary problems, HIGA-MO is able to travel along adjacent fronts towards a better front (w.r.t. the dominance relationship) once it finds a local front in general. As a result, fronts that are connected to the Pareto fronts, but which actually are not Pareto dominant themselves, will be uncovered. Consequently, both characteristics basically measure the coverage of the Pareto fronts (for

Table 3 – Top 10 strongest (positive or negative) correlations among the algorithm characteristics.

	Characteristic 1	Characteristic 2	Correl.
1	`HIGA-MO.pop_glob.conn.front`	`HIGA-MO.pop_glob.single.front`	0.982
2	`HIGA-MO.cov_glob.conn.front`	`SLS.cov_glob.conn.front`	0.969
3	`HIGA-MO.pop_glob.conn.front`	`HIGA-MO.pop_loc.single.front`	0.935
4	`HIGA-MO.pop_loc.single.front`	`SLS.pop_loc.single.front`	0.932
5	`HIGA-MO.pop_glob.single.front`	`HIGA-MO.pop_loc.single.front`	0.930
6	`SLS.pop_glob.conn.front`	`SLS.pop_glob.single.front`	0.892
7	`SLS.pop_glob.conn.set`	`SLS.pop_glob.single.set`	0.888
8	`HIGA-MO.cov_glob.single.front`	`SLS.cov_glob.single.front`	0.886
9	`HIGA-MO.pop_glob.single.front`	`SLS.pop_loc.single.front`	0.880
10	`HIGA-MO.pop_glob.conn.front`	`SLS.pop_loc.single.front`	0.873

Table 4 – Ten strongest correlations between algorithm and problem characteristics.

	Characteristic 1	Characteristic 2	Correl.
1	`prob.count.les`	`HIGA-MO.cov_loc.single.set`	-0.828
2	`prob.count.sets`	`HIGA-MO.cov_loc.conn.set`	-0.818
3	`prob.conn_pf.count_abs`	`HIGA-MO.pop_loc.single.set`	-0.811
4	`prob.count.sets`	`HIGA-MO.cov_loc.single.set`	-0.805
5	`prob.conn_ps.count_abs`	`HIGA-MO.pop_loc.single.set`	-0.791
6	`prob.count.layers`	`HIGA-MO.cov_loc.single.set`	-0.772
7	`prob.count.les`	`SLS.cov_loc.single.set`	-0.765
8	`prob.count.sets`	`SLS.cov_loc.conn.set`	-0.745
9	`prob.count.les`	`SLS.cov_loc.conn.set`	-0.741
10	`prob.conn_ps.length_rel`	`SLS.pop_glob.single.set`	0.732

HIGA-MO). Interestingly, SLS also shows a strong positive correlation (0.892) between these two characteristics, but it is caused by a completely different behavior: the likeliness of SLS actually finding a global efficient set is – as already described for the four exemplary problems – similar to the one of SLS finding a local efficient set that is connected to any of the global efficient sets. Note that in contrast to HIGA-MO, the naïve SLS does not profit from these connections and instead converges to those multi-objective local optima – remember, it is just a local search algorithm and not a global optimizer. Further details about the ten strongest correlations between pairs of algorithm characteristics can be found within Table 3.

Aside from detecting possible redundancy among the characteristics or explaining certain algorithmic behavior, it is of interest to find out which problem characteristics might cause specific algorithm characteristics – independent of influences from any of the other characteristics. Thus, we are interested in the strongest correlations between problem and algorithm characteristics as listed within Table 4. Not very surprisingly, the nine strongest correlations between problem and algorithm characteristics are – without any exceptions – negative and seven out of these nine pairs actually state that an increasing number of sets, fronts or layers leads to a reduced coverage of global or local sets/fronts. The two exceptions to this are correlations based on the percentage of HIGA-MO's final population that is located in the proximity of any local efficient set (`HIGA-MO.pop_loc.single.set`) and the number of fronts/sets that have a connection to the Pareto sets/fronts (`prob.conn_ps.count_abs` and `prob.conn_pf.count_abs`), i.e., the more fronts/sets are connected to the Pareto fronts/sets, the less likely HIGA-MO will find any of the fronts/sets in general. Looking at the strongest *positive* correlation between problem and algorithm characteristics, it is plausible that an increase in the relative length of the Pareto set (`prob.conn_ps.length_rel`) causes a higher percentage of SLS' final population to be located near the Pareto set (`SLS.pop_glob.single.set`).

