

# Effects of 1-Greedy $\mathcal{S}$ -Metric-Selection on Innumerably Large Pareto Fronts

Nicola Beume<sup>1</sup>, Boris Naujoks<sup>1</sup>, Mike Preuss<sup>1</sup>,  
Günter Rudolph<sup>1</sup>, and Tobias Wagner<sup>2</sup>

<sup>1</sup> Department of Computer Science (LS11), TU Dortmund University, Germany  
{nicola.beume,boris.naujoks,mike.preuss,guenter.rudolph}@tu-dortmund.de

<sup>2</sup> Institute of Machining Technology (ISF), TU Dortmund University, Germany  
wagner@isf.de

**Abstract.** Evolutionary multi-objective algorithms (EMOA) using performance indicators for the selection of individuals have turned out to be a successful technique for multi-objective problems. Especially, the selection based on the  $\mathcal{S}$ -metric, as implemented in the SMS-EMOA, seems to be effective. A special feature of this EMOA is the greedy  $(\mu + 1)$  selection. Based on a pathological example for a population of size two and a discrete Pareto front it has been proven that a  $(\mu + 1)$ - (or 1-greedy) EMOA may fail in finding a population maximizing the  $\mathcal{S}$ -metric. This work investigates the performance of  $(\mu + 1)$ -EMOA with small fixed-size populations on Pareto fronts of innumerable size. We prove that an optimal distribution of points can always be achieved on linear Pareto fronts. Empirical studies support the conjecture that this also holds for convex and concave Pareto fronts, but not for continuous shapes in general. Furthermore, the pathological example is generalized to a continuous objective space and it is demonstrated that also  $(\mu + k)$ -EMOA are not able to robustly detect the globally optimal distribution.

## 1 Introduction

The main question addressed in this work is concerned with the general suitability of a 1-greedy evolutionary multi-objective algorithm (EMOA) for the approximation of continuous Pareto fronts, which consist of an innumerable number of Pareto optimal solutions. As a 1-greedy EMOA, we denote a steady-state  $(\mu + 1)$ -EMOA that replaces only one individual by greedily selecting the  $\mu$  best ones according to a preference relation (in the style of the definitions of  $k$ -greediness by Zitzler et al. [ZTB08]). The question is of special interest, since—for some time now—we advocate the use of an EMOA that adheres to the 1-greedy scheme using the  $\mathcal{S}$ -metric or dominated hypervolume as preference relation, namely the SMS-EMOA [BNE07]. In contrast to other EMOA (e.g. NSGA-II [DPAM02]), it accepts only one new individual per generation in order to monotonically improve the quality of the Pareto front approximation. Naturally, one can ask if exchanging only one individual at a time is sufficient to avoid getting stuck in non-optimal configurations. However, past experience with the SMS-EMOA has

nourished the belief that this algorithm is capable of coping with all practically relevant situations, although a general proof in either direction is missing. This kind of general proof, even if restricted to continuous Pareto fronts, is sophisticated, unless impossible. Thus, the aim of this paper is to gradually phase this task using both case-related formal proofs and empirical studies.

Recently, a simple discrete counter-example has been provided, which proved that the 1-greedy scheme based on the dominated hypervolume can fail [ZTB08] (cf. Sec. 2.3). However, the example is extreme in many aspects: It employs a population of only two individuals on a Pareto front of four points. Thus, we would like to know to what extent this phenomenon occurs in more realistic scenarios. We are interested in continuous Pareto fronts and show that the discrete counter example can easily be extended into the (piecewise) continuous domain, with the essential property still holding: For most initializations, a 1-greedy EMOA will fail to obtain the optimal distribution of points on that Pareto front. To further investigate the cause of failure, we optimize the  $\mathcal{S}$ -metric value of the population directly using a (1, 5)- and a (5, 10)-CMA-ES (Covariance Matrix Adaptation Evolution Strategy [HO01]). Our results show that not only 1-greedy EMOA, but also non-elitist EMOA with  $\lambda > \mu$  can fail with high probability. This indicates that the problem is indeed very hard since the local optimum is a strong attractor for any kind of optimizer. Studying the structure of the problem, we give generalizing conjectures on the interrelationship of the Pareto front and greediness.

Given the successful applications and assuming that the failure on the mentioned counter examples stem from the extreme constitution of the Pareto front, we investigate the properties on connected simpler shapes. For linear Pareto fronts, it is proved that a 1-greedy hypervolume selection scheme is sufficient to reach the optimal distribution of points with respect to the dominated hypervolume. Regarding convex Pareto fronts, we show that the problem of maximizing the hypervolume with a given number of points is not concave, otherwise the 1-greediness would hold directly (cf. Sec. 4.1). However, the concavity is not a necessary condition for 1-greediness. We perform empirical studies on Pareto fronts of different curvature, which demonstrate that even a simplified SMS-EMOA reaches the global optimum showing that the problem is solvable by a 1-greedy EMOA. Furthermore, these studies give counter-intuitive insights on the optimal distributions of the points and their corresponding hypervolume contributions, i.e., the amount that is disjointly dominated by a point and is lost when the point is removed [BNE07].

The paper is structured as follows. In section 2 the basic definitions, which are used in this paper, are provided and the discrete pathological example for 1-greedy indicator-based EMOA is recapitulated. The continuous variant of the example is derived and, together with another problem with disconnected fronts, empirically studied in section 3. Afterwards, we focus on continuous Pareto fronts by analyzing and empirically studying connected fronts of different curvature in section 4. For simple cases, also formal proofs are provided. Finally, the paper is summarized and the important results are concluded in section 5.

