Multimodality in Multi-objective Optimization – More Boon than Bane?

Christian Grimme(✉), Pascal Kerschke, and Heike Trautmann

Information Systems and Statistics, ERCIS, University of Münster,
Leonardo-Campus 3, 48149 Münster, Germany
{christian.grimme,kerschke,trautmann}@uni-muenster.de

Abstract. This paper addresses multimodality of multi-objective (MO) optimization landscapes. Contrary to common perception of local optima, according to which they are hindering the progress of optimization algorithms, it will be shown that local efficient sets in a multi-objective setting can assist optimizers in finding global efficient sets. We use sophisticated visualization techniques, which rely on gradient field heatmaps, to highlight those insights into landscape characteristics. Finally, the MO local optimizer MOGSA is introduced, which exploits those observations by sliding down the multi-objective gradient hill and moving along the local efficient sets.

Keywords: Multi-objective optimization · Multimodality · Fitness landscapes · Basins of attraction · Local search · Gradients

1 Introduction

In single-objective (SO) continuous optimization, multimodality of the problem landscape is a crucial factor determining problem hardness. It is well-known that solvers might get trapped in local optima or at least require a large computational budget to (repeatedly) escape from the latter [22]. We will show that, counter-intuitively, MO optimizers do not necessarily face the same challenges. Contrarily, the existence of local efficient sets is potentially beneficial for sliding towards the global optimum along them. For this purpose a sophisticated visualization technique based on gradient field heatmaps, using the cumulated lengths of the normalized (approximated) gradients of both objectives towards the respective attracting local efficient set, is proposed. Respective figures reveal ridges and basins of attractions of local efficient sets - both in decision, as well as in objective space.

Basically, the gained insights can be exploited in two different ways: First, we pave the ground for retrieving as much information about the problems' landscape characteristics as possible with the potential of deriving informative exploratory landscape features for MO optimization problems, which is a rather new research field with only few results so far. Secondly, we introduce MOGSA, a (local) MO optimizer (MOO), which builds upon the straightforward idea of

K. Deb et al. (Eds.): EMO 2019, LNCS 11411, pp. 126–138, 2019.
https://doi.org/10.1007/978-3-030-12598-1_11

sliding down the MO gradient hill towards the global efficient set by exploiting properties of local efficient sets. Experiments show that common benchmark sets almost exclusively show similarities in enabling the algorithm to exploit the multimodal problem nature. MOGSA has the potential to even outperform competitive state-of-the-art MO optimizers and to be efficiently hybridized with other MOO approaches. Section 2 gives an overview of related work followed by a detailed description of our proposed visualization approach in Sect. 3. MOGSA is conceptually introduced in Sect. 4 together with preliminary experimental results. Conclusions are drawn in Sect. 5.

2 Related Work

SO continuous optimization insights are often directly transferred to MOO: *Multimodality implies the existence of traps for local search methods in SO, which prevent global convergence. Thus, multimodality in MO must be challenging for finding the global efficient set.* This assumption is the more astonishing, as almost no insights into the landscapes of continuous MO problems exist. It is restricted to very few general visualization techniques [29] and an early approach of [11]. These, however, only provide limited information on locality or landscape features. To the authors' best knowledge, only recent own work [18,20] provides insights into MO landscapes. Theoretical and empirical results [13,16,21] imply, that local optima in multimodal MOO are not necessarily traps for optimizers but following combined gradient directions can strongly support MOO. This is supported by many works (e.g., [14,24–26]), in which gradient-based directed search methods are applied to MOO problems. Note that for combinatorial MOO, analogies from SO landscape analysis (modality, ruggedness, correlation, and plateaus) are often used for making abstract problem features accessible [6,9,23,32]. However, due to the high dimensionality of combinatorial problems, a visual landscape representation is usually not possible or helpful.

3 A Gradient-Based Methodology for Visualizing Multi-objective Landscapes

Formal Preliminaries. Let $f : \mathcal{X} \to \mathbb{R}^p$ a vector valued function and define $x \in \mathcal{X}$ dominates $x' \in \mathcal{X} \Leftrightarrow f(x) \leq f(x') \wedge f(x) \neq f(x')$. The challenge of a MOO problem (MOP) is to find all points in $\mathcal{X}$ that are not dominated by other points in $\mathcal{X}$. This so-called *efficient set* $\mathcal{X}^*$ has an image $f(\mathcal{X}^*)$ w.r.t. all objectives which is called the *Pareto front*. In the following, we are considering unconstrained continuous minimization MOPs, that is $\mathcal{X} = \mathbb{R}^d$ for some dimension d.

