Hypervolume-based Metaheuristics for Multiobjective Optimization

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

der Technischen Universität Dortmund
an der Fakultät für Informatik

von

Nicola Beume

Dortmund
2011

Tag der mündlichen Prüfung:
20.12.2011

Dekanin:
Prof. Dr. Gabriele Kern-Isberner

Gutachter:
Prof. Dr. Günter Rudolph, Technische Universität Dortmund
Prof. Dr. Christian Igel, University of Copenhagen

Abstract

The purpose of multiobjective optimization is to find solutions that are optimal regarding several goals. In the branch of vector or Pareto optimization all these goals are considered to be of equal importance, so that compromise solutions that cannot be improved regarding one goal without deteriorating in another are Pareto-optimal.
A variety of quality measures exist to evaluate approximations of the Pareto-optimal set generated by optimizers, wherein the hypervolume is the most significant one, making the hypervolume calculation a core problem of multiobjective optimization. This thesis tackles that challenge by providing a new hypervolume algorithm from computational geometry and analyzing the problem's computational complexity.
Evolutionary multiobjective optimization algorithms (EMOA) are state-of-the-art methods for Pareto optimization, wherein the hypervolume-based algorithms belong to the most powerful ones, among them the popular SMS-EMOA. After its promising capabilities have already been demonstrated in first studies, this thesis is dedicated to deeper understand the underlying optimization process of the SMS-EMOA and similar algorithms, in order to specify their performance characteristics. Theoretical analyses are accomplished as far as possible with established and newly developed tools. Beyond the limitations of rigorous scrutiny, insights are gained via thorough experimental investigation. All considered problems are continuous, whereas the algorithms are as well applicable to discrete problems.
More precisely, the following topics are concerned. The process of approaching the Pareto-optimal set of points is characterized by the convergence speed, which is analyzed for a general framework of EA with hypervolume selection on several classes of bi-objective problems. The results are achieved by a newly developed concept of linking single and multiobjective optimization. The optimization on the Pareto front, that is turning the population into a set with maximal hypervolume, is considered separately, focusing on the question under which circumstances the steady-state selection of exchanging only one population member suffices to reach a global optimum. We answer this question for different bi-objective problem classes.
In a benchmarking on so-called many-objective problems of more than three objectives, the qualification of the SMS-EMOA is demonstrated in comparison to other EMOA, while also studying their cause of failure. Within the mentioned examinations, the choice of the hypervolume's reference point receives special consideration by exposing its influence. Beyond the study of the SMS-EMOA with default setup, it is analyzed to what extent the performance can be improved by parameter tuning of the EMOA anent to certain problems, focusing on the influence of variation operators. Lastly, an optimization algorithm based on the gradient of the hypervolume is developed and hybridized with the SMS-EMOA.

Contents

Preface

Motivation

As known from every day life, people typically do not only have just one wish but several—that may even contradict. In the context of optimization, this means that a solution to a problem shall not just be optimized regarding one goal but fulfill several demands simultaneously. So, considering a real-world problem as a single-objective one does in the best case stem from modesty or priority setting but may be inadequate and a sign of too abstract modeling and over-simplification. The field of multiobjective optimization contrarily provides adequate tools to handle real-world problems canonically by respecting multiple objectives at once.
By the approach of Pareto or vector optimization the solution's quality is not a single function value but given by a vector containing one value for each objective. As all objectives are treated to be of equal importance, the resolution of a multi-objective problem is a set of Pareto-optimal solutions, containing those solutions which cannot be improved in one objective without worsening in another, thus each representing an optimal compromise. The Pareto-optimal set is typically too large to be captured completely or can practically not be exactly attained in continuous objective spaces. Thus, optimizers can only be expected to achieve an approximation of the Pareto-optimal set. To this end, it is an important task to compare the quality of such approximations. Several quality measures exist, wherein the hypervolume indicator is the most significant and suitable one, making it a key aspect of multiobjective optimization. The hypervolume measures the size of the objective space consisting of worse points than the considered set, so that capturing the objective space up to the optimal set corresponds to its maximization. The space is bounded by a reference point which is a parameter of the hypervolume. The hypervolume calculation is time-consuming since the computation time grows exponentially with the dimension of the objective space.

> The hypervolume computation as the core problem of multiobjective optimization is tackled in this thesis by analyzing the hypervolume complexity, giving a simple hypervolume algorithm, and dealing with the choice of its reference point.

Evolutionary multiobjective optimization algorithms (EMOA) developed to state-of-the-art methods for Pareto optimization. Evolutionary algorithms (EA) are

randomized search heuristics that are inspired by the notion of the natural evolution as a process of stepwise improvement through variation and selection. They are especially suitable for complex, badly understood real-world problems since they do not require any problem information such as gradients but handle the problem as a black box. Due to their effectiveness and universal applicability, they gained more and more acceptance in industrial applications, especially the EMOA for multiobjective optimization.
In principle, EMOA can consist of operators developed for single-objective optimization, except for the selection. The selection decides about which solutions to keep or to discard for the next iteration. This requires that the solutions are comparable, which is not directly possible in the multiobjective case since the solutions are vectors. When one solution is better in one and worse in another objective than a second solution, they cannot be sorted. The recent approach is to invoke quality measures to evaluate the current approximation of the EMOA, resulting in the so-called indicator-based EMOA.
Among these algorithms is the SMS-EMOA, where SMS stands for S-metric (a synonym for hypervolume) selection. As it is agreed that an optimizer shall generate an approximation of the optimal set with a high hypervolume value, the idea to directly aspire the hypervolume maximization within the EMOA seems natural and doing this in the selection solves the problem of how to handle incomparable solutions: A solution's quality is evaluated by its contribution to the hypervolume of the current approximation. This is a single-objective problem enabling the solutions to be sortable.
Nowadays, the SMS-EMOA is a popular and accepted algorithm which is in the research community acknowledged by nearly 300 citations (due to Google Scholar (2011)) of the initial conference publication and the journal article, and for example in Germany reflected by its recommendation in the guidelines of the Association of German Engineers (VDI) VDI-Fachbereich Bionik (2011). Yet, although it took just one day to develop SMS-EMOA and it has been competitive right from the start, it is not obvious why and it took years to understand its optimization process.

> This thesis is dedicated to the deeper understanding of the optimization process underlying hypervolume-based EMOA like the SMS-EMOA and to specify their performance characteristics.

Overview and Format

The introductory Chapter 1 gives fundamental concepts on a basic level also amendable for readers unfamiliar with the topic. The two main chapters and each section start with specific and technically detailed preliminaries for the respective subject.

Chapter 2 deals with the calculation of the hypervolume, starting with a comprehensive overview of its properties including results from literature in simplified notation and additional insights (Sec. 2.1). Our main contributions, the lower bound of the problem complexity and the upper bound as an algorithm are given in the Sections 2.2 and 2.3.

The characteristics of hypervolume-based EMOA are studied in Chapter 3. It focuses on the SMS-EMOA and starts by recapitulating its concept and results of previous studies, next to an overview of other EMOA considered in comparison (Sec. 3.1). First, the convergence towards the Pareto front is analyzed separately (Sec. 3.2), followed by a study of the selection behavior on the Pareto front (Sec. 3.3). In Sec. 3.4, the whole process of the SMS-EMOA and other optimizers is benchmarked on problems with many objectives. The performance influence of tuning the parameters—especially the variation—of SMS-EMOA is studied in Section 3.5. Finally, a hybridization of the SMS-EMOA and a hypervolume gradient technique is presented (Sec. 3.6).

Chapter 2 and each section in Chapter 3 close with a summary of the key results, and a discussion on their impact and remaining open problems as conclusions. The final Chapter 4 recapitulates in a similar way the whole thesis while taking a broader view.

Nearly all results of this dissertation have already been released in publications I (co-)authored. References to these papers are marked with a star (e.g. Beume et al. (2010)*). Parts of these peer-reviewed publications are reproduced here literally. For precise referencing and complaisant readability, these parts are indicated by blue font and the cited original publication is referenced in the beginning of the respective section. A reference is given at the end of each published theorem, for example as Beume et al. (2010)* if the statement has not been changed and by (cf. Beume et al. (2010)*) in case the result has been altered or extended. For figures and tables, the component's name in blue font (e.g. Fig. 2.10) indicates that it has been adopted unchanged.

Throughout this thesis the term *hypervolume* replaces the deprecated synonym *S-metric*.

For easier readability, vectors are only marked as transposed in case of vector calculations where it is important whether a vector is a row vector or a column vector. A column vector is denoted by $\mathbf{x}$ and its transposed row vector by $\mathbf{x}^\top$. The terms *vector* or *point* are used synonymously. Likewise, we consequently distinguish between sets and multisets in formal statements but often use the term *set* in descriptive text for convenience. The abbreviations EA and EMOA shall represent the singular as well as the plural form.

List of Underlying Publications

The publications are listed in their order of appearance in this thesis. My contribution to each joint publication of k authors can be quantified as a fraction of about $1/k$.

1. N. Beume, C. M. Fonseca, M. López-Ibáñez, L. Paquete, and J. Vahrenhold. On the complexity of computing the hypervolume indicator. *IEEE Transactions on Evolutionary Computation*, 13(5):1075–1082, 2009.

   The lower bound developed in this paper, conjointly with Carlos M. Fonseca and Jan Vahrenhold, is described in Section 2.2, and a preposed lemma in Section 2.1.

2. N. Beume. S-metric calculation by considering dominated hypervolume as Klee's measure problem. *Evolutionary Computation*, 17(4):477–492, 2009.

   The results are reproduced in Section 2.3.

3. N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.

   This article summarizes the highlights of our publications Emmerich et al. (2005)*, Naujoks et al. (2005a)*, Naujoks et al. (2005b)* and is briefly reviewed in Section 3.1.

4. N. Beume, B. Naujoks, and G. Rudolph. SMS-EMOA: Effektive evolutionäre Mehrzieloptimierung. *at-Automatisierungstechnik*, 56(7):357–364, 2008.

   The article is briefly reviewed in Section 3.1. It is an extension of:

   N. Beume, B. Naujoks, and G. Rudolph. Mehrkriterielle Optimierung durch evolutionäre Algorithmen mit S-Metrik-Selektion. In R. Mikut and M. Reischl, editors, *Proc. of the 16. Workshop Computational Intelligence*, pages 1–10. Universitätsverlag Karlsruhe, Karlsruhe, 2006. **Young Author Award**.

5. N. Beume, M. Laumanns, and G. Rudolph. Convergence rates of (1+1) evolutionary multiobjective algorithms. In R. Schaefer et al., editors, *Proc. of the 11th International Conference on Parallel Problem Solving from Nature – PPSN XI: Part I*, volume 6238 of *Lecture Notes in Computer Science*, pages 597–606. Springer, Berlin, 2010. **Best Student Paper Award**.

   The results are contained in Section 3.2.

6. N. Beume, M. Laumanns, and G. Rudolph. Convergence rates of SMS-EMOA on continuous bi-objective problem classes. In H.-G. Beyer and W. B. Langdon, editors, *Proc. of the 11th ACM SIGEVO workshop on Foundations*

of Genetic Algorithms (FOGA 2011), pages 243–252. ACM, New York, NY, 2011.

The results are as well contained in Section 3.2.

7. N. Beume, B. Naujoks, M. Preuss, G. Rudolph, and T. Wagner. Effects of 1-greedy S-metric-selection on innumerably large pareto fronts. In M. Ehrgott et al., editors, *Proc. of the 5th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2009)*, volume 5467 of *Lecture Notes in Computer Science*, pages 21–35. Springer, Berlin, 2009.

   Section 3.3 is based on this publication.

8. T. Wagner, N. Beume, and B. Naujoks. Pareto-, aggregation-, and indicator-based methods in many-objective optimization. In S. Obayashi et al., editors, *Proc. of the 4th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, volume 4403 of *Lecture Notes in Computer Science*, pages 742–756. Springer, Berlin, 2007.

   Section 3.4 is based on this publication.

9. S. Wessing, N. Beume, G. Rudolph, and B. Naujoks. Parameter tuning boosts performance of variation operators in multiobjective optimization. In R. Schaefer et al., editors, *Proc. of the 11th International Conference on Parallel Problem Solving from Nature – PPSN XI: Part I*, volume 6238 of *Lecture Notes in Computer Science*, pages 728–737. Springer, Berlin, 2010.

   Section 3.5 is based on this publication.

10. M. Emmerich, A. Deutz, and N. Beume. Gradient-based/evolutionary relay hybrid for computing pareto front approximations maximizing the S-Metric. In T. Bartz-Beielstein et al., editors, *Proc. of the 4th International Workshop on Hybrid Metaheuristics (HM 2007)*, volume 4771 of *Lecture Notes in Computer Science*, pages 140–157. Springer, Berlin, 2007.

    The results are reproduced in Section 3.6 and a lemma is preposed to Section 2.1.

11. P. Koch, O. Kramer, G. Rudolph, and N. Beume. On the hybridization of SMS-EMOA and local search for continuous multiobjective optimization. In F. Rothlauf, editor, *Proc. of the 11th Annual conference on Genetic and Evolutionary Computation (GECCO 2009)*, pages 603–610. ACM, New York, NY, 2009.

    The results are briefly reviewed in Section 3.6.

12. T. Voß, N. Beume, G. Rudolph, and C. Igel. Scalarization versus indicator-based selection in multi-objective CMA evolution strategies. In J. Wang

et al., editors, *Proc. of the 2008 IEEE Congress on Evolutionary Computation (CEC 2008) within the 2008 IEEE World Congress on Computational Intelligence*, pages 3041–3048. IEEE Press, Piscataway, NJ, 2008.

The results are briefly reviewed in Section 3.6.

Software

The described algorithms are available in several programming languages at:

`http://www.hypervolume.org`

Acknowledgments

First of all, I would like to express my gratitude to my supervisor Günter Rudolph for sharing his expertise in uncountable professional discussions, his strong interest in my work, and helpful support of this thesis. I thank Christian Igel for agreeing to examine this thesis and his helpful comments. Thanks to my co-authors of the publications that underlie this document for the fruitful cooperations, namely André Deutz, Michael Emmerich, Carlos M. Fonseca, Christian Igel, Patrick Koch, Oliver Kramer, Marco Laumanns, Manuel López-Ibáñez, Boris Naujoks, Luís Paquete, Mike Preuß, Günter Rudolph, Jan Vahrenhold, Thomas Voß, Tobias Wagner, and Simon Wessing, as well as further co-authors of publications not included here. Moreover, I am grateful for the friendly and inspiring atmosphere at the Chair of Algorithm Engineering (LS11). Finally, I heartily thank my beloved husband Thorsten for accompanying and supporting me throughout this work, and for everything else.

1 Introduction

The fundamental concepts of the discipline are presented on a basic level amendable to non-familiar readers, referring to subsequent chapters where the knowledge is required and further literature. The basic definitions of multiobjective optimization follow (Sec. 1.1), next to the ideas of how to measure the quality of approximations of the optimal set (Sec. 1.2). Evolutionary algorithms are introduced as suitable optimizers for multiobjective optimization (Sec. 1.3). In Section 1.4, we distinguish the two computation models used in this work, and describe our approach of experimental test scenarios in Section 1.5.

1.1 Multiobjective Optimization

Multiobjective or multicriteria optimization can be considered as a canonical approach to solve real-world problems. As experienced in every day life, a solution to a problem shall not just be optimized regarding one goal but fulfill several demands. Two main approaches can be distinguished, namely the *a priori* and the *a posteriori* approaches. The a priori techniques (cf. Miettinen (1999)) require a weighting of the problem's objectives prior to the optimization. Summing up the weighted objectives results in a single-objective function with the desired solution as the optimum. Thereby, user preferences can be respected by the optimizer in order to generate a solution with the favored properties. A drawback is that it is usually hard for the user to express her wishes in weights, and to specify the intended ratio before knowing which compromises are possible among objectives. Moreover, not all compromises of objectives are reachable with the described weighting.
Contrarily, a posteriori approaches (cf. Miettinen (1999)) like the *vector* or *Pareto optimization*[1] consider all objectives to be of equal importance, thus all compromise solutions that cannot be improved in one objective without worsening in another are considered to be optimal, thus resulting in a set of optimal solutions. The optimizer approximates the optimal set, so that the user (or decision maker) can decide based on this knowledge which compromise solution fits best to her interests. This approach is used in this thesis, where we deal with the generation of the approximation but not with the decision of a solution. The basic concepts are detailed in the following definitions. For a comprehensive introduction and overview, see e.g. Miettinen (1999), Deb (2001), Coello Coello et al. (2002), or Ehrgott (2005).

[1] Note that the nomenclature in the literature is ambiguous: it is also common to treat multiobjective optimization and Pareto optimization as synonyms.

The following definitions allow us to distinguish between sets and multisets where necessary.

Definition 1.1 *Let $A = \{\{a^{(1)}, a^{(2)}, \ldots, a^{(m)}\}\}$ denote a multiset of size m, so A may contain copies, i.e., $a^{(i)} = a^{(j)}$, is possible with $i \neq j$ and $i, j \in \{1, \ldots, m\}$, or otherwise A is a set.*

Definition 1.2 *The operator* set *denotes the transformation of a multiset into a set by removing copies, while keeping one element of each subset of equal elements.*

We consider unconstrained multiobjective optimization problems in real-valued spaces.

Definition 1.3 *An unconstrained multiobjective optimization problem is defined as $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$, with $f : \mathbb{R}^n \to \mathbb{R}^d$, $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_d(\mathbf{x}))$, $f_i : \mathbb{R}^n \to \mathbb{R}, i \in \{1, \ldots, n\}$.*
Let the domain of f be denoted as $S \subset \mathbb{R}^n$ and the co-domain be the multiset $Z \in \mathcal{M}(\mathbb{R}^d)$ with elements from $\mathbb{R}^d$.

We notate optimization problems as minimization problems. They can be transferred into equivalent maximization problems since $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = -\max_{\mathbf{x}\in\mathbb{R}^n}\{-f(\mathbf{x})\}$.
The following relations among vectors are applied to relate solutions of multiobjective problems to each other in order to compare their quality.

Definition 1.4 *Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ denote two d-dimensional points (or vectors).*

1. $\mathbf{a} \preceq \mathbf{b}$, $\mathbf{a}$ *weakly dominates* $\mathbf{b} :\Longleftrightarrow$
   $\forall i \in \{1, \ldots, d\} : a_i \leq b_i$

2. $\mathbf{a} \prec \mathbf{b}$, $\mathbf{a}$ *dominates* $\mathbf{b} :\Longleftrightarrow$
   $\mathbf{a} \preceq \mathbf{b}$ *and* $\mathbf{a} \neq \mathbf{b}$, *i.e.,* $\exists i \in \{1, \ldots, d\} : a_i < b_i$

3. $\mathbf{a} \prec\!\prec \mathbf{b}$, $\mathbf{a}$ *strictly dominates* $\mathbf{b} :\Longleftrightarrow$
   $\forall i \in \{1, \ldots, d\} : a_i < b_i$

4. $\mathbf{a} \| \mathbf{b}$, $\mathbf{a}$ *and* $\mathbf{b}$ *are incomparable*: $\Longleftrightarrow$
   neither $\mathbf{a} \preceq \mathbf{b}$ *nor* $\mathbf{b} \preceq \mathbf{a}$.

5. $\mathbf{a} \preceq \mathbf{b} \iff \mathbf{b} \succeq \mathbf{a}$, *analogously for* $\succ, \succ\!\succ$

6. $\mathbf{a} \not\preceq \mathbf{b} \iff \mathbf{a} \preceq \mathbf{b}$ *is not true, i.e.,* $\mathbf{b} \preceq \mathbf{a}$ *or* $\mathbf{a} \| \mathbf{b}$.
   $\mathbf{a} \not\prec \mathbf{b} \iff \mathbf{a} \prec \mathbf{b}$ *is not true, i.e.,* $\mathbf{a} = \mathbf{b}$ *or* $\mathbf{b} \preceq \mathbf{a}$ *or* $\mathbf{a} \| \mathbf{b}$.
   $\mathbf{a} \not\prec\!\prec \mathbf{b} \iff \mathbf{a} \prec\!\prec \mathbf{b}$ *is not true, i.e.,* $\mathbf{a} \preceq \mathbf{b}$ *or* $\mathbf{b} \preceq \mathbf{a}$ *or* $\mathbf{a} \| \mathbf{b}$.
   $\mathbf{a} \nparallel \mathbf{b} \iff \mathbf{a} \| \mathbf{b}$ *is not true, i.e.,* $\mathbf{a} \preceq \mathbf{b}$ *or* $\mathbf{b} \preceq \mathbf{a}$.

Lemma 1.5 *A hierarchy of dominance relations exists by definition:*

$$\mathbf{a} \prec\prec \mathbf{b} \quad \Rightarrow \quad \mathbf{a} \prec \mathbf{b} \quad \Rightarrow \quad \mathbf{a} \preceq \mathbf{b}.$$

The minimal or best elements regarding the Pareto dominance are considered anent to an optimization problem or within sets or multisets.

Definition 1.6 *The minimal elements of a multiobjective optimization problem $f : S \to Z$ regarding the partial order induced by the Pareto dominance are called* Pareto-optimal. *Formally, we denote the multiset of optima in the objective space as $PF(f) \subseteq Z$ as the Pareto front*

$$PF(f) = \{\{\mathbf{v} \in Z \mid \nexists \mathbf{a} \in Z : \mathbf{a} \prec \mathbf{v}\}\},$$

and the set of optimizers $PS(f) \subseteq S$ in the search space as the Pareto set

$$PS(f) = \{\mathbf{x} \in S \mid \nexists \mathbf{y} \in S : f(\mathbf{y}) \prec f(\mathbf{x})\} = \{\mathbf{x} \in S \mid f(\mathbf{x}) \in PF(f)\}.$$

Definition 1.7 *A set $A \subset \mathbb{R}^d$ of points that are pairwisely incomparable regarding the $\prec$ relation, formally $\forall \mathbf{a}, \mathbf{b} \in A : \mathbf{a} \not\prec \mathbf{b}$ and $\mathbf{b} \not\prec \mathbf{a}$, is called a non-dominated set or an antichain regarding $\prec$.*
A non-dominated multiset can be generated from an arbitrary multiset $B \subset \mathbb{R}^d$ by removing dominated elements:

$$ndms(B) = \{\{\mathbf{b} \in B \mid \nexists \mathbf{a} \in B : \mathbf{a} \prec \mathbf{b}\}\},$$

and a non-dominated set can be generated by additionally removing copies:

$$nds(B) = \mathsf{set}\{\{\mathbf{b} \in B \mid \nexists \mathbf{a} \in B : \mathbf{a} \prec \mathbf{b}\}\}.$$

Note that the term *non-dominated set* is commonly used in literature assuming that copies in sets are unlikely in continuous spaces, whereas the term *non-dominated multiset* is newly introduced here.

The nadir point of a set constitutes of the worst coordinate values in the set, respectively in each dimension. It describes the smallest upper bound of a set, in the sense that it is the point which is weakly dominated by each point of the set and there is no point dominating this one with this property. Analogously the ideal point is the largest lower bound of a multiset, i.e., the point that weakly dominates each element of the multiset. The basic definitions above are illustrated in Fig. 1.1.

Definition 1.8 *The nadir point* **nad** *of a multiset $A \in \mathcal{M}(\mathbb{R}^d)$ describes the smallest upper bound of the multiset:*

$$\mathbf{nad}(A) := (z_1, \ldots, z_d), \text{ with } z_i = \max\{a_i \mid \mathbf{a} \in A\}, i \in \{1, \ldots, d\}.$$

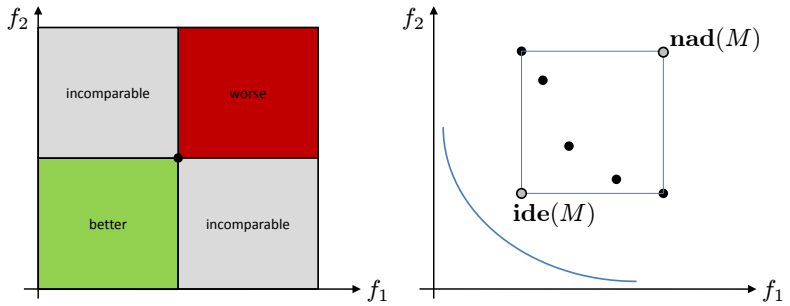


Fig. 1.1: Left: Dominance relation of a point in a 2-dimensional space. Right: An example of a Pareto front (blue line), an approximating set M (black points), and its nadir and ideal point.