## 2 Hypervolume Selection and Greediness

### 2.1 Definitions of Greediness

Zitzler et al. [ZTB08] denote a preference relation as *k-greedy* if

1. for any given set, there exists a finite number of iterations resulting in the optimal set regarding the preference relation, and
2. there is a sequence of improving populations per iteration when exchanging  $k$  elements of a population at most.

We denote an EMOA as *greedy* if the selection is performed greedily according to a preference relation, i.e., the best population regarding the preference is selected. Greediness implies elitist selection and  $k$  is related to the number of offspring, i.e., the number of possible changes in the population per iteration. Thus, a *k-greedy* EMOA performs a  $(\mu + k)$  selection scheme regarding a pre-defined preference relation. The problem of finding a population for a given optimization problem which is optimally composed regarding the indicator of the preference relation is termed *k-greedy solvable* if from any initial population there exists an improving path to the optimum which can be traversed by changing at most  $k$  element of the population per iteration. Note that the selection allows that any problem is  $\mu$ -greedy solvable for a  $(\mu + k)$ -EMOA with  $k \geq \mu$  assuming that all search points are sampled with positive probability. A problem is *local k-greedy solvable* if the optimum can be obtained by exchanging only with neighboring solutions in the objective space. If a problem is local  $k$ -greedy solvable, this implies that it also is  $k$ -greedy solvable. We concentrate on 1-greediness regarding the hypervolume indicator thus study if the population which achieves the highest possible hypervolume value given a fixed population size and a reference of the  $\mathcal{S}$ -metric is reachable by replacing at most one individual per generation by selecting the subset of size  $\mu$  which obtains the highest  $\mathcal{S}$ -metric value among all those  $\mu + 1$  subsets.

### 2.2 Considered Test Functions

For the experimental investigation of greedy EMOA, a set of academic minimization problems is considered. This set contains the simple test functions T1-T4, which have a continuous concave or convex Pareto front, where the sign of the second derivative with respect to the first objective does not change. Test function T5 changes its curvature from concave to convex, while still being connected and continuously differentiable. Note that T1 and T4 describe the same Pareto front. The decision variable  $x \in [0, 1]$  is bounded to an interval, in which  $f_1(x)$  increases and  $f_2(x)$  decreases in order to allow only non-dominated individuals.

T1: $f_1(x) = x^2, \quad f_2 = (1 - x)^2$	Schaffer [Sch85], convex
T2: $f_1(x) = x, \quad f_2 = 1 - x$	DTLZ1 [DTLZ02], convex
T3: $f_1(x) = \sin((\pi/2)x), \quad f_2 = \cos((\pi/2)x)$	DTLZ2 [DTLZ02], concave
T4: $f_1(x) = x, \quad f_2(x) = (1 - x^{1/\alpha})^\alpha, \quad \alpha = 2$	convex

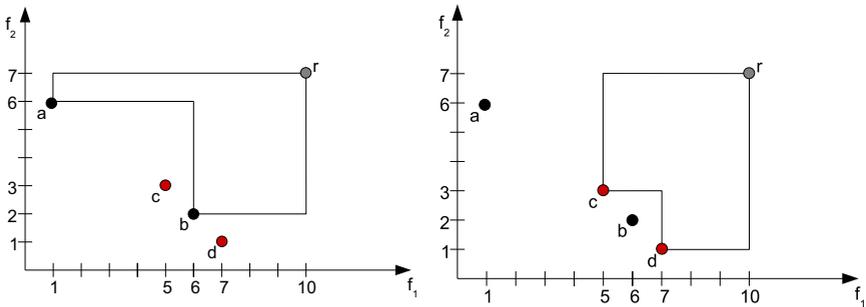
- T5:  $f_1(x) = x, \quad f_2(x) = (1 - x^{1/\alpha})^\alpha, \quad \alpha = 3x/2 + 1/2$  concave-convex  
 T6:  $f_1(x) = x, \quad f_2(x) = \begin{cases} -(1/8)x + 6.125 & x < 5 \\ -x + 8 & x \geq 5 \end{cases}$   
 T7:  $f_1(x) = x, \quad f_2(x) = 1 - \sqrt{x} - x \cdot \sin(10\pi x)$  ZDT3 [ZT98]

For the experiments on T1-T4, the reference point applied in the selection of the EMOA is fixed to  $\mathbf{R} = (2, 2)^T$  (superscript  $T$  denotes transposition). The two test functions T6 and T7 are multi-modal with respect to the hypervolume of the population. T6 is the continuous conversion of a pathological example given by Zitzler et al. [ZTB08] (cf. Sec. 2.3). Its decision variable  $x \in [1, 7]$  is bounded and  $\mathbf{R} = (10, 7)^T$  is used. T7 is defined in the domain of  $x \in [0, 1]$  being the only function, for which not all  $x$  are Pareto optimal leading to a disconnected Pareto front of five convex parts.

### 2.3 Pathological Example for a Finite Pareto Front

Zitzler et al. [ZTB08] proved that, in general, a 1-greedy EMOA is not able to obtain the set which covers the maximal dominated hypervolume. They showed this by a counter example in a two-dimensional objective space with a Pareto front consisting of four points as reproduced in Fig. 1, where the algorithm shall optimize the distribution of a population of two individuals. When the population is initialized with the two points  $a$  and  $b$ , the global optimum formed by the points  $c$  and  $d$  is unreachable for a 1-greedy EMOA. Any combination of either  $a$  or  $b$  with a different point leads to a worse hypervolume value and is therefore not accepted. Thus, the set  $\{a, b\}$  is a local optimum of the hypervolume maximization. The example can easily be extended to a higher number of objectives by choosing all additional coordinates as 1, since multiplying by 1 does not change the hypervolume values.