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich
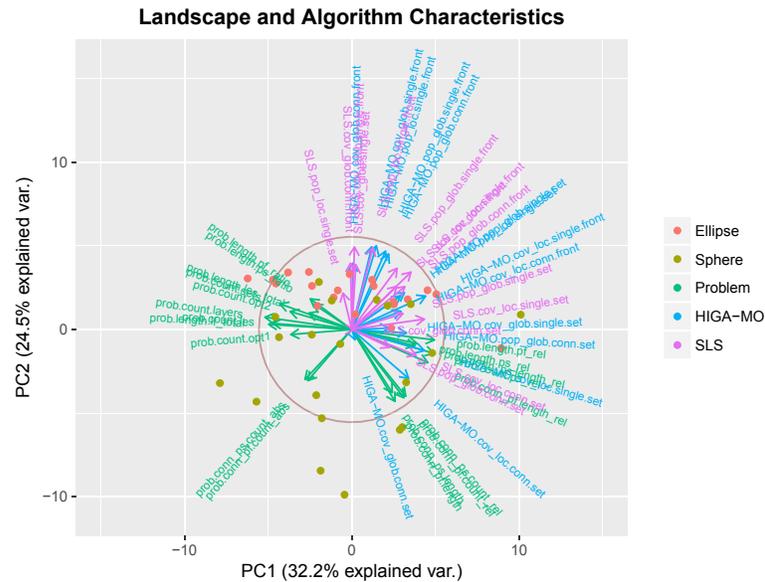

Figure 11 – Biplot of a principal component analysis based on the correlation matrix of all 20 problem (prefix `prob`) and 28 algorithm characteristics – 14 per algorithm (`HIGA-MO` or `SLS`).

These findings are also supported by Fig. 11, which displays a *biplot* (Gabriel, 1971; Gower and Hand, 1995) of a *principal component analysis* (PCA, Härdle and Simar (2015)) based on the correlation of all 48 characteristics. The goal of such a PCA is to reduce the number of dimensions of the original data set. This is achieved by representing the data by so-called *principal components* (PCs), which basically are a linear combination of the variables from the original data set. Each of these PCs is constructed in such a way that it explains the highest amount of variance (and thus information) within the hyperplane that is orthogonal to all the previously constructed PCs. Consequently, the first PC explains the most variance of the original data set, the second PC the second most, etc. As shown within Fig. 11, the two PCs derived from a PCA based on all characteristics already explain 55% – and thus, more than half – of the variance of all 48 characteristics. As the name of the figure ("biplot") indicates, it actually summarizes two figures within one: (a) a projection of the 40 benchmark problems onto the (hyper)plane that is constructed by the first two PCs, showing problems with sphere-shaped peaks as green dots and instances with ellipse-shaped peaks as red dots and (b) the explanatory input of each characteristic on the first two PCs as arrows – colored according to the group that they represent: problem characteristics are shown as green arrows, whereas the algorithm characteristics are colored in blue (HIGA-MO) and pink (SLS). Note that arrows of characteristics, which perfectly explain at least one of the two PCs, would actually touch the shown circle.