Definition 1.9 *The ideal point* **ide** *of a multiset* $A \in \mathcal{M}(\mathbb{R}^d)$ *describes the largest lower bound of the multiset:*

$$\mathbf{ide}(A) := (z_1, \ldots, z_d), \text{ with } z_i = \min\{a_i \mid \mathbf{a} \in A\}, i \in \{1, \ldots, d\}.$$

Next, the dominance relation among vectors is transferred to sets or multisets of vectors. The relations $\prec$ and $\prec\prec$ are defined counter-intuitively in literature, so they are omitted here as we do not need them.

Definition 1.10 *Let* A, B *denote two non-empty finite multisets with elements from* $\mathbb{R}^d$.

1. $A \preceq B$, *A weakly dominates* $B :\iff$
   $\forall \mathbf{b} \in B : \exists \mathbf{a} \in A : \mathbf{a} \preceq \mathbf{b}$.

2. $A \parallel B$, *A and B are incomparable* $:\iff$
   neither $A \preceq B$ *nor* $B \preceq A$.

It follows a classification of relations.

Definition 1.11 *A binary relation* R *is a preorder (or quasi-order) over a set* M, *if* $\forall a, b, c \in M$ *it fulfills:*

1. $a \mathsf{R} a$ *(reflexivity)*

2. $a \mathsf{R} b \ \wedge \ b \mathsf{R} c \ \Rightarrow \ a \mathsf{R} c$ *(transitivity)*

Definition 1.12 *A binary relation* R *is a partial order (or antisymmetric preorder) over a set* M, *if it is a preorder over* M *and* $\forall a, b \in M$ *it fulfills:*

$$a \mathsf{R} b \ \wedge \ b \mathsf{R} a \ \Rightarrow \ a = b \quad \textit{(antisymmetry)}$$

Definition 1.13 *A binary relation* R *is a strict (or irreflexive) partial order over a set* M*, if* $\forall a, b, c \in M$ *it fulfills transitivity as in Def. 1.11 and:*

1. $a \not\mathsf{R}\ a$ *(irreflexivity)*
2. $a\ \mathsf{R}\ b\ \Rightarrow\ b \not\mathsf{R}\ a$ *(asymmetry)*

With the definitions above, we characterize the dominance relations among vectors, sets, or multisets to gain a deeper understanding of the relations.

Corollary 1.14 *The relation* $\preceq$ *is a partial order over* $\mathbb{R}^d$*.*

Proof. $\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^d$ hold with $i \in \{1, \ldots, d\}$:

1. Reflexivity: $\mathbf{a} \preceq \mathbf{a}$, since $\forall i : a_i \leq a_i$.
2. Transitivity: $\mathbf{a} \preceq \mathbf{b} \wedge \mathbf{b} \preceq \mathbf{c} \Rightarrow \mathbf{a} \preceq \mathbf{c}$, since $\forall i : a_i \leq b_i \wedge b_i \leq c_i \Rightarrow a_i \leq c_i$.
3. Antisymmetry: $\mathbf{a} \preceq \mathbf{b} \wedge \mathbf{b} \preceq \mathbf{a} \Rightarrow \mathbf{a} = \mathbf{b}$, as $\forall i : a_i \leq b_i \wedge b_i \leq a_i \Rightarrow a_i = b_i$.

□

Corollary 1.15 *The relations* $\prec$ *or* $\prec\!\!\prec$ *form a strict (or irreflexive) partial order over* $\mathbb{R}^d$*.*

Proof. Transitivity holds as in Cor. 1.14.2. $\forall \mathbf{a} \in \mathbb{R}^d$ hold with $i \in \{1, \ldots, d\}$:

1. Irreflexivity: $\mathbf{a} \not\prec \mathbf{a}$ and $\mathbf{a} \not\prec\!\!\prec \mathbf{a}$.
   Since $\nexists i : a_i < a_i$, neither $\mathbf{a} \prec \mathbf{a}$ nor $\mathbf{a} \prec\!\!\prec \mathbf{a}$ hold.
2. Asymmetry: $\mathbf{a} \prec \mathbf{b} \Rightarrow \mathbf{b} \not\prec \mathbf{a}$ and $\mathbf{a} \prec\!\!\prec \mathbf{b} \Rightarrow \mathbf{b} \not\prec\!\!\prec \mathbf{a}$.
   Since $\mathbf{a} \prec \mathbf{b}$ with $\exists i : a_i < b_i$ holds, so $\mathbf{b} \prec \mathbf{a}$ with $\forall i : b_i \leq a_i$ is not true. Analogously, since $\mathbf{a} \prec\!\!\prec \mathbf{b}$, with $\forall i : a_i < b_i$ holds, so $\mathbf{b} \prec\!\!\prec \mathbf{a}$ with $\forall i : b_i < a_i$ is not true. □

Corollary 1.16 *The relation* $\preceq$ *is a preorder over subsets from* $\mathbb{R}^d$ *as well as over* $\mathcal{M}(\mathbb{R}^d)$*, the set of all multisets with elements from* $\mathbb{R}^d$*, but over both not a partial order.*

Proof. 1. and 2. are proved for multisets, holding as well for sets, whereas the negative property 3. is proved for sets, holding as well for multisets. $\forall$ multisets $A, B, C \in \mathcal{M}(\mathbb{R}^d)$ hold with $i \in \{1, \ldots, d\}$:

1. Reflexivity: $A \preceq A$, since $\forall \mathbf{a} \in A : \exists \mathbf{a} \in A : \mathbf{a} \preceq \mathbf{a}$.
2. Transitivity: $A \preceq B \wedge B \preceq C \Rightarrow A \preceq C$.
   It holds $\forall \mathbf{b} \in B : \exists \mathbf{a} \in A : \mathbf{a} \preceq \mathbf{b}$, so $\forall i : a_i \leq b_i$, and $\forall \mathbf{c} \in C : \exists \mathbf{b} \in B : \mathbf{b} \preceq \mathbf{c}$, so $\forall i : b_i \leq c_i$. It follows $\forall \mathbf{c} \in C : \exists \mathbf{a} \in A : \mathbf{a} \preceq \mathbf{c}$, so $\forall i : a_i \leq c_i$.

3. No antisymmetry: $A \preceq B \wedge B \preceq A \not\Rightarrow A = B$.
   Antisymmetry does not hold for $\preceq$ on sets including dominated points. Let $A, B \subset \mathbb{R}^d$ with $B = A \cup \{\mathbf{a}\}$ and $A \preceq \{\mathbf{a}\}$, $\mathbf{a} \notin A$. Then mutually weak dominance holds while B contains an additional dominated point. As an additional argument, antisymmetry does not hold for $\preceq$ on multisets containing copies. Let $A, B \in \mathcal{M}(\mathbb{R}^d)$ with $B = A \cup \{\mathbf{a}\}$ and $\mathbf{a} \in A$. Then mutually weak dominance holds while the sets are not equal. □

Corollary 1.17 *The relation $\preceq$ is a partial order over the non-dominated sets of $\mathcal{M}(\mathbb{R}^d)$.*

Proof. The proofs of reflexivity and transitivity are analogous to Cor. 1.16. $\forall$ multisets $A, B, C \in \mathcal{M}(\mathbb{R}^d)$, thus $nds(A), nds(B), nds(C) \subset \mathbb{R}^d$ hold with $i \in \{1, \ldots, d\}$:

1. Reflexivity: $nds(A) \preceq nds(A)$.

2. Transitivity: $nds(A) \preceq nds(B) \wedge nds(B) \preceq nds(C) \;\Rightarrow\; nds(A) \preceq nds(C)$.

3. Antisymm.: $nds(A) \preceq nds(B) \wedge nds(B) \preceq nds(A) \;\Rightarrow\; nds(A) = nds(B)$.
   Proof by contradiction: Assume $nds(A) = nds(B)$ does not hold. Then $\exists \mathbf{a} \in A$ with $\mathbf{a} \notin B$. Since $nds(B) \preceq nds(A)$ there $\exists \mathbf{b} \in B : \mathbf{b} \preceq \mathbf{a}$ with $\mathbf{a} \neq \mathbf{b}$. Then, since $nds(A) \preceq nds(B)$ there $\exists \mathbf{c} \in A : \mathbf{c} \preceq \mathbf{b}$. Due to transitivity with $\forall i : c_i \leq b_i \leq a_i$ and $\mathbf{a} \neq \mathbf{c}$ as $\mathbf{a} \neq \mathbf{b}$, it follows $\mathbf{c} \prec \mathbf{a}$. This contradicts with the definition of $nds(A)$. □

1.2 Quality Assessment

To evaluate the quality of an approximation to Pareto-optimal points, the approaching is normally considered in the objective space towards the Pareto front. Approximating the Pareto set is an even more ambitious aim as one point in the objective space may have multiple preimages in the search space so that the Pareto set is larger. It is commonly agreed that the Pareto front approximation is the primary goal, although the Pareto set is as well desired, e.g. to have several different solutions with equal quality to choose from.
The definitions in Section 1.1 allow to relate approximations to each other, e.g. a set that dominates another is a better approximation. Yet, the typical case is that the approximations are incomparable. Since the Pareto dominance is not a total order, not all points are comparable. A point located in the middle of a 2-dimensional objective space dominates 1/4 of it, is dominated by 1/4, and is incomparable to 1/2 of the objective space. So it is comparable with 1/2 of the space, and this fraction decreases exponentially with increasing dimension of the objective space

to 2^{-d+1} in d dimensions. As the comparability of sets requires the comparability of their elements, approximations are typically incomparable so that the concept of Pareto dominance is insufficient. Yet a quality assessment is desired to evaluate the performance of optimizers. So an ordering is established artificially with the help of *quality measures or indicators* based on quality aspects additional to the dominance.
A quality measure or indicator is a mapping of a multiset with elements from $\mathbb{R}^d$ to $\mathbb{R}$, a definition that leaves arbitrary freedom which quality aspect to represent. The following informal properties are esirable and often demanded from quality indicators, cf. (Deb, 2001, Ch. 8.1).

Axioms 1.18 *A quality measure or indicator $\alpha : \mathcal{M}(\mathbb{R}^d) \to \mathbb{R}$ shall respect the following aspects.*

1. *Convergence: The approximation shall minimize the distance to the Pareto front.*
2. *Diversity: The approximation shall be spread along the whole Pareto front and be well distributed.*
3. *Cardinality: The size of the approximating set shall be appropriate.*

The last point is discussed contrarily. Of course a larger set approximates an infinitely large set better than a smaller. Despite that, it is agreed that the aim of an optimizer is not to generate the whole Pareto front because of too much required resources, but to determine a good representing set.
The indicators we consider are so-called unary measures as they evaluate exactly one set of points. There also exist measures to evaluate two or more sets relative to each other. This has the drawback that the results are not comparable with sets that have not been invoked in that comparison. So, unary measures are more suitable for benchmarks as they enable to comprise earlier or supplementary studies.

The total order the quality measures establish shall be conform with the Pareto dominance relations as far as possible. This demand is specified with the terms of completeness and compatibility regarding a relation, cf. Zitzler et al. (2003). Concretely, for a comparable pair of sets with e.g. $A \preceq B$ and $B \npreceq A$ the quality measure shall evaluate the pair such that A is not inferior to B.

Definition 1.19 *Let $A, B \in \mathcal{M}(\mathbb{R}^d)$ denote finite multisets. A quality measure $\alpha : A \to \mathbb{R}$ is complete regarding a relation $\mathsf{R} \subseteq A \times B$ if: $\forall A, B : A \mathsf{R} B \Rightarrow \alpha(A)$ better than $\alpha(B)$.*

The total order induced by the measure shall not contradict with the partial order of the relation. The better set regarding the measure shall as well be better or at least not worse regarding the relation.

Definition 1.20 *Let $A, B \in \mathcal{M}(\mathbb{R}^d)$ denote finite multisets. A quality measure $\alpha : A \to \mathbb{R}$ is* compatible *regarding a relation $\mathsf{R} \subseteq A \times B$ if: $\forall A, B : A \mathsf{R} B \Leftarrow \alpha(A)$ better than $\alpha(B)$. It is* weakly compatible *regarding R if: $\forall A, B : A \mathsf{R} B \Leftarrow \alpha(A)$ not worse than $\alpha(B)$.*

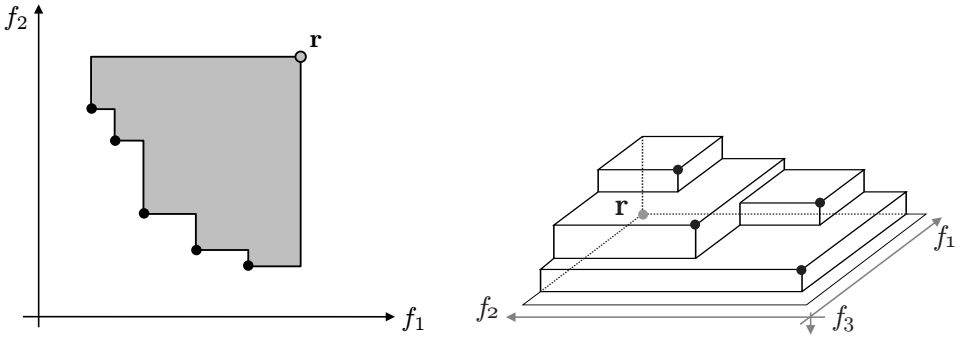


Fig. 1.2: Left: Illustration of the 2-dimensional hypervolume (gray area) of a set (black points), bounded by the reference point **r**. Right: Hypervolume illustration in a 3-objective space.

Several quality indicators exist and are in usage. Many concentrate either mainly on the diversity or on the convergence. A challenge is to develop measures that evaluate the quality without requiring the knowledge of the Pareto front. Typically convergence measures calculate the distance of the approximation to the Pareto front and are therefore only applicable for test problems where the user is equipped with this knowledge.

The *hypervolume (or S-metric)* by Zitzler and Thiele (1998) is the most significant and accepted quality measure in multiobjective research. It measures the size of the space containing points that are dominated by at least one member of the considered set, and is bounded by a reference point (cf. Fig. 1.2). The hypervolume respects all aspects of Axioms 1.18, does not require knowledge of the Pareto front, and is the only known measure which fulfills completeness and compatibility w. r. t. the Pareto dominance relations in the best possible way. A formal definition and detailed properties are given in Chapter 2 followed by the author's contributions.

It is recommended to use several indicators to evaluate the quality of approximation in order to emphasize different aspects. Further indicators used in our experimental analyses are described in Chapter 3. A comprehensive introduction to quality indicators can be found in Knowles et al. (2006), and an in-depth overview of quality indicators and their properties is given in Zitzler et al. (2003) and Zitzler et al. (2008a).

1.3 Evolutionary Algorithms

Evolutionary Algorithms (EA) can be classified according to several methodologies. EA are *randomized direct search meta-heuristics for black-box optimization.* We explain the terms in an informal way to just sketch the concepts. Search heuristics (see e.g. Schwefel (1995)) do not algorithmically calculate an optimal solution but iteratively try valid solutions of unknown quality in order to stepwisely proceed to high quality solutions of the optimization problem. The search space is formally defined so that arbitrary solutions can be generated. Many approaches require information on the gradient of the problem to decide about promising search directions. Contrarily, the term 'direct' says that no information about the problem is required except for the function values obtained by point-wise evaluation. This concept suggest these methods for complex problem with unknown properties. Black-box optimization as well expresses this by assuming that nothing is known about the problem so that information can only be gained by executing the black-box with a solution to evaluate the solutions quality. The term 'heuristic' says that although thorough strategies are performed, no guaranty of a certain resulting quality of the solutions can be given as this demands assumptions on the problem class. The term 'meta' is an addition to express that the algorithms are not limited to a certain problem class but can optimize arbitrary problems. It as well alludes that the methods may conjoin several different techniques, which is also termed 'hybrid metaheuristic'.

Evolutionary algorithms moreover belong to methods subsumed under the umbrella term *Computational Intelligence.* These have in common to be inspired by natural archetypes, and also include fuzzy systems, artificial neural networks, swarm intelligence, and artificial immune systems (see e.g. Engelbrecht (2007); Konar (2005) for an overview). Evolutionary algorithms are inspired by the Darwinian notion of the evolution as a process of reproduction and selection by survival of the fittest. The strongest individuals adapted best to the environment shall pass their positive characteristics to the next generation by their genetic information so that a species is improved by evolution. Modern EA still base on the concept of iterative improvements but do not try to imitate biological processes. The nature is used as an valuable source of inspiration, whereas the techniques are then embossed according to the latest knowledge of computer science, mathematics, and statistics.

Historically, evolutionary algorithms had different names due to different communities. Nowadays, *evolutionary algorithm (EA)* or *evolutionary computing (EC)* are generic terms subsuming the former directions of genetic algorithms (GA) evolutionary strategies (ES) as well as estimation of distributions algorithms (EDA), evolutionary programming (EP), and genetic programming (GP), whereas the three latter differ more in their concepts and are also considered as independent fields. An introduction and overview can be found in Bäck et al. (1997); Eiben and Smith (2007) and recent theoretical results in Auger and Doerr (2011).

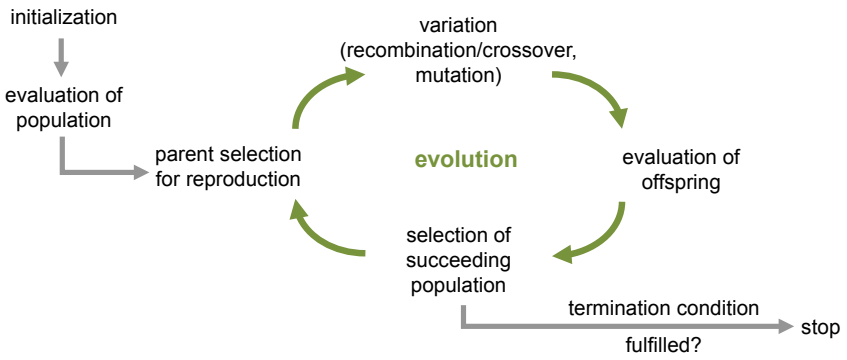


Fig. 1.3: Operators and evolution cycle of an EA.

Basic Principles

The main components and progression of EA are as follows. The search space has to be encoded such that solutions consist of a vector of parameter values, which is called the representation of solutions. The parameter vector consists of so-called decision variables and is also termed genome according to the biological paradigm. It is, besides the solutions quality and possibly other attributes, stored in a solution or individual. The optimization starts with the initialization of a set of individuals, denoted as population, which is maintained and sought to be optimized throughout the run of the algorithm (cf. Fig. 1.3, Alg. 1.1). The population is mostly initialized uniformly at random in the given search space. The individuals are then evaluated to determine their solution quality or fitness by a black-box solver.

Then a loop or generation is repeated: From the population, individuals are selected to become parents of new individuals, a process called parent or mating selection, which is often performed uniformly at random or in a tournament modus where individuals are compared pairwise according to their fitness until 'winners' are determined. New individuals or offspring are generated from the parents by probabilistic variation. Recombination or crossover operators combine the information of typically two individuals to create a new one. Afterwards mutation operators slightly alter the offspring. Crossover is optional, alternatively a copy of a parent is generated and then only changed by mutation.

Having generated and evaluated an offspring population, a subset of the current individuals is discarded to keep only the most promising ones, based on their objective or fitness values. If we discard the whole preceding population, this is called comma selection, denoted by (μ, λ) for a population of size μ and λ offspring per generation. A so-called plus selection filters the best individuals from both old and offspring population, denoted as $(\mu + \lambda)$. In both concepts the population is resized to μ individuals in each generation. After the selection of the subsequent population, the generation is completed. The evolutionary loop has no defined end since the algorithm is not able to verify a solution as optimal. Thus, an aborting

Algorithm 1.1: (μ+1)-EA or (μ+1)-EMOA

```
choose μ individuals from ℝ^n for initial population P^(0), t ← 0
evaluate population P^(0)
repeat
    select 2 parents from P^(t)
    create offspring Y^(t) by recombination of parents
    mutate offspring Y^(t)
    evaluate offspring Y^(t)
    select worst individual Z^(t) from P^(t) ∪ Y^(t)
    update population to P^(t) ← P^(t) ∪ Y^(t) \ Z^(t)
    t ← t + 1
until termination criterion fulfilled
```

condition has to be defined by the user, typically according to resource limitation or the witnessed progress.
Into any component user expertise (if available) may be included in the optimization process to bias the random decisions to known good directions. Seemingly every day, new EA appear, however the state-of-the EA for single-objective optimization are the variants of the CMA-ES by Hansen (2006), Hansen and Ostermeier (2001), and Hansen et al. (2003).

EA for Multiobjective Optimization

Evolutionary Multiobjective Optimization Algorithms (EMOA), also called *Multiobjective Evolutionary Algorithms (MOEA)*, are perfectly qualified for the task of generating an approximation of the Pareto front due to the population concept of optimizing a set of points. In principle, the operators of EA for single-objective optimization can also be used for multiobjective optimization, except for the selection. So the example algorithm in Alg. 1.1 is an EA as well as an EMOA. There is no necessary difference for the variation or other operators considering the search space only, whereas of course dedicated operators may be more effective. Only the selection unavoidably has to deal with the multi-dimensional objective space. Here again, the common concepts—like tournament selection, plus or comma selection etc.—for selecting among the best individuals could be adopted. Yet the problem is the typical absence of a total order among the individuals, so that there is no sorting the selection can base on. So, for EMOA—and as well in modern EA—the fitness value does not simply equate the objective function values, as the fitness incorporates additional information relative to the current population. Appropriate totally ordered fitness values the selection can base on are in EMOA generated as follows. Many modern EMOA have a two-stage selection process: Firstly, the indi-

viduals are classified as far as possible regarding the dominance or measures based upon it. Examples are non-dominated sorting which performs a kind of ranking based on the transitivity of the dominance, other measures e.g. count the number of dominating individuals. Secondly, an additional measure is applied to enforce an order among the individuals that are still incomparable after the first stage. To this end, diversity measures or quality indicators are used.

The indicator-based EMOA use a quality indicator for the selection at least as a second stage measure. The SMS-EMOA follows this principle by first performing non-dominated sorting on the population and one offspring, and then for the worst subset, the hypervolume is determined and the individual with the least contribution to the hypervolume is denoted as the worst one and is discarded.

The clearly most popular EMOA, NSGA-II by Deb et al. (2002a), is nowadays outdated and no longer recommended, yet it still serves as a reference optimizer in benchmarks, next to the classic algorithm SPEA2 by Zitzler et al. (2002). A modern EMOA is e.g. ϵ-MOEA by Deb et al. (2005a), and state-of-the-art EMOA are GDE3 by Kukkonen and Lampinen (2007), or MSOPS and MSOPS-II by Hughes (2003, 2007), as well as the hypervolume-based EMOA, with IBEA by Zitzler and Künzli (2004), MO-CMA-ES by Igel et al. (2007); Voß et al. (2010), and SMS-EMOA by Beume et al. (2007)* being the most popular algorithms.

Besides the classic introductions by Deb (2001) or Coello Coello et al. (2002), an overview of modern EMOA and related techniques can be found in Abraham and Goldberg (2005), Knowles et al. (2007), Branke et al. (2008), and recent theoretical insights in Brockhoff (2010b). The EMOA invoked in this thesis are described in Chapter 3, especially detailing the SMS-EMOA and its characteristics.

1.4 Computation Models

Standard Algorithmic Model

The computation model we consider for the hypervolume calculation in Chapter 2 is the standard model for algorithmic complexity, i.e., also for computational geometry, which is the context we consider the hypervolume in. The calculation time is measured as the number of elementary calculation steps. All steps are assumed to have equal costs (uniform cost model) and we do neither consider space nor resources for communication in distributed systems. An algorithm is considered as a bounded-degree algebraic decision tree that is traversed depending on the input. For a more in-depth exposition, see e.g. Preparata and Shamos (1988, Sec. 1.4).

Runtimes are given w. r. t. the size of a set of vectors m, without the dimension of the vectors d, assuming d is a constant, i.e., $d = O(1)$ w. r. t. m.

This model is as well used to express the computational resources of an evolutionary algorithm per generation without counting the function evaluations.