We define two cases of MOP local optimality [13]. First, the usual definition of locality in neighborhood $B_\varepsilon(x) = \{x' \in \mathbb{R}^d | \, ||x - x'|| \leq \varepsilon\}$: A *local efficient point* $x \in \mathbb{R}^d$ is a point for which exists $\varepsilon > 0$ such that no $x' \in B_\varepsilon(x)$ dominates x. Further, $x \in \mathbb{R}^d$ is called a *strictly local efficient point*, if there exists $\varepsilon > 0$

such that x dominates all $x' \in B_\varepsilon(x) \setminus \{x\}$. We consider a set $A \subseteq \mathbb{R}^d$ to be *connected* iff there do not exist two open and *disjoint* subsets $U_1, U_2 \subseteq \mathbb{R}^d$ such that $A \subseteq (U_1 \cup U_2)$, $(U_1 \cap A) \neq \emptyset$, and $(U_2 \cap A) \neq \emptyset$. Further, let $B \subseteq \mathbb{R}^d$. A subset $C \subseteq B$ is a *connected component* of B iff $C \neq \emptyset$ is connected, and there exists no connected set D with $D \subseteq B$ such that $C \subset D$. Consequently, each connected component consisting of locally efficient points is called *local efficient set*.

The transfer of strict local efficiency to sets needs another definition. A local efficient set $\mathcal{X}_L$ is called *strictly local efficient set*, iff there exists an $\varepsilon > 0$ such that the environment set $E_\varepsilon(\mathcal{X}_L) = \{x' \in \mathbb{R}^d \setminus \mathcal{X}_L \mid \exists x \in \mathcal{X}_L : ||x - x'|| \leq \varepsilon\}$ is not empty and each point in $E_\varepsilon(\mathcal{X}_L)$ is dominated by at least one point in $\mathcal{X}_L$, see Fig. 1. The distinction between local and strict local efficient sets allows for a fine-grained analysis, where only the latter describes (rare) local traps for MOO algorithms.

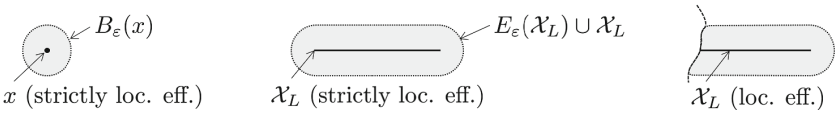

Fig. 1. Schematic examples illustrating the definitions of a strictly local efficient point (left) and set (center), as well as of a non-strictly local efficient set (right).

Idea of Gradient-Based Visualization for Multi-objective Landscapes. For visually relating decision and objective space in MOO, a utility function based on the aggregated MO gradient is proposed in [18] and a necessary condition in [17] leading to the following concept:

Let the objectives be continuously differentiable in $\mathbb{R}^d$ and $x^* \in \mathcal{X}$ be a local efficient point of $\mathcal{X}$. Then, there exists a vector $\nu \in \mathbb{R}^p$ with $0 \leq \nu_i$, $i = 1, \ldots, p$, and $\sum_{i=1}^{p} \nu_i = 1$, such that $\sum_{i=1}^{p} \nu_i \nabla f_i(x^*) = 0$. For a local efficient point and a suitable weighting vector, gradients for all objectives cancel each other out. In the special case of a bi-objective problem, gradients become anti-parallel and only differ in length. By normalizing the SO gradients, the then *normalized multi-objective gradient* becomes zero when a local efficient point is reached. Otherwise, the length and the direction of the normalized multi-objective gradient provides information on the attraction area and closeness of a local efficient point or set. For visualization purposes, we compute the discretized path of a given point to a (local) efficient point following the MO gradient direction. Therefore, the search space is divided into a grid of discrete points and then the combined bi-objective gradient is computed for each of the grid points. The accumulated length of the path towards the local efficient point is considered as utility value that determines the "height" of the respective decision vector.

If the MO gradient ∇f is unknown, it can be approximated as g by means of its SO gradients ∇f_j, $j = 1, \ldots, p$ via $g_j = \nabla f_j^{(t)} =$

$\sum_{i=1}^{d} \frac{f_j(x^{(t)}+\delta\cdot e_i)-f_j(x^{(t)}-\delta\cdot e_i)}{2\cdot\delta}$ $(*)$, where f_j is the j-th (single) objective, δ is a (small) step-size and e_i is the d-dimensional unit vector. Using MO gradient paths on all grid points, the MO landscape can be visualized in a two-dimensional heatmap as exemplarily shown in Fig. 2.

Visual Inspection of State-of-the-Art Benchmarks for MOP. First, we illustrate the concept for a highly multimodal MOP of the very recent *bi-objective black-box optimization benchmark* (BBOB) [4,30]: bi-objective problems are constructed by combining multimodal SO functions of the BBOB set [15].