As one can see, the characteristics basically form the same two groups as within the visualization of the correlation matrix: the eleven problem characteristics, which formed the smaller cluster in Fig. 10 are pointing in the negative direction of PC1, whereas the remaining characteristics once again form a second (bigger) cluster. Also, some "sub-clusters", such as `prob.conn_ps.count_abs` and `prob.conn_pf.count_abs` (bottom left of the biplot), are even more visible than within the correlation matrix plot (topleft). While the majority of problem characteristics – with the exception of problem characteristics describing the connectedness towards the Pareto fronts/sets (`prob.conn_*`) – are mostly horizontally aligned and
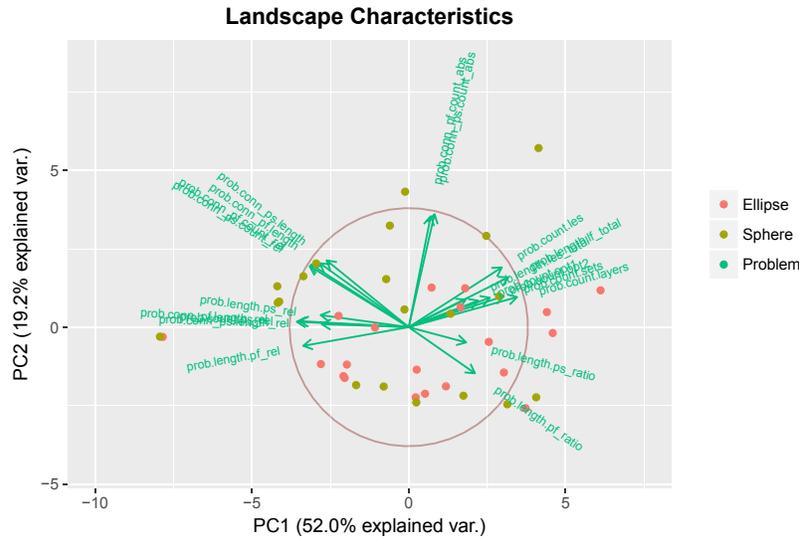
Figure 12 – Biplot resulting from a principal component analysis, which has been applied to the 20 landscape characteristics for each of the 40 benchmark problems.
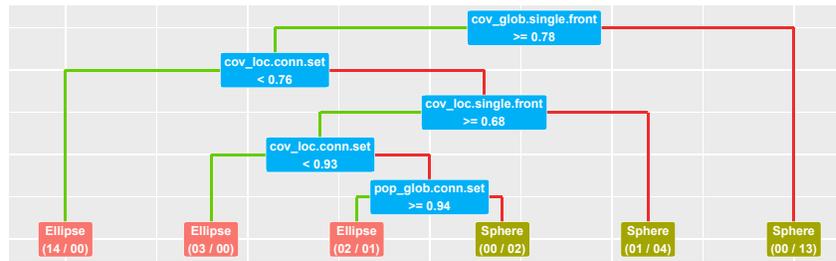

Figure 13 – Classification tree for separating the benchmark problems with different peak shapes using the algorithm characteristics of HIGA-MO.

thus only have a small influence on the second PC, the latter is mainly influenced by algorithm characteristics. Furthermore, although the problems seem to be separable by the shape of their peaks, they are mainly distinguished by the second PC, i.e., rather by the algorithm than the problem characteristics. This is due to the different algorithm concepts and thus their different behavior w.r.t. the degree of multimodality. These differences tend to be larger for ellipse-shaped peaks.

When looking at the biplots that are purely based on the 20 landscape characteristics or on the 28 algorithm characteristics, the aforementioned discoveries become even more obvious. Although, in case of the landscape-based PCA (Fig. 12), the first two PCs explain roughly 70% of the variance, they are not able to clearly separate problems with sphere-shaped peaks from the ones with ellipse-shaped peaks. Instead, it rather can be used to group the landscape characteristics – into two (or maybe three) clusters. The left group, with the exception of the two characteristics that measure the total lengths of the fronts and sets that are connected to the Pareto fronts/sets, comprises characteristics of the proportion of counts or lengths of the corresponding fronts/sets to the number/lengths of all fronts/sets (indicated by the suffix `rel`), whereas the remaining characteristics form a second cluster. Eventually, one could divide the latter

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

once again, given that `prob.conn_ps.count_abs` and `prob.conn_pf.count_abs`, i.e., the number of local efficient sets or local fronts that are connected to the Pareto sets/fronts, form a potential third cluster. Therefore, we can conclude that the chosen landscape characteristics manage to cover different aspects of the problem landscape.

In contrast to that, the PCA based on the algorithm characteristics (for either one of the two algorithms), is able to split the data according to the shape of the peaks. In case of the stochastic local search algorithm, the knowledge of `SLS.pop_loc.single.front` is already sufficient to correctly classify 95% of the problems[13]. For HIGA-MO, it is much harder to find a classification model of comparable accuracy. For instance, Fig. 13 displays a classification tree, which satisfies this condition – but it requires five algorithm characteristics.