Model for Black-Box Optimization

For black-box optimization, a dedicated complexity measure, the *black-box complexity*, has been developed to express optimization times required to solve a problem class by means of a randomized direct search heuristic. It is assumed that the operations of an algorithm are neglectable compared to the time required for a function evaluation of a real-world problem. Hence, the black-box complexity only gives the number of function evaluations until an optimal function value is found. See Wegener (2005, Ch. 9.) for exact definitions and a more in-depth exposition.
In theory on evolutionary algorithms for single-objective optimization, this measure is common practice. As usually, upper bounds for the complexity of a problem are gained by algorithms that solve the problem in that time. It depends on the problem and the optimizer how sound the measure actually is. In multiobjective optimization, the black-box complexity is not necessarily suitable since the calculations of an evolutionary algorithm may be more demanding, e.g. due to hypervolume calculation, so that their resources only fall behind time-consuming function evaluations.
In Chapter 3 we state the runtime of algorithms as resources per generation w. r. t. the standard algorithmic model, and optimization times in black-box complexity.

1.5 Experimental Analyses

Experimental Studies

Experimental studies are a research discipline with a documented tradition of hundreds of years, and are more or less the only approach in natural sciences. It could be surmised that computer experiments as not necessary due to the mathematical-based system and algorithms. However, the resulting random processes are far too complex to calculate exactly what is happening. As current theoretical tools are insufficient to answer many relevant research questions, experiments are the only ways to gain insights. Moreover, they are also used to validate that mathematically calculated facts can also be observed when actually running the algorithms, which is not evidently since theoretical analyses as well make assumptions to abstract from reality.
In this work we perform experiments in order to experience how well the theoretical bound of the runtime fits with the real runtime of the hypervolume algorithms in Section 2.3.2. When studying the EMOA in Chapter 3, rigorous analyses are performed as far as possible, but the processes are often too complex for current tools, so that experimental studies are performed instead. An overview of state-of-the-art methods of experimental analyses is given in Bartz-Beielstein et al. (2010a). For the documentation of experiments, we mainly follow Preuss (2007), which formulates the application of common guidelines of experiments in natural sciences to

computer experiments in the area of evolutionary computation. Tests on statistical significance are performed when it seems necessary in order to validate the results.

Academic Test Problems

To benchmark and study certain aspects of EMOA, academic test problems are used. These have the advantage that the problem is well understood and the optima are known as it has been designed clearly structured. They shall represent typical characteristics of real-world problems paradigmatically. In our studies, the test set is respectively chosen according to the posed research question, and with popular test problems in order to establish comparability with similar studies. Note that all problem functions are real-valued and to be minimized.
The ZDT family by Zitzler et al. (2000) comprises 2-objective continuous functions, the discrete ZDT5 is excluded here like in most studies from literature. The Pareto front of ZDT1 is convex, ZDT2's is concave, ZDT3's consists of five convex parts. ZDT4 has the same Pareto front as ZDT1 but the functions are multimodal. The Pareto front of ZDT6 is a subset of ZDT2's and the points are distributed asymmetrically. The Pareto set is equal for all ZDT function, with all except the first decision variable equaling zero. The problem family is often used for an initial benchmark of a new optimizer; we apply it in Sections 3.3–3.6.
The DTLZ function family by Deb et al. (2002b) is another popular problem set with scalable number of objectives. Like the ZDT problems, they contain multimodal problems based on sine and cosine functions. DTLZ1 has a linear Pareto front (e.g. a plane in case of three objectives) and DTLZ2 and DTLZ3 have the same convex Pareto front, more precisely a section of a hypersphere. The runtime of hypervolume algorithms is benchmarked on subsets of the linear and convex DTLZ Pareto fronts (Section 2.3.2). To study and benchmark EMOA, the DTLZ problems and variations of it are applied in Sections 3.3–3.5.
Moreover, other test problems as well as a real-world problem are used in our studies, specified in the respective sections, including the test suite of the CEC 2007 competition, organized by Huang et al. (2007), for the study in Section 3.5.

Parameter Handling

Undoubtedly, different optimizers perform different on certain problem classes. Since meta-heuristics, such as EA, are equipped with several parameters that allow to change their behavior, an EA or EMOA can be seen as a plurality of algorithms and the choice of parameter values deserves a closer look.
Two major forms of parameter settings are to be distinguished: parameter control and parameter tuning. *Parameter control* deals with the handling of parameter values within the EA during its execution. The main concepts are static or adaptive operation, thereby especially dynamic over time or self-adaptive subjected to the evolution process (see Michalewicz et al. (2007) for an overview).

Parameter tuning is the process of finding good values for the external parameters that are set before the algorithm starts (see Bartz-Beielstein et al. (2010a) for an overview of tuning methods). These may function as the initial values of internal adaptive parameter control methods. Finding the best parameterized EA for a problem is an optimization problem by itself. Yet, the goal of the optimization process is not obvious and naturally a multiobjective problem, while demands are e.g. fast progress and convergence to a global optimum. Further requirements arise from the stochasticity of EA, e.g. low variance and high reliability of the results' qualities. We deal with parameter tuning in Section 3.5.

2 Hypervolume Calculation

Calculating the dominated hypervolume is a core problem in multi-objective optimization as it is the standard quality measure to evaluate approximations of the Pareto front. After an overview of the problem's properties, we deal with the computational resources required for the hypervolume calculation. Taking the view of computational geometry, the best known lower bound is proved with tools of complexity theory (Sec. 2.2, Beume et al. (2009a)*). By algorithm engineering, we derive a simple algorithm realizing a hardly worse upper bound than the best known one (Sec. 2.3, Beume (2009)*).

2.1 Problem Properties

The hypervolume is a unary quality measure that maps a set to a scalar value. This way, an arbitrary number of sets can be compared using the order induced by the hypervolume values. To simplify matters, we consider one or two sets in the following.
After providing basic definitions, we survey properties in simple notation with formal proofs. An emphasis is the relation of the preorder of the dominance relation and the total order according to the hypervolume. The choice of the reference point and its influence are discussed. Results for the distribution of points within a set of maximal hypervolume are presented detailing our contribution for linear Pareto fronts. We give an overview of exact as well as of approximation algorithms.

Basic Definitions

The dominated hypervolume describes the size of the space dominated by a set—or more general a multiset—of points. It has first been suggested by Zitzler and Thiele (1998) for performance assessment.

Definition 2.1 *Let* $M = \{\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(m)}\}\} \in \mathcal{M}(\mathbb{R}^d)$, $d \geq 2$, $m \in \mathbb{N}$ *be a finite multiset, and* $\mathbf{r} \in \mathbb{R}^d$ *indicate the* reference point. *The* dominated hypervolume *(or* S-metric*) is defined as the quantity*

$$\begin{aligned} H(M, \mathbf{r}) &:= \mathsf{Leb}\left(\{\mathbf{u} \in \mathbb{R}^d \mid \exists \mathbf{v} \in M : \mathbf{v} \preceq \mathbf{u} \preceq \mathbf{r}\}\right) \\ &= \mathsf{Leb}\left(\bigcup_{i=1}^{m} [\mathbf{v}^{(i)}, \mathbf{r}]\right). \end{aligned} \tag{2.1}$$

Leb *denotes the Lebesgue measure in* $\mathbb{R}^d$*. The d-dimensional interval* $[\mathbf{v}^{(i)}, \mathbf{r}]$ *describes the hypercuboid spanned by* $\mathbf{v}^{(i)}$ *and* $\mathbf{r}$*, indicating the space dominated by* $\mathbf{v}^{(i)}$ *and bounded by* $\mathbf{r}$*. In the following, we denote* $H_s(M, \mathbf{r}) := \{\mathbf{u} \in \mathbb{R}^d \mid \exists \mathbf{v} \in M : \mathbf{v} \preceq \mathbf{u} \preceq \mathbf{r}\}$

An illustration of the 2-dimensional dominated hypervolume is given in Fig. 2.1 (left). For convenience, we use the popular term *hypervolume* in the following for dominated hypervolume. The older term S-metric is misleading as the measure is not a metric in the mathematical sense. In the context of multiobjective optimization, the hypervolume is sought to be maximized.

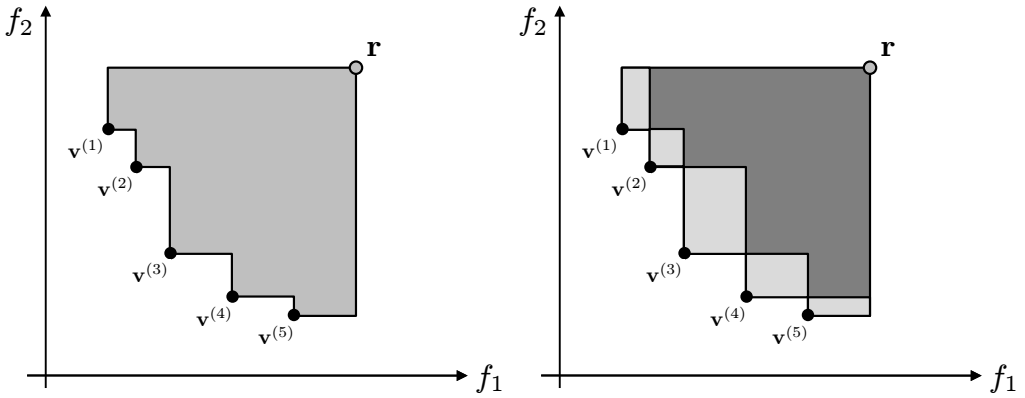


Fig. 2.1: Left: Illustration of the 2-dimensional hypervolume (gray area) of a set (black points), bounded by the reference point $\mathbf{r}$. The axes f_1, f_2 indicate our notion of the 2-dimensional space as the image of a multiobjective problem. Right: The hypervolume contribution of each point is indicated in light gray.

More general definitions of the hypervolume exist, which we do not use here but briefly mention. Instead of one reference point, a reference set can be used to bound the dominated space (cf. Zitzler et al. (2010)). Zitzler et al. (2007) introduced the *weighted hypervolume* by using a weight distribution to emphasize certain regions of the objective space and thereby realize user preferences. They showed that the weighted hypervolume indicator has the same properties w. r. t. the dominance relation as the original one.

In order to optimize the composition of a set, it is desirable to quantify the value of a single point within that set. To this end, we define the hypervolume contribution of a point to a set as the hypervolume that is exclusively dominated by that single point and thus gets lost when the point is excluded from the set (Fig. 2.1, right). Points with higher hypervolume contribution are preferred over others in the set. Note that the hypervolume contribution is typically calculated w. r. t. a set of non-dominated points. Dominated points or copies in multisets have a contribution of zero but they may reduce the contribution of non-dominated points by sharing a region with them which is then not associated as their exclusive contribution.

Definition 2.2 *The hypervolume contribution of the point* $\mathbf{v} \in M$ *to the hypervolume of the multiset* $M \in \mathcal{M}(\mathbb{R}^d)$ *is defined as*

$$H(\mathbf{v}, M, \mathbf{r}) = H(M, \mathbf{r}) - H(M \setminus \{\mathbf{v}\}, \mathbf{r}).$$

Throughout this section it becomes obvious that the hypervolume respects the Axioms 1.18. Points closer to the Pareto front have a higher hypervolume, the hypervolume of a set increases with additional non-dominated points, so that spreading along the whole Pareto front gets rewarded. Next, we present basic properties of the hypervolume.

Corollary 2.3 *Let* $A, B \in \mathcal{M}(\mathbb{R}^d)$ *be finite multisets. Then,* $\forall A, B, \forall \mathbf{v} \in \mathbb{R}^d, \forall \mathbf{r} \in \mathbb{R}^d$ *holds:*

1. $H(A, \mathbf{r}) = 0 \iff \nexists \mathbf{a} \in A : \mathbf{a} \prec\prec \mathbf{r}$
2. $H(A, \mathbf{r}) > 0 \iff \exists \mathbf{a} \in A : \mathbf{a} \prec\prec \mathbf{r}$
3. $H(\mathbf{v}, A, \mathbf{r}) = 0 \iff \mathbf{v} \not\prec\prec \mathbf{r}$ *or* $\exists \mathbf{a} \in A : \mathbf{a} \preceq \mathbf{v}$
4. $H(\mathbf{v}, A, \mathbf{r}) > 0 \iff \mathbf{v} \prec\prec \mathbf{r}$ *and* $\nexists \mathbf{a} \in A : \mathbf{a} \preceq \mathbf{v}$
5. *Sets with higher hypervolume are preferred over other sets.*
6. *Points with higher hypervolume contribution are preferred over others in the set.*
7. *Let* M *be a set of elements from the co-domain* $Z \subset \mathbb{R}^d$ *of a multiobjective optimization problem* $f : \mathbb{R}^n \to \mathbb{R}^d$*. The hypervolume of the set* M *is maximal among the subsets of* Z *iff it contains the Pareto front of* f*.*

Proof. The hypervolume as well as the hypervolume contribution are either positive or zero; negative values are not defined.
1., 2.: The hypervolume is positive iff there is at least one point that strictly dominates the reference point. This implies that the hypercuboid spanned by the point and the reference point is full-dimensional. Weaker forms of dominance are not sufficient as they allow equal coordinate values of both points which leads to a hypervolume of zero.
3., 4.: The hypervolume contribution of a point $\mathbf{v}$ is positive iff $\mathbf{v}$ strictly dominates the reference point (for the same reason as in 2.), and $\mathbf{v}$ is not weakly dominated by any point of A. For all points in or on the boundary of the space dominated by A holds that they are at least weakly dominated by one point of A. Thus $\mathbf{v}$ indeed lies outside this area, so there is a portion of hypervolume dominated by $\mathbf{v}$ but not dominated by any other point in A, and it is full-dimensional. Formally, the point can be worsened in each dimension by $\varepsilon > 0$ (sufficiently small) to $\mathbf{v}'$. For $\mathbf{v}'$ holds $\mathbf{v} \prec\prec \mathbf{v}' \prec\prec \mathbf{r}$, and $\nexists \mathbf{a} \in A, \mathbf{a} \neq \mathbf{v} : \mathbf{a} \preceq \mathbf{v}'$. The hypervolume contribution of $\mathbf{v}$ is

the Lebesgue measure of the set of these $\mathbf{v}'$.
Otherwise, the hypervolume contribution is zero (4.).
5., 6.: Regarding the hypervolume and the hypervolume contribution higher values are better.
7.: See Fleischer (2003, Theorem 1.) for a proof. The fact also becomes evident in the proof of Theorem 2.4.4. □

Ordering according to the Hypervolume

As a quality measure, the hypervolume creates a total order among sets which are only preordered regarding the Pareto dominance relation. The implications below directly follow from the fact that the dominated space of A includes B, if $A \preceq B$. Intuitively, the implications mean that the total order induced by the hypervolume does not contradict with the preorder of the dominance relation, and whenever sets are comparable regarding the dominance relation, the hypervolume indicates this. The following collection of characteristics is partly redundant in the sense that some observation follow from others, yet it seems worthwhile to present facts from different points of view.

Theorem 2.4 *Let the dominance relation $\preceq$ induce a preorder and the hypervolume $H(\cdot, \mathbf{r})$ induce a total order among the finite non-empty multisets $A, B \in \mathcal{M}(\mathbb{R}^d)$. Then, $\forall A, B, \forall \mathbf{b} \in \mathbb{R}^d, \forall \mathbf{r} \in \mathbb{R}^d$ holds:*

1. $A = B \quad \Rightarrow \quad H(A, \mathbf{r}) = H(B, \mathbf{r})$
2. $A \preceq B \quad \Rightarrow \quad H(A, \mathbf{r}) \geq H(B, \mathbf{r})$
3. $A \preceq \{\mathbf{b}\} \quad \Rightarrow \quad H(A \cup \{\mathbf{b}\}, \mathbf{r}) = H(A, \mathbf{r})$
4. $A \not\preceq \{\mathbf{b}\}$ *and* $\mathbf{b} \prec\prec \mathbf{r} \quad \Rightarrow \quad H(A \cup \{\mathbf{b}\}, \mathbf{r}) > H(A, \mathbf{r})$
5. *For all sets A, B of mutually incomparable points:*
   $A \preceq B$ *and* $A \neq B \quad \Rightarrow \quad \forall \mathbf{r} : (\mathbf{nad}(A \cup B) \prec\prec \mathbf{r}) : H(A, \mathbf{r}) > H(B, \mathbf{r})$
6. $A \preceq B$ *and* $B \not\preceq A \quad \Rightarrow \quad \forall \mathbf{r} : (\mathbf{nad}(A \cup B) \prec\prec \mathbf{r}) : H(A, \mathbf{r}) > H(B, \mathbf{r})$
7. $B \not\preceq A$, *i.e.,* $(A \preceq B \wedge A \neq B) \vee A \| B \quad \Leftarrow \quad \exists \mathbf{r} \in \mathbb{R}^d : H(A, \mathbf{r}) > H(B, \mathbf{r})$
8. $A \preceq B$ *and* $B \not\preceq A \quad \Leftarrow \quad \forall \mathbf{r} : (\exists \mathbf{v} \in A \cup B : \mathbf{v} \prec\prec \mathbf{r}) : H(A, \mathbf{r}) > H(B, \mathbf{r})$

Proof. **1.:** Equal sets of course yield equal hypervolume values, as obvious by replacing A by B, whereas this does not hold the other way round.
2.: $A \preceq B$ implies $\forall \mathbf{b} \in B : \mathbf{b} \in H_s(A, \mathbf{r})$, so $H(A, \mathbf{r})$ cannot be smaller than $H(B, \mathbf{r})$.
3. The implication says that a dominated point does not add anything to the hypervolume of a set. The left hand side implies that $\mathbf{b}$ either lies inside the

hypervolume of A ($\mathbf{b} \in H_s(A, \mathbf{r})$), or $\mathbf{b}$ is outside such that it does not strictly dominate the reference point ($\mathbf{b} \not\prec\prec \mathbf{r}$). In both cases, $\mathbf{b}$ does not have a positive contribution to A (cf. Corollary 2.3.3.).
4. $A \not\preceq \{\mathbf{b}\}$ means that $\nexists \mathbf{a} \in A : \mathbf{a} \preceq \mathbf{b}$. This combined with $\mathbf{b} \prec\prec \mathbf{r}$ is the condition for a positive contribution to a set (cf. Corollary 2.3.4.). Thereby $\mathbf{b}$ has a positive contribution to $A \cup \{\mathbf{b}\}$ so the hypervolume of this set is larger than that of A. This property allows for a proof of Corollary 2.3.7.: The hypervolume of a set can be improved as long as there is a non-dominated point to include. Therefore the Pareto front is the set with the maximal hypervolume among all sets of solutions of a multiobjective problem.
5. $A \preceq B$ implies $H(A, \mathbf{r}) \geq H(B, \mathbf{r})$ (cf. 2.). $A \neq B$ implies $\exists \mathbf{a} \in A : \mathbf{a} \notin B$. This $\mathbf{a}$ has a positive contribution to A, if Corollary 2.3.4 is fulfilled. This positive contribution added to the hypervolume of A makes the hypervolume of A indeed larger than the hypervolume of B. Given the preconditions Corollary 2.3.4 is automatically fulfilled, since $\mathbf{a}$ is neither a copy, nor dominated within A, and $\mathbf{a} \prec\prec \mathbf{r}$. With toned down additional precondition, 5. reads:
$A \preceq B$ and $nds(A) \neq nds(B) \;\Rightarrow\; \forall \mathbf{r}$ with $\mathbf{nad}(A \cup B) \prec\prec \mathbf{r} : H(A, \mathbf{r}) > H(B, \mathbf{r})$.
6. According to the proof of 5., 6. holds when the non-dominated subsets of A and B are different. This is the case for $A \preceq B \wedge B \not\preceq A$.
7. $B \preceq A$ implies $\forall \mathbf{a} \in A : \exists \mathbf{b} \in B : \mathbf{b} \preceq \mathbf{a}$ and so $A \subseteq H_s(B, \mathbf{r})$ causing $H(A, \mathbf{r}) \leq H(B, \mathbf{r})$ (cf. 2.). This contradicts with the right hand side, so, the opposite must be true. Since 2. holds for all reference points, it suffices to observe $H(A, \mathbf{r}) > H(B, \mathbf{r})$ for one reference point to conclude a contradiction.
8. If $A \preceq B$ is not true, then $\exists \mathbf{b} \in B : \nexists \mathbf{a} \in A : \mathbf{a} \preceq \mathbf{b}$. Choose the reference point such that it is strictly dominated by this $\mathbf{b}$ but not by A: $r_i = b_i + \varepsilon$ for $i = 1, \ldots, d$ and $\varepsilon > 0$ sufficiently small. Then $H(B, \mathbf{r}) > 0$ while $H(A, \mathbf{r}) = 0$. This contradicts with the right hand side, so the opposite must be true. $B \preceq A$ implies $H(B, \mathbf{r}) \geq H(A, \mathbf{r})$ due to 2. This contradicts with the right hand side, hence $B \not\preceq A$ holds. □

Zitzler et al. (2003) observed 5. and 7. (with less precise specification of the reference point), whereas 5. is named $\triangleright$-completeness, and 7. $\not\triangleright$-compatibility (cf. Definition 1.19, 1.20). In Zitzler et al. (2008a), a property resembling 5. is named *strict monotonicity*. Zitzler et al. (2010, Th. 3.2) term the hypervolume a *refinement* of the dominance relation due to a property resembling 5. (cf. Zitzler et al. (2010, Th. 3.1)). They observed 3. and 4. (cf. (Zitzler et al., 2010, Th. 3.2)) and showed that both in tandem imply 5.
The hypervolume and its variations are the only unary indicators that feature both property 5. and 7. (cf. Zitzler et al. (2007)). Implication 7. cannot be stronger as proved in (Zitzler et al., 2003, Th. 1). Intuitively, it is clear that from a greater hypervolume value, it cannot be followed that the sets are comparable regarding the dominance relation. This would erase the case of incomparability and make the dominance relation a total order, which is obviously not true.

The following lemma says that for two points, their ordering regarding the hypervolume is always equivalent to their ordering according to their hypervolume contribution when combined to a set (cf. Fig. 2.2). This lemma is used in Section 3.2 and has been described informally in Beume et al. (2011)*.

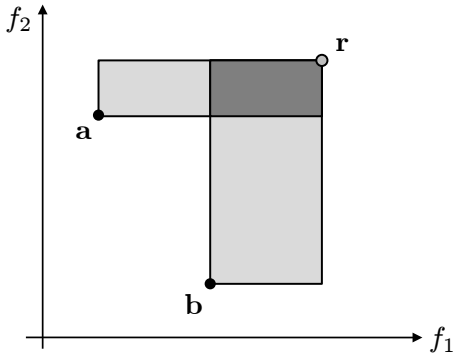


Fig. 2.2: The light gray rectangles depict the hypervolume contribution of the points. The dark gray area is dominated by both points, so its value added to each contribution results in the absolute hypervolume of each point. The order induced by both measures is equal.

Lemma 2.5 *Let* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$. *Then* $\forall d \geq 2, \mathbf{r} \in \mathbb{R}^d$ *with* $\mathbf{nad}(\mathbf{a}, \mathbf{b}) \prec\prec \mathbf{r}$ *holds:*

$$H(\mathbf{a}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) \bullet H(\mathbf{b}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) \iff H(\{\mathbf{a}\}, \mathbf{r}) \bullet H(\{\mathbf{b}\}, \mathbf{r}),$$

with $\bullet$ *being the same relation out of* $\{<, >, =\}$. *(Beume et al. (2011)*).*

Proof. The hypervolume dominated by two points consists of the points' contributions and a part dominated by both points (Fig. 2.2). The hypervolume of each point is equivalent to its hypervolume contribution plus the hypervolume dominated by both points. Since this value is equal for both points, it does not affect the order of the points induced by their contributions. Formally, we distinguish three cases: (i) Let $\mathbf{a} = \mathbf{b}$. Then $H(\{\mathbf{a}\}, \mathbf{r}) = H(\{\mathbf{b}\}, \mathbf{r})$ and $H(\mathbf{a}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) = H(\mathbf{b}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) = 0$. (ii) Let $\mathbf{a} \prec \mathbf{b}$. Then $H(\{\mathbf{a}\}, \mathbf{r}) > H(\{\mathbf{b}\}, \mathbf{r})$ as $\mathbf{b} \in H_s(\{\mathbf{a}\}, \mathbf{r})$ and $H(\mathbf{a}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) > H(\mathbf{b}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) = 0$. (iii) Let $\mathbf{a} \| \mathbf{b}$ with $a_1 < b_1$ and $a_2 > b_2$. With $q := (r_1 - b_1)(r_2 - a_2)$, then $H(\{\mathbf{a}\}, \mathbf{r}) = (b_1 - a_1)(r_2 - a_2) + q = H(\mathbf{a}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) + q$ and $H(\{\mathbf{b}\}, \mathbf{r}) = (r_1 - b_1)(a_2 - b_2) + q = H(\mathbf{b}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) + q$. Thus $H(\{\mathbf{a}\}, \mathbf{r}) > H(\{\mathbf{b}\}, \mathbf{r}) \iff H(\mathbf{a}, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) > H(b, \{\mathbf{a}, \mathbf{b}\}, \mathbf{r})$, and analogously for the other relations and vice versa roles of $\mathbf{a}$ and $\mathbf{b}$. □

Choice of the Reference Point

The ranking of sets according to the hypervolume allows for their comparison. To make this comparison fair, it is common practice to choose the same reference point

for all sets, though it is not proven to be the best concept. For incomparable sets regarding the dominance relation, the ranking due to the hypervolume depends on the choice of the reference point. An example for two sets with different rankings depending on the reference point is given in Figure 2.3.