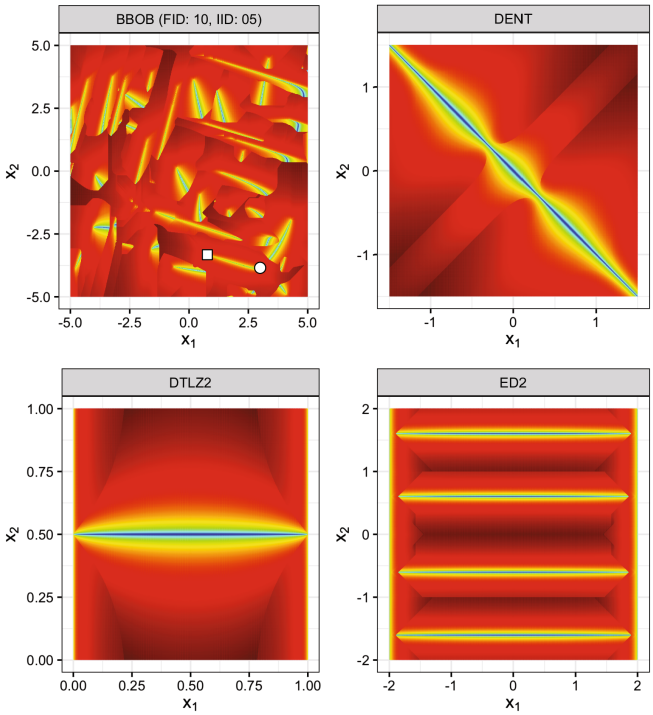

Fig. 2. Visualization of some exemplary and representative MOPs. The box and circle (top left sub-figure) denote the global SO optima for the SO functions the BBOB instance consists of. (Color figure online)

We also consider DENT [12] with a partly convex and concave Pareto-front, while the global efficient set has a rather simple structure. DTLZ2 [8], which has a completely concave Pareto-front, is focused followed by ED2 [10] similarly constructed as DTLZ2 but allowing specific adjustment of concavity and locality.

Figure 2 depicts the respective 2D decision space of the considered problems. Red colored areas denote a "long" distance (w.r.t. gradient descent) to the respective local efficient set. (Strictly) Local efficient sets are colored in blue and usually surrounded by green to yellow areas which denote "small" distances.

We observe known properties of the problems in decision space, like strong multimodality for the BBOB problem or rather simple structures of efficient sets for the remaining problem instances. Based on comprehensive visual inspections of all well-known MO benchmark sets, two crucial observations could be made:

Basins of attraction: We can visually inspect the basins of attraction for all local efficient sets based on the MO gradient on the discretized decision space. Each basin can be considered as funnel towards its local efficient set. We may imagine a "MO ball" following the MO gradient: then, the ball - rolling down this funnel - will finally reach the local efficient set.

Ridges due to superposition of basins: Basins of attraction superpose each other and as a consequence expose ridges that cut basin funnels and local efficient sets (see e.g., BBOB in Fig. 2). A superposition describes the abrupt change from one basin of attraction towards another one. Following definitions in Sect. 3, only local efficient sets are subject to superposition and ridges. Strict local efficient sets are - by definition - not superposed. Thus, non-strict local efficient sets offer a sure path towards the superposing basin of attraction (if we just walk along the local efficient set itself), while strictly efficient sets can be considered traps for gradient steered optimization.

Therefore, local efficient sets are not necessarily traps for gradient-based descent methods in MOO in contrast to the strict counterpart. In fact, not strict local efficient sets offer a path to neighbouring basins of attraction and can be exploited for finally reaching the global (strict) efficient set. Even more interesting, strict local efficient sets are (empirically) rare for the benchmarks we investigated so far (an example will be shown at the end of Sect. 4).

4 Exploiting Multimodality for Efficient Optimization

Superposition of basins of attraction results in ridges between adjacent basins, as well as abruptly cutted local efficient sets as schematically depicted in Fig. 3. We therefore propose a MOO algorithm, which exploits this superposed structure by "sliding" from basin to basin until it reaches a global efficient set.

An Optimization Algorithm that Slides Through Local Optima. MOGSA, a ***m****ulti-****o****bjective* ***g****radient* ***s****liding* ***a****lgorithm*, consists of (multiple repetitions of) the following two phases detailed below: (1) follow the MOP's MO gradient until a (locally or globally) efficient point was found, (2) explore the corresponding efficient set by following the gradients of the MOP's SO components.

Find a local (or global) efficient point: At first, MOGSA performs a local search by sliding down the MO gradient landscape as described in Algorithm 1. Given an initial individual[1], the MOP's SO gradients are approximated (line 2). Next,

[1] If no initial point is given, it will be sampled randomly within the search space.

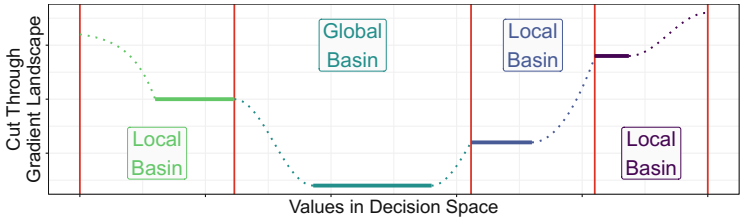

Fig. 3. Schematic view on the superimposed structure of basins of attraction. The vertical red lines represent the ridges distinguishing adjacent basins from each other, solid horizontal lines (within each basin) illustrate the respective efficient sets and the dotted lines represent the path towards the attracting efficient set. (Color figure online)

