Pareto Set and EMOA Behavior for Simple Multimodal Multiobjective Functions

Mike Preuss, Boris Naujoks, and Günter Rudolph

Universität Dortmund, Lehrstuhl für Algorithm Engineering
44221 Dortmund, Germany
{mike.preuss, boris.naujoks, guenter.rudolph}@uni-dortmund.de
http://ls11-www.cs.uni-dortmund.de

Abstract. Recent research on evolutionary multiobjective optimization has mainly focused on Pareto fronts. However, we state that proper behavior of the utilized algorithms in decision/search space is necessary for obtaining good results if multimodal objective functions are concerned. Therefore, it makes sense to observe the development of Pareto sets as well. We do so on a simple, configurable problem, and detect interesting interactions between induced changes to the Pareto set and the ability of three optimization algorithms to keep track of Pareto fronts.

1 Introduction

In recent years, *evolutionary multiobjective optimization* (EMO) [1,2] has developed from a marginal into one of the most actively pursued areas within *evolutionary computation* (EC). Many new algorithms and measures have been suggested, and, with them, concepts like Pareto set and Pareto front have entered the common EC vocabulary [3]. Increasing interest in multiobjective techniques has even evoked new theoretical approaches that employ multiple objectives to simplify an originally singleobjective problem [4]. However, most of the current EMO research concentrates on processes observed in the objective space, which consists of the possibly obtainable value combinations of the considered objective functions. Undoubtedly, approximating the Pareto front well is the final aim of EMO algorithms (EMOAs), and the Pareto set distribution may be of minor interest for estimating their performances. Nevertheless, for improving these algorithms, as well as for attaining guidelines on which of the solutions contained in the approximated Pareto set shall eventually be implemented in a real-world situation, a well-founded understanding of Pareto set distributions is supposed to be a major advantage.

Research on singleobjective algorithms largely focuses on population behavior in the decision space, or simplified models thereof, e.g. using basins of attraction as means of abstraction [5]. Especially for multimodal problems, numerous techniques have been invented to prevent the populations from converging to a single point too soon. Some of these, as crowding [6], are also applied in EMOAs. But diversity maintenance is only sought in objective space, to ensure good coverage

T.P. Runarsson et al. (Eds.): PPSN IX, LNCS 4193, pp. 513–522, 2006.
© Springer-Verlag Berlin Heidelberg 2006

of the Pareto front. However, for at least one of the objective functions being multimodal, it is clear that this coverage cannot normally be achieved when the whole population is clustered around one local minimum of this function. We thus conjecture that a) there are situations—and these are not uncommon—where the Pareto set does not share the aspired nice properties of the received Pareto front the user normally focusses the attention on, and b) that diversity maintenance is not only needed in objective but also in decision space for successfully treating *multiobjective optimization problems* (MOPs): The product designer is mainly interested in a thorough covering of the Pareto front for maximum wide scope in selecting solutions according to the (conjectured) customers' desires. This is the situation which contemporary EMOAs are designed for. But the product engineer is mainly interested in a thorough covering of the Pareto set since it is important to know if a certain design can be realized by different parameters of the production process: Solutions may differ in sensitivity or in shorter tooling times and the like. Evidently, contemporary EMOAs are not geared toward product engineers yet.

These both sides of one medal (Pareto front in the objective space, Pareto set in the decision space) and the conjunction between them has not been studied in detail before. Only few theoretical results for special classes of search spaces and multiobjective functions were presented before, cf. Ehrgott [7]. But the handled cases are restricted in a way that no generalization can be foreseen. Some effort has been made in the development of test functions not only with regard to a nice behaving Pareto front, but also with aspired properties in the decision space, cf. Okabe et al. [8]. Zhou et al. [9] propose a specialized EA to implicitly handle and profit from regularities in the objective as well as in the decision space. Such regularities stem from the test functions proposed by Okabe et al. [8] and cannot be expected generally.