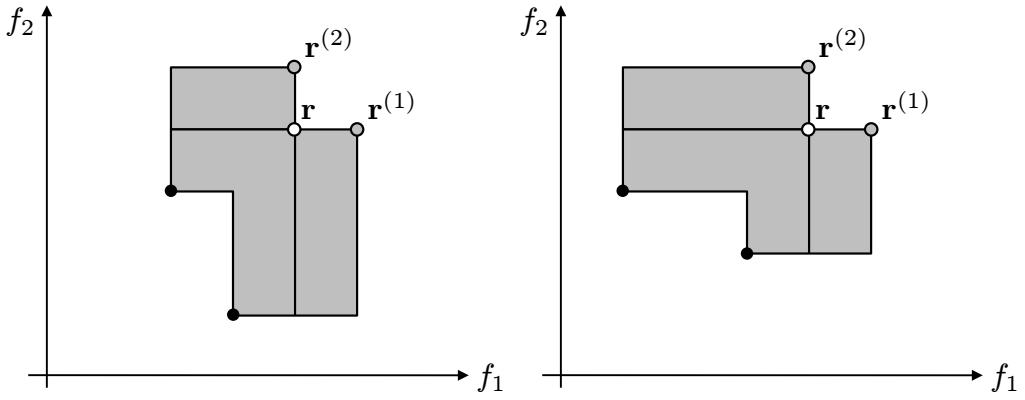


Fig. 2.3: Example for two different ranking of sets, depending on the reference point. The set on the left is ranked higher regarding $\mathbf{r}^{(1)}$, the set on the right regarding $\mathbf{r}^{(2)}$. Both sets have equal hypervolume values regarding $\mathbf{r}$.

The question how the reference point should be chosen is as yet unanswered. Auger et al. (2009a) calculated for some 2-dimensional functions how the reference point has to be chosen such that sets of fixed size with a maximal hypervolume include points on the boundary of the Pareto front. However, the boundary solutions may or may not be of interest. A common strategy to choose the reference point is to position it close to the expected nadir point of the sets. If the reference point is in large distance to the sets, the problem of numerical imprecision increases, and it may become hard to distinguish the hypervolume values of the sets.

Within the SMS-EMOA, the concept of the adaptive reference point is introduced. The idea is to 'neutralize' the objective that is not bounded by a point within the set. To this end, the adaptive reference point is chosen such that it consists of the nadir point added by a vector of ones. Note that points that are the worst ones in the population regarding an objective function have a difference to the reference point of exactly 1 w. r. t. that worst objective (Fig. 2.4). Note that the ordering of sets may differ w. r. t. the adaptive reference point compared to a fixed reference point, just like between two fixed reference points as discussed above.

Definition 2.6 *The* adaptive reference point *of a multiset $M \in \mathcal{M}(\mathbb{R}^d)$ is defined as*

$$\mathbf{r}(M) := \mathbf{nad}(M) + (1, \ldots, 1)^\top, \textit{i.e.}, \ \forall i \in \{1, \ldots, d\} : r_i = nad(M)_i + 1 \quad (2.2)$$

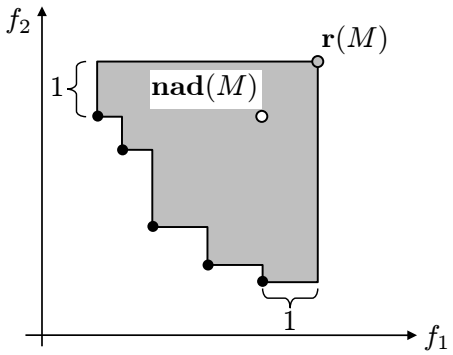


Fig. 2.4: Illustration of the adaptive reference point $\mathbf{r}(M)$. It differs by 1 from the worst coordinates of the set M in each dimension.

Distribution of Points

To compare sets of equal cardinality, it is worth knowing the maximal hypervolume value that can be reached by a fixed number of points located on a certain Pareto front. Optimal distributions have up to date only been calculated for classes of 2- or 3-dimensional problems, see Auger et al. (2009a,b, 2010), Brockhoff (2010a). Emmerich et al. (2007)*, Beume et al. (2009a)*, Friedrich et al. (2009), and Auger et al. (2009a) showed that for linear Pareto fronts, the optimal distribution are such that the distance between neighboring points are equal. Auger et al. (2009a) proved the most general result for all linear connected Pareto fronts, covering the other results. Our proofs are detailed in the following.

Lemma 2.7 *The 2-dimensional hypervolume $H(A, \mathbf{r})$ of m points, including $(0, q)$ and $(q, 0)$, on the straight line $g(y) = -y + q$ w. r. t. the reference point $\mathbf{r} = (q, q)$ is maximal if and only if the points are equally spaced. (Beume et al. (2009a)*, Emmerich et al. (2007)*).*

Proof. We first present a geometric proof based on Beume et al. (2009a, Lemma 1)* and afterwards a proof based on Emmerich et al. (2007, Sec. 3)* using analytic arguments.
(1) We prove the statement by a contradiction. Let us assume a set A with maximal hypervolume while not all points in A are equally spaced. Then there exist three consecutive points $\mathbf{a}, \mathbf{b}, \mathbf{c}$ in A, in sorted f_1-order, that are not equally spaced, i.e., $b_1 - a_1 \neq c_1 - b_1$. Note that for the purpose of this proof we do not need to actually find these points; it is sufficient to know that they exist. Without loss of generality, we have the situation depicted in Fig. 2.5 (right). The hypervolume contribution of $\mathbf{b}$ is the red rectangle, or $(c_1 - b_1) \cdot (a_2 - b_2)$. For all points on the straight line $g(y)$, the perimeter of the rectangle they span with the origin is constant: $2(y + g(y)) = 2q$. Likewise, the perimeter of the rectangle of the hypervolume contribution of $\mathbf{b}$ $H(\mathbf{b}, A, \mathbf{r})$ spanned by $\mathbf{b}$ and bounded by $\mathbf{a}, \mathbf{c}$ is constant for

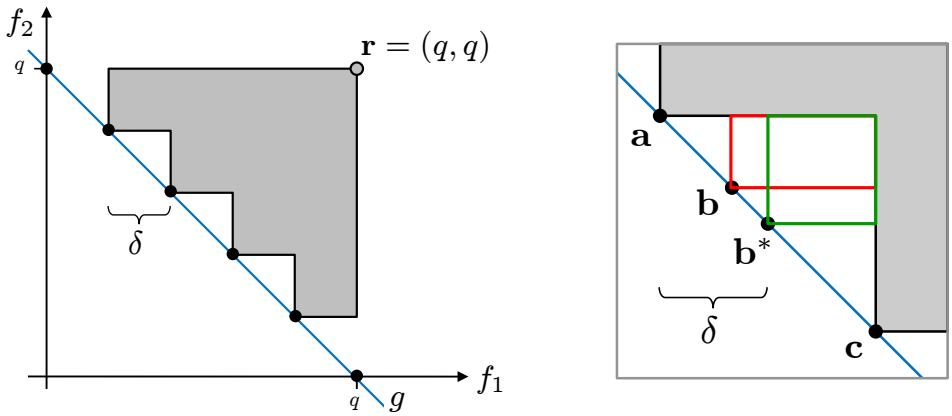


Fig. 2.5: The hypervolume of a linear function is maximal if the points are equally spaced, here with inter-point distance δ. The right excerpt details the approach of the geometric proof (1).

arbitrary position of $\mathbf{b}$ on $g(y)$. For a given perimeter, the area of a rectangle is maximal if and only if it is a square. This results in the unique optimal position of $\mathbf{b} = \mathbf{b}^*$ with $b_1^* - a_1 = c_1 - b_1^*$. So, the hypervolume of A with $\mathbf{b}^*$ instead of $\mathbf{b}$ is larger, i.e., we can move $\mathbf{b}$ to $\mathbf{b}^*$, making $\mathbf{a}, \mathbf{b}, \mathbf{c}$ equally spaced, while increasing the hypervolume. This is the desired contradiction. Conversely, we see that for an equally spaced set of points, every three consecutive points are equally spaced, so the hypervolume contribution of each point is a square. Again, trying to make any three consecutive points non-equally spaced results in a decrease of the contribution of the middle point and the claim follows.

(2) We reformulate the maximization of the hypervolume for the theorem setup as minimizing the area in between the dominated hypervolume and the straight line $g(y)$. Let $u_1, \ldots, u_{m-1}$ denote the lengths of the intervals between the first coordinates of the points. The sum of the squares $\sum_{i=1}^{m-1} u_i^2$ is twice the area we want to minimize. The constraints $\sum_{i=1}^{m-1} u_i = 1$, $\forall i : 0 \leq u_i$, and $u_{m-1} = 1 - \sum_{i=1}^{m-2} u_i$ assure that the points are distributed over the interval $[0, q]$ while occupying its boundaries. In combination, the minimization problem $h = \sum_{i=1}^{m-2} u_i^2 + (1 - \sum_{i=1}^{m-2} u_i)^2$ is yielded. Computing the partial derivatives of h results in $\frac{\partial h}{\partial u_j} = 2u_j - 2(1 - \sum_{i=1}^{m-2} u_i)$ where $j = 1, \ldots m-2$. Each of these partial derivatives has a value of zero at $u_1 = \frac{1}{m-1}, \ldots, u_{m-2} = \frac{1}{m-1}$ and at this point the minimum occurs, as verified by a positive definite Hessian matrix. Translations back to the original problem result in $u_1 = \frac{1}{m-1}, \ldots, u_{m-2} = \frac{1}{m-1}$ and $u_{m-1} = \frac{1}{m-1}$. Hence, the points maximizing the hypervolume are equidistant (with two occupying the end points), and this is the only distribution of point yielding the maximal hypervolume value for this setup.

□

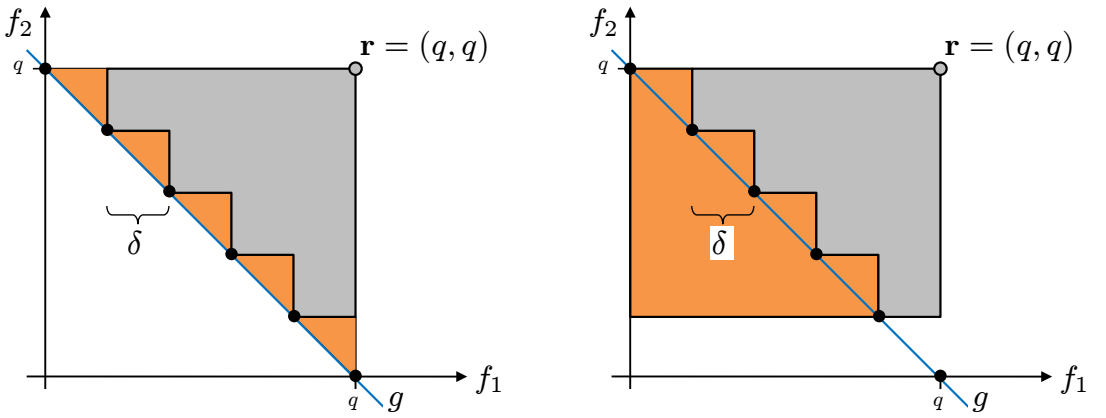


Fig. 2.6: Illustration of the calculation of the maximal hypervolume for the setup in Lemma 2.7. The points are equally spaced with inter-point distance $\delta = q/(m-1)$. The calculation is given in equation (2.3) (left) and (2.4) (right).

The maximal hypervolume for the setup of Lemma 2.7 can be calculated in two easy geometric ways. The first is considering the square spanned by the origin and the reference point. Halving it results in the triangle of the reference point and the assumed Pareto front g. Subtracting the triangles between the hypervolume and g (Fig. 2.6, left) yields in combination the following equation.

$$hv_{max} = \frac{1}{2}q^2 - (m-1)\frac{1}{2}\left(\frac{q}{m-1}\right)^2 = \frac{1}{2}q^2 - \frac{q^2}{2(m-1)} \tag{2.3}$$

$$= \frac{1}{2} \cdot (m-1)\frac{q}{m-1} \cdot (m-2)\frac{q}{m-1} = \frac{q^2(m-2)}{2(m-1)} \tag{2.4}$$

Equation (2.4) is geometrically interpreted as the bounding rectangle of the hypervolume with side lengths of $(m-1)$, rep. $(m-2)$ times the inter-point distance $\delta = q/(m-1)$ (Fig. 2.6, right).

Algorithms and Problem Complexity

The hypervolume of a single point $\mathbf{v} \in \mathbb{R}^d$ with $\mathbf{v} \preceq \mathbf{r}$ results in

$$H(\{\mathbf{v}\}, \mathbf{r}) = \prod_{i=1}^{d}(r_i - v_i). \tag{2.5}$$

Hypervolume algorithms typically expect a set of pairwise incomparable points as their input. This demands a preprocessing that filters out copies and dominated points. The runtime of detecting the subset of non-dominated points is $O(m \log m)$ for $d = 2, 3$ and $O(m(\log m)^{d-2})$ for $d \geq 4$ according to Kung et al. (1975). Filtering copies is possible within these runtimes.

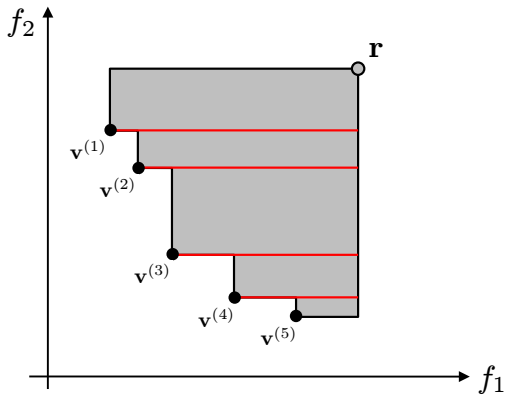


Fig. 2.7: Illustration of the hypervolume calculation due to equation (2.6). The set $\{\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(5)}\}$ is vertically partitioned in rectangles.

In 2-dimensional spaces ($d = 2$), the hypervolume can easily be calculated as the sum of areas of rectangles. A rectangle is spanned by a point in M and bounded in one dimension by a neighboring point, and in the other by the reference point (cf. Fig. 2.7). Sorting $M = \{\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(m)}\}$ ascending regarding the first component, i.e., $v_1^{(1)} < v_1^{(2)} < \ldots < v_1^{(m)}$ it also holds $v_2^{(1)} > v_2^{(2)} > \ldots > v_2^{(m)}$. Then the hypervolume of M with $M \preceq \mathbf{r}$ is

$$H(M, \mathbf{r}) = (r_1 - v_1^{(1)})\,(r_2 - v_2^{(1)}) + \sum_{i=2}^{m} (r_1 - v_1^{(i)})\,(v_2^{(i-1)} - v_2^{(i)}). \tag{2.6}$$

Partitioning horizontally instead of vertically works analogously.
For $d = 3$ Fonseca et al. (2006) give an asymptotically optimal algorithm with a runtime of $\Theta(m \log m)$. Generalized to higher dimensions its upper bound is $O(m^{d-2} \log m)$.
The first hypervolume algorithm HSO has been described independently by Zitzler (2001) and Knowles (2002). In the worst case of HSO, the space gets partitioned into $\binom{m+d-2}{d-1}$ cuboids (cf. While et al. (2006)), resulting in a runtime of $O(m^{d-1})$. While et al. (2005) developed heuristics for HSO, which reorder the input such that the worst possible case is avoided, while the resulting runtime is unknown. Fleischer (2003) developed an algorithm called LebMeasure whose upper bound equals the one of HSO but has been published claiming a polynomial runtime, proved wrong by While (2005). Beume and Rudolph (2006a)* and independently Fonseca et al. (2006) observed that the hypervolume is a special case of *Klee's measure problem* (cf. Klee (1977)), for which the algorithm by Overmars and Yap (1991) gives an upper bound of $O(m^{d/2} \log m)$. In Beume (2009)*, this algorithm is adapted to the special case of the hypervolume and simplified, while preserving the runtime, as described in detail in Section 2.3. The algorithm in Overmars and Yap (1991) has been slightly improved by Chan (2010) such that the smallest known upper

bound for the hypervolume is $O(m^{d/2}2^{O(\log^* m)})$ (with $\log^*$ denoting the iterated logarithm).
Bringmann and Friedrich (2008) show by reduction from the problem MON-CNF (satisfiability of monotone boolean formula given in conjunctive normal form) that the hypervolume computation is $\#P$-hard (see Wegener (2005) for an overview of complexity classes). This means that no algorithm with a polynomial runtime can exist, assuming $P \neq NP$. The only concrete lower bound of the hypervolume is given by Beume et al. (2009a)* as $\Omega(m \log m)$ for $d \geq 2$ as described in detail in Section 2.2.

The proven hard complexity motivates the development of approximation algorithms (see e.g. Vazirani (2004) for an introduction and overview of approximation algorithms).
It is easy to verify whether a point lies inside or outside the dominated hypervolume. This way, the hypervolume can be estimated by Monte Carlo sampling (see Kroese et al. (2011) for an introduction and overview of Monte Carlo methods) as done by Bader and Zitzler (2011), Bringmann and Friedrich (2008, 2009a), and Everson et al. (2002).
The approach in Bringmann and Friedrich (2008) is the first FPRAS (fully polynomial-time random approximation scheme, cf. Vazirani (2004)), i.e., the expected runtime is polynomial regarding the input and the failure probability. With probability $1-\delta$, it provides an additive ε-approximation of the hypervolume in runtime $O(\log(1/\delta)md/\varepsilon^2)$.
For the calculation of the hypervolume contribution, Bringmann and Friedrich (2009b) give an algorithm based on the one by Overmars and Yap (1991) with runtime $O(m^{d/2} \log m + m^\lambda)$ for the contribution of all subsets of size λ to a set of size $m = \mu + \lambda$. This way, the best subset of size μ is detected by discarding the worst subset of size λ, as required in a $(\mu + \lambda)$ selection scheme of a multiobjective evolutionary algorithm. By reduction from #MON-CNF, they show in Bringmann and Friedrich (2009a) that the problem of identifying the point with minimal hypervolume contribution of a set is $\#P$-hard, and approximating it by a factor of $2^{d^{1-\varepsilon}}$ is NP-hard for any ε. Nevertheless, the authors give a practical approximation algorithm with runtime $\Omega(dm^2)$ which seems to be the typical runtime in practice.
The hypervolume shall measure the quality of an approximation of the Pareto front. Its suitability for this task can be evaluated by the concept of the approximation factor from theory of approximation algorithms. Bringmann and Friedrich (2010a) consider bi-objective problems with the Pareto front consisting of a monotonically decreasing, upper semi-continuous function. Considering a set of size m approximating a Pareto front, Bringmann and Friedrich (2010a) show that the optimal multiplicative approximation factor is $1 + \Theta(1/m)$. The same holds for sets of m solutions with maximal hypervolume (w. r. t. a Pareto front), whereas the constant hidden in the Landau notation can be large. Bringmann and Friedrich (2010b) show that the additive approximation factor differs from the optimal one by a

factor of at most $m/(m-2)$, whereas such a good factor cannot be achieved regarding the multiplicative approximation. (Note that the proofs of the proposed theorems are currently unpublished.) Therefore Friedrich et al. (2011) suggest the *logarithmic hypervolume* whereas the input is logarithmized before the hypervolume calculation, yielding a near optimal multiplicative approximation factor.

2.2 Lower Bound

Knowing that the hypervolume problem belongs to the complexity class $\#P$ describes the limits of the runtime only in a rather abstract way. Beyond that, we are interested in the complexity given as a concrete lower bound for all problem instances that is especially sound for instances of realistic size. We prove the highest known concrete lower bound of $\Omega(m \log m)$ for calculating the hypervolume of m points by reduction from the so-called Uniform Gap problem. This bound is sharp for the 2- and the 3-dimensional version of the hypervolume, so that the problem complexity is $\Theta(m \log m)$ for $d \in \{2, 3\}$ and $\Omega(m \log m)$ for $d \geq 4$. The proof is detailed after the description of the reduction method.

A lower bound of the runtime of a problem says that the worst case instance of the problem cannot be solved faster by any algorithm in the computation model (the standard model in algorithmics, see Section 1.4).
Lower bounds do not have to be developed from scratch but can be transferred from one problem to another, i.e., by considering the relative complexity. A tool of this kind is the method of *reduction*. It was originally described for decision problems as formal languages to be accepted or rejected by a Turing machine. We use the definition of reduction as 'algorithmically no more difficult than' as formulated by Wegener (2005, Def. 4.1.1).

Definition 2.8 *A problem $\mathcal{A}$ is algorithmically no more difficult than a problem $\mathcal{B}$ if there is an algorithm solving $\mathcal{A}$ that makes use of an algorithm solving $\mathcal{B}$ and has the following properties, with $p(m), q(m), r(m)$ denoting polynomials with respect to the input size m:*

1. *The runtime of the algorithm for $\mathcal{A}$, not counting the calls to the algorithm for $\mathcal{B}$, is $\leq p(m)$.*

2. *The number of calls to the algorithm solving $\mathcal{B}$ is $\leq q(m)$.*

3. *The input size for each call of the algorithm for $\mathcal{B}$ is $\leq r(m)$.*

Then, if the algorithm for $\mathcal{B}$ has the runtime $t_{\mathcal{B}}(m)$, the algorithm for $\mathcal{A}$ has the runtime $t_{\mathcal{A}}(m) \leq p(m) + q(m) \cdot t_{\mathcal{B}}(r(m))$, and we denote $\mathcal{A}$ as polynomial-time Turing reducible to $\mathcal{B}$, written as $\mathcal{A} \leq_T \mathcal{B}$.

The algorithm for $\mathcal{A}$ has to solve each problem instance of $\mathcal{A}$. Depending on its input, it generates input instances for $\mathcal{B}$. The solutions to $\mathcal{B}$ returned by the algorithm for $\mathcal{B}$ are afterwards interpreted to generate a valid solution for $\mathcal{A}$. The problem instances for $\mathcal{B}$ could be encoded subproblems of $\mathcal{A}$. In the simplest case, the complete input instance of $\mathcal{A}$ is transformed to one input instance of $\mathcal{B}$, and the solution of $\mathcal{B}$ is retransformed accordingly.

If a lower bound for $\mathcal{A}$ is already known, then the runtime of the solver of $\mathcal{B}$ including the transformation cannot be smaller than this lower bound. This way the lower bound of $\mathcal{A}$ can be transferred to $\mathcal{B}$ by rearranging the inequality in Def. 2.8.

$$\frac{t_{\mathcal{A}}(m) - p(m)}{q(m)} \leq t_{\mathcal{B}}(r(m)), \quad q(m) \geq 1 \tag{2.7}$$

By this reduction two problems can be shown to belong to the same complexity class. For more dignified statements, the polynomials $p(m), q(m), r(m)$ of the transformation have to be small.
A reduction from the problem *Uniform Gap (UG)* to the dominated hypervolume $H(A, \mathbf{r})$ is performed to transfer a lower bound of UG to $H(A, \mathbf{r})$. In the concept of the reduction method described above, UG is the analyzed problem $\mathcal{A}$, and the hypervolume (HV) is the problem of interest $\mathcal{B}$.

Definition 2.9 *The following decision problem is called Uniform Gap (UG): Given a set* $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\} \subset \mathbb{R}$ *and* $\delta \in \mathbb{R}$ *decide whether the* $x^{(i)}, i \in \{1, \ldots, m\}$ *can be permuted by a permutation* σ *into a non-decreasing sequence such that the gap between each pair of consecutive elements equals* δ*, i.e.,*
$\forall 1 \leq i \leq m-1 : x^{(\sigma(i))} + \delta = x^{(\sigma(i+1))}$.