the MO gradient[2] at the current position is computed by summing up the normalized SO gradients (line 3). Note that the length of the individual's combined gradient already provides information on its location w.r.t. its attracting efficient set. If it is large (i.e., its length is close to two), both SO gradients point into similar directions, whereas a length of zero indicates opposing and hence offsetting gradients. In the latter case, the respective individual is regarded as being locally efficient. In order to account for numerical imprecision, MOGSA considers all individuals whose MO gradients have a length less than $\gamma = 10^{-6}$ as locally efficient and stops its downhill search once such a point was found (lines 4–6). Otherwise, it performs a gradient-descent step using the gradient length scaled by σ_1 (line 7). If it was too short (i.e., less than $\varepsilon = 10^{-6}$), the algorithm reached a dead end. Such a dead end could for instance occur, if MOGSA evaluates a point on the boundary of the feasible space and its gradient is pointing towards the infeasible area. In such a scenario, MOGSA leaves this dead end by restarting from an unexplored region of the MOP's feasible space (lines 8–10).

Once the algorithm has successfully performed three consecutive gradient steps (lines 11–18), it computes the angle between the three "youngest" individuals to detect whether the latter two are located on the same or opposite side of the attracting efficient set. In the former case, MOGSA continues its downhill search. However, in the latter scenario, i.e., $\boldsymbol{x}^{(t)}$ and $\boldsymbol{x}^{(t+1)}$ are located on opposite sides, it performs an interval bisection procedure, which exploits the individual's closeness to the efficient set in order to quickly converge to the efficient point. For this purpose, we modified the classical bisection method [5] such that the interval will be split according to the ratio of the lengths of the gradients, i.e., $x^{(t+2)} = x^{(t)} + (x^{(t+1)} - x^{(t)}) \cdot ||\boldsymbol{g}^{(t)}||/||\boldsymbol{g}^{(t+1)}||$, rather than simply at its center.

Explore Efficient Set: Once a locally efficient individual was found by Algorithm 1, it starts its exploration phase from there (see Algorithm 2). MOGSA computes the SO gradients (line 2) and follows the (normalized) gradient of the first objective scaled by σ_2 (line 4) as long as the step size is at least $\varepsilon = 10^{-6}$

[2] Note that the current implementation of MOGSA only enables the optimization of bi-objective problems.

Algorithm 1. Find local efficient point

1: **Require:**
   a) MOP $\boldsymbol{f} : \mathcal{X} \to \mathbb{R}^p$ with $\mathcal{X} \subseteq \mathbb{R}^d$,
   b) starting individual $\boldsymbol{x}^{(t)} \in \mathcal{X}$,
   c) step-size $\delta = 10^{-6}$ for grad.-approx.,
   d) maximum gradient length $\gamma = 10^{-6}$ of a locally efficient individual,
   e) scaling factor $\sigma_1 = 1$ for step-size,
   f) maximum difference $\varepsilon = 10^{-6}$ between individuals to be considered identical.
2: Approximate single-objective gradients $\boldsymbol{g}_j^{(t)}$, currently only for $j \in \{1, 2\}$, using Eqn. $(*)$
3: Combine normalized single-obj. gradients:
$$\boldsymbol{g}^{(t)} = \boldsymbol{g}_1^{(t)}/||\boldsymbol{g}_1^{(t)}|| + \boldsymbol{g}_2^{(t)}/||\boldsymbol{g}_2^{(t)}||$$
4: **If** $||\boldsymbol{g}^{(t)}|| < \gamma$ **then**
5: $\boldsymbol{x}^{(t)}$ is locally efficient $\rightsquigarrow$ exit algorithm
6: **end if**
7: Do gradient step: $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} + \sigma_1 \cdot \boldsymbol{g}^{(t)}$ (if $\boldsymbol{x}^{(t+1)} \notin \mathcal{X}$, place it on boundary)
8: **If** $||\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t+1)}|| \leq \varepsilon$ **then**
9: restart, i.e., draw an alternative $\boldsymbol{x}^{(t+1)}$ (using *Optimal Augmented Latin Hypercube Sampling*) and proceed to step 2
10: **end if**
11: **If** $\boldsymbol{x}^{(t+1)}$ is at least 3^{rd} element since last restart **then**
12: comp. $\omega = \angle(\boldsymbol{x}^{(t+1)} - \boldsymbol{x}^{(t)}, \boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)})$
13: **If** $\omega \leq 90°$ **then**
14: individuals are approaching efficient set from same side $\rightsquigarrow$ proceed to step 2
15: **else**
16: $\boldsymbol{x}^{(t)}$ and $\boldsymbol{x}^{(t+1)}$ are located on opposite sides of efficient set $\rightsquigarrow$ perform weighted interval bisection between them
17: **end if**
18: **end if**
19: **Return** archive of visited points