Note which aspects are necessary to make the problem hard: The reference point is chosen such that the objective values are weighted asymmetrically. Thus, the points on the right have a high hypervolume contribution though being quite close to each other. Furthermore, the second coordinate of the point  $a$  is

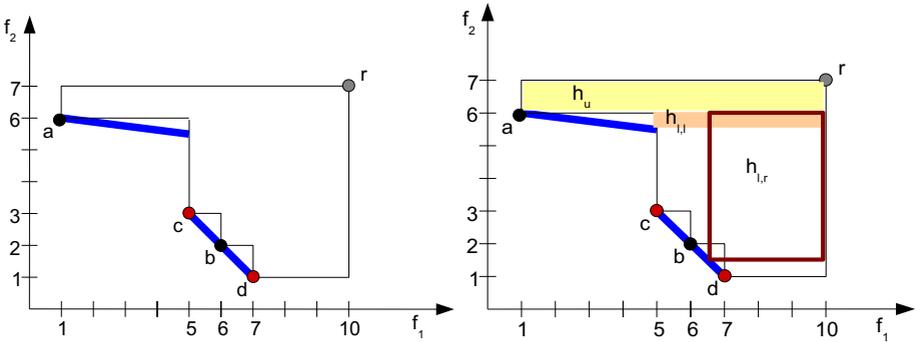


**Fig. 1.** Pathological example for 1-greedy EMOA with hypervolume-based selection. Points  $c$  and  $d$  are optimal but the population is initialized with  $a$  and  $b$  which form a local optimum.

positioned close enough to the reference point to avoid an optimal distribution which includes this point.

### 3 Pathological Examples with Innumerable Pareto Fronts

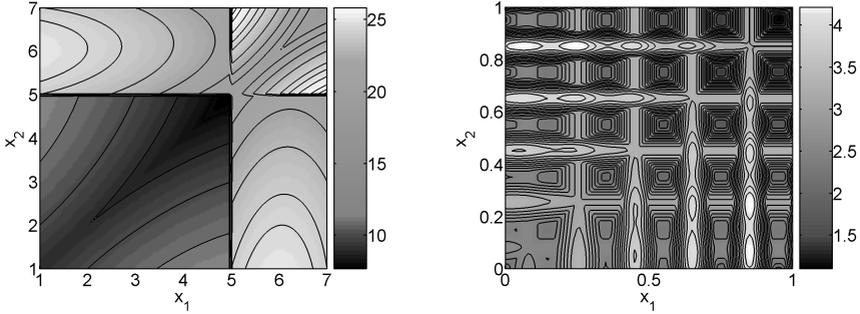
The discrete example of section 2.3 may easily be extended to the (piecewise) continuous case by connecting the points of the original configuration by two line segments as shown in Fig. 2 (cf. T6 in Sec. 2.2). The slope of the right segment results in  $m_2 = -1$  and for the left one  $m_1 = -1/8$  is chosen, to correctly transfer the situation of the discrete case in terms of the optimality properties of the different point distributions. For  $m_1 < -0.2$ , point  $a$  is no longer part of a local optimum and the basin of attraction is shifted to the right. For reasons of simplification, we further on discuss the problem as a two-dimensional parameter optimization problem, whose parameters are the two  $x$ -coordinates of the two search points on the Pareto front.



**Fig. 2.** Left: Conversion of the discrete pathological example for 1-greedy EMOA to the continuous space with a disconnected Pareto front of two linear segments (T6). Still the points  $c$  and  $d$  are optimal, whereas  $a$  and  $b$  form a local optimum. Right: Dissection of the hypervolume for one point fixed at  $a$  and one moving on either the left ( $h_{l,l}$ ) or right ( $h_{l,r}$ ) line (cf. Sec. 3.1 for details).

In the sense of parameter optimization, one may speak of multi-modality, and the property of 1-greediness translates to the possibility to execute a successful line search parallel to the coordinate axes. Thus, we can have multi-modality while still being able to do a successful step out of a local optimum by moving in parallel to one of the coordinate axes towards a better point. This especially is the case for multi-modal but separable hypervolume landscapes, such as shown for the problem T7 in the right plot of Fig. 3, where it is often necessary to cross large areas of worse values, so that the function is not local 1-greedy solvable.

The hypervolume landscape of problem T6 is depicted in the left plot of Fig. 3. The contour lines indicate that a 1-greedy EMOA is not able to leave the local optimum. A  $\mu$ -greedy scheme would not encounter this principal difficulty as it allows for steps in any direction. However, it also faces the problem of locating



**Fig. 3.** Hypervolume landscapes for two individuals on the disconnected test functions T6 (left,  $\mathbf{R} = (10, 7)^T$ ) and T7 (right,  $\mathbf{R} = (2, 2)^T$ )

a good area, which may be also difficult. The analysis of both 1-greedy and  $\mu$ -greedy approaches on T6 is performed in the next subsections.

### 3.1 Proof: T6 Is Not 1-Greedy Solvable Regarding the $\mathcal{S}$ -Metric

In the following, we denote the two points of our population by their  $x$ -coordinates so that a population is  $x = (x_1, x_2)$ . It can be shown that for T6, the global optimum at  $x = (5, 7)$  can indeed not be reached by doing only 1-greedy steps from the starting point  $x = (1, 6)$ . To accomplish this, we have to look at the different cases resulting from fixing one point and moving the other over the allowed interval. In each case, the resulting hypervolume shall not exceed 25, which is the value for  $x = (1, 6)$ . Otherwise, a 1-greedy successful step would be possible.