For Uniform Gap, a lower bound of $\Omega(m \log m)$ is known (see e.g. Preparata and Shamos (1988, Th. 6.16)). Obviously, the problem can be solved by sorting and calculation of the distances in time $O(m \log m)$, resulting in a complexity of $\Theta(m \log m)$.

Theorem 2.10 *Solving* $H(A, \mathbf{r})$*, with* $A \subset \mathbb{R}^d, |A| = m, \mathbf{r} \in \mathbb{R}^d$ *has a time complexity of* $\Omega(m \log m)$*. (Beume et al. (2009a)*).*

Proof. We show that UG is polynomial-time reducible to the 2-dimensional ($d = 2$) hypervolume (HV2). Thereby, the lower bound of $\Omega(m \log m)$ immediately follows for the 2-dimensional hypervolume due to Def. 2.8. The relating property of the problems is that the hypervolume of linearly positioned points is maximal if the points are equally spaced, as proved in Lemma 2.7. We argue that the reduction to higher-dimensional versions of the hypervolume works analogously.
Algorithm 2.1 solves UG by solving HV2 as follows. Let the input for UG be given as $A = \{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\} \subset \mathbb{R}$ and $\delta \in \mathbb{R}$. First the minimal and maximal element of A are determined in time $\Theta(m)$. Then, from each number in A, a 2-dimensional point is constructed by the number and its negation: $\mathbf{y}^{(i)} = (x^{(i)}, -x^{(i)})$. Thereby, the points lie on a line with slope -1 in the second and/or forth quadrant. By adding the minimal and maximal values respectively (Alg. 2.1, line 4), the set of point is shifted to the first quadrant with the minimal

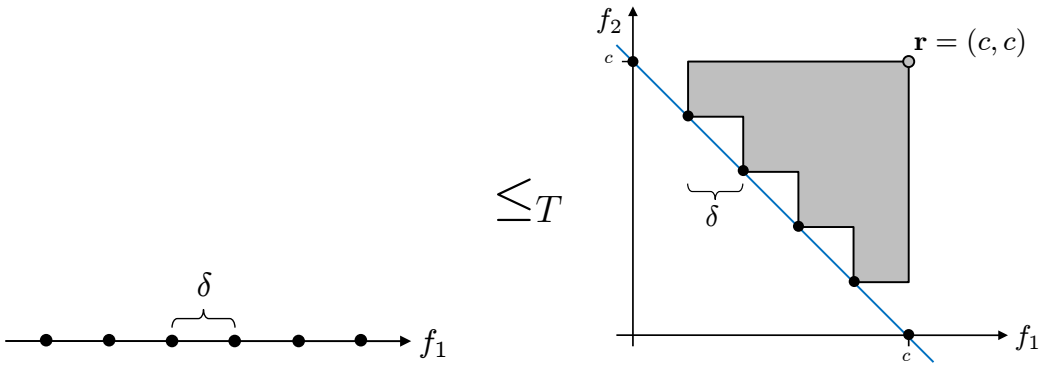


Fig. 2.8: Transformation from UG to HV2. The input numbers of UG are transformed in 2-dimensional coordinates and normalized such that the points lie on a diagonal in the first quadrant with the boundary point on the axes. Only if all neighbors in the original set have distance δ, the hypervolume is maximal.

Algorithm 2.1: UGsolver(A, δ)

input : set of m numbers $A = \{x^{(i)}, \ldots, x^{(m)}\} \subset \mathbb{R}$, distance $\delta \in \mathbb{R}$
output : Is the input an instance of UG? (true or false)

$c_{\max} = \max\{x^{(i)} \mid x^{(i)} \in A\}$
$c_{\min} = \min\{x^{(i)} \mid x^{(i)} \in A\}$
for $i \in \{1, \ldots, m\}$ **do** /* transformation to 2 dimensions */
  $\mathbf{y}^{(i)} \leftarrow (x^{(i)} + c_{\min}, \quad -x^{(i)} + c_{\max})$
$c = c_{\max} - c_{\min}$
$\mathrm{hv}_{\max} = \frac{1}{2}c^2 - \frac{c^2}{2(m-1)}$ /* acc. to (2.3) */
$\mathrm{hv} = H(\{\mathbf{y}^{(i)}, \ldots, \mathbf{y}^{(m)}\}, (c, c))$ /* call hypervolume algorithm for $d = 2$ */
if $\mathrm{hv} == \mathrm{hv}_{\max}$ **then**
  return true /* $(A, \delta) \in$ UG */
else
  return false /* $(A, \delta) \notin$ UG */

and maximal elements on the coordinate axes. The value $c = c_{\max} - c_{\min}$ describes the length of the interval enclosing the points (Fig. 2.8). Then, it is checked whether the transformed input corresponds to the setup of Lemma 2.7.

The line, the points lie on, can be considered as a Pareto front $g(y) = -y + c$. Then, we know by Lemma 2.7 that the hypervolume of points on a linear Pareto front is maximal if the points are equally spaced along the Pareto front. Thus, the hypervolume value of equally spaced points is unique, i.e., there is no other set of points yielding this hypervolume value.

The maximal hypervolume of uniformly spaced points on g is calculated according to equation (2.3) w. r. t. the reference point $\mathbf{r} = (c, c)$ in time $\Theta(1)$. We call a hypervolume algorithm for $d = 2$ to solve $H(\{\mathbf{y}^{(i)}, \ldots, \mathbf{y}^{(m)}\}, (c, c))$ and simply check, whether the returned value equals hv_{max}. In case of equality, UG is solved as $(A, \delta) \in UG$, and otherwise $(A, \delta) \notin UG$.

The runtime of Algorithm 2.1 without the hypervolume algorithm is linear ($p(m) = \Theta(m)$). The hypervolume algorithm is called exactly once ($q(m) = 1$), whereas the input size is $r(m) = \Theta(m)$. Equation (2.7) results in $t_{\mathcal{A}}(m) - \Theta(m) \leq t_{\mathcal{B}}(r(m)) \Rightarrow \Omega(m \log m) \leq t_{\mathcal{B}}(r(m))$, and we proved $UG \leq_T HV2$, with a lower bound of $\Omega(m \log m)$ for $HV2$.

The reduction to higher-dimensional versions of the hypervolume works by embedding the described 2-dimensional construction as the first coordinates and setting all other coordinates to 1, i.e., $\mathbf{y}^{(i)} = (x^{(i)} + c_{\min}, \quad -x^{(i)} + c_{\max}, 1, \ldots, 1)$. A hypervolume algorithm for the d-dimensional input is called with reference point $(c, c, 1, \ldots, 1)$. As multiplication by 1 does no change, the hypervolume value equals the one of the 2-dimensional version and is checked against $\text{hv}_{\max}$ which is calculated as before. Thus, it holds $UG \leq_T HV$ for arbitrary $d \geq 2$. The input size for HV is enlarged by factor d compared to the input of UG. This results in $\Omega(m \log m) - \Theta(dm) \leq t_{\mathcal{B}}(dm)$, thus $\Omega(m \log m)$ for HV with $d = o(\log m)$, and the same lower bound for $d = \Omega(\log m)$ for reading the input.

□

2.3 Upper Bound

The hypervolume is a geometric measure, so it appears to be obvious to investigate whether the community of computational geometry may provide a solution. The search for a similar problem resulted in the so-called *Klee's Measure Problem*.

Definition 2.11 *Let* $M = \{[\mathbf{a}^{(1)}, \mathbf{b}^{(1)}], [\mathbf{a}^{(2)}, \mathbf{b}^{(2)}], \ldots, [\mathbf{a}^{(m)}, \mathbf{b}^{(m)}]\} \subset \mathbb{R}^d$, $d \in \mathbb{N}^+$, $m \in \mathbb{N}$ *be a set of* m *d-dimensional intervals, each spanning a hypercuboid. Klee's Measure Problem (KMP) is the joined fair size (or measure) of the hypercuboids w. r. t. the Lebesgue measure* Leb *defined as:*

$$\mathrm{KMP}(M) = \mathsf{Leb}\left(\bigcup_{i=1}^{m} [\mathbf{a}^{(i)}, \mathbf{b}^{(i)}]\right). \tag{2.8}$$

Originally, the problem has been formulated for simple intervals in Klee (1977). Bentley (1977) generalized it to d-dimensions, where the d-dimensional intervals become axis-parallel hypercuboids, and we formulate the size here w. r. t. the Lebesgue measure. Note that, independently of its dimension, a hypercuboid is completely defined by two corners on a space diagonal, whereas the interval bounds give the 'lower left' and the 'upper right' corner. The similarity to the hypervolume is obvious in the chosen presentation as both measures calculate the size of axis-aligned hypercuboids (cf. Equation (2.1)). Beume (2006)* and independently Fonseca et al. (2006) observed the hypervolume to be a special case of the KMP.

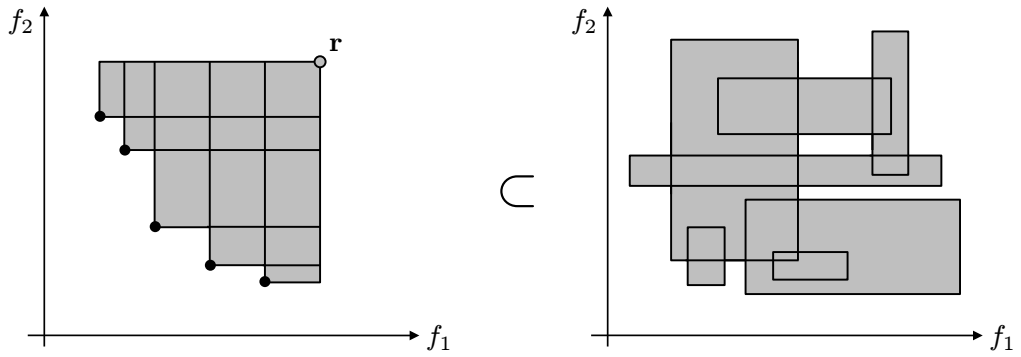


Fig. 2.9: Left: Illustration of the hypervolume consisting of overlapping hypercuboids (rectangles in 2 dimensions) spanned with the reference point. Right: Example of an instance of Klee's Measure Problem (KMP) with arbitrarily positioned rectangles.

To convert the input for the hypervolume into an input for the KMP, each point is taken as the lower bound of an interval and the reference point is the upper bound within all intervals.
While the hypercuboids of KMP may be positioned arbitrarily, the hypercuboids of the hypervolume problem are aligned in a special way (Fig. 2.9). All hypercuboids

share the reference point as their common upper bound, i.e., the upper right corner. If the input consists of mutually incomparable points only, then no hypercuboid is completely contained within others—a fact that corresponds to the observation that each non-dominated point has a positive contribution.
The KMP has a lower bound of $\Omega(m \log m)$ for $d \geq 1$ shown by Fredman and Weide (1978). The algorithm by Overmars and Yap (1991) with a runtime of $O(m^{d/2} \log m)$ has been the best known one until recently Chan (2010) yielded a slight improvement to $O(m^{d/2} 2^{O(\log^* m)})$ with $\log^*$ denoting the iterated logarithm. The new idea is a subdivision of certain hypercuboids. With the described transformation of the hypervolume to the KMP, the following theorem immediately holds.

Theorem 2.12 *The hypervolume of a set of m points in d dimensions can be calculated in time $O(m^{d/2} 2^{O(\log^* m)})$. (Beume (2009)*).* □

The algorithm described in the following is based on the one by Overmars and Yap (1991) and adapted to the special case of the hypervolume, i.e., the special alignment of the hypercuboids is utilized for simplifications.
In the following description of the algorithm, we use the vocabulary of geometry detached from the context of partially ordered points. For convenience, we omit the prefix 'hyper' and talk of 'volume' or 'cuboid' in arbitrary dimensions. A point is denoted to *cover* a region if it weakly dominates the region's lower boundary, thus the region is completely contained in the cuboid induced by the point. A point *partially covers* a region if its induced cuboid intersects the region.

2.3.1 Hypervolume Algorithm adopted from KMP Algorithm

The following presentation is mainly transferred literally from (Beume, 2009, Sec.4-5)*. The algorithm by Overmars and Yap (1991) is a sweep-line algorithm that uses a specific data structure to calculate a $(d-1)$-dimensional volume and performs a sweep along the remaining dimension to get the d-dimensional measure. For the partitioning of the $(d-1)$-dimensional space into regions, a data structure called *orthogonal partition tree* is used, i.e., a binary space partition tree whose splitting lines are extensions of the axis-parallel cuboids. An example of a 3-dimensional non-dominated set is shown in Figure 2.10, and its corresponding orthogonal partition tree managing 2-dimensional slices in Figure 2.11.
The significant idea of Overmars and Yap's algorithm is to not partition the space into empty and covered regions, but stopping the partitioning as soon as a region contains a grid structure of rectangles, called *trellis* in the following. In a trellis, the cuboids that intersect the region, cover it completely in each of the $(d-1)$ dimensions except one. An example of this structure is shown in Figure 2.12. A cuboid that does not cover the i^{th} dimension completely is called an *i-pile*. For each dimension i, the 1-dimensional KMP of the projection of the i-piles on the i^{th}

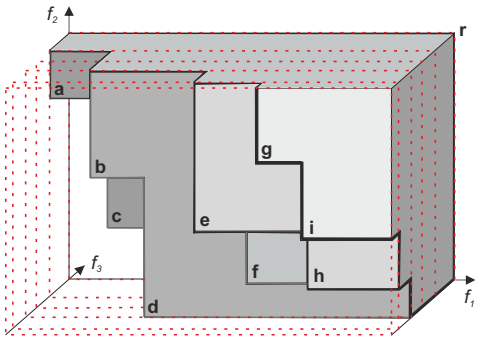


Fig. 2.10: The figure displays a non-dominated set of nine 3-dimensional points. The weakly dominated volume is bounded by the reference point **r**. Along the d^{th} coordinates, the d-dimensional space is cut into $(d-1)$-dimensional slices (dashed lines), which are stored in the orthogonal partition tree. The d-dimensional volume is calculated by computing the $(d-1)$-dimensional volume with the help of the orthogonal partition tree and sweeping along the slices in dimension d.

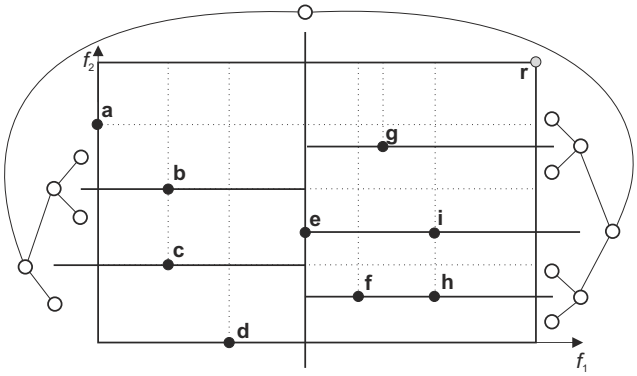


Fig. 2.11: Illustration of the 2-dimensional orthogonal partition tree for a 3-dimensional KMP. The non-dominated set of Figure 2.10 is projected on the first two dimensions. The lines show the partitioning of the 2-dimensional space, which is upper bounded by the gray reference point **r** and lower bounded by the contained points. The dotted lines adumbrate their induced weakly dominated cuboids. The orthogonal partition tree is depicted, whereas the nodes are placed alongside their associated region. The sweep is performed along the third dimension f_3.

coordinate axis is solved. Thereby the exact position of the cuboids is neglected. Let K_i denote the value of the 1-dimensional KMP of the i-piles, and L_i denote the size of the region in dimension i, respectively. The contained volume of the region

is calculated by the inclusion-exclusion principle (cf. Overmars and Yap (1991)) in constant time, assuming d is a constant:

$$\sum_{1\leq a\leq d-1}(-1)^{a+1}\left(\sum_{1\leq j_1<\ldots<j_a\leq d-1}\left(\prod_{1\leq i\leq a}K_{j_i}\prod_{l\in\{j_1,\ldots,j_a\}\wedge l\neq j_i}L_l\right)\right). \tag{2.9}$$

For clarification, we consider a 3-dimensional KMP with 2-dimensional volume in the regions. Then the volume is calculated as: $L_1K_2 + L_2K_1 - K_1K_2$.
[...] Overmars and Yap (1991) describe two variants of their algorithm. In the first version, the orthogonal partition tree is build up completely in a preprocessing step and the sweep is performed afterwards, inserting beginning cuboids into the data structure and removing enclosed ones. In the other variant, the data structure is build on the fly by splitting the current node if necessary. By recursing on the left child before the right, the partition tree is traversed in pre-order and the sweep is simulated whenever a leaf node is reached. This technique refers back to Edelsbrunner and Overmars (1985) and is called *streaming*. The orthogonal partition tree requires $O(m^{d/2})$ storage, whereas the streaming variant works with linear space as only one node is considered at one time. Thus it is to be preferred, easier to implement, and even more efficient because some special cases can be handled easier. Here, the algorithm based on the streaming variant and adapted to the hypervolume calculation (cf. Algorithm 2.2) is described in detail with remarks to differences to the original one by Overmars and Yap (1991).
The main procedure of the algorithm has the following parameters.

`double[][] region` The current region is represented by a two-dimensional array containing the vectors of the lower bounds and the upper bounds.

`list points` Points whose induced cuboids partially or completely cover `region` are stored in a list `points`.

`int split` The dimension at which `region` is cut to generate two child regions is called `split`.

`double cover` The value of the d^{th} coordinate of the first cuboid that covers the parent node's region is stored in `cover`.

Inputs of the algorithm are a set of non-dominated points and a reference point, thereby the cuboids are represented indirectly. The reference point `r`, the initial size m of the input set, and the dimension d are assumed to be known globally.
Before the main procedure `HOY` starts, the list of points is sorted ascending according to the d^{th} component of the vectors. This sorting will be maintained stable in all recursive calls of `HOY`. The procedure is initially called with the whole $(d-1)$-dimensional space as `region`, the non-dominated input set as `points`, `split`$= 1$ and `cover` as the d^{th} coordinate of the reference point `r`. A small example with $m = 9$ points in $d = 3$ dimensions is pictured in Figure 2.11.

Algorithm 2.2: HOY(region, points, split, cover)

```
1  coverNew = cover; coverIndex = 1; allPiles = true; bound = -1
                                                  /* is the region completely covered? */
2  while (coverNew == cover  &&  coverIndex != points.length) do
3      if covers(points[coverIndex], region) then
4          coverNew = points[coverIndex][d]
5          volume += getMeasure(region) * (cover - coverNew)
6      else coverIndex++
7  while (points[coverIndex][d] == coverNew  &&  coverIndex > 1) do coverIndex−−
8  if coverIndex == 1 then return
                                                  /* do the cuboids form a trellis? */
9  for i = 1 to coverIndex - 1 do
10     if checkPile(points[i], region) == -1 then allPiles = false
11 if allPiles then
                                                  /* calculate volume by sweeping along dimension d */
12     i = 1;  for j = 1 to d - 1 do trellis[j] = r[j]
13     repeat
14         current = points[i][d]
15         repeat
16             pile = getPile(points[i], region)
17             if points[i][pile] < trellis[pile] then trellis[pile] = points[i][pile]
18             i++
19             if i < coverIndex - 1 then next = points[i][d] else next = coverNew
20         until current != next
21         volume += measure(trellis, region) * (next - current)
22     until next == coverNew
23 else
                                                  /* split region in two children regions */
24     repeat
25         intersect = ∅; nonIntersect = ∅
26         for i = 1 to coverIndex - 1 do
27             intersection = intersects(points[i], region, split)
28             if intersection == 1 then add(points[i][split], intersect)
29             if intersection == 0 then add(points[i][split], nonIntersect)
30         if intersect ≠ ∅ then bound = median(intersect)
31         else if nonIntersect.length > √m then bound = median(nonIntersect)
32         else split++
33     until bound != -1
                                                  /* recurse on the two children regions */
34     regionC = region; regionC[1][split] = bound; pointsC = ∅
35     for i = 1 to coverIndex - 1 do
36         if partCovers(points[i], regionC) then move(points[i], pointsC)
37     if pointsC ≠ ∅ then HOY(regionC, pointsC, split, coverNew)
38     reinsert(pointsC, points);
39     regionC = region; regionC[0][split] = bound; pointsC = ∅
40     for i = 1 to coverIndex - 1 do
41         if partCovers(points[i], regionC) then move(points[i], pointsC)
42     if pointsC ≠ ∅ then HOY(regionC, pointsC, split, coverNew)
43     reinsert(pointsC, points)
```

The algorithm recursively splits the region, whereas the two resulting regions correspond to the children nodes within the binary tree. The splitting ends when the region contains a trellis, thus a leaf node is reached and the volume can be calculated. The procedure `HOY` consists of three parts. First it is checked if a cuboid covers `region`. If the remaining cuboids form a trellis, their hypervolume is calculated. Otherwise the region is further partitioned and the volume is calculated in recursive calls.
The d^{th} coordinate of the first covering point is saved as `coverNew` and the corresponding index in `points` as `coverIndex`. The volume is increased by the region's complete $(d-1)$-dimensional volume multiplied with the distance of `coverNew` to `cover`. Since the list `points` is sorted according to the d^{th} coordinate component, the points behind `coverIndex` do not add volume. These and the point itself are discarded in the remainder of this call of the procedure by considering `points` only to index (`coverIndex`-1). The remaining points are still required on higher levels of recursion. It may occur that points in front of `coverIndex` have the same d^{th} coordinate as the point at `coverIndex`. These points must be discarded, hence the additional while-loop reduces `coverIndex` if necessary. If `coverIndex`$=1$, `HOY` is aborted because no points are left. In the original description, only covering cuboids are removed. We added here, that covered cuboids are also discarded.
In the second part of `HOY`, the algorithm checks if the induced cuboids form a trellis. If so, the sweeping along the d^{th} dimension is performed to calculate the contained volume. The points with the first d^{th} coordinate (equal values may occur) are considered and $(d-1)$ 1-dimensional KMP are solved for them. The $(d-1)$-dimensional volume is calculated by the inclusion-exclusion-principle according to Eq. (2.9) and multiplied with the distance to the next d^{th} coordinate. This is done for all consecutive d-boundaries. The last distance in dimension d is calculated as difference to `coverNew`.

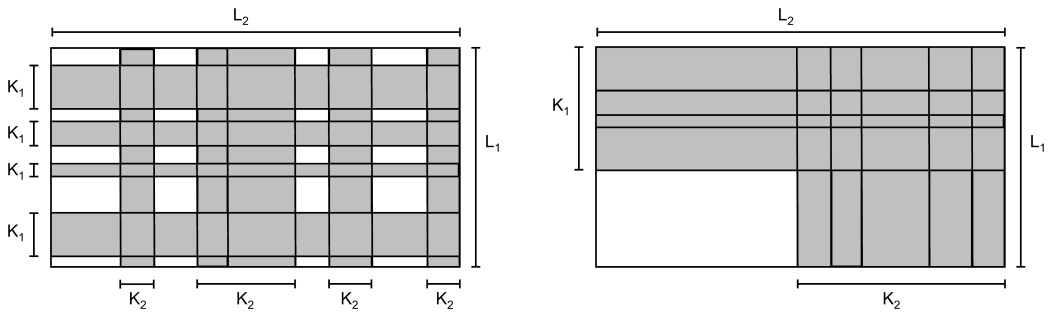


Fig. 2.12: The left figure shows an example of a 2-dimensional trellis for the general 3-dimensional KMP. The structure on the right arises for the specific problem of calculating the hypervolume, whenever the condition of a trellis is fulfilled.

To solve a 1-dimensional KMP on piles, Overmars and Yap invoke a segment tree to calculate the union of the 1-dimensional intervals. For the special case of the

hypervolume calculation, this can be done significantly faster and segment trees are not necessary. In case the cuboids fulfill the condition of a trellis, they actually form an even simpler structure. An example is shown in Figure 2.12 (right). Since each cuboid extends to the reference point in each dimension, no upper bounds of cuboids are contained inside of the current region. A region may only contain lower boundaries and the remainder of the region is covered from thereon. Thus, only the minimal i^{th} coordinate of the i-piles has to be identified. The result of the 1-dimensional KMP is the difference of this value to the region's upper bound, respectively in each dimension. The minimal values are stored in a $(d-1)$-dimensional array called `trellis`. Cuboids that become active during the sweep procedure are checked if they undercut the current values in `trellis`. Then `trellis` is updated in constant time by just one comparison. The update of the originally applied segment tree takes time $O(\log m)$. This factor is saved on by this adapted algorithm.