2 Aims and Methods

Our approach is constructive; on a minimalist bimodal bicriteria test problem, we study structural changes of true Pareto set and Pareto front on a set of targeted modifications. These are derived both analytically and empirically, the latter employing grid-based and stochastic enumerators. Furthermore, we observe how different EMOAs cope with the original problem and the changes. More detailed, we try to answer the following questions:

- How do Pareto set distributions change when the problem is modified? Are there unexpected outcomes?
- Are different EMOAs able to cover Pareto set and Pareto front well?
- Are there consistent similarities/dissimilarities in the behavior of different EMOAs due to problem modifications that hint to distinct capabilities of these?

3 A Simple Test Problem: TWO-ON-ONE

To deepen the insight in behavior and structure of Pareto sets mapping onto Pareto fronts, we define the plainest bimodal/unimodal test problem we could think of. It consists of a polynomial function f_1 of degree four with two optima, and the sphere function f_2, which is of degree two:

$$f = (f_1, f_2) : \mathbb{R}^2 \to \mathbb{R}^2 : f_1(x_1, x_2) = x_1^4 + x_2^4 - x_1^2 + x_2^2 - cx_1x_2 + dx_1 + 20,$$
$$f_2(x_1, x_2) = (x_1 - k)^2 + (x_2 - l)^2$$

The level (niveau) of the optima of f_1 can be adjusted smoothly via parameter d. With d positive, the optimum in the positive x_1 domain is lifted up in comparison to the optimum in the negative x_1 domain. Consequently, the former becomes a local optimum, while the latter remains a global optimum (asymmetric optima). With parameter $c = 0$, both minimizers are located on the x_1 axis, but for increasing c, their connecting line is rotated counterclockwise, until its gradient is nearly 1.

The function f_2 is unimodal, the location of its minimizer determined by parameters k and l. For $k = l = 0$ it is located in the origin, right between the minimizers of the bimodal function f_1. By variation of k and l the minimizer is moved away from the connecting line of the minimizers of f_1. Next to changing the Pareto front, this also effects the Pareto set.

Table 1. Parameter setting for the five cases of TWO-ON-ONE, c was set to 10

Cases		1	2	3	4	5
Parameters	d	0	0	0.25	0.25	0.25
	k	0	1	0	1	0
	l	0	0	0	0	1

In order to allow for a theoretical analysis of the problem, five parameter settings have been fixed (Table 1, Figure 1). These result in different placements of the minimizers in search space, two for the symmetric case (both optima of f_1 identical), three for the asymmetric case (optima distinct). While all these settings are expected to lead to ordinary (generic) Pareto fronts, the Pareto sets are expected to look more complex.

The coordinates of the minimizers of f_1 can be determined analytically to

$$(x^*_{1_{1,2}}, x^*_{2_{1,2}}) = \left(\pm\frac{1}{2}\sqrt{(\sqrt{101}+1)}, \pm\frac{1}{20}(\sqrt{101}-1)\sqrt{(\sqrt{101}+1)} \right) .$$

In cases 1 and 2, both optima of f_1 are on the same level, ensured by $d = 0$. In 1, the minimizer of the sphere function is located in the origin, where f_1 has a saddle point. In 2, the optimum is moved right on the x_1 axis by one unit.

Cases 3 and 4 repeat the same configuration, with asymmetric optima of f_1; the global optimum resides in the negative x_1 and x_2 domain, and the local