Let the population start with  $x_1 = 1$  and  $x_2 = 6$ , thus the locations  $a$  and  $b$ . If  $x_1$  stays at  $x_1 = 1$  then  $x_2$  can either move on the right or the left line. We can always compute the total hypervolume as sum of the volume beyond dominated by  $(x_1 = 1, 6)$  ( $h_u = 9$ , cf. Fig. 2, left), and the contribution of the point at  $x_2$ . For  $x_2$  moving on the right line, the contribution  $h_{l,r}(x_2)$  is:

$$h_{l,r}(x_2) = (10 - x_2) (6 - (-x_2 + 8)) = -x_2^2 + 12x_2 - 20. \quad (1)$$

The maximum of this upside-down parabola obtained by standard calculus is at  $x_2 = 6$ . That is, for moving on the right line, there is no better point than  $x_2 = 6$ , which leads to a total hypervolume of 25. If  $x_2$  moves on the left line segment, its contribution  $h_{l,l}(x_2)$  is:

$$h_{l,l}(x_2) = (10 - x_2) \left( 6 - \left( -\frac{x_2}{8} + \frac{6}{8} \right) \right) \quad (2)$$

This negative parabola has its optimum at  $x_2 = \frac{11}{2}$ , which is unreachable in our scenario as the largest  $x_2$  value is still below 5. The resulting total hypervolume for this case is, thus, smaller than 13. When fixing  $x_2 = 6$ , and moving  $x_1$ , we can either situate it on the left or the right line segment. For the former, the whole hypervolume evaluates to:

$$h_l(x_1) = (10 - x_1) \left( 7 - \left( -\frac{x_1}{8} + \frac{6}{8} \right) \right) + 4 \left( \frac{6}{8} \right) - \frac{x_1}{8} - 2 = -\frac{x_1^2}{8} - \frac{x_1}{8} + \frac{25}{4} \quad (3)$$

This parabola has its maximum at  $x_1 = -\frac{1}{2}$ , the best allowed point is  $x_1 = 1$ , for which the hypervolume is 25. The latter case considers  $x_1$  moving on the right line segment. Here, we define a helper function  $h_{r,r}(y, z)$ , which computes the hypervolume of any two points on that line, using the fact that the slope is  $m_2 = -1$  so that the desired value is the difference of a large rectangle through both points and the reference point and a small rectangle with both points as diagonal corners:

$$h_{r,r}(y, z) = (10 - y)(7 - (z + 8)) - (z - y)^2 = -y^2 - z^2 + yz + y + 10z - 10. \quad (4)$$

For  $x_1 < 6$  ( $h_{r,r}(x_1, 6)$ ) or respectively  $x_1 > 6$  ( $h_{r,r}(6, x_1)$ ) this leads to

$$h_{r,r}(x_1, 6) = -x_1^2 + 7x_1 + 14 \quad \text{and} \quad h_{r,r}(6, x_1) = -x_1^2 + 16x_1 - 40 \quad (5)$$

with the maximum values at

$$\arg \max_{x_1} (h_{r,r}(x_1, 6)) = \frac{3}{2} \quad \text{and} \quad \arg \max_{x_1} (h_{r,r}(6, x_1)) = 8$$

with corresponding largest attainable values  $h_{r,r}(5, 6) = 24$  and  $h_{r,r}(6, 7) = 23$ . Consequently, there is no 1-greedy move from  $x = (1, 6)^T$  resulting in at least the same hypervolume value of 25.  $\square$

### 3.2 Experiment: How $\mu$ -Greedy Solvable Is T6?

*Pre-experimental planning:* We consider a 1-greedy and a  $\mu$ -greedy single-objective evolutionary algorithm (EA) moving on the Pareto front only (resembling, e.g., SMS-EMOA and NSGA-II). So, the search space is the Pareto set and the EA directly maximize the  $\mathcal{S}$ -metric value of the population. A (1, 5)- and a (5, 10)-CMA-ES are added to the set of algorithms. These do not have existing EMOA counterparts, but shall be tested to see if moving with even more degrees of freedom (non-elitist selection and a surplus of offspring) pays off. In our first runs, we observed that the standard set of termination criteria as well as standard boundary treatment (by quadratic penalties) deteriorate the performance of the CMA-ES. The termination criteria make it stop too early, when there is still a good chance to obtain the optimal solution of  $x = (5, 7)$ , and the boundary treatment hinders coming near to it. Both have been switched off hereafter.

*Task:* We expect that the  $\mu$ -greedy EA performs significantly better than the 1-greedy EA in terms of success rates.

*Setup:* All four algorithms are run 100 times per mutation step size (0.1 and 0.5) allowing up to 5000 evaluations and a minimum hypervolume value of 25.9 is regarded as success. The start points are scattered uniformly at random over the allowed domain (1 to 7).

**Table 1.** Success rates (100 repeats) of different algorithm types for detecting the globally optimal distribution of two individuals on the T6 Pareto front. The initial mutation step size as the only free parameter is tested at 0.1 and 0.5. Only the CMA-ES variants adapt it through the run.

Mutation step size	1-greedy EA	$\mu$ -greedy EA	(1, 5)-CMA-ES	(5, 10)-CMA-ES
0.1	12%	15%	55%	100%
0.5	15%	13%	61%	100%

*Results/Visualization:* The results are given in Table 1 by means of success rates.

*Observations:* Table 1 documents that the 1-greedy EA indeed fails, but so also does the  $\mu$ -greedy EA. The CMA-ES solves the problem in more than half of the runs. The effective run length (until stagnation) is very short for the 1-greedy and  $\mu$ -greedy EA, usually below 1000 evaluations. The CMA-ES often takes much longer. At the same time, it can be observed that it pushes the internally adapted mutation step sizes to very high values.