If the cuboids do not form a trellis, the region is split in two and the algorithm proceeds on the two emerged regions. The partitioning aspires that no points are contained inside of a region. To this end, the dimension that is cut by the splitting hyperplane is to be determined. As the cuboids are axis-parallel, the i^{th} coordinate of a point induces a so-called *i-boundary* that is a hyperplane which cuts through the i^{th} coordinate axis and is parallel to all others. The sub-procedure `intersects` detects whether a given point induces a `split`-boundary inside of the region. Points that additionally induce an i-boundary with $i <$`split` are stored in a list `intersect`, the others in `nonIntersect`. By recursion, the region will be split along each of the `split`-boundaries of the points in `intersect`. In each call of `HOY`, the median of these `split`-boundaries is chosen as the splitting hyperplane. This choice takes time $O(coverIndex)$. If `intersect` is empty, but `nonIntersect` contains more than $\sqrt{m}$ `split`-boundaries, the region is split along the median of these. The list `intersect` is especially empty for `split`=1. Thus, the points are partitioned into subsets of size $O(\sqrt{m})$ while splitting across the first dimension. If `intersect` is empty and there are not more than $\sqrt{m}$ `split`-boundaries in `nonIntersect`, `split` is increased and the search for the splitting line is tried again, beginning with the sub-procedure `intersects`.

In the example of Figure 2.11, the space is split once along the median 1-boundary. Afterwards, each region contains no more than $\sqrt{9} = 3$ 1-boundaries and `split` is increased. The left region is split along the median 2-boundary of those points that establish a 1-boundary within the region. Concerning the left child region, the point **d** is a 1-pile and no further partitioning is required. The right child region is split again because the point **b** is located inside of it.

Knowing the splitting line, the left child region is defined accordingly. Points that partially cover the child's region are sent down to recursion, together with the child region itself, the split value, and the value of `coverNew` of the current region. Afterwards, the points are reunited with the list `points` and the recursion on the right child's region is performed analogously.

Note that points are never copied, but moved from `points` to other lists if necessary. Thus, recursion does not cause any increase of storage, since each point is stored at only one place at one time. Invoking pointers to the elements in `points` would also be possible as their amount of storage is marginal. All lists of points are sorted, since this is done in the pre-processing step. Whenever a list is to be reunited with `points`, this can be done in linear time, whereas the sorting is maintained.
[...] Details on the implementation of the sub-procedures invoked during `HOY` (Algorithm 2.2) are described in a technical report (cf. Beume and Rudolph (2006b)*).

2.3.2 Theoretical vs. Experimental Runtime

[...] The two variants—the classical one and the streaming technique (cf. Section 2.3.1)—of the algorithm by Overmars and Yap (1991) have the same run time. Actually the same operations are done, though in different order. The authors describe the analysis for the variant which completely builds the orthogonal partition tree before the sweep. The pre-processive sorting requires $O(m \log m)$ time. It is shown that a cuboid is stored in $O(m^{(d-2)/2})$ leaves of the partition tree since the partitioning ensures that this is an upper bound for the number of partially covered regions.
The proof of the upper bound of the runtime is based on the number of partially covered regions per cuboid. Here, the explanation that this number does not exceed $O(\sqrt{m}^{d-2})$ is given. Recall that a cuboid partially covers a region if an i-boundary cuts through the region. This characteristic is illustrated in Figure 2.13. A region that is generated by a splitting through dimension i is termed an *i-partition*. There are $O(\sqrt{m}^{i-1})$ $(i-1)$-partitions. The i-boundary of a cuboid intersects an $(i-1)$-partition at most once and thereby cuts through one of its i-partition. Thus an i-boundary intersects $O(\sqrt{m}^{i-1})$ i-partitions. When the partitioning is done concerning the remaining $(d-1-i)$ dimensions, each i-partition is subdivided into $O(\sqrt{m}^{d-1-i})$ $(d-1)$-partitions. The cuboid's i-boundary cuts $O(\sqrt{m}^{i-1}) \cdot O(\sqrt{m}^{d-1-i}) = O(\sqrt{m}^{d-2})$ $(d-1)$-partitions, which corresponds to the number of leafs that contain the cuboid.
The contained volume within these leaves has to be updated when the cuboid is inserted or removed from the orthogonal partition tree during the sweep. Thus, over all cuboids there are $O(m^{(d-2)/2} \cdot m) = O(m^{d/2})$ updates. Updating means computing the measure in the trellis for each step of the sweeping. This originally takes time $O(\log m)$ with the help of the segment trees, but the adapted algorithm (Algorithm 2.2) computes an update in constant time. The number of inner nodes influenced by a cuboid is bounded by the number of influenced leaves times the depth of the tree, so updating these results in a run time of $O(m^{d/2} \log m)$.

To get a notion of the actual performance of the presented algorithm (termed HOY), it is applied to common test sets provided by While et al. (2006) and its

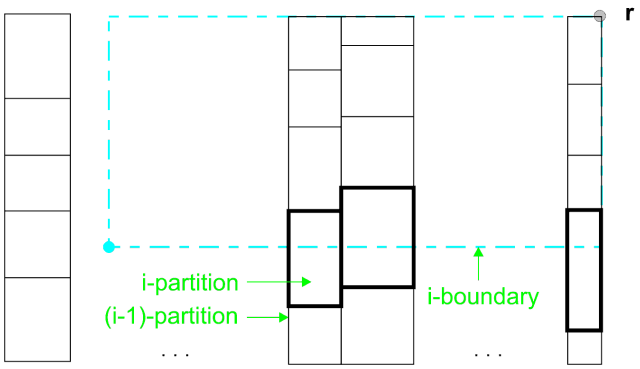


Fig. 2.13: Illustration of the number of intersected regions. The dashed lines adumbrate the induced weakly dominated hypercuboid. The columns show $(i-1)$-partitions with their contained i-partitions. The bold i-partitions are intersected by the i-boundary of the hypercuboid.

runtime is considered in comparison to algorithms by Fonseca et al. (2006) (termed FPL). The question we want to pose is which algorithm is the fastest on these test functions.

HOY has been implemented in C++ and the FPL algorithms in C. The programs were compiled with G++-4.2, GCC-4.2 respectively, with optimization level 3 and run under Ubuntu 8.04.1 on a single core Intel Pentium IV 3.4GHz processor with 1MB of cache, which is almost the same hardware setup as used in the study by Fonseca et al. (2006). To allow comparability we choose the test family DTLZ-Sphere (corresponding to DTLZ2, see Deb et al. (2002b)) and Random which were also invoked in the previous study and additionally DTLZLinear (DTLZ1). The data sets are random samples of these Pareto fronts of different size with three to nine objectives. The reference point $\mathbf{r}$ for the hypervolume calculation is chosen as $(2,\ldots,2)^\top, (11,\ldots,11)^\top, (1,\ldots,1)^\top$, respectively, so that all Pareto optimal points dominate it. For each setting (dimension and size fixed) 10 sample sets are processed. Source code of the algorithms are available at the authors' homepages.

The variants FPL3 and FPL4 have been considered since they outperform variants 1 and 2 in the earlier study. FPL4 shows a slightly better performance than FPL3 with lower variance, so only the results of FPL4 are shown here. They resemble the results in Fonseca et al. (2006) where objectives have been read in reverse order to comply with the preceding study by While et al. (2006). Since FPL4 outperforms the HSO in that study and no implementation of the HSO with the permutation heuristics by While et al. (2005) is available, HSO is not considered here.

The experimental results for the sets with $d \in \{3,4,5,6\}$ are shown using Box-and-Whisker plots in Fig. 2.14. Each set is processed twice: in original order and with objectives in reverse order, whereas the longer runtime corresponds to the latter.

On the sets, both algorithm perform similar for $d < 5$ but FPL4 is a bit faster. For $d \in \{5,6\}$ FPL4 is highly sensitive to the order of the objectives in the input. On

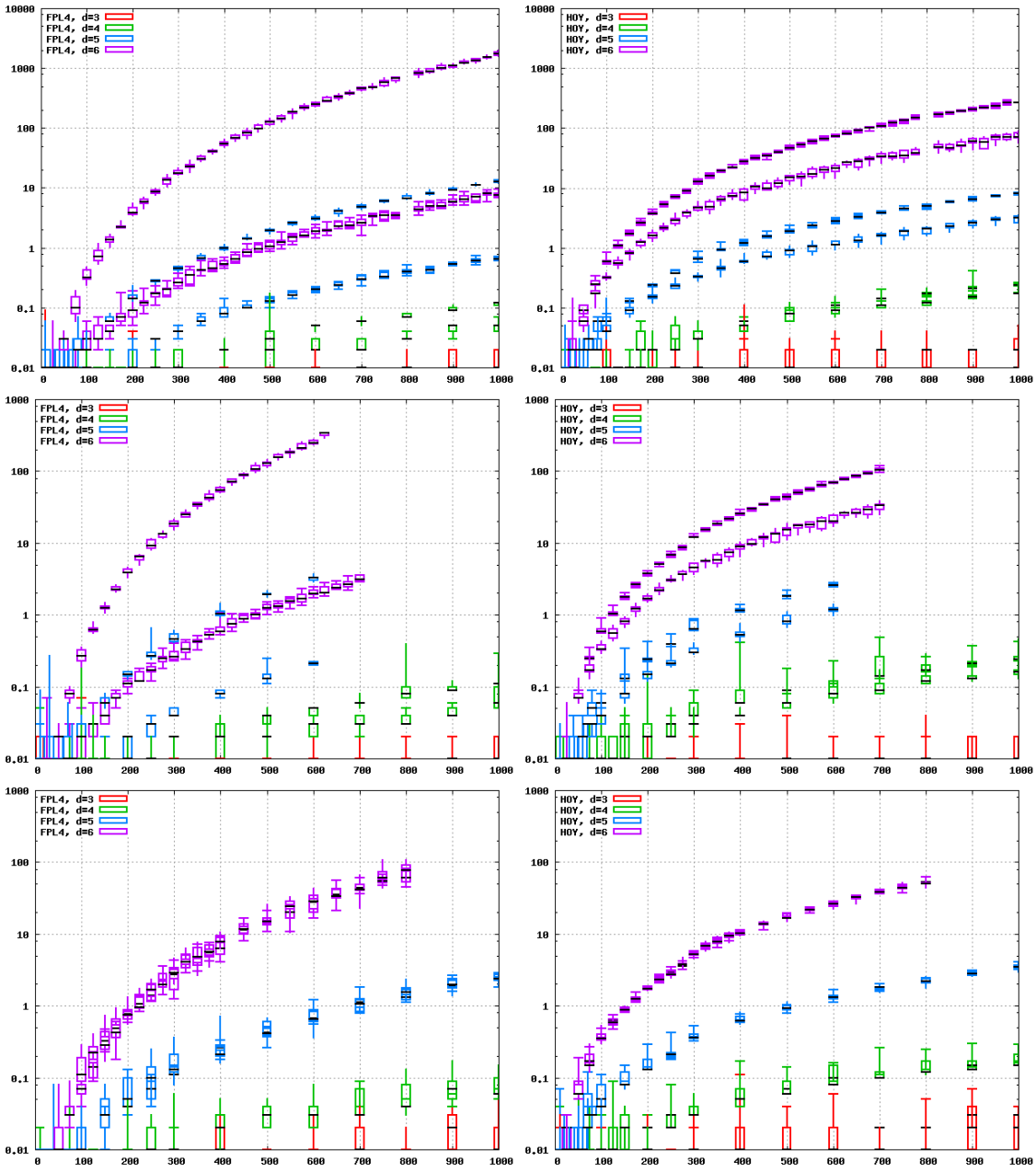


Fig. 2.14: Results of FPL 4 (left) and HOY (right) on the test set DTLZSphere (top), DTLZLinear (middle), and Random (bottom), with the number of points on the horizontal axis and the computation time in seconds on the logarithmic vertical axis. The objectives in the input set are in the original order for the lower runtimes and in reverse order for the higher ones, respectively. The DTLZSphere data set with 800 6-dimensional points is corrupt so the results have been left out.

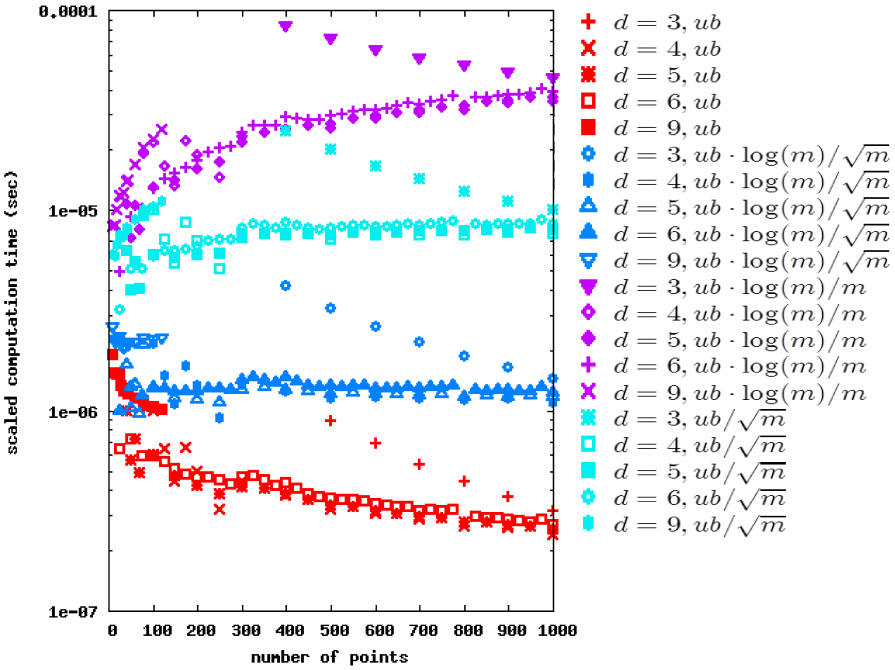


Fig. 2.15: Results of HOY on the test set DTLZSphere divided by different functions as potential upper bounds of the runtime.

the spherical and the linear Pareto fronts (Fig. 2.14, top, middle) there is partly a hundredfold magnification of the runtime. HOY's runtime does not deviate that much due to permutations of the input so that its runtime lies between FPL4's slow and its fast runs. This might be due to the different techniques of dividing the space into sub-problems. On the randomly generated fronts no significant difference can be observed for different input orders. Further studies on different permutations emphasize these observations.

Although HOY has a lower worst case runtime, this advantage seem not to strike through for low dimensional sets of the considered size. Larger sets seem to be required to reveal the lower magnitude of the runtime resembling the asymptotic behavior. For the 5- and 6-dimensional data HOY performs better than the slow runs of FPL4 also for small sets.

The indifference of permuting the random sets might be due to the lack of structure so that no adversely constellation is formed. The questions arise how worst case instances might be structured and if they occur at all in the practical application scenario of the algorithms. To get insight to this, other sets featuring different properties shall be considered in future studies. Both algorithms might be improved by the heuristics of While et al. (2005), where it is expected that FPL gains more since there seemed to be easy constellations for it. However, it is unclear how much speed-up can be gained. The HOY serves as a robust method regarding the order of the input.

To visualize the order of the measured runtime, we consider the longest observed runtime on DTLZSphere (sets in reverse order) and divide it by an asymptotic upper bound of $m^{d/2}$ (denoted as ub) as shown in Fig. 2.15 (red symbols). The graph is not a constant but decreases alluding that the concrete runtime is lower. The blue curve with an additional factor of $\log(m)/\sqrt{m}$ seems to be an adequate estimate for this instance. The concrete constant of the runtime is irrelevant here and just reflects processor speed. Small 3-dimensional sets seemed to fall out of the schema, so it is assumed that their runtime is dominated by constants that lose influence on the magnitude for larger sets.

2.4 Conclusions

This chapter recapitulates the state-of-the-art on hypervolume research detailing our contributions on distributions of points, and lower and upper bounds of the runtime. In Section 2.1, properties of the hypervolume are presented in very simple notation (waiving additional symbols from literature) to make the statements easily readable. An overview of the research on multiobjective optimization with an emphasis on the hypervolume is given in Brockhoff (2010b).

Though Bringmann and Friedrich (2008) proved the problem complexity as $\#P$-hard, a concrete lower bound is yet interesting as the complexity class does not give a sound insight in small problem instances. We give the computational complexity (matching lower and upper) for the highly practical problem instances of $d = 2$ and $d = 3$. It might be negatively surprising that the 2-dimensional problem cannot be solved faster than $\Omega(m \log m)$ time or positively surprising that the 3-dimensional hypervolume is of equal complexity. The true problem complexity is expected to be larger than the lower bound for higher dimensions. Further dimension-specific complexity results would be of interest.

The result regarding the distribution of points is a byproduct of two publications. Recently, the topic gained attention to deepen the understanding of the structure of sets with maximal hypervolume.

Identifying the hypervolume as a special case of Klee's measure problem established a new basis for theoretical analyses. This insight reduced the best known upper bound dramatically from $O(m^{d-1})$ to $O(m^{d/2} \log m)$. Our simplified algorithm, adapted to solving the hypervolume, made this insight amenable to and utilizable for the community of multiobjective optimization. It led to the current best upper bound of $O(m^{d/2} 2^{O(\log^* m)})$, and enabled and initiated further advance, like algorithms for the hypervolume contribution.

The future of the hypervolume calculation seems to lie in the approximation, at least when integrated into optimization algorithms to solve high-dimensional problems. Further results, regarding the approximation factor of the hypervolume for $d > 2$ are desirable. Yet faster exact algorithms are of course interesting as well.

A promising open problem is the efficient update of data structures maintaining the hypervolume. When calculated internally of an evolutionary multiobjective optimization algorithm, the slight changing of the set from one generation to the next could be sped up.

3 Characteristics of Hypervolume-based EMOA

Hypervolume-based EMOA are state-of-the-art methods for multiobjective optimization. Our examinations are centered around certain features of the SMS-EMOA (S-metric selection EMOA), whereas the results hold for a broader class of algorithms as we generalize the aspects we focus on.
The following Section 3.1 introduces basic properties of the SMS-EMOA in comparison to other popular EMOA. Section 3.2 analyzes how fast EMOA converge towards the Pareto front. Assuming that the population is located on the Pareto front, we analyze whether the maximal hypervolume is reached by distributing the points (Sec. 3.3). Afterwards, the performance of the SMS-EMOA is demonstrated especially for problems with more than three objectives, while other EMOA fail (Sec. 3.4). Section 3.5 is dedicated to the parameter tuning of the SMS-EMOA to adapt it to application problems. Finally, a hybrid SMS-EMOA is presented that invokes gradients of the hypervolume in Section 3.6.

3.1 Concepts and Properties

Quality indicators are applied to evaluate and compare the approximations of the Pareto front generated by optimizers. Among these, the hypervolume is the most accepted one and considered to be significant and fair. Since it is agreed that the optimization goal is to reach a high hypervolume value (besides possibly other quality values), the idea to directly aspire the hypervolume maximization within the optimization process suggests itself. So, in hypervolume-based EMOA, the hypervolume is used within the selection operator as a single-objective substitute function. This way, a total order is established in the population, which solves the problem of comparing individuals that are incomparable regarding the dominance relation. In the following, we give an overview of hypervolume-based EMOA with a focus on the SMS-EMOA, next to related EMOA considered in this chapter.

SMS-EMOA

The SMS-EMOA (S-metric Selection EMOA) developed by Emmerich et al. (2005)*, Beume et al. (2007)* is a well-accepted hypervolume-based EMOA. Not all operators of this EMOA are specified as it is mainly a selection concept. The

Algorithm 3.1: $(\mu + 1)$-SMS-EMOA

```
choose initial population P^(0) of size μ from ℝ^n, set t = 0
repeat
    choose parents for variation
    create offspring Y^(t) from parents by variation
    evaluate offspring Y^(t)
    Q^(t) ← (P^(t) ∪ Y^(t))
    {F_1, ..., F_v} ← non-dominated-sorting(Q^(t))
    if |F_v| > 1 then
        calculate adaptive reference point w. r. t. Q^(t)
        calculate hypervolume contributions of points in F_v
        remove element with smallest contribution from P^(t)
    else remove single element in F_v from P^(t)
    increase generation counter t
until termination criterion fulfilled
```

other operators, namely initialization, parent selection, and the variation can be chosen problem-specific—no special operator is required for the success of the hypervolume-based selection. Yet, it is detailed below which operators are usually applied, after the description of the selection. An overview of the SMS-EMOA is given in Algorithm 3.1.
The selection is performed in a steady-state scheme, i.e., $(\mu+1)$, where exactly one individual is generated per generation and a plus selection is performed to choose the best succeeding population. This is the most direct way to incorporate the information of the new offspring in the next iteration, and contrasts the popular $(\mu + \mu)$ selection of other EMOA.
After the offspring is generated, non-dominated sorting is performed for the population unified with the offspring. This sorting partitions the individuals according to their objective vectors into non-dominated sets or anti-chains.

Definition 3.1 *Non-dominated sorting is a hierarchical partitioning of a multiset F with elements from $\mathbb{R}^d$ into partitions (or fronts) $F_1, ..., F_v$ such that the following properties hold:*

1. $F_1 = ndms(F)$.
2. $\forall i$ *with* $2 \leq i \leq v$: $\quad F_i = ndms(F \setminus \bigcup_{1 \leq j < i} F_j)$.
3. $ndms(F_v) \setminus F_v = \emptyset$.

The last aspect implies that partitioning according to 1. and 2. results in exactly v partitions. The partitioning bases on Dilworth's decomposition theorem (cf.

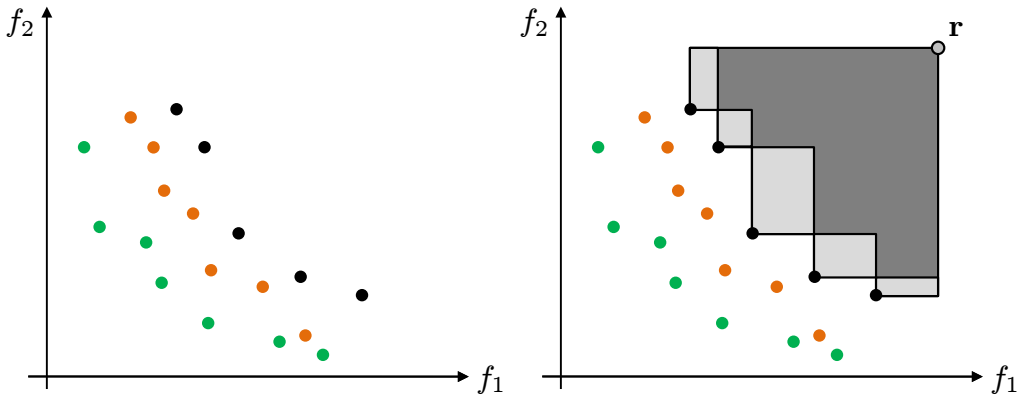


Fig. 3.1: Selection of SMS-EMOA. Left: Non-dominated sorting of $Q^{(t)}$ into fronts F_1, F_2, F_3. Right: For the points of the worst front, the hypervolume w. r. t. the adaptive reference point and their hypervolume contributions are calculated (light gray rectangles).

Dilworth (1950)) saying that a set of size h (in the case of dominance a transition chain of $h-1$ dominance relations) can be partitioned into h antichains or non-dominated sets. A first algorithmic description for the application in the field of evolutionary computation has been formulated by Goldberg (1989) roughly as follows: The non-dominated multiset of F is determined, resulting in the first front F_1. Then this subset is removed from F and the non-dominated multiset of $F \setminus F_1$ is determined. This multiset is removed and the process iterates until the set F is empty, see Fig. 3.1 (left) for an example. The usage of multisets instead of sets causes that copies are contained in the same front. The non-dominated sorting was made popular by the application in the NSGA and NSGA-II as described below. The runtime is $O(\mu \log^{d-1} \mu)$ due to Jensen (2003) for a population of size μ in a d-dimensional objective space.