Summarizing the previous findings, one can say that the problem characteristics capture different properties of a problem's landscape than the shape of its peaks, whereas the algorithm characteristics indicate that the peak-shape actually influences the behavior of HIGA-MO and SLS. Moreover, the benchmark problems cover various landscape characteristics and degrees of multimodality. By correlating problem and algorithm characteristics the basic algorithm behaviors could be explained and conceptual differences were detected. Certainly, for both algorithms the problem gets harder with an increasing number of (local) fronts, but in general HIGA-MO is able to cope with it much better. The next step will be the construction of exploratory landscape features based on our findings, which will be able to assess the multimodality aspects prior to optimization and thereby allow to build algorithm selection models. For this purpose also a larger set of (state-of-the-art) optimization algorithms will have to be applied in a systematic way.

9 Discussion and Outlook

This paper provides concepts for thoroughly understanding multimodality in the context of multi-objective optimization problems, both theoretically as well as experimentally. A specifically designed benchmark set constructed by means of a sophisticated multiple peaks generator is introduced and used as a testbed for contrasting two conceptually different search strategies, i.e. Hypervolume gradient ascent and stochastic local search. Mixed spheres and elliptically shaped variants reveal different degrees of multimodality and levels of problem hardness. Problems' landscapes and specifically basins of attractions are visualized based on a scalar combination of gradients in order to substantially increase problem understanding which is clearly enhanced by deriving local front specifications analytically.

Moreover, algorithm characteristics are introduced 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. Those are related to respective problem characteristics in a systematic way using multivariate analytical techniques. Obviously, problem characteristics reflect much more information than just the number of peaks and the spherical or elliptical structure. Basic algorithm behaviors could be explained and conceptual differences detected in that certainly the problem gets harder with an increasing number of (local) fronts, but in general HIGA-MO is able to cope with it much better.

Therefore, the basis for systematically constructing multi-objective Exploratory Landscape Features is formed which has huge potential w.r.t. algorithm benchmarking, selection and design, also for higher dimensional problems. A thorough and system-

---

[13]19 out of 20 sphere-shaped problems have a value of at most 0.74, whereas the same amount of ellipse-shaped problems has a value of greater or equal to 0.82.

atic benchmark of state-of-the-art multi-objective optimizers will be conducted while simultaneously extending the problem generator to varying problem topologies. Moreover, specific design of optimizers addressing both diversity in decision space as well as multimodality of the landscape will be focused.

References

Bischl, B., Mersmann, O., Trautmann, H., and Preuss, M. (2012). Algorithm Selection Based on Exploratory Landscape Analysis and Cost-Sensitive Learning. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation (GECCO)*, pages 313 – 320. ACM.

Bossek, J. (2017). smoof: Single and Multi-Objective Optimization Test Functions. *The R Journal*. R-package version 1.5.

Daolio, F., Liefooghe, A., Verel, S., Aguirre, H., and Tanaka, K. (2016). Problem Features vs. Algorithm Performance on Rugged Multi-Objective Combinatorial Fitness Landscapes. *Evolutionary Computation Journal (ECJ)*.

Deb, K., Thiele, L., Laumanns, M., and Zitzler, E. (2005). Scalable Test Problems for Evolutionary Multiobjective Optimization. In Abraham, A., Jain, L., and Goldberg, R., editors, *Evolutionary Multiobjective Optimization*, Advanced Information and Knowledge Processing, pages 105 – 145. Springer.

Ehrgott, M. (2005). *Multicriteria Optimization*. Springer, 2nd edition.

Emmerich, M. T. M. and Deutz, A. H. (2012). Time complexity and zeros of the hypervolume indicator gradient field. In Schütze, O., Coello Coello, C. A., Tantar, A.-A., Tantar, E., Bouvry, P., Del Moral, P., and Legrand, P., editors, *EVOLVE – A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation III*, volume 500 of *Studies in Computational Intelligence (SCI)*, pages 169 – 193. Springer.

Emmerich, M. T. M., Deutz, A. H., and Beume, N. (2007). 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., and Sampels, M., editors, *Proceedings of the 4th International Workshop on Hybrid Metaheuristics (HM)*, volume 4771 of *Lecture Notes in Computer Science (LNCS)*, pages 140 – 156. Springer.