*Discussion:* The most surprising fact is surely that also the  $\mu$ -greedy EA fails. It seems that the small basin of the global optimum is hard to find, even if it is possible to move there. A larger mutation step size could help in jumping out of the vicinity of the local optimum, but it also scatters search steps over a larger area. Furthermore, the attractor at (1, 6) is much stronger than expected. Most runs end here, even if started at far distant points. The CMA-ES uses a very interesting strategy by enlarging the mutation rates. It is finally able to generate offspring over the whole domain of the problem, thereby degenerating (by learning) to a random search. Presumably, this is necessary to hinder premature convergence to the point  $x_2 = 6$ . Eventually, some points are placed in the vicinity of the global optimum. Therefore, increasing the number of evaluations most likely leads to higher success rates.

From the in-run distribution of the individuals, it is clear that the (5, 10)-CMA-ES manages to place some of the 5 individuals in each basin of attraction after some generations. Thus, it approximates the global optimum quite well. However, a population of more than one parent would translate back to a multi-population EMOA.

Summarizing, it shall be stated that although the function is of course  $\mu$ -greedy solvable,  $\mu$ -greedy EA without additional features like step-size adaptation have roughly the same chance of getting to the global optimum like 1-greedy EA. Note that the same applies to the original discrete example presented by Zitzler et al. [ZTB08], where, however, a much lower number of points to jump to exists. This means that, where the discrete example does not pose a problem to a  $\mu$ -greedy scheme, the continuous one does.

### 3.3 More Points and a Strong Geometrical Argument

The provided example of a non 1-greedy function is fragile: Moving the reference point from  $\mathbf{R} = (10, 7)^T$  towards symmetry makes it 1-greedy solvable

again. Also note that the whole construction breaks down when going to a three-dimensional problem. Empirical tests with EMOA using a population with size  $\mu = 3$  show that the optimal distribution  $x = (1, 5, 7)$  will always be obtained, regardless of the chosen method (1-greedy or  $\mu$ -greedy).

Continuing this line of thought, it is of course possible to build a problem that is also misleading for 1-greedy algorithms with population sizes of three. In fact, reducing the sizes of the basins of attraction in the hypervolume landscape would be a move towards this goal (see Fig. 3). However, such a problem will also become increasingly difficult for a  $\mu$ -greedy algorithm, as empirically shown in section 3.2.

*Conjecture 1.* Continuously defined functions, which are not 1-greedy solvable for large population sizes ( $\mu \gg 3$ ) are not generally considerably easier for  $\mu$ -greedy algorithms.

One may however pose the question if these non 1-greedy solvable functions have to be defined piecewise. From Fig. 3, we may deduct that piecewise definition here is just a matter of construction and not a necessary condition. There is no reason withstanding creation of a continuous and even continuously differentiable non 1-greedy solvable function (so that the boundaries between pieces become flat) except that its analytical formulation may be much more complicated. Remember that, for non 1-greedy solvability, we only have to establish that from *one* point, line searches in all dimensions fail. This leads us to the following conjecture:

*Conjecture 2.* Non 1-greedy solvable, continuously differentiable functions can be constructed for any finite population size.

## 4 $\mathcal{S}$ -Metric Properties on Continuously Differentiable Pareto Fronts

This section analyzes the convergence of 1-greedy EMOA to the distribution maximizing the  $\mathcal{S}$ -metric for two special cases in the first part. Afterwards, the properties of 1-greedy EMOA on differently shaped Pareto fronts are empirically studied.

### 4.1 Theoretical Analysis on Linear and Convex Pareto Fronts

Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be a bi-objective function to be minimized. We assume that the Pareto front  $f(X^*)$  associated with the Pareto set  $X^* \subset \mathbb{R}^2$  is a Jordan arc with parametric representation

$$f(X^*) = \left\{ \begin{pmatrix} s \\ \gamma(s) \end{pmatrix} : s \in [0, 1] \subset \mathbb{R} \right\}, \quad (6)$$

where  $\gamma : [0, 1] \rightarrow \mathbb{R}$  is twice continuously differentiable. Let  $y^{(1)}, \dots, y^{(\mu)} \in f(X^*)$  be distinct objective vectors on the Pareto front. According to (6), we have  $y^{(i)} = (s_i, \gamma(s_i))^T$  for  $i = 1, \dots, \mu$ . As a consequence, the  $\mathcal{S}$ -metric or dominated hypervolume of the points  $y^{(1)}, \dots, y^{(\mu)}$  is given by

$$H(s) = (r_1 - s_1)[r_2 - \gamma(s_1)] + \sum_{i=2}^{\mu} (r_1 - s_i)[\gamma(s_{i-1}) - \gamma(s_i)] \quad (7)$$

with reference point  $\mathbf{R} = (r_1, r_2)^T$ ,  $0 \leq s_1 < s_2 < \dots < s_\mu \leq 1 \leq r_1$  and  $r_2 \geq \gamma(0)$ .