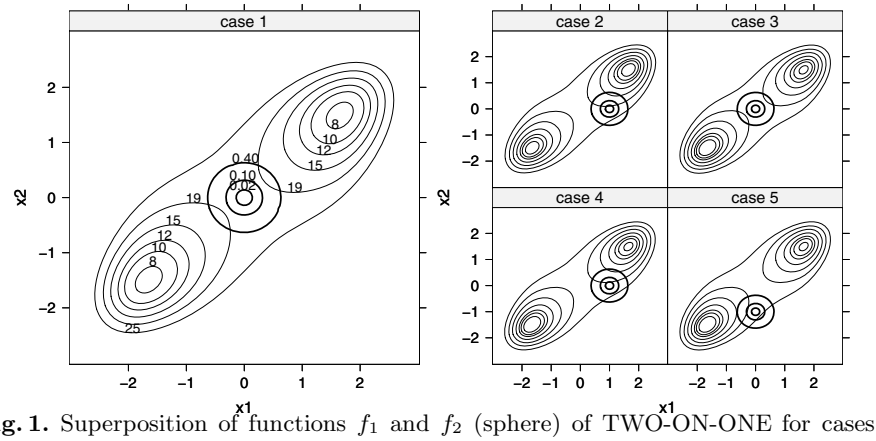

Fig. 1. Superposition of functions f_1 and f_2 (sphere) of TWO-ON-ONE for cases 1 to 5. The optima of f_1 are symmetrical (equal fitness) in cases $1 + 2$, asymmetrical in $3+4+5$, with the right minimizer shifted slightly upwards and the left one downwards.

optimum in the positive domain. It is expected that the Pareto set now establishes a connection between the global optima of f_1 and of f_2. This is due to the solutions in the global optimum of f_1 being mapped to the extremal part of the Pareto front. Consequently, solutions from the local optimum of f_1 may be lost.

The same situation is expected for case 5, which is similar, except for a movement of the minimizer of f_2 towards the global minimizer of f_2, whereas in case 4, it is brought nearer to the local minimizer.

4 Experimental Investigation of Pareto Sets

Our expectation is that for all symmetrical cases, the Pareto sets consist of two curves, connecting either peak with the minimizer of the sphere function f_2. For the asymmetric cases, it seems reasonable that Pareto sets contain only points on a curve between the global minimizers of f_1 and f_2. But this expectation stems from thought experiments rather than empirical or analytical facts.

We employ two simple tools, a grid based and a stochastic enumerator, for obtaining a first, rough impression of structure and location of the Pareto sets. Either one samples points from a given interval and keeps a list of the Pareto-optimal solutions found. Tried points are either taken from a pre-specified grid or determined randomly. As we shall see, it sometimes makes sense to use both, as the obtained results can subtly differ.

Experiment 1: Determine Pareto sets and fronts of TWO-ON-ONE.

Pre-experimental planning: First experiments were performed with a grid-based enumerator only. They revealed an unexpectedly wide Pareto set (Fig. 2, left). We thus additionally sampled by means of a stochastic enumerator.

Task: Find location of the Pareto sets, detect deviations from the expected.

Setup: For each of the 5 cases specified in Table 1, we sample points in the interval $x_1, x_2 \in [-3, 3]$. The grid-based enumeration consists of $300 \times 300 = 90,000$ points each, the stochastic enumeration of $500,000$ points each. The difference is intended as we hope for a better resolution with the latter method, to shed light on the bar-shaped artifacts. All non-dominated points are archived.

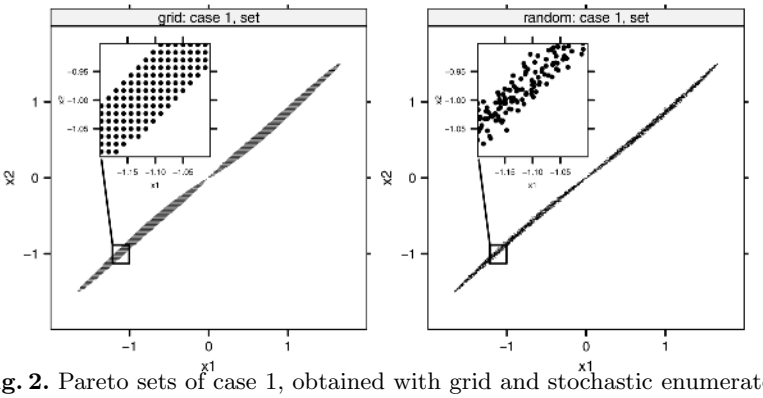

Fig. 2. Pareto sets of case 1, obtained with grid and stochastic enumerator

Experimentation/Visualization: Figures 2 and 3 show the most interesting of the obtained Pareto sets and fronts. All others largely comply with the previously stated expectations and are omitted due to space limitations.