Fonseca, C. M. M. d. (1995). *Multiobjective Genetic Algorithms with Application to Control Engineering Problems*. PhD thesis, University of Sheffield.

Friendly, M. (2002). Corrgrams: Exploratory Displays for Correlation Matrices. *The American Statistician*, 56(4):316 – 324.

Gabriel, K. R. (1971). The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika*, 58(3):453 – 467.

Gower, J. C. and Hand, D. J. (1995). *Biplots*, volume 54. Chapman and Hall \CRC.

Gray, L. (2003). A Mathematician Looks at Wolfram's New Kind of Science. *Notices*, 50(2):200 – 211.

Härdle, W. K. and Simar, L. (2015). *Applied Multivariate Statistical Analysis*. Springer, 4th edition.

Jobson, J. D. (1992). *Applied Multivariate Data Analysis: Volume II: Categorical and Multivariate Methods*. Springer.

Kerschke, P. (2017). *flacco: Feature-Based Landscape Analysis of Continuous and Constrained Optimization Problems*. R-package version 1.7.

Kerschke, P. and Grimme, C. (2017). An Expedition to Multimodal Multi-Objective Optimization Landscapes. In Trautmann, H., Rudolph, G., Kathrin, K., Schütze, O., Wiecek, M., Jin, Y., and Grimme, C., editors, *Proceedings of the 9th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, volume 10173 of *Lecture Notes in Computer Science (LNCS)*, pages 329 – 343. Springer.

Kerschke, P., Preuss, M., Wessing, S., and Trautmann, H. (2015). Detecting Funnel Structures by Means of Exploratory Landscape Analysis. In *Proceedings of the 17th Annual Conference on Genetic and Evolutionary Computation (GECCO)*, pages 265 – 272. ACM.

Kerschke, P., Preuss, M., Wessing, S., and Trautmann, H. (2016a). Low-Budget Exploratory Landscape Analysis on Multiple Peaks Models. In *Proceedings of the 18th Annual Conference on Genetic and Evolutionary Computation (GECCO)*, pages 229 – 236. ACM.

Kerschke, P., Wang, H., Preuss, M., Grimme, C., Deutz, A. H., Trautmann, H., and Emmerich, M. T. M. (2016b). Towards Analyzing Multimodality of Multiobjective Landscapes. In Handl, J., Hart, E., Lewis, P. R., López-Ibánez, M., Ochoa, G., and Paechter, B., editors, *Proceedings of the 14th International Conference on Parallel Problem Solving from Nature (PPSN XIV)*, volume 9921 of *Lecture Notes in Computer Science (LNCS)*, pages 962 – 972. Springer.

P. Kerschke, H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, M. T. M. Emmerich

Kozera, R., Noakes, L., and Klette, R. (2003). *External versus Internal Parameterizations for Lengths of Curves with Nonuniform Samplings*, pages 403–418. Springer Berlin Heidelberg, Berlin, Heidelberg.

Laforge, F. O., Roslund, J., Shir, O. M., and Rabitz, H. (2011). Multiobjective Adaptive Feedback Control of Two-Photon Absorption Coupled with Propagation Through a Dispersive Medium. *Physical Review A*, 84(1):013401–1 – 013401–10.

Liefooghe, A., Verel, S., Daolio, F., Aguirre, H., and Tanaka, K. (2015). A Feature-Based Performance Analysis in Evolutionary Multiobjective Optimization. In Gaspar-Cunha, A., Henggeler Antunes, C., and Coello Coello, C. A., editors, *Proceedings of the 8th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, volume 9019 of *Lecture Notes in Computer Science (LNCS)*, pages 95 – 109. Springer.

Maulana, A., Jiang, Z., Liu, J., Bäck, T. H. W., and Emmerich, M. T. M. (2015). Reducing complexity in many objective optimization using community detection. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC)*, pages 3140 – 3147. IEEE.

Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., and Rudolph, G. (2011). Exploratory Landscape Analysis. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation (GECCO)*, pages 829 – 836, New York, NY, USA. ACM.