Algorithm 2. Explore efficient set

1: **Require:**
   a) MOP $\boldsymbol{f} : \mathcal{X} \to \mathbb{R}^p$ with $\mathcal{X} \subseteq \mathbb{R}^d$,
   b) starting individual $\boldsymbol{x}^{(0)} \in \mathcal{X}$,
   c) step-size $\delta = 10^{-6}$ for grad.-approx.,
   d) maximum length $\gamma = 10^{-6}$ of a (local efficient) individual's gradient,
   e) scaling factor $\sigma_2 = 1$ for step-size,
   f) maximum difference $\varepsilon = 10^{-6}$ between individuals to be considered identical.
2: **Initialize**, i.e., set $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(0)}$ and approx. single-obj. gradients $\boldsymbol{g}_j^{(t)}$ (for $j \in \{1, 2\}$)
3: **Explore set from $\boldsymbol{x}^{(0)}$ in direction of $\boldsymbol{g}_1$:**
4: $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} + \sigma_2 \cdot (\boldsymbol{g}_1^{(t)}/||\boldsymbol{g}_1^{(t)}||)$ (if $\boldsymbol{x}^{(t+1)} \notin \mathcal{X}$, place it on boundary)
5: **If** $||\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t+1)}|| \leq \varepsilon$ **then**
6: no step performed $\rightsquigarrow$ proceed to step 19
7: **end if**
8: Approx. $\boldsymbol{g}_1^{(t+1)}$ and $\boldsymbol{g}_2^{(t+1)}$ using Eqn. $(*)$
9: **If** $\left(||\boldsymbol{g}_1^{(t+1)}|| \leq \gamma\right)$ or $\left(||\boldsymbol{g}_2^{(t+1)}|| \leq \gamma\right)$
10: **then**
11: found single-objective optimum $\rightsquigarrow$ proceed to step 19
12: **end if**
13: compute angle $\alpha = \angle(\boldsymbol{g}_1^{(t)}, \boldsymbol{g}_1^{(t+1)})$
14: compute angle $\beta = \angle(\boldsymbol{g}_1^{(t+1)}, \boldsymbol{g}_2^{(t+1)})$
15: **If** $(\alpha > 90°)$ or $(\beta < 90°)$ **then**
16: left efficient set $\rightsquigarrow$ proceed to step 19
17: **end if**
18: still in efficient set $\rightsquigarrow$ proceed to step 4
19: **Explore set from $\boldsymbol{x}^{(0)}$ in direction of $\boldsymbol{g}_2$:**
20: analog to steps 4-18, but using exchanged gradients $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$
21: **Return** archive of visited points

(lines 5–7). These steps are repeated until MOGSA has reached the local optimum of the first objective (lines 9–12) or even left the efficient set (lines 13–17). The latter can have two reasons: (i) it left the efficient set, but remains in the same basin of attraction (indicated by an angle of more than 90° between two consecutive gradients of the first objective), or (ii) it left the basin of attraction and crossed the ridge to an adjacent basin (indicated by an angle of less than 90° between the two single-objective gradients in the current individual). Once MOGSA finished exploring one part of the efficient set (by following the first objective), it explores the set once more (starting in the same initial individual), but this time follows the second objective (lines 19–20). If at least one of the two exploration phases stopped because of a crossed ridge, the respective efficient set can not be globally efficient as it is apparently superimposed by another basin of attraction. In such a case, MOGSA again executes Algorithms 1 and 2 - starting

from the individual belonging to the adjacent, and thus more promising basin of attraction. However, in case neither end of the efficient set has been cut by a ridge, a strictly local efficient set - and thus likely a globally efficient Pareto set - was found.

Comparison of MOGSA and State-of-the-Art MOO Algorithms. Figure 4 illustrates the different search behaviors of MOGSA (top row), NSGA-II [7] (middle row) and SMS-EMOA [1] (bottom) with default parameter settings. The traces of their optimization paths are shown in the decision spaces of two exemplary MOPs: an instance of the bi-objective BBOB (left column) and DTLZ2 (right). For both problems, MOGSA was executed until it terminated