Whenever the  $\mathcal{S}$ -metric is concave, the 1-greedy selection scheme of the SMS-EMOA with fixed reference point and a population of  $\mu$  individuals on a continuous front will not get stuck prematurely since it is sufficient to move a single variable  $s_i$  at each iteration towards ascending values of the  $\mathcal{S}$ -metric in order to reach its maximum. Furthermore, we are going to use the result that a twice differentiable function is concave if and only if its Hessian matrix is negatively semidefinite. Partial differentiation of (7) leads to

$$\frac{\partial H(s)}{\partial s_1} = \gamma(s_1) - r_2 + (s_1 - s_2) \gamma'(s_1) \quad (8)$$

$$\frac{\partial H(s)}{\partial s_i} = \gamma(s_i) - \gamma(s_{i-1}) + (s_i - s_{i+1}) \gamma'(s_i) \quad (i = 2, \dots, \mu - 1) \quad (9)$$

$$\frac{\partial H(s)}{\partial s_\mu} = \gamma(s_\mu) - \gamma(s_{\mu-1}) + (r_1 - s_\mu) \gamma'(s_\mu) \quad (10)$$

and finally to

$$\frac{\partial^2 H(s)}{\partial s_i \partial s_{i-1}} = -\gamma'(s_{i-1}) \quad (i = 1, \dots, \mu - 1) \quad (11)$$

$$\frac{\partial^2 H(s)}{\partial s_i \partial s_i} = 2\gamma'(s_i) + (s_i - s_{i+1}) \gamma''(s_i) \quad (i = 1, \dots, \mu) \quad (12)$$

$$\frac{\partial^2 H(s)}{\partial s_i \partial s_{i+1}} = -\gamma'(s_i) \quad (i = 2, \dots, \mu) \quad (13)$$

with  $s_{\mu+1} := r_1$ . Other second partial derivatives are zero. Thus, the Hessian matrix  $\nabla^2 H(s)$  of the  $\mathcal{S}$ -metric as given in (7) is a tridiagonal matrix.

**Linear Pareto Front.** Suppose that  $\gamma(s) = ms + b$  is a *linear function*. Then,  $\gamma(\cdot)$  is strongly monotone decreasing with  $\gamma'(s) = m < 0$  and  $\gamma''(s) = 0$  for all  $s \in (0, 1)$ . In this case, the Hessian matrix reduces to a tridiagonal matrix with identical diagonal entries  $2m < 0$  and identical off-diagonal entries  $-m > 0$ . Recall that a square matrix  $A$  is weakly diagonal dominant if  $|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$  for all  $i$  and that a weakly diagonal dominant matrix is negatively definite if all diagonal entries are negative. It is easily seen that these conditions are fulfilled. As a result, we have proven:

**Theorem 1.** *If the Pareto front of a bi-objective minimization problem is linear, then the  $\mathcal{S}$ -metric or dominated hypervolume of  $\mu$  distinct points on the Pareto front is a strictly concave function.  $\square$*

From this result, we can deduce that it is sufficient to move a single point at a time for reaching the maximal  $\mathcal{S}$ -metric value in the limit. Next, we try to generalize this result.

**Convex Pareto Front.** Suppose that  $\gamma(\cdot)$  is a *convex function*. Then,  $\gamma(\cdot)$  is strongly monotone decreasing with  $\gamma'(s) < 0$  and  $\gamma''(s) > 0$  for all  $s \in (0, 1)$ . Again, the Hessian matrix is tridiagonal, but the criterion of diagonal dominance of the Hessian does not always hold. Actually, the Hessian is not negatively semidefinite in general. This is easily seen from a counter-example: Let  $\gamma(s) = (1 - \sqrt{s})^2$  (T4,  $\alpha = 2$ ) with  $\gamma'(s) = 1 - 1/\sqrt{s} < 0$  and  $\gamma''(s) = \frac{1}{2}s^{-\frac{3}{2}} > 0$  for  $s \in (0, 1)$  and reference point  $\mathbf{R} = (1, 1)^T$ . Consider three points on the Pareto front with  $s_1 = (\frac{1}{10})^2$ ,  $s_2 = (\frac{19}{20})^2$ ,  $s_3 = (\frac{20}{21})^2$  leading to the Hessian matrix

$$\nabla^2 H(s) = \begin{pmatrix} -\frac{1857}{4} & 9 & 0 \\ 9 & -\frac{326392}{3024819} & \frac{1}{19} \\ 0 & \frac{1}{19} & -\frac{2461}{16000} \end{pmatrix},$$

whose leading principal minors are  $\Delta_1 < 0$ ,  $\Delta_2 < 0$  and  $\Delta_3 > 0$  indicating that the Hessian matrix with this particular choice of points  $s_1, s_2, s_3$  is not negatively semidefinite. On the other hand, if  $s = (\frac{1}{100}, \frac{1}{4}, \frac{4}{9})^T$ , it is easily verified that  $\Delta_1 < 0$ ,  $\Delta_2 > 0$  and  $\Delta_3 < 0$  indicating that the Hessian matrix is negatively definite in this particular case. In summary, the Hessian matrix is indefinite and we have proven:

**Theorem 2.** *If the Pareto front of a bi-objective minimization problem is convex, then the  $\mathcal{S}$ -metric or dominated hypervolume of  $\mu$  distinct points on the Pareto front is not a concave function in general.  $\square$*

However, this result does imply neither that there are no convex fronts with concave  $\mathcal{S}$ -metric nor that the 1-greedy selection scheme of the SMS-EMOA gets necessarily stuck on convex fronts.

## 4.2 Empirical Results of SMS-EMOA on Connected Pareto Fronts

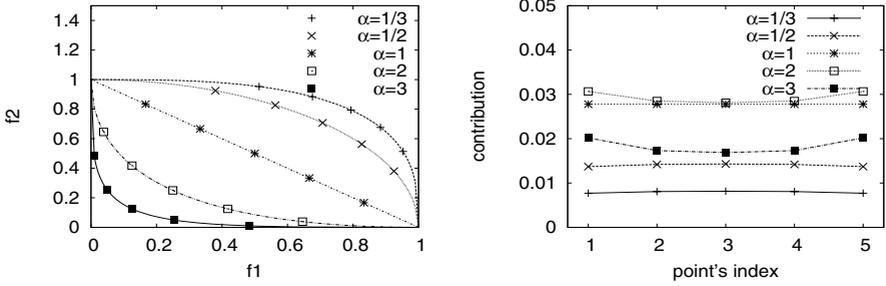
In this section, it is empirically analyzed whether the 1-greedy SMS-EMOA can robustly obtain the  $\mathcal{S}$ -metric-optimal distribution of points for the approximation of piecewise continuous Pareto fronts with different curvature (convex to concave). It is assumed that the population has already arrived *on* the Pareto front and performs only local refinements. Thereby, the local 1-greediness as defined in section 2.1 of the considered test functions is empirically analyzed. Recall that a local 1-greedy solvable problem is also 1-greedy solvable. To accomplish this, a comprehensive study on the set of simple test functions T1-T5 is conducted (cf. Sec. 2.2). Due to the results in section 3.3, we restrict our analyses on small populations.