Observations: The figures clearly show that neither grid nor stochastic sampling produces a clear-cut picture of the true Pareto sets. Roughly, case 1 yields a smeared areal, propeller-like structure (Fig. 2) instead of the expected single curve. However, the Pareto set appears narrower under stochastic enumeration.

For case 4 (sphere function f_2 moved towards local optimum of f_1), the Pareto front splits into two parts at $f_1 \approx 8$, as visualized in Fig. 3. Accordingly, the Pareto set breaks up into two distinct fragments. Note that no connection exists between the location of the sphere and the global optimum of f_1. At the left edge of its right part, the grid-based approximated Pareto set reveals a strange curl which is not visible in the stochastically approximated Pareto set. Pareto fronts of cases 4 and 5 both contain pieces of very low point densities, at $17 < f_1 < 19$ in the former, and $15 < f_1 < 17$ in the latter case.

Discussion: We regard the obtained Pareto set approximations for case 1 as rather misleading, and analytical investigation in §5 supports this view. However, considering the amount of sampled points ($90\,k$ and $500\,k$), and taking into account that the latter (stochastically approximated) Pareto set is much tighter, one may conclude even from our empirical data that the true Pareto set indeed is most likely located on a curve and non-areal. The enumerators are probably misguided by the huge difference in gradients of f_1 and f_2 in direction of the connecting line between the two optima of f_1 and orthogonal to it. Following from that, any EMOA will experience the same situation: Practically identical values of the objective functions can have a large set of preimages and thus spread in search space.

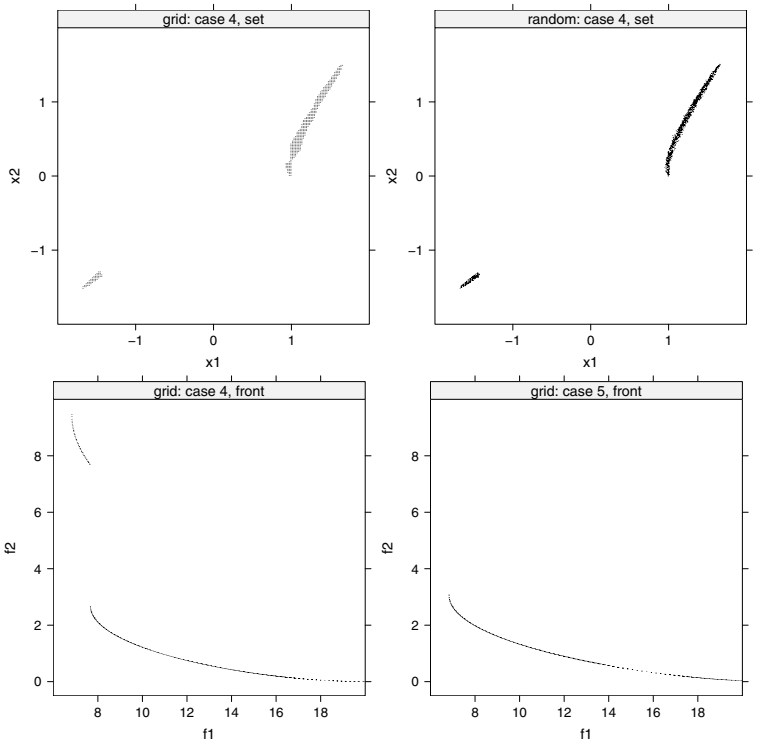

Fig. 3. Surprising Pareto sets and fronts for cases 4 and 5

Results obtained for case 4 show that contrary to our expectation, by far the larger Pareto set portion resides in the range between the local optimum of f_1 and the optimum of f_2. Only where function values for f_1 are better than may be attained at the local optimum, points from the left fragment can enter the Pareto set, resulting in a stepped Pareto front. The curl found near $x_1 = 1$ seemingly corresponds to the low density part of the Pareto front which must be located in proximity of the sphere center as values for f_2 are near 0. The stochastic Pareto set approximation is again tighter than the grid-based one, leading to the conjecture that the true Pareto set is non-areal as in case 1.