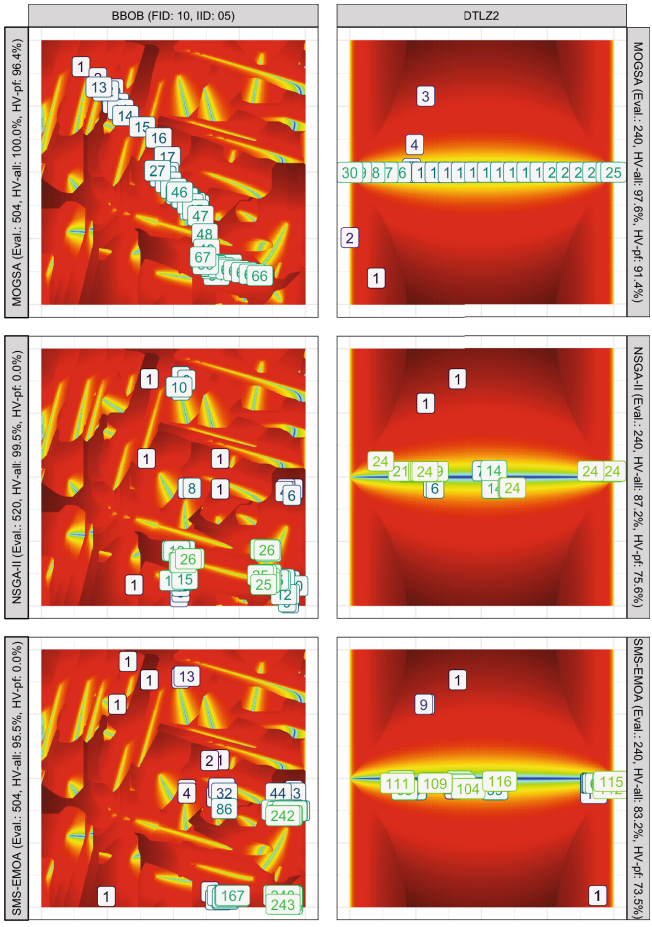

Fig. 4. Exemplary comparison of the search behavior of MOGSA (top row), NSGA-II (middle) and SMS-EMOA (bottom) in the search space of two popular MOPs: an instance from the bi-objective BBOB (FID 10, IID 5; left column) and DTLZ 2 (right).

successfully, i.e., after only 504 resp. 240 function evaluations including gradient approximation while its two contenders were then executed with the same budget. Note that the MOPs were created using the R-package `smoof` [3], and for the competing MOEAs (i.e., NSGA-II and SMS-EMOA) `ecr2` [2] was used.

The deceptive structure of DTLZ2 initially lured MOGSA towards its boundaries (see top right image of Fig. 4). However, once it reached the boundaries, it immediately restarted and quickly converged towards the global efficient set, which it then explored very efficiently. Given the rather small amount of function evaluations needed by MOGSA, we restricted the population sizes of NSGA-II and SMS-EMOA to $\mu = 5$ individuals such that they were able to run for a reasonable number of generations. Although both solvers approached the efficient set, neither of them was able to actually explore it nearly as precise or evenly distributed (in the decision space) as MOGSA. This is also supported by the corresponding covered hypervolumes (HV). As the latter strongly depend on their reference points, we provide two HV values per pair of optimizer and MOP. HV-all uses the nadir of *all* individuals from the archive of points evaluated when constructing the heatmap or by any of the three optimizers, whereas HV-pf uses the nadir based on all individuals along the (theoretical) Pareto front. All HV values are shown as ratios of the maximum achievable HV - i.e., the HV based on all individuals, which were used for identifying the nadir of HV-all.

Even in case of much more complex landscapes (left column of Fig. 4), MOGSA is able to successfully maneuver through the basins of attraction towards the Pareto set. Note that the depicted trace does not simply display a positive outlier. Out of ten runs, in which we executed MOGSA from ten randomly chosen starting points, MOGSA outperformed its contenders - within the considered budget - in nine (NSGA-II) and ten runs (SMS-EMOA), respectively.

5 Discussion and Conclusion

Our proposed gradient field heatmaps visualize interaction effects among different objectives in the search space providing a new perspective on popular benchmark problems by revealing interesting properties such as basins of attraction (similar to the idea of cell mapping, see, e.g., [19]), ridges between them, local and/or global efficient sets, etc. This results in important insights into the structure of MO landscapes, which in turn can be used for algorithm design - as we successfully demonstrated with the design of MOGSA in Sect. 4.

Note that certain structures found in search space (such as the attraction basins) are also visible in the objective space. As a result, our method is not restricted to MOPs, whose search *and* objective space both are 2D. Instead, it allows to illustrate any MOP for which at least *one* of the two spaces is 2D.

As indicated earlier, there exist MOPs, whose landscapes contain local optima that are true MO traps. However, throughout our experiments with more than a hundred of different MOPs, we only encountered one single landscape (displayed in Fig. 5), which displayed such a “malign” strictly local efficient set. Despite possessing only *one* Pareto front in the objective space, it contains

two cut-free efficient sets in the decision space. The *local* front - spanning from the pink square to the cyan circle - is entirely dominated by the Pareto front. However, as both ends of the local front are nondominated in its close neighborhood, MOGSA is unable to leave it towards an improving attractor. In future work, we intend to analyze (a) the causes for such traps, and (b) whether real-world applications (in contrast to artificial problems) possess such traps more often.

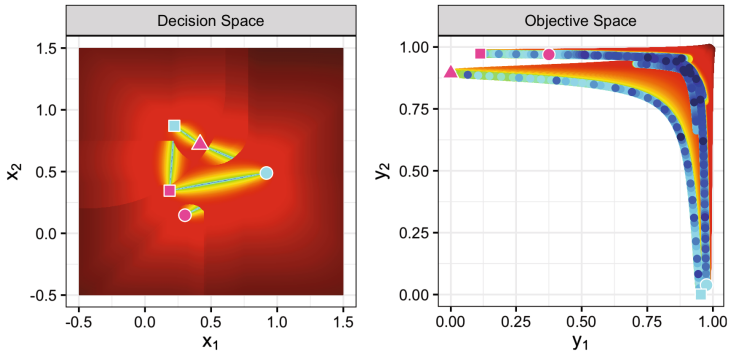

Fig. 5. Example of a MOP with a true MO trap (ranging from pink square to cyan circle). The problem combines two multi-sphere problems created using MPM2 [33]. (Color figure online)

Within our exploratory proof-of-concept experiments, MOGSA reached the global Pareto front much faster in terms of function evaluations - despite the costs for *approximating* the gradient - than NSGA-II and SMS-EMOA. This effect would increase even further, if the exact gradient was accessible. Without a doubt MOGSA's performance strongly depends on its parametrization - especially on its step-size related factors. If they are set too high, the algorithm jumps across entire basins, whereas short steps increase the number of iterations and thus function evaluations. Therefore, a future aspect is the development of sophisticated step-size adaptation mechanisms, which ideally adjust the parameters automatically to the landscape at hand.