After the partitioning, the worst individuals are in the worst front F_v. Does F_v only contain one element, then this is discarded, the population size is thereby reduced to size μ and the generation completed. In case that F_v contains more than one individual—which all are incomparable to each other—the hypervolume is invoked to establish an ordering. To this end, the adaptive reference point (cf. Def. 2.6) is determined relative to $Q^{(t)}$ (cf. Algorithm 3.1, line 9) thus has in each coordinate the maximal value of the population plus one. Then the hypervolume of $H(F_v, \mathbf{r})$ is calculated as well as the hypervolume contribution of each element of F_v to this value. A naive way to do this is to calculate the contribution canonically according to its definition as the set's total hypervolume minus the set without that point. Using the algorithm by Bringmann and Friedrich (2009b) the hypervolume contributions are calculated in one algorithm run, resulting in a runtime of $O(\mu^{d/2} \log \mu)$, which is as well the total runtime of SMS-EMOA per generation without resources for function evaluations.

Algorithm 3.2: SBX and PM variation

input : parents $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}$ from population $P^{(t)}$, parameters η_c, η_m
output : offspring $\mathbf{x}^{(o)}$

/* SBX */

choose u uniformly from $[0, 1[$

if $u \leq 0.5$ **then** $\beta_i = 2u^{\frac{1}{\eta_c+1}} - 1$ **else** $\beta_i = \left(\frac{1}{2(1-u)}\right)^{\frac{1}{\eta_c+1}}$

$x_i^{(o1)} = 0.5(\ (1+\beta_i)x_i^{(1)} + (1-\beta)x_i^{(2)}\ )$

$x_i^{(o2)} = 0.5(\ (1-\beta_i)x_i^{(1)} + (1+\beta)x_i^{(2)}\ )$

$\mathbf{x}^{(o')} \leftarrow$ choose uniformly $\mathbf{x}^{(o1)}$ or $\mathbf{x}^{(o2)}$ /* 1 offspring in SMS-EMOA */

/* PM */

set boundaries $x_i^{\min}$ and $x_i^{\max}$ for each decision variable x_i

choose u uniformly from $[0, 1]$

if $u < 0.5$ **then** $\delta_i = (2u)^{\frac{1}{\eta_m+1}} - 1$ **else** $\delta_i = 1 - (2(1-u))^{\frac{1}{\eta_m+1}}$

$x_i^{(o)} = x_i^{(o')} + \delta_i(x_i^{\max} - x_i^{\min})$

In the following, the typical specification of the remaining operators is described. The initialization is done uniformly at random, i.e., the initial individuals start at random points within the search space. The parent selection is performed uniformly as well, i.e., from the population two individuals are chosen (with replacement) as parents, contrarily to the binary tournament used in most other EMOA.

As variation operators, the simulated binary crossover (SBX) and the polynomial mutation (PM) by Deb and Agrawal (1995) and Deb and Goyal (1996) are often used. SBX tries to imitate one-point crossover on binary strings where the offspring is created by adopting the values from one parent until the index of the crossover point, and beyond the values from the other parent. Details are given in Algorithm 3.2 according to the description in Deb (2001). The spread factor β determines the distance of the offspring to the parents. SBX always creates two offspring, so for the SMS-EMOA only one of them is chosen randomly and then mutated, whereas for other EMOA both offspring are used.

External parameters of SBX and PM are the distribution indexes $\eta_c, \eta_m \in \mathbb{R}^+$, respectively, controlling the variance. For our experiments, our implementation bases on the NSGA-II implementation (cf. Deb et al. (2005b)) which uses two additional parameters p_c and p_m. The parameter p_c is the application probability of SBX, so if it is not performed, the parents are copied unchanged. The parameter p_m gives the probability to perform PM per decision variable, if it is not applied the variable is left unchanged. The implementation of SBX and PM in NSGA-II slightly differs in other aspects from Algorithm 3.2, e.g. SBX is only performed with a probability of 0.5 per variable.

Variants of SMS-EMOA

The SMS-EMOA on bi-objective problems can operate without a reference point by always keeping the boundary solutions, i.e., those without a worse neighbor in one objective. The SMS-EMOA has been introduced with this concept in Emmerich et al. (2005)*, whereas the adaptive reference point has firstly been suggested in Naujoks et al. (2005a)*. However, the usage of the adaptive reference point is recommended.
Several variants of the SMS-EMOA with slightly different selection concepts have been presented in Naujoks et al. (2005a)*. They are not considered in his thesis but briefly mentioned for completeness. One additional selection criterion is the number of dominating points of a point $\mathbf{v}$ in a population $P^{(t)}$:

$$d(\mathbf{v}, P^{(t)}) = \{\mathbf{a} \in P^{(t)} \mid \mathbf{a} \prec \mathbf{v}\}. \tag{3.1}$$

After non-dominated sorting has been performed, the selection is such that the hypervolume is only invoked when the population and the offspring form a non-dominated set, i.e., the non-dominated sorting resulted in a single front. Otherwise, in case of dominated individuals, the number of dominating point is used to determine a worst individual among those of the last front. For each point, the number of points dominating it is counted, and the point with the highest number is considered to be the worst. It is a kind of diversity measure as it assumes that point within an areas of many better ones are dispensable, whereas points in less explored regions are kept. This measure is moreover faster than the hypervolume calculation. It is expected that the dominating points measure mainly comes into operation in the early generations. After some iterations, the population typically is a non-dominated set and the situation is changed only slightly by the offspring that may be incomparable, dominated or dominating. So the cheaper measure is applied in the beginning of the optimization and the more time-consuming hypervolume when more precision is required close to the Pareto front. Thereby the algorithm's runtime is reduced and performance studies showed that this version performs similar to the original SMS-EMOA (cf. Beume (2006); Naujoks et al. (2005a); Wessing and Naujoks (2010)*). The non-dominated sorting can also be left out in this version of the SMS-EMOA, making the algorithm conceptually simpler, while still maintaining its effectiveness (cf. Beume (2006)*).
Wessing and Naujoks (2010) studied variants with offsets for the adaptive reference point different from the vector of ones. Yet, no clear conclusions could be gained.
Two variants of the SMS-EMOA with dynamic population size have been considered in Naujoks et al. (2005a)*. The idea is that a small population shall be able to move faster (anent to iterations) towards the Pareto front. Starting with a small initial population size, only the non-dominated individuals are kept while all dominated individuals are discarded. The population then grows and develops into a larger non-dominated set. One variant is to define a certain maximal population size and to discard one individual via the hypervolume selection in case that the size

is exceeded. A second variant is to switch to a selection with fixed population size, i.e., the original selection as soon as the maximal population size is reached. Thereby dominated individuals may survive in the population and the idea is that this may increase the diversity and simplify the spreading along the Pareto front.
The SMS-EMOA has moreover been considered in a meta-model-assisted framework in Emmerich et al. (2005)*, Beume et al. (2007)*, Naujoks et al. (2005b)*. A Kriging model is applied to pre-evaluate individuals and to evaluate only promising individuals exactly. This shall reduce the amount of time-consuming evaluations of real-word problems.

Properties

The hypervolume of the population tends to grow throughout the optimization, but it is not monotonically increasing. It may happen that the population's hypervolume decreases compared to the previous generation, when this process is observed by calculating the hypervolume w. r. t. a fixed reference point. This is due to the internally adaptive reference point which alters the optimization goal.
The SMS-EMOA aims at optimizing the hypervolume of its population by keeping the individuals with highest contribution to this measure, which shall minimize the loss caused by the selection. Yet, this approach is just a surrogate of what is actually aspired. The maximization of the points' hypervolume contributions is not equivalent to the maximization of their total hypervolume. This is obvious for a continuous Pareto front: The limit values of the distances between points are zero, and so are their hypervolume contributions as their dominated regions overlap almost completely but the hypervolume is maximal. So, the hypervolume contribution is an effective concept but not equivalent to aiming at optimal positions of all points.
Brockhoff et al. (2008) performed a theoretical runtime analyses of an algorithm called SIBEA which corresponds to the SMS-EMOA framework. They considered two academic test problems in binary spaces. For the problem LOTZ (leading ones, trailing zeros) a runtime of $O(\mu n^2)$ has been shown, while for the more complex one, the authors indicate to be unsure about the correctness of the proof. This is the only runtime analysis of hypervolume-based EMOA, so that our convergence rates in Section 3.2 are the only results for continuous search spaces.

Applications

The SMS-EMOA has been applied to aerodynamic test problems, namely the 2-dimensional NACA in Emmerich et al. (2005)*, Naujoks et al. (2005b)*, Beume et al. (2007)* and the 3-dimensional RAE 2822 in Naujoks et al. (2005a)*, Naujoks et al. (2005b)*, Beume et al. (2007)*, Beume et al. (2008)* which are as well described in Section 3.5. The article Beume et al. (2008)* is the first presentation of the SMS-EMOA in a German journal. It reviews the previous work and describes

applications to real-world problem. Next to the aerodynamic problems, there are several application to vehicle routing problems, an optimization in chemical engineering, and a gear shaft optimization. Further applications of the SMS-EMOA are e.g. on surface reconstruction by Wagner et al. (2007b), water distribution networks by Reehuis et al. (2011), software architectures by Li et al. (2011), job scheduling in grid computing by v. d. Kuijl et al. (2010), and molecular control using femtosecond lasers techniques by Klinkenberg et al. (2010).

Other Hypervolume-based Methods

We give a brief overview of other methods using the hypervolume.

IBEA With the IBEA (Indicator-based EA) by Zitzler and Künzli (2004) a framework of applying a quality indicator in the selection has been developed, thereby it gave the name to a class of EMOA. Variants with the hypervolume as indicator (IBEA_{HD}) and the additive ϵ indicator (IBEA_{ϵ^+}) are popular. Contrarily to the SMS-EMOA, a binary quality indicator is invoked for a pairwise comparison of individuals. We consider the mentioned two variants of IBEA in Section 3.2 and in the benchmarking of Section 3.4.
The IBEA performs a binary tournament for mating, and after arbitrary variation operators, the following indicators are used for the fitness assignment and a $(\mu+\mu)$ selection. For two points, $I_{\epsilon+}(\mathbf{a}, \mathbf{b})$ calculates the minimal distance ϵ by which $\mathbf{a}$ can be or needs to be moved in each direction in order to weakly dominate $\mathbf{b}$:

$$I_{\epsilon+}(\mathbf{a}, \mathbf{b}) = \min_{\epsilon \in \mathbb{R}} \{\forall i \in \{1, \ldots, d\} : f_i(\mathbf{a}) - \epsilon \leq f_i(\mathbf{b})\}. \tag{3.2}$$

The indicator value is negative or zero for $\mathbf{a} \prec \mathbf{b}$, so the smaller the better.
For two points, I_{HD} is defined as

$$I_{HD}(\mathbf{a}, \mathbf{b}) = \begin{cases} H(\{\mathbf{b}\}, \mathbf{r}) - H(\{\mathbf{a}\}, \mathbf{r}) & \text{if } \mathbf{a} \prec \mathbf{b} \\ H(\{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) - H(\{\mathbf{a}\}, \mathbf{r}) & \text{otherwise.} \end{cases} \tag{3.3}$$

The definition is similar to the hypervolume contribution in the latter case, but also defined for dominated ones. The indicator values are subsumed in the fitness function

$$F(\mathbf{a}) = \sum_{\mathbf{b} \in P \setminus \{\mathbf{a}\}} -e^{-I(\mathbf{b}, \mathbf{a})/\kappa} \tag{3.4}$$

including a scaling factor κ. The fitness is to be maximized. The worst individual is discarded iteratively, while updating the fitness values after each reduction of the population. The runtime is stated as $O(\mu^2)$ in Zitzler and Künzli (2004).

MO-CMA-ES The CMA-ES (Covariance Matrix Adaptation Evolution Strategy) by Hansen and Ostermeier (2001), Hansen et al. (2003), Hansen (2006) is the state-of-the-art EA for single-objective optimization. In EC, the mutation of a genome has traditionally been done by adding a normally distributed vector. In order to efficiently explore complex search landscapes, a distortion of the distribution is desired, which the CMA-ES realizes by the adaptation of the covariance matrix of the normal distribution. The multiobjective version of the CMA-ES, the MO-CMA-ES by Igel et al. (2007), Suttorp et al. (2009), Voß et al. (2010), results in a powerful EMOA. The MO-CMA-ES is not included in our benchmark on many-objective problems as it has not been developed till then.
MO-CMA-ES combines the selection of the SMS-EMOA with the variation from the CMA-ES. One parent is selected uniformly at random from the population, it is copied to generate an offspring, which is then mutated according to the parent's strategy variables. The selection is performed according to the non-dominated sorting and the hypervolume like in the SMS-EMOA. If the offspring survives, i.e., is not the worst individual, its step size and Cholesky factor are updated (Suttorp et al. (2009, Procedure `updateCholesky`)), followed by an update of the parent's step size (see Voß et al. (2010)). The Cholesky factor determines the shape of the distribution used for future mutations of this individual's offspring. The covariance matrix is not directly calculated but given implicitly by the Cholesky factor. Starting from the unit covariance matrix it is stepwisely adapted, aiming at the approximation of the inverse of the Hessian.
Depending on the dimensions of the problem and the algorithm parameterization, the runtime of the MO-CMA-ES is either dominated by the covariance matrix adaptation, requiring $\Theta(n^2)$ or by the hypervolume computation for the selection as $O(\mu^{d/2} \log \mu)$.

ESP A first hypervolume-based EMOA was ESP (Evolution Strategy with Probabilistic mutation) by Huband et al. (2003), which uses a rough estimation of the hypervolume contribution for selection. For a point $\mathbf{y}$ in a set M, it is defined as

$$\mathrm{H}_{ESP}(\mathbf{y}, M) \;=\; \prod_{i=1}^{d} \min\{y'_i - y_i \;\mid\; y_i < y'_i \;,\;\; \mathbf{y}' \in M\}, \tag{3.5}$$

which corresponds to the product of the distances to the neighboring points. The boundary points are assigned a value of infinity. In a 2-dimensional space, this measure of the inner points indeed equals the hypervolume contribution, but for higher dimensions, it is an approximation of unknown quality. The selection is performed in a $(\mu + \mu)$ scheme.

HypE Bader and Zitzler (2011) developed the hypervolume-based EMOA HypE (hypervolume estimation), which uses an estimation of the hypervolume via Monte Carlo sampling. They exploit the fact that it can be decided efficiently whether a point is located inside or outside the dominated hypervolume.

SPAM A more general framework than IBEA for respecting several preferences, e.g. expressed by quality indicators, in a hierarchical process is SPAM (Set Preference Algorithm for Multiobjective optimization) by Zitzler et al. (2008b). It is specified which properties indicator shall have in order to refine the order with each stage. The hypervolume is highly qualified to be applied in the framework due to its strong properties.

Archiving Strategies The selection according to a quality measure is not only applied within populations but also for maintaining an external archive (of an arbitrary multiobjective optimizer) of fixed size as formulated by Knowles and Corne (2003). They describe a reduction procedure $\texttt{AA}_{reduce}$ for updating an archive and also suggest the hypervolume contribution as a selection criterion.

Swarm Algorithms Following the principals from hypervolume-based EMOA, analogous techniques have also been developed for swarm algorithms, see e.g. Mostaghim et al. (2007), or artificial immune systems by Pierrard (2011) and may as well be integrated into other methods.

Related EMOA

We give an overview of other EMOA studied in this chapter, which do not invoke the hypervolume.

NSGA-II The NSGA-II (Non-dominated Sorting Genetic Algorithm) by Deb et al. (2002a) is clearly the most popular EMOA and has been used in hundreds of applications. Its original version is nowadays no longer recommended as it is outperformed by modern EMOA. Our studies contributed to this notion by performance benchmarks, revealing its drawbacks especially in case of many objectives as analyzed in Section 3.4. We also refer to NSGA-II in Section 3.2.

NSGA-II was the archetype of SMS-EMOA, so they have in common the non-dominated sorting, whereas the hypervolume contribution is invoked instead of the crowding distance as the secondary selection criterion, and the NSGA-II uses a $(\mu + \mu)$ selection. The parents are determined through a binary tournament, and SBX and PM are applied as variation operators. The selection process starts with non-dominated sorting of the population and the μ offspring. The first fronts are included as long as they completely fit into the next population. From the front which can only partly be included due to the population size, the best individuals are determined with the diversity measure *crowding distance*. It is defined as

half the perimeter of the hyperrectangle spanned by the neighboring points of the considered front M:

$$\begin{aligned} cd(\mathbf{a}, M) &= \sum_{i=1}^{d} \quad \min\{b_i - a_i \mid a_i < b_i \ , \ \mathbf{b} \in M\} \\ &\quad + \quad \min\{a_i - c_i \mid c_i < a_i \ , \ \mathbf{c} \in M\}, \end{aligned} \tag{3.6}$$

and for boundary points as infinity.
NSGA-II has a runtime of $O(\mu \log^{d-1} \mu)$, see Jensen (2003), per generation without counting the resources for function evaluations.

SPEA2 SPEA2 (Strength Pareto EA 2) by Zitzler et al. (2001, 2002) as well belongs to the outdated algorithms nowadays, supported by our demonstration that it is not qualified for many objectives (see Section 3.4). We also refer to it in Section 3.2.
In addition to the population $P^{(t)}$, SPEA2 uses an archive $\overline{P}^{(t)}$ of mainly non-dominated individuals, both of fixed size. Parents are chosen via binary tournament selection from the archive $\overline{P}^{(t)}$. After arbitrary variation operators and evaluation, the fitness is determined as follows. A strength value of each point is defined as the number of points it dominates. For each point in $(P^{(t)} \cup \overline{P}^{(t)})$, the strength values of its dominating points are subsumed to its raw fitness

$$raw(\mathbf{a}) = \sum_{\mathbf{b} \text{ with } \mathbf{b} \prec \mathbf{a}} |\{\mathbf{c} \mid \mathbf{b} \prec \mathbf{c}\}|, \quad \mathbf{a}, \mathbf{b}, \mathbf{c} \in (P^{(t)} \cup \overline{P}^{(t)}). \tag{3.7}$$

Note that each non-dominated point has a raw fitness of zero. A k-nearest neighbor method (see e.g. Silverman (1986)) is additionally used. Roughly, the inverse of the distance to the kth nearest neighbor, denoted as $\mathbf{v}_{\mathbf{a}}^k$ anent to point $\mathbf{a}$, is used as a density measure:

$$den(\mathbf{a}) = (\mathbf{v}_{\mathbf{a}}^k + 2)^{-1}. \tag{3.8}$$

The fitness $raw(\mathbf{a}) + den(\mathbf{a})$ is to be minimized. The best individuals are copied to the archive for the next generation. In case of equal fitness values, a k-nearest neighbor distance with increasing k is used as a further criterion.
The runtime of SPEA2 is $O(\mu^3)$, with $|P^{(t)}| = |\overline{P}^{(t)}| = \mu$, in the worst case due to the eventually elaborate k-nearest neighbor method, and typically $O(\mu^2 \log \mu)$ per generation.

ϵ-MOEA A more modern and very popular EMOA is the ϵ-MOEA, based one the at that time innovative concept of ϵ-dominance by Laumanns et al. (2002), which means that the dominance relation among points is coarsened to the dominance among boxes containing these points. The algorithm has been proposed by Laumanns et al. (2002) and further developed by Deb et al. (2005a). We consider

the latter version of the ϵ-MOEA in the benchmark on many-objective problems in Section 3.4.
The ϵ-MOEA manages a population $P^{(t)}$ of fixed size and an archive of non-dominated solutions $E^{(t)}$. From the population, a parent is determined via binary tournament selection w. r. t. the dominance relation, i.e., the non-dominated individual wins the comparison or in case of mutually non-dominance, an individual is selected at random. From this parent and another one chosen uniformly at random from the archive, one offspring is created by variation, typically using SBX and PM. Then a steady-state selection is performed on the population. A dominated offspring is rejected, whereas a dominating offspring replaces one of the concerned population members chosen randomly. If the offspring and the population members are mutually non-dominated, the offspring replaces a population member chosen uniformly at random. If the offspring has been included in the population it is checked whether it is also admitted to the archive using the concept of ϵ-dominance as follows. The objective space is divided into a grid of hyperboxes, whose size depend on the parameter vector $\epsilon = (\epsilon_1, \ldots, \epsilon_d)$. Each archive member is attributed with a vector $\mathbf{eb}$ giving its position in the objective space in terms of indexes of ϵ-boxes as:

$$\mathbf{eb}(\mathbf{a})_i = \lfloor (a_i - f_i^{\min})/\epsilon_i \rfloor, \quad i \in \{1, \ldots, d\} \tag{3.9}$$

with $\mathbf{a}$ denoting its objective vector, and $f_i^{\min}$ the minimum possible objective value of objective f_i. Based on these $\mathbf{eb}$ vectors, the dominance is checked, i.e., if the offspring is dominated, it is rejected; if it is non-dominated, it is added to the archive. If it has the same $\mathbf{eb}$ vector as an archive member, i.e., is contained in the same ϵ-box, the decision is performed according to the dominance as usually. If both points are non-dominated, then the one with the closer Euclidean distance to the $\mathbf{eb}$ vector, i.e., the lower left corner of the box, replaces the other, so that at most one point is contained in each ϵ-box. Thereby the archive size is upper bounded by the number of non-dominated ϵ-boxes and enforces a good distribution of the points. The archive (not the population) is the actual output of the algorithm.
The runtime per generation is with $O(|E|)$ extremely low. The parameterization is a bit difficult since the ϵ vector determines the size of the archive, i.e., the output set.

Besides the algorithms above, several others exist and new EMOA are steadily developed. Outstanding from the mass are MSOPS and MSOPS-II by Hughes (2005, 2007) due to their successful aggregation concept. MSOPS is studied and described in detail in Section 3.4, including an advancement. GDE3 by Kukkonen and Lampinen (2007) showed very good performance in competitions, where the success is attributed to its variation by differential evolution (DE), which is studied in Section 3.5.

3.2 Convergence Rates

Convergence rates express how fast an algorithm approaches a certain state, e.g. having found an optimal solution. Therefore, these are strong descriptions of an algorithm's performance, but hard to gain for evolutionary algorithms. Convergence rates have mainly been calculated for simple EA, but rarely for state-of-the-art EA like presented in the following for the SMS-EMOA on bi-objective continuous problems and briefly for other EMOA, based on the publications Beume et al. (2010)* and Beume et al. (2011)*.

The key idea of the proofs is to show equivalence between a multiobjective algorithm and a single-objective EA. In EMOA, incomparable points w. r. t. the dominance relation are typically compared regarding a single-objective substitute function, like the hypervolume in the SMS-EMOA. Selecting the points with best values regarding that substitute function in the EMOA results in the same decisions as the selection of a single-objective EA, which uses this substitute function (or a similar function resulting in the same ranking of points) as its fitness function. Thus, the selection of incomparable points w. r. t. the dominance relation is equal in both algorithms. We show that the EA and the EMOA are *algorithmically equivalent* by proving identical selection and assuming all other operators to be equal. This is a new method to relate single- and multi-objective EA. A remarkable aspect is that in our proofs the algorithmic equivalence holds for *any* bi-objective problem without demands on the objective functions as we are considering the selection in the objective space only. Thus, any future result regarding the single-objective EA on a function class including the substitute function holds for the SMS-EMOA as well.

Having proved the algorithmic equivalence, we transfer known results for the single-objective EA on function classes including the substitute function to the SMS-EMOA, and thereby gain first convergence results for a state-of-the-art EMOA. These results apply to certain specifications of the algorithms, i.e., certain mutation operators. In cases where no convergence result for the single-objective case is known, the method is still helpful as it suffices to analyze the more amendable single-objective EA instead of the EMOA. The contributions of this work are not only the proved results but also the proof technique of relating single-objective EA to multi-objective ones.

Related work and basic definitions are given in the following subsection. The remainder of the section is structured regarding the selection scheme of the SMS-EMOA, including different concepts for the choice of the reference point.

In Section 3.2.2, the SMS-EMOA is considered with its original adaptive reference point. The transformation is easy when the selection is performed on a set of two points, so we can handle the cases of (1+1) and (1,2) selection, but no larger sets. For the (1+1)-SMS-EMOA and the (1,2)-SMS-EMOA a linear convergence rate is proved. It is thereby the only EMOA without an explicit weighting of the objectives

reaching this speed. Due to the limitation of the approach, Section 3.2.3 discusses a new selection concept using pairwise comparisons of points, to base again on the easy case of selecting among two points. Section 3.2.4 analyzes the SMS-EMOA using a reference point that is fixed throughout the optimization process. Here, a convergence rate is proved that is new for the single-objective EA as well. Finally, other popular EMOA (NSGA-II, SPEA2, IBEA) are related to the analyzed cases of the SMS-EMOA to show similarities and differences (Sec. 3.2.5). This section is closed with a summary and directions of future work (Sec. 3.2.6).