Two more conclusions may be drawn from the case 4 results. Firstly, search space distances between optima of separate objective functions play a major role for the composition of Pareto sets, and secondly, it is necessary to keep the population of EMOAs spread over several local optima of the treated objective functions during an optimization run.

5 Analytical Derivation of Pareto Sets

The Pareto set for case 1 can be derived analytically but its analytic expression is too complex and space-consuming to be presented here. Instead, we suggest the linear approximation

$$\left(x_1, \frac{\sqrt{101}-1}{10}\, x_1\right) \quad \text{for} \quad x_1 \in \left[-\frac{1}{2}\sqrt{\sqrt{101}+1}, \frac{1}{2}\sqrt{\sqrt{101}+1}\right]$$

whose deviation from the true convex-concave curve is less than 0.045 for all x_1 above. In any case, the Pareto set is a 1-dimensional connected set and not an areal set of higher dimension as the output of the grid or stochastic enumerator might suggest (see fig. 2).

As can be seen from the symmetry $f(x_1, x_2) = f(-x_1, -x_2)$ the entire Pareto front can be built solely by positive (or negative) points of the Pareto set. Thus, it may happen that an EMOA approximating the Pareto front quite well with regard to the S-metric has found only points in the decision space with, say, positive components. As a consequence, a good value for the S-metric tells only half the story.

The Pareto sets of the other cases are also amenable to an analytic solution but the expressions are far away from being manageable easily. This observation is quite counter-intuitive given the pretended simple expressions and structural design of the objective functions.

6 Behavior of EMOAs on TWO-ON-ONE

Whereas Pareto sets and fronts of problem TWO-ON-ONE have been explored in §4 and determined analytically in §5, we now turn to the behavior of different EMOAs in a second experiment. Note that it is not intended to argue in favor of or against any algorithm here, but rather to detect possible differences.

Algorithms. We invoke two standard techniques next to a new development within the field. The Pisa framework[1] is used to conduct the referred optimization runs. Here, the TWO-ON-ONE problem has been implemented as a variator, which can be optimized with respect to different objectives and multiple selectors. Among the set of available selectors, NSGA-II and SPEA2 are chosen, because these appear to be the currently most well-known and commonly used algorithms in the field [1,2]. Additionally, the more recent SMS-EMOA [10,11] is tested within this framework. The SMS-EMOA was designed for featuring a performance measure, namely the hypervolume or S-metric, as secondary ranking criterion in a NSGA-II like manner. The additional effort for a third algorithm in the study seems to be justified, because the SMS-EMOA was found to spread solutions more nicely over Pareto-fronts than the other two algorithms. This aspired behavior is purchased by a runtime of $O(\mu \log \mu + \mu^{(d/2+1)} \log \mu)$ of the SMS-EMOA, with μ denoting the population size and d the number of objectives (cp. Beume [12]). In contrast, the runtime of NSGA-II and SPEA2 is quadratic in the population size and polynomial in the number of objectives.

[1] PISA - Platform and Programming Language Independent Interface for Search Algorithms, ETH Zurich, `www.tik.ee.ethz.ch/pisa/`

Measures. To detect differences in algorithm behaviors on the most interesting cases, we define two simple measures. For case 1, we measure if the resulting population P is fairly distributed over the left and right wings of the Pareto set by taking the fraction on the less crowded wing into account:

$$\text{fair}(P) = \frac{1}{2} - \frac{\min(|\{\text{individual} \in P : x_1 < 0\}|, |\{\text{individual} \in P : x_1 \geq 0\}|)}{|P|} \quad (1)$$

For case 4, we are interested in the fraction of points in proximity of the global optimum of f_1, corresponding to the search points in the left half of the search space:

$$\text{left}(P) = \frac{|\{\text{individual} \in P : x_1 < 0\}|}{|P|} \quad (2)$$