Given a reasonable parametrization, MOGSA converges rather quickly and hence only consumes a small part of the budget. To further improve the algorithm's performance, one could invest additional budget into further runs of the optimizer as these allow to (a) avoid running into MO traps, and (b) detect different parts of the global Pareto front (if existent). Given the deterministic search behavior, one should restart from unexplored regions as MOGSA would otherwise quickly move into already visited basins. The chances of starting "far away" from previously seen areas, can be increased by applying sophisticated sampling mechanisms such as *optimal augmented latin hypercube sampling* [28].

Currently, the results were only shown for 2D problems, however, the algorithm's concept is transferable to higher dimensional search and objective spaces.

Adapting it to larger search spaces is straightforward as this simply requires the gradient approximation in further dimensions - resulting in two additional function evaluations per individual and search space dimension. In case of larger objective spaces, it might be beneficial to first optimize w.r.t. two objectives and once a strictly local efficient set was found, MOGSA could travel along the adjacent fronts (by following different objectives) until a cut-free simplex of global efficient points (comprising the global optimum) is found. In order to keep the optimizer competitive in larger search spaces - and hence diminishing the effect of the curse of dimensionality - future work could deal with alternative, i.e., cheaper, approaches for approximating the gradients [27].

Although there have been numerous works in the past, which also make use of gradients for MO optimization purposes, MOGSA is the first one that explicitly exploits the problem's multimodal structure from a search space point of view.

As our proposed algorithm is a local search algorithm, it might be promising to hybridize it with other EMOAs or - given that MOGSA focuses on the search space - with optimizers that are able of walking along the fronts in the objective space [25]. Moreover, investigating our findings from a theoretical point of view is of central interest for future work.

Our work revealed multiple further open issues w.r.t. benchmarking, problem characterization, and algorithm selection. For instance, all existing MO benchmarks should be compared thoroughly (e.g., using our visualization approach). The gained insights can then be used to group the MOPs in an appropriate way, similar to the five groups of BBOB [15]. Once important properties of MO landscapes have been identified, landscape features (see, e.g., [31] or [20]), which (a) 'measure' these different properties, and (b) can later on be used to perform algorithm selection on a portfolio of complementary, powerful MO algorithms, can be designed. Such a complementary portfolio requires an extensive benchmark of competitive state-of-the-art algorithms on a manifold of MOPs.

References

1. Beume, N., Naujoks, B., Emmerich, M.T.M.: SMS-EMOA: multiobjective selection based on dominated hypervolume. EJOR **181**(3), 1653–1669 (2007)
2. Bossek, J.: ecr 2.0: a modular framework for evolutionary computation in R. In: Proceedings of GECCO Companion, pp. 1187–1193. ACM (2017)
3. Bossek, J.: smoof: single- and multi-objective optimization test functions. R J. (2017). https://journal.r-project.org/archive/2017/RJ-2017-004/
4. Brockhoff, D., Tran, T.D., Hansen, N.: Benchmarking numerical multiobjective optimizers revisited. In: Proceedings of GECCO, pp. 639–646. ACM (2015)
5. Burden, R.L., Faires, D.J.: Numeric Analysis, 3rd edn. Prindle, Weber & Schmidt Publishing Company, Boston (1985)
6. Daolio, F., Liefooghe, A., Verel, S., Aguirre, H.E., Tanaka, K.: Global vs local search on multi-objective NK-landscapes: contrasting the impact of problem features. In: Proceedings of GECCO, pp. 369–376. ACM (2015)
7. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE TEVC **6**(2), 182–197 (2002)