Miettinen, K. (1998). *Nonlinear Multiobjective Optimization*, volume 12 of *International Series in Operations Research & Management Science*. Springer.

Nicolaou, C. A. and Brown, N. (2013). Multi-Objective Optimization Methods in Drug Design. *Drug Discovery Today: Technologies*, 10(3):e427 – e435.

Paquete, L., Chiarandini, M., and Stützle, T. (2004). *Pareto Local Optimum Sets in the Biobjective Traveling Salesman Problem: An Experimental Study*, pages 177–199. Springer Berlin Heidelberg, Berlin, Heidelberg.

Preuss, M., Naujoks, B., and Rudolph, G. (2006). Pareto Set and EMOA Behavior for Simple Multimodal Multiobjective Functions. In Runarsson, T. P., Beyer, H.-G., Burke, E. K., Merelo-Guervós, J. J., Whitley, L. D., and Yao, X., editors, *Proceedings of the 9th International Conference on Parallel Problem Solving from Nature (PPSN IX)*, volume 4193 of *Lecture Notes in Computer Science (LNCS)*, pages 513 – 522. Springer.

Rice, J. R. (1976). The Algorithm Selection Problem. *Advances in Computers*, 15:65 – 118.

Srinivas, N. and Deb, K. (1994). Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation Journal (ECJ)*, 2(3):221 – 248.

Stadler, P. F. and Flamm, C. (2003). Barrier Trees on Poset-Valued Landscapes. *Genetic Programming and Evolvable Machines*, 4(1):7 – 20.

Tantar, E., Dhaenens, C., Figueira, J. R., and Talbi, E.-G. (2008). A Priori Landscape Analysis in Guiding Interactive Multi-Objective Metaheuristics. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC)*, pages 4104 – 4111. IEEE.

Ulrich, T., Bader, J., and Thiele, L. (2010). *Defining and Optimizing Indicator-Based Diversity Measures in Multiobjective Search*, pages 707 – 717. Springer.

Verel, S., Liefooghe, A., and Dhaenens, C. (2011). Set-Based Multiobjective Fitness Landscapes: A Preliminary Study. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation (GECCO)*, pages 769 – 776, New York, NY, USA. ACM.

Verel, S., Liefooghe, A., Jourdan, L., and Dhaenens, C. (2013). On the Structure of Multiobjective Combinatorial Search Space: MNK-Landscapes with Correlated Objectives. *European Journal of Operational Research (EJOR)*, 227(2):331 –342.

Wang, H., Deutz, A. H., Bäck, T. H. W., and Emmerich, M. T. M. (2017a). Hypervolume Indicator Gradient Ascent Multi-objective Optimization. In Trautmann, H., Rudolph, G., Kathrin, K., Schütze, O., Wiecek, M., Jin, Y., and Grimme, C., editors, *Proceedings of the 9th International Conference on Evolutionary Multi-Criterion Optimization (EMO)*, volume 10173 of *Lecture Notes in Computer Science (LNCS)*, pages 654 – 669. Springer.

Wang, H., Ren, Y., Deutz, A. H., and Emmerich, M. T. M. (2017b). On Steering Dominated Points in Hypervolume Indicator Gradient Ascent for Bi-Objective Optimization. In Schütze, O., Trujillo, L., Legrand, P., and Maldonado, Y., editors, *NEO 2015: Results of the Numerical and Evolutionary Optimization Workshop*, volume 663 of *Studies in Computational Intelligence (SCI)*, pages 175 – 203. Springer.

Wessing, S. (2015). *Two-Stage Methods for Multimodal Optimization*. PhD thesis, Technische Universität Dortmund.

Wessing, S. (2016). *optproblems: Infrastructure to Define Optimization Problems and Some Test Problems for Black-Box Optimization*. python package version 0.9.

Zadorojniy, A., Masin, M., Greenberg, L., Shir, O. M., and Zeidner, L. (2012). Algorithms for Finding Maximum Diversity of Design Variables in Multi-Objective Optimization. *Procedia Computer Science*, 8(Supplement C):171 – 176. Special Issue on Conference on Systems Engineering Research.

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation Journal (ECJ)*, 8(2):173 – 195.