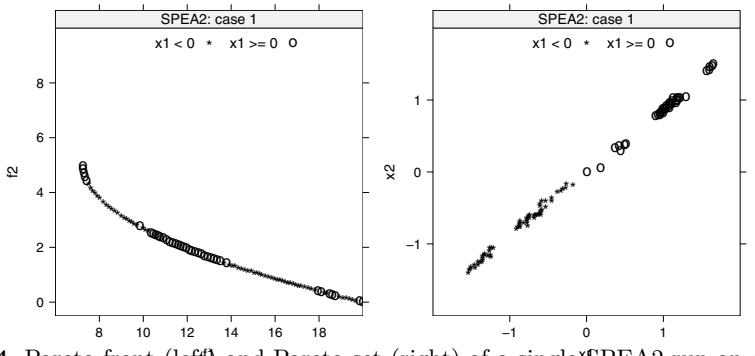

Fig. 4. Pareto front (left) and Pareto set (right) of a single SPEA2 run on case 1

Experiment 2: Search Space Behavior of EMOAs on function TWO-ON-ONE.

Pre-experimental planning: First runs indicated that results on cases 2, 3, and 5 are comparable for all three algorithms. We thus focused on cases 1 and 4. For case 1 it was found that at least 50 runs are necessary to get a detailed picture of differences in measure fair(P), for case 4, 20 runs seemed sufficient.

Task: Detect differences in the obtained Pareto sets and fronts that may be related to test problem properties. We employ bootstrap permutation tests with 49,999 replicates and significance level 5% for the measured data.

Setup: The decision space was limited to $f_1, f_2 \in [-50, 50]$, thereby enclosing the region around the optima of f_1 and f_2, and a certain amount of space the algorithms have to bypass to get there. All three algorithms, NSGA-II, SPEA2, and SMS-EMOA, are run with a population size of 100 for 30,000 evaluations, otherwise utilizing default parametrizations. For case 1 and 4, 50 and 20 runs are performed, respectively.

Experimentation/Visualization: Figure 4 depicts a typical outcome for case 1. More extreme population distributions with almost all individuals on one wing

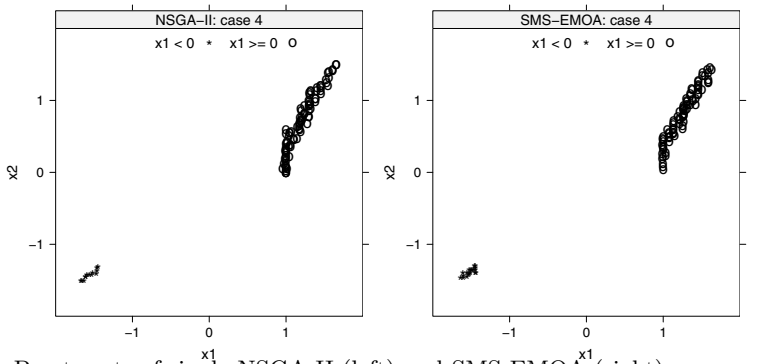

Fig. 5. Pareto sets of single NSGA-II (left) and SMS-EMOA (right) runs on case 4

of the Pareto set also happen. In figure 5, resulting Pareto sets for two different algorithms are presented, again, typical runs are chosen.

Observations: Figure 4 demonstrates that for symmetric optima, the Pareto front often contains large chunks of points originating from the proximity of different minimizers. Accordingly, the approximated Pareto sets show corresponding clouds of points, unevenly distributed over the true Pareto set. For case 4, Figure 5 shows that the algorithms are able to spread their populations over both important parts of the Pareto set. However, the shape of the clouds near the global minimizer of f_1 is different: NSGA-II often forms lines of points in that region, whereas the SMS-EMOA rather builds areal structures.

Discussion: For case 1, hypothesis testing reveals a slight difference (p-value 0.071) between NSGA-II and SMS-EMOA and a strong one (p-value 0.030) between NSGA-II and SPEA2. SMS-EMOA and SPEA2 may be considered behaving relatively similar (p-value 0.427). NSGA-II covers both wings of the Pareto set more evenly on average, its *fair*-measure is 0.110, compared to 0.149 and 0.152 for SPEA2 and SMS-EMOA, respectively.