8. Deb, K., Thiele, L., Laumanns, M., Zitzler, E.: Scalable test problems for evolutionary multiobjective optimization. In: Abraham, A., Jain, L., Goldberg, R. (eds.) Evolutionary Multiobjective Optimization, pp. 105–145. Springer, London (2005)
9. Ehrgott, M., Klamroth, K.: Connectedness of efficient solutions in multiple criteria combinatorial optimization. EJOR **97**(1), 159–166 (1997)
10. Emmerich, M.T.M., Deutz, A.H.: Test problems based on Lamé superspheres. In: Obayashi, S., Deb, K., Poloni, C., Hiroyasu, T., Murata, T. (eds.) EMO 2007. LNCS, vol. 4403, pp. 922–936. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-70928-2_68
11. da Fonseca, C.M.M.: Multiobjective genetic algorithms with application to control engineering problems. Ph.D. thesis, University of Sheffield (1995)
12. Gerstl, K., Rudolph, G., Schtze, O., Trautmann, H.: Finding evenly spaced fronts for multiobjective control via averaging Hausdorff-measure. In: 2011 8th International Conference on Electrical Engineering, Computing Science and Automatic Control, pp. 1–6 (2011). https://doi.org/10.1109/ICEEE.2011.6106656
13. Grimme, C., Kerschke, P., Emmerich, M.T.M., Preuss, M., Deutz, A.H., Trautmann, H.: Sliding to the global optimum: how to benefit from non-global optima in multimodal multi-objective optimization. In: Proceedings of LeGO (2018, accepted)
14. Grimme, C., Lepping, J., Papaspyrou, A.: Adapting to the habitat: on the integration of local search into the predator-prey model. In: Ehrgott, M., Fonseca, C.M., Gandibleux, X., Hao, J.-K., Sevaux, M. (eds.) EMO 2009. LNCS, vol. 5467, pp. 510–524. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01020-0_40
15. Hansen, N., Finck, S., Ros, R., Auger, A.: Real-parameter black-box optimization benchmarking 2009: noiseless functions definitions. Technical report, INRIA (2009)
16. Jin, Y., Sendhoff, B.: Connectedness, regularity and the success of local search in evolutionary multi-objective optimization. In: Proceedings of the IEEE CEC, vol. 3, pp. 1910–1917. IEEE (2003)
17. John, F.: Extremum problems with inequalities as subsidiary conditions. In: Studies and Essays, Courant Anniversary Volume, pp. 187–204. Interscience (1948)
18. Kerschke, P., Grimme, C.: An expedition to multimodal multi-objective optimization landscapes. In: Trautmann, H., et al. (eds.) EMO 2017. LNCS, vol. 10173, pp. 329–343. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54157-0_23
19. Kerschke, P., et al.: Cell mapping techniques for exploratory landscape analysis. In: Tantar, A.-A., et al. (eds.) EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V. AISC, vol. 288, pp. 115–131. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07494-8_9
20. Kerschke, P., et al.: Towards analyzing multimodality of continuous multiobjective landscapes. In: Handl, J., Hart, E., Lewis, P.R., López-Ibáñez, M., Ochoa, G., Paechter, B. (eds.) PPSN 2016. LNCS, vol. 9921, pp. 962–972. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45823-6_90
21. Kerschke, P., et al.: Search dynamics on multimodal multi-objective problems. Evol. Comput. 1–33 (2018). https://doi.org/10.1162/evco_a_00234
22. Preuss, M.: Multimodal Optimization by Means of Evolutionary Algorithms. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-07407-8. https://www.springer.com/de/book/9783319074061
23. Rosenthal, S., Borschbach, M.: A concept for real-valued multi-objective landscape analysis characterizing two biochemical optimization problems. In: Mora, A.M., Squillero, G. (eds.) EvoApplications 2015. LNCS, vol. 9028, pp. 897–909. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16549-3_72

24. Schütze, O., Hernández, V.A., Trautmann, H., Rudolph, G.: The hypervolume based directed search method for multi-objective optimization problems. J. Heuristics **22**(3), 273–300 (2016)
25. Schütze, O., Martín, A., Lara, A., Alvarado, S., Salinas, E., Coello, C.A.: The directed search method for multi-objective memetic algorithms. Comput. Optim. Appl. **63**(2), 305–332 (2016)
26. Schütze, O., Sanchez, G., Coello Coello, C.A.: A new memetic strategy for the numerical treatment of multi-objective optimization problems. In: Proceedings of GECCO, pp. 705–712. ACM (2008)
27. Spall, J.C.: Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Trans. Autom. Control **37**(3), 332–341 (1992)
28. Stein, M.: Large sample properties of simulations using latin hypercube sampling. Technometrics **29**, 143–151 (1987)
29. Tušar, T., Filipič, B.: Visualization of Pareto front approximations in evolutionary multiobjective optimization: a critical review and the prosection method. IEEE TEVC **19**(2), 225–245 (2015)
30. Tušar, T., Brockhoff, D., Hansen, N., Auger, A.: COCO: the bi-objective black box optimization benchmarking (bbob-biobj) test suite. arXiv preprint (2016)
31. Ulrich, T., Bader, J., Thiele, L.: Defining and optimizing indicator-based diversity measures in multiobjective search. In: Schaefer, R., Cotta, C., Kołodziej, J., Rudolph, G. (eds.) PPSN 2010. LNCS, vol. 6238, pp. 707–717. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15844-5_71
32. Verel, S., Liefooghe, A., Jourdan, L., Dhaenens, C.: On the structure of multiobjective combinatorial search space: MNK-landscapes with correlated objectives. Eur. J. Oper. Res. **227**(2), 331–342 (2013)
33. Wessing, S.: Two-stage methods for multimodal optimization. Ph.D. thesis, Technische Universität Dortmund (2015). http://hdl.handle.net/2003/34148