*Pre-experimental planning:* The SMS-EMOA selection operator discards the individual with the lowest hypervolume contribution, i.e., the amount that gets lost when the individual is removed since the hypervolume part is disjointly dominated by that individual. In order to provide a deeper understanding of this selection, the areas of individuals, which enter the population, are visualized in Fig. 5 for a given approximation using exemplarily the one-dimensional test function T1 defined in Sec. 2.2. It can be seen that the areas of success are directly adjacent to the solutions of the current approximation, which would be discarded instead. They therefore indicate the direction in which this solution should be shifted. Fig. 5 has been created based on T1, but the same fact holds for T2-T5. Intuitively, one may assume that the hypervolume contributions of individuals tend to equal values for all points of an optimally distributed set since, otherwise, a solution can move closer to the point with a higher contribution.

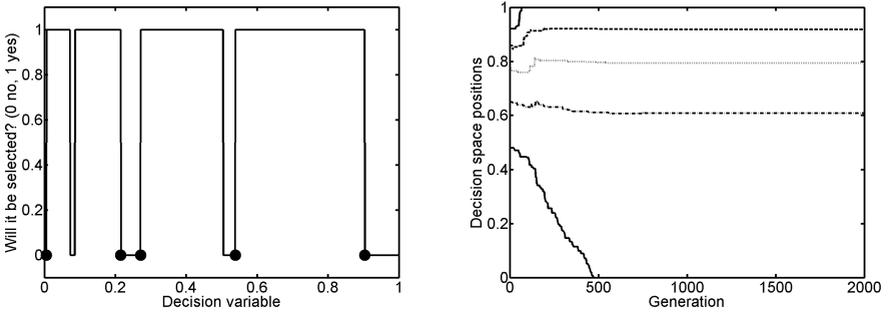
In order to investigate this conjecture, we compute these hypervolume contributions for the analytically determined optimal distributions of populations of five individuals, which are shown in Fig. 4 (left). For computing these distributions, test function T4 is considered with  $\alpha \in \{1/3, 1/2, 1, 2, 3\}$  resulting in two concave fronts for  $\alpha < 1$ , convex fronts for  $\alpha > 1$ , and a linear front for  $\alpha = 1$ . Furthermore, a reference point  $\mathbf{R} = (1.0, 1.0)^T$  positioned exactly at the boundaries of the Pareto front is used. Due to the construction of T4, the distribution is symmetrical to the bisecting line. Thus, the central point of the population lies exactly on this line. Since the hypervolume of the population of SMS-EMOA monotonically increases, the population will tend to these optimal distributions in case of a successful optimization.

The right part of Fig. 4 shows the corresponding hypervolume contributions sorted with respect to the first objective. The contribution values are symmetrical to the point in the middle. It can be observed that the contribution values tend to grow with increasing  $\alpha$ , when  $\alpha < 3$ . On the concave Pareto front, the point in the middle (in the knee region) has the highest contribution and the contribution values are decreasing when going to the boundaries. On the convex Pareto front, the values decrease from the boundaries to the middle, so the point in the knee region has the lowest contribution. On linear fronts, the distribution obtaining the maximal hypervolume value is the set of equally spaced points as proved by Beume et al. [BFLI<sup>+</sup>07]. Only in this case, the contributions of all points are equal. Therefore, our first intuition was misleading.

To further investigate the effect of single local refinements, a local search SMS-EMOA, which uses only Gaussian mutations of single individuals of the current population with small stepsize  $\sigma = 0.01$  to generate new candidates for selection, has been implemented. Fig. 5 plots the run of the decision space variables of this local search (5+1)-SMS-EMOA on T4 with  $\alpha = 1/3$ , when a fixed reference point  $\mathbf{R}' = (2, 2)^T$  is chosen. As starting positions, the optimal distributions for the closer reference point  $\mathbf{R} = (1.0, 1.0)^T$  are used. It can be seen that the algorithm is able to guide the solutions from the old to the new optimal positions. A closer look on the resulting population yields that the contribution of the points at the boundaries depends on the choice of the reference point. For the new reference



**Fig. 4.** Optimal positions of points on Pareto fronts of different curvature (left) and the corresponding hypervolume contributions of the points (right). The reference point  $\mathbf{R} = (1.0, 1.0)$  is chosen.



**Fig. 5.** Left: Acceptance of solutions depending on the decision space variable  $x$  for test function T1 and a randomly initialized population. The decision space variables of this population are indicated by black dots. Right: The run of the sorted decision space variables, which result in the population, over the generations of the local search (5 + 1)-SMS-EMOA on the concave test function T4 with  $\alpha = 1/3$ , when a fixed reference point  $\mathbf{R}' = (2, 2)^T$  is chosen.

point  $\mathbf{R}'$ , which is situated at a greater distance to the Pareto front, the optimal points move closer to the boundaries and the contributions near the boundaries grow<sup>1</sup>. As a consequence, the other points follow these extremal solutions to cover the resulting distance. The optimal distribution does emerge.