All three algorithms cope surprisingly well with case 4. Here, hypothesis tests hint to a similarity between SMS-EMOA and NSGA-II (p-value 0.468) and sharp distinction between SPEA2 and SMS-EMOA, and SPEA2 and NSGA-II, both p-values 0.001. As indicated by the histograms, NSGA-II and SMS-EMOA both place more points near the global optimizer, their *left*-measures are 0.169 and 0.167, respectively. SPEA2 only puts 13.1% of its final population there. Unfortunately, we are currently not able to explain what makes the algorithms behave differently in this respect.

7 Summary and Outlook

The main message of the work presented here is our belief in the fact that a neat covering of the Pareto front is not sufficient for meeting the needs of all clients that may use EMOAs. Therefore, future versions of EMOAs should also take into account a proper covering of the Pareto set. Evidently, contemporary EMOAs

cannot deliver this kind of behavior. For this purpose we need an effective measure for assessing the quality of a solution set in decision space—similarly to the S-metric in objective space.

To follow this avenue we have to deepen our understanding of EMOA behavior in the decision space, which may be quite counter-intuitive as our seemingly simple test problem has revealed. Obviously, EMOAs are easily confronted with strangely shaped approximate Pareto sets as the ones obtained from our empirical approaches to determine the true Pareto set. We attribute this behavior to scaling issues between orthogonal gradients and discretized computer representation of real values, but this assessment can only be preliminary. The important point is that EMOAs have no means of detecting 'real' Pareto set shapes, they have to cope with their inexact counterparts. These problems are currently not reflected in algorithm design. Furthermore, we are convinced that a thorough analysis of the interaction between Pareto front and Pareto set will eventually lead to new insights, new search operators, and even better EMOAs.

References

1. Deb, K.: Multi-Objective Optimization using Evolutionary Algorithms. Wiley, Chichester, UK (2001)
2. Coello Coello, C.A., Van Veldhuizen, D.A., Lamont, G.B.: Evolutionary Algorithms for Solving Multi-Objective Problems. Kluwer, New York (2002)
3. Coello, C.A.C.: Evolutionary multi-objective optimization: a historical view of the field. IEEE Computational Intelligence Magazine **1**(1) (2006) 28–36
4. Neumann, F., Wegener, I.: Minimum spanning trees made easier via multi-objective optimization. In Beyer, H.G., ed.: Genetic and evolutionary computation conference (GECCO), ACM Press, New York (2005) 763–769
5. Preuss, M., Schönemann, L., Emmerich, M.: Counteracting genetic drift and disruptive recombination in $(\mu \dot{+} \lambda)$-ea on multimodal fitness landscapes. In Beyer, H.G., ed.: Genetic and evolutionary computation conference (GECCO), ACM Press, New York (2005) 865–872
6. De Jong, K.A.: An analysis of the behavior of a class of genetic adaptive systems. PhD thesis, University of Michigan (1975)
7. Ehrgott, M.: Multicriteria Optimization. 2 edn. Springer, Berlin (2005)
8. Okabe, T., Jin, Y., Olhofer, M., Sendhoff, B.: On Test Functions for Evolutionary Multi-objective Optimization. In Yao, X., et al., eds.: Parallel Problem Solving from Nature (PPSN), Springer, Berlin (2004) 792–802
9. Zhou, A., Zhang, Q., Jin, Y., Tsang, E., Okabe, T.: A model-based evolutionary algorithm for bi-objective optimization. In: Congress on Evolutionary Computation (CEC), IEEE Press, Piscataway NJ (2005) 2568–2575
10. Emmerich, M., Beume, N., Naujoks, B.: An EMO algorithm using the hypervolume measure as selection criterion. In Coello, C.A.C., et al., eds.: Evolutionary Multi-Criterion Optimization (EMO), Springer, Berlin (2005) 62–76
11. Naujoks, B., Beume, N., Emmerich, M.: Multi-objective optimisation using S-metric selection: Application to three-dimensional solution spaces. In: Congress on Evolutionary Computation (CEC), IEEE Press, Piscataway NJ (2005) 1282–1289
12. Beume, N.: Hypervolumen-basierte Selektion in einem evolutionären Algorithmus zur Mehrzieloptimierung. Diploma thesis, University of Dortmund (2006)