The following experiment is conducted to empirically support the arising assumption that even a local search-based  $(\mu+1)$  SMS-EMOA is able to approximate a set of optimally distributed Pareto-optimal points for continuous problems with convex and concave shaped Pareto fronts.

*Task:* Check the hypothesis that the local search SMS-EMOA is able to approximate the optimally distributed subset of the Pareto front of the given continuous

<sup>1</sup> For  $\mathbf{R}'$  the sorted contributions are (0.61, 0.015, 0.016, 0.015, 0.61).

test problems T1-T5 for fixed population sizes  $\mu \in \{1, \dots, 6\}$  with an accuracy limited by the step size  $\sigma = 0.01$ .

*Setup:* For each test function T1-T5 and population sizes  $\mu \in \{1, \dots, 6\}$ , approximations for the  $\mathcal{S}$ -metric-optimal distribution are globally calculated by the MATLAB implementation of the (5,10)-CMA-ES [HO01], where no limit on the function evaluations, but a lower limit on  $\sigma_i$  ( $i = 1, \dots, \mu$ ) of  $10^{-12}$  is specified. The successful application of this algorithm for the calculation of optimal distributions, even for multi-modal hypervolume landscapes, has already been shown in section 3.2. For each configuration 10,000 runs of the local search SMS-EMOA are performed using different random initializations and the results after  $\mu \cdot 1,000$  generations are compared to the approximations found by the global optimization of the CMA-ES. A run is denoted as failed when the hypervolume of the found approximation is below 99% of the approximated optimal one.

*Observations:* The local search SMS-EMOA detects the optimal distribution in all runs for the convex and concave test function T1-T5 except for 10% of the initializations on the concave-convex Pareto front T5 for  $\mu = 1$ . When the initial solution is situated close to the left border ( $x < 0.1$ ), the local search SMS-EMOA converges to the left border, which indicates the optimum for the concave part of the Pareto front, instead of detecting the globally optimal position in the inflection point.

*Discussion:* Based on thorough experimentation, it can be assumed that even a local search SMS-EMOA robustly detects the globally optimal distribution in cases where the sign of the second derivative with respect to the first objective does not change. However, due to emerging effects of the local refinements and their interaction, for higher population sizes, this result seems to hold also for concave-convex Pareto fronts.

## 5 Conclusions

In this paper, we have investigated how a 1-greedy EMOA performs on different kinds of continuous Pareto fronts using both formal proofs and empirical analyses. So far, only an artificial discrete problem existed to show that a 1-greedy EMOA with hypervolume selection is not able to obtain the set covering the maximal hypervolume. We have shown that this problem can be converted to the continuous space while preserving its important properties. Thereby, it has been demonstrated that the local optimum is not only a singularity, but has an actual attractor, which makes the problem also hard for  $\mu$ -greedy EMOA.

Furthermore, it has been shown that even a local-search-based 1-greedy EMOA successfully detects the globally optimal distribution for most connected continuous Pareto front types. Failures have only been observed for very small population sizes and we therefore think that the risk of not being 1-greedy decreases with increasing population size. First hints on possible explanations have been provided. Additionally, we have proven that the hypervolume is 1-greedy on linear Pareto fronts and formalize the necessary condition of 1-greediness in general.

In this work, we only consider situations, in which the points are located exactly on the Pareto front, which is not realistic for continuous spaces. We claim that the risk of getting stuck in a local optimum decreases when the population is not close to the Pareto front since there are more improving directions. Future work shall further investigate the influence of the reference point on the properties of the distribution of points and the convergence to the distribution obtaining the optimal  $\mathcal{S}$ -metric value.

## Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the Collaborative Research Center 'Computational Intelligence' (SFB 531), the SFB/TR TRR 30, and grant no. RU 1395/3-2. We also acknowledge support by the German *Federal Ministry of Economics and Technology (BMW)*.

## References

- [BFLI<sup>+</sup>07] Beume, N., Fonseca, C.M., López-Ibáñez, M., Paquete, L., Vahrenhold, J.: On the Complexity of Computing the Hypervolume Indicator. Technical Report CI-235/07, Reihe CI, SFB 531 (2007)
- [BNE07] Beume, N., Naujoks, B., Emmerich, M.: SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research* 181(3), 1653–1669 (2007)
- [DPAM02] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6(2), 182–197 (2002)
- [DTLZ02] Deb, K., Thiele, L., Laumanns, M., Zitzler, E.: Scalable Multi-objective Optimization Test Problems. In: Proc. of the 2002 Congress on Evolutionary Computation (CEC 2002), vol. 1, pp. 825–830. IEEE Press, Piscataway (2002)
- [HO01] Hansen, N., Ostermeier, A.: Completely Derandomized Self-Adaptation in Evolution Strategies. *IEEE Computational Intelligence Magazine* 9(2), 159–195 (2001)
- [Sch85] Schaffer, J.D.: Multiple objective optimization with vector evaluated genetic algorithms. In: Grefenstette, J.J. (ed.) Proc. 1st Int'l. Conf. Genetic Algorithms (ICGA), pp. 93–100. Lawrence Erlbaum, Mahwah (1985)
- [ZT98] Zitzler, E., Thiele, L.: Multiobjective optimization using evolutionary algorithms - A comparative case study. In: Eiben, A.E., Bäck, T., Schoenauer, M., Schwefel, H.-P. (eds.) PPSN 1998. LNCS, vol. 1498, pp. 292–301. Springer, Heidelberg (1998)
- [ZTB08] Zitzler, E., Thiele, L., Bader, J.: On Set-Based Multiobjective Optimization. Technical Report 300, Computer Engineering and Networks Laboratory, ETH Zurich (February 2008)