3.2.1 Preliminaries

Related Work

The following part is mainly transferred literally from Beume et al. (2010)*.
Convergence properties of EMOA are yet not well understood. More recently, theory concentrated on the convergence or runtime of simple EMOA on special discrete problems, considering whether and how quickly the Pareto set is reached. For the case of a continuous search space $\mathbb{R}^n$ only a few results exist for specialized algorithms, the first obtained by Rudolph (1998). He showed that a multi-objective (1+1)-EA that accepts incomparable points with probability $\frac{1}{2}$ converges with probability 1 to the Pareto set if the step size is chosen proportional to the distance to the Pareto set, while two other step size concepts fail. Hanne (1999) considered stochastic convergence of EMOA with different selection schemes, the possibilities of temporary fitness deterioration, and on problems with unattainable solutions. A recent subject of interest has been whether a certain distribution on the Pareto front can be obtained that is optimal regarding specified preferences.
Despite these advances, the convergence rate in continuous space remains a neglected topic. Teytaud (2007) shows that the convergence rate scales badly with increasing number of objectives entailing that any comparison-based EMOA performs hardly better than random search for a large number of objectives. Also a general lower bound for the convergence time is given.

Convergence

The following part is mainly transferred literally from Beume et al. (2011)*.

Definition 3.2 *Let X be a random variable and (X_t) a sequence of random variables defined on a probability space $(\Omega, \mathcal{A}, P)$. Then (X_t) is said to*
(a) converge completely to X, if for any $\epsilon > 0$

$$\lim_{t\to\infty} \sum_{i=1}^{t} Pr(|X_i - X| > \epsilon) < \infty;$$

(b) converge almost surely or with probability 1 to X, if

$$Pr(\lim_{t\to\infty} |X_t - X| = 0) = 1;$$

(c) converge in probability to X, if for any $\epsilon > 0$

$$\lim_{t\to\infty} Pr(|X_t - X| > \epsilon) = 0;$$

(d) converge in mean to X, if

$$\lim_{t\to\infty} E(|X_t - X|) = 0. \qquad \square$$

The velocity of approaching a limit is expressed by the *convergence rate*.

Definition 3.3 *Let $(Z_k : k \geq 0)$ be a non-negative random sequence. The sequence is said to converge geometrically fast in mean (in probability, w.p. 1) to zero if there exists a constant $q > 1$ such that the sequence $(q^k Z_k : k \geq 0)$ converges in mean (in probability, w.p. 1) to zero. Let $q^* > 1$ be supremum of all constants $q > 1$ such that geometrically fast convergence is still guaranteed. Then $c = 1/q$ is called the* convergence rate. *A sequence with geometrically fast convergence is synonymously denoted to have a* linear convergence rate. $\square$

Let $\rho(\cdot)$ denote a function that measures the performance of an EA's population X_k and ρ^* the target value. If the sequence $(Z_k)_{k\geq 0}$ defined by $Z_k = |\rho(X_k) - \rho^*|$ converges (in any mode mentioned above) to zero with a certain convergence rate, then the EA approaches the target performance value with this rate.
For example, let $\rho(X_k)$ be the best objective function value of the population at generation $k \geq 0$ of a single-criterion EA and ρ^* be the global minimum of the objective function. If Z_k converges to zero then the EA converges to the global minimum. Similarly, let $\rho(X_k)$ be the dominated hypervolume of population X_k and ρ^* the maximal dominated hypervolume in the multi-objective scenario then the population converges to the maximum dominated hypervolume if $Z_k \to 0$ as $k \to \infty$. For the convergence analysis in the Sections 3.2.2 and 3.2.3 we consider the convergence of the population towards the Pareto front. Thereby, $\rho(X_k)$ denotes the distance to a certain point on the Pareto front.
Note that we analyze algorithms w. r. t. their black-box-complexity (cf. Section 1.4), i.e., we consider the number of function evaluations and express convergence rates in this measure.

Convexity

Definition 3.4 *A set $S \subseteq \mathbb{R}^n$ is said to be* convex *if $\xi\,\mathbf{x} + (1-\xi)\,\mathbf{y} \in S$ for all $\mathbf{x}, \mathbf{y} \in S$ and $\xi \in [0,1]$. A function $f : S \to \mathbb{R}$ is termed*
(a) convex *if* $\quad f(\xi\,\mathbf{x} + (1-\xi)\,\mathbf{y}) \leq \xi\, f(\mathbf{x}) + (1-\xi)\, f(\mathbf{y})$,
(b) strictly convex *if* $\quad f(\xi\,\mathbf{x} + (1-\xi)\,\mathbf{y}) < \xi\, f(\mathbf{x}) + (1-\xi)\, f(\mathbf{y})$,

(c) strongly convex *if*

$$f(\xi\,\mathbf{x} + (1-\xi)\,\mathbf{y}) \leq \xi\, f(\mathbf{x}) + (1-\xi)\, f(\mathbf{y}) + \frac{L}{2}\,\xi\,(1-\xi)\,\|\mathbf{x}-\mathbf{y}\|^2,$$

(d) (K,Q)-strongly convex *if it is strongly convex and*

$$\frac{K}{2}\,\xi\,(1-\xi)\,\|\mathbf{x}-\mathbf{y}\|^2 \leq \xi\, f(\mathbf{x}) + (1-\xi)\, f(\mathbf{y}) - f(\xi\,\mathbf{x} + (1-\xi)\,\mathbf{y}) \leq \frac{L}{2}\,\xi\,(1-\xi)\,\|\mathbf{x}-\mathbf{y}\|^2 \quad (3.10)$$

with $K, L \in \mathbb{R}^+, 0 < K \leq L < \infty$ *and* $Q = L/K$. □

Definition 3.4.(a) says that a function is convex if its epigraph (the set of points lying on or above its graph) is convex. For convex functions any local minimum is a global one, i.e., minima have equal function values. Strictly convex functions are convex with a unique minimizer. Strongly convex functions are a subclass fulfilling a tighter bound of the inequality, whereas for (K,Q)-strongly convex functions the relation of the terms is bounded from two sides. The inequalities become more precise with increasing values of the parameters K and L.
Finally, notice that $f(\cdot)$ is termed *concave* if $-f(\cdot)$ is convex.

Definition 3.5 *A symmetric quadratic matrix A with eigenvalues $\{\{\nu_1, \ldots, \nu_n\}\}$ is*

1. *positive semidefinite iff* $\nu_i \geq 0, \forall i \in \{1, \ldots, n\}$
2. *positive definite iff* $\nu_i > 0, \forall i \in \{1, \ldots, n\}$
3. *negative semidefinite iff* $\nu_i \leq 0, \forall i \in \{1, \ldots, n\}$
4. *negative definite iff* $\nu_i < 0, \forall i \in \{1, \ldots, n\}$
5. *indefinite iff* $\exists \nu_i < 0$ *and* $\exists \nu_j > 0$, *with* $i, j \in \{1, \ldots, n\}$.

Theorem 3.6 *(cf. e.g. Hiriart-Urruty and Lemaréchal (2001, Th. 4.3.1)) Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and Q denote its Hessian matrix. Then it holds:*

1. *f is convex* $\iff$ *Q is positive semidefinite*
2. *f is strictly convex* $\impliedby$ *Q is positive definite*
3. *f is concave* $\iff$ *Q is negative semidefinite*
4. *f is strictly concave* $\impliedby$ *Q is negative definite.* □

Lemma 3.7 *Let $f_1(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + a_0$ and $f_2(\mathbf{x}) = \mathbf{b}^\top \mathbf{x} + b_0$ be linear functions with $\mathbf{a}, \mathbf{b}, \mathbf{x} \in \mathbb{R}^n$, with $a_0, b_0 \in \mathbb{R}$, and $n \geq 2$. The dominated hypervolume for a fixed reference point $\mathbf{r} \in \mathbb{R}^2$ is a concave function if the matrix $\mathbf{a}\,\mathbf{b}^\top$ is negative semidefinite. (Beume et al. (2011)*).*

Proof. Notice that the dominated hypervolume

$$\begin{aligned} H(\{f(\mathbf{x})\}, \mathbf{r}) &= [r_1 - f_1(\mathbf{x})]\,[r_2 - f_2(\mathbf{x})] \\ &= [r_1 - (\mathbf{a}^\top \mathbf{x} + a_0)]\,[r_2 - (\mathbf{b}^\top \mathbf{x} + b_0)] \\ &= [(r_1 - a_0) - \mathbf{a}^\top \mathbf{x}]\,[(r_2 - b_0) - \mathbf{b}^\top \mathbf{x}] \\ &= (r_1 - a_0)\,(r_2 - b_0) - [(r_1 - a_0)\,\mathbf{b} + (r_2 - b_0)\,\mathbf{a}]^\top \mathbf{x} + \mathbf{a}^\top \mathbf{x} \cdot \mathbf{b}^\top \mathbf{x} \\ &= (r_1 - a_0)\,(r_2 - b_0) - [(r_1 - a_0)\,\mathbf{b} + (r_2 - b_0)\,\mathbf{a}]^\top \mathbf{x} + \mathbf{x}^\top (\mathbf{a}\,\mathbf{b}^\top) \mathbf{x} \end{aligned}$$

is a quadratic form which is concave iff $\mathbf{a}\,\mathbf{b}^\top$ is negative semidefinite. □

Example: $\mathbf{a}\mathbf{b}'$ is negative semidefinite if $\mathbf{b} = -\mathbf{a}$ since $\mathbf{a}\mathbf{a}^\top$ is positive semidefinite. Lemma 3.7 is used in Section 3.2.4.

Definition 3.8 *A symmetric positive semidefinite matrix A has bounded bandwidth $\kappa \geq 1$ iff for all eigenvalues $\nu_1 \leq \nu_2 \leq \ldots \leq \nu_n < \infty$ holds: $\exists \kappa \geq 1 : \forall i = 1, \ldots, n : \; \nu_i \in [\nu_1, \kappa \cdot \nu_1]$, with κ being a constant.*
A function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A \mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$ is denoted to have bounded bandwidth iff its matrix A has bounded bandwidth.

Note that the bounded bandwidth cannot be fulfilled if one but not all eigenvalues are zero. So it can only hold for the zero matrix or positive definite matrices. Consider a small counter example where the eigenvalues cannot be bounded by a constant. In the diagonal matrix A the diagonal entries correspond to the eigenvalues which all are positive, so the matrix is positive definite. Then, there is no constant κ such that the interval $[1, \kappa \cdot 1]$ contains the largest eigenvalue n:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & n \end{pmatrix}, \quad \{\{1, 1, n\}\} \in [1, \kappa \cdot 1] \quad \Rightarrow \; \lightning \kappa \text{ is a constant.}$$

Lemma 3.9 *Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A \mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$ be a quadratic convex function with bounded bandwidth, and A is not the zero matrix. Then, its matrix A is positive definite and has bounded bandwidth.*

Proof. As f is convex, it follows that A is positive semidefinite, so $\forall i = 1, \ldots, n : \nu_i \geq 0$. The condition $\exists \kappa \geq 1 : \forall i = 1, \ldots, n : \nu_i \in [\nu_1, \kappa \cdot \nu_1]$ of the bounded bandwidth is only fulfilled if A is the zero matrix (which we exclude from consideration) or if $\forall i = 1, \ldots, n : \nu_i > 0$ holds, and thus A is even positive definite, which implies that f is actually strictly convex. □

Algorithmical Setup

We consider the SMS-EMOA not only with the standard $(\mu+1)$-selection but also other schemes of the $(\mu \dot{+} \lambda)$ framework. A new selection scheme with a tournament selection based on pairwise comparisons is explained in detail in Section 3.2.3. Several mutation operators are considered in order to calculate convergence rates, or respectively transfer known results. We do not provide convergence rates for recombination.
The reference point is mainly chosen following the concept of the adaptive reference point. Recall that points that are the worst ones in the population regarding an objective function have a distance to the adaptive reference point of exactly 1 w. r. t. that worst objective (cf. Fig. 3.2). This fact is exploited to gain the convergence result described in Section 3.2.2.1.
From Lemma 2.5 we know that in case of rating two points by their absolute hypervolume, the resulting order is the same as if the points were rated by their hypervolume contribution. We refer to this lemma to argue that both measures behave equally in case of selecting among two points.

Algorithmic Equivalence

The following part is mainly transferred literally from Beume et al. (2011)*.
We exploit that convergence results for one EA [...] can be transferred to another in case the algorithms behave similar. Therefore, we introduce the following definition of *algorithmic equivalence*:

Definition 3.10 *Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be two stochastic sequences of states generated by two evolutionary algorithms A_X and A_Y. The EA A_X and A_Y are called* algorithmically equivalent *if $\forall t \geq 0\colon X_t \overset{d}{=} Y_t$ holds for their associated state sequences, i.e., X_t and Y_t have the same distribution for all $t \geq 0$.* □

In particular, two algorithms are algorithmically equivalent if both use the same probability distribution for probabilistic decisions, and their deterministic decisions are equal. We use the term in this way by assuming equal probabilistic operators (initialization, variation) and showing that the deterministic selection operators produce the same output in case of the same input for certain classes of problems. We do not consider the computational resources of the EA operators but only the input-output-behavior, thus the state of the EA's population.

3.2.2 SMS-EMOA with Adaptive Reference Point

3.2.2.1 (1+1)-SMS-EMOA

We start our analysis with simple versions of algorithms and show the algorithmic equivalence of the (1+1)-SMS-EMOA (Alg. 3.4) and the (1+1)-EA (Alg. 3.3). The

Algorithm 3.3: (1+1)-EA for single-objective minimization

1 choose individual $X^{(0)} \in \mathbb{R}^n$, set $t = 0$
2 **repeat**
3 | create offspring $Y^{(t)}$ from $X^{(t)}$ by variation
4 | evaluate offspring $Y^{(t)}$
5 | **if** $f(Y^{(t)}) \leq f(X^{(t)})$ **then** /* selection */
6 | | $X^{(t+1)} = Y^{(t)}$
7 | **else** $X^{(t+1)} = X^{(t)}$
8 | increment t
9 **until** termination criterion fulfilled

Algorithm 3.4: (1+1)-SMS-EMOA for 2-objective minimization

1 choose individual $X^{(0)} \in \mathbb{R}^n$, set $t = 0$
2 **repeat**
3 | create offspring $Y^{(t)}$ from $X^{(t)}$ by variation
4 | evaluate offspring $Y^{(t)}$
5 | **if** $f(Y^{(t)}) \preceq f(X^{(t)})$ **or** /* selection */
| | $\big(f(Y^{(t)}) \parallel f(X^{(t)})$ **and** $H(\{f(Y^{(t)})\}, \mathbf{r}^{(t)}) \geq H(\{f(X^{(t)})\}, \mathbf{r}^{(t)})\big)$ with
| | $\mathbf{r}^{(t)} = \big(\max\{f_1(X^{(t)}), f_1(Y^{(t)})\} + 1,\ \max\{f_2(X^{(t)}), f_2(Y^{(t)})\} + 1\big)$
6 | **then**
7 | | $X^{(t+1)} = Y^{(t)}$
8 | **else** $X^{(t+1)} = X^{(t)}$
9 | increment t
10 **until** termination criterion fulfilled

algorithms actually differ only in line 5. Both algorithms initialize one individual, create one offspring by variation from the parent individual and select the better individual for the proceeding generation. For the algorithmic equivalence, all operators apart from the selection can be chosen arbitrarily, whereas we presume that they are chosen to be identical for both algorithms.

The single-objective selection operator of the (1+1)-EA compares the function values of the two individuals and keeps the offspring if its value is not worse, i.e., not greater than the parent's in case of minimization. The (1+1)-SMS-EMOA checks whether the individuals' function vectors are comparable regarding the Pareto-dominance relation, and keeps the offspring if it weakly dominates the parent. In case of incomparable individuals, the dominated hypervolume is invoked as a selection criterion. First, the adaptive reference point is calculated depending on both points. Then, the offspring is kept if its hypervolume is not worse. According

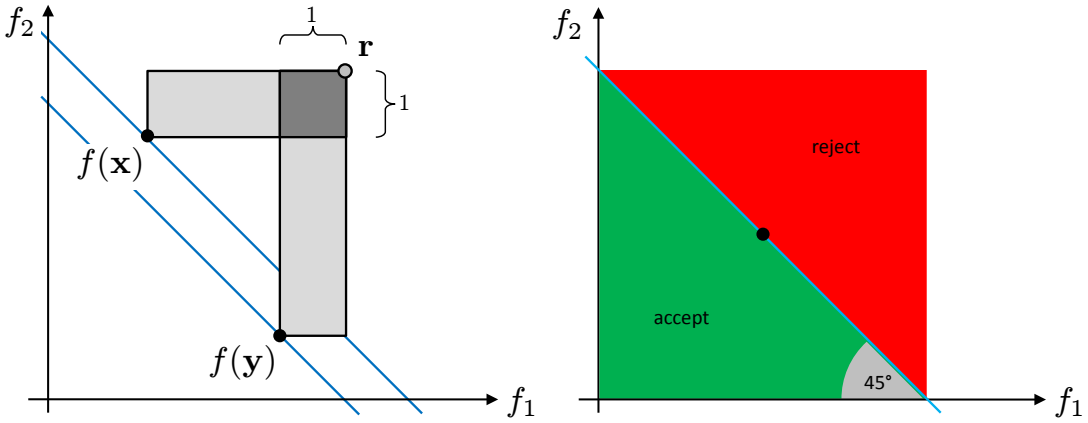


Fig. 3.2: Left: Hypervolume of the points $f(\mathbf{x})$ and $f(\mathbf{y})$ w. r. t. the adaptive reference point $\mathbf{r}$, with the hypervolume contributions shaded in light gray. Right: Regions of acceptance or rejection for the sum of objectives. The 45° lines are isolines of f^s.

to Lemma 2.5, this decision always corresponds to choosing the point with higher hypervolume contribution to the set of two points.
In the following, we show that the single- and the multi-objective selection operators perform identical decisions in case of certain objective functions. No preconditions for the bi-objective problem are required. The single-objective problem shall be the (evenly weighted or unweighted) sum f^s of the functions of the bi-objective problem as defined below.

Definition 3.11 *For a bi-objective minimization problem* $f : \mathbb{R}^n \to \mathbb{R}^2$, $\min\{f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x})) \mid \mathbf{x} \in \mathbb{R}^n\}$, *the corresponding single-objective minimization problem* $f^s : \mathbb{R}^n \to \mathbb{R}$ *with* $f^s(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ *is named* unweighted or evenly weighted sum. □

Figure 3.2 illustrates the relation of the hypervolume and the f^s. Our considerations lead to the following theorem.

Theorem 3.12 *The (1+1)-SMS-EMOA with adaptive reference point applied to a bi-objective minimization problem* $f : \mathbb{R}^n \to \mathbb{R}^2, \min\{f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x})) \mid \mathbf{x} \in \mathbb{R}^n\}$ *is algorithmically equivalent to a (1+1)-EA applied to the minimization of the single-objective function* $f^s : \mathbb{R}^n \to \mathbb{R}$ *with* $f^s(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$, *presumed all operators apart from the selection are identical in both algorithms. (Beume et al. (2010)*).*

Proof. Since all operators apart from the selection are assumed to be identical for both algorithms, only the algorithmic equivalence of the selection operators remains to be shown. To this end, it suffices to show that for the same pair of parent and offspring, the algorithms select the same individual for survival. This

equivalent behavior is proved by two implications: (1) Whenever the (1+1)-SMS-EMOA accepts the offspring, then the (1+1)-EA accepts it as well. (2) Whenever the (1+1)-EA accepts the offspring, the (1+1)-SMS-EMOA does so—or: whenever the (1+1)-SMS-EMOA rejects the offspring, the (1+1)-EA does so. The proof of the second implication is omitted since showing the latter formulation works analogously to the proof of the first implication which is presented in detail in the following.

For the acceptance of the offspring in the (1+1)-SMS-EMOA, we distinguish the cases of comparable (1.a) and incomparable (1.b) individuals.

(1.a) A comparable offspring $\mathbf{y}$ is accepted if it weakly dominates its parent $\mathbf{x}$, i.e., $f(\mathbf{y}) \preceq f(\mathbf{x})$, which is shown to imply the acceptance by the (1+1)-EA, i.e., $f^s(\mathbf{y}) \leq f^s(\mathbf{x})$.

$$\begin{aligned} f(\mathbf{y}) \preceq f(\mathbf{x}) &\Leftrightarrow f_1(\mathbf{y}) \leq f_1(\mathbf{x}) \ \wedge \ f_2(\mathbf{y}) \leq f_2(\mathbf{x}) \\ &\Rightarrow f_1(\mathbf{y}) + f_2(\mathbf{y}) \leq f_1(\mathbf{x}) + f_2(\mathbf{x}) \\ &\Leftrightarrow f^s(\mathbf{y}) \leq f^s(\mathbf{x}) \end{aligned} \tag{3.11}$$

(1.b) Let the offspring and the parent be incomparable, i.e., $f(\mathbf{y}) \parallel f(\mathbf{x})$. Then the offspring is accepted if its hypervolume value is greater than the parent's, which is shown to imply the acceptance by the (1+1)-EA as well. Since $f(\mathbf{y}) \parallel f(\mathbf{x})$ we have to distinguish two cases: $f_1(\mathbf{x}) > f_1(\mathbf{y}) \wedge f_2(\mathbf{x}) < f_2(\mathbf{y})$ or $f_1(\mathbf{x}) < f_1(\mathbf{y}) \wedge f_2(\mathbf{x}) > f_2(\mathbf{y})$. We consider the first case; the proof for the second one works analogously.

Let $f_1(\mathbf{x}) > f_1(\mathbf{y}) \wedge f_2(\mathbf{x}) < f_2(\mathbf{y})$. According to Algorithm 3.4, the adaptive reference point is $\mathbf{r} = (f_1(\mathbf{x}) + 1, f_2(\mathbf{y}) + 1)^\top$. Recall that the hypervolume of a single point $\mathbf{v} \in \mathbb{R}^2$ with $\mathbf{v} \preceq \mathbf{r}$ (which holds for the adaptive reference point by definition) is $H(\{\mathbf{v}\}, \mathbf{r}) = (r_1 - v_1)(r_2 - v_2)$ according to Eq. (2.5)). Inserting the reference point yields the following equations, whereas one factor reduces to 1 for each point due to concept of the adaptive reference point.

$$\begin{aligned} H(\{f(\mathbf{x})\}, \mathbf{r}) &= \Big(f_1(\mathbf{x}) + 1 - f_1(\mathbf{x})\Big)\Big(f_2(\mathbf{y}) + 1 - f_2(\mathbf{x})\Big) \\ &= f_2(\mathbf{y}) - f_2(\mathbf{x}) + 1 \\ H(\{f(\mathbf{y})\}, \mathbf{r}) &= \Big(f_1(\mathbf{x}) + 1 - f_1(\mathbf{y})\Big)\Big(f_2(\mathbf{y}) + 1 - f_2(\mathbf{y})\Big) \\ &= f_1(\mathbf{x}) - f_1(\mathbf{y}) + 1 \end{aligned}$$

Due to Lemma 2.5 is does not matter whether we consider absolute or contributive hypervolume values. In fact, the hypervolume contributions match the equations above without the term +1. The offspring is accepted if

$$\begin{aligned} H(\{f(\mathbf{y})\}, \mathbf{r}) \geq H(\{f(\mathbf{x})\}, \mathbf{r}) &\Rightarrow f_1(\mathbf{x}) - f_1(\mathbf{y}) + 1 \geq f_2(\mathbf{y}) - f_2(\mathbf{x}) + 1 \\ &\Leftrightarrow f_1(\mathbf{x}) - f_1(\mathbf{y}) \geq f_2(\mathbf{y}) - f_2(\mathbf{x}) \\ &\Leftrightarrow f_1(\mathbf{x}) + f_2(\mathbf{x}) \geq f_1(\mathbf{y}) + f_2(\mathbf{y}) \\ &\Leftrightarrow f^s(\mathbf{x}) \geq f^s(\mathbf{y}), \end{aligned}$$

and the acceptance by the (1+1)-SMS-EMOA implies the acceptance by the (1+1)-EA as desired. The analogous proof of the other case of incomparable points completes implication (1) that whenever the (1+1)-SMS-EMOA accepts the offspring so does the (1+1)-EA. The analogous proof of implication (2) in the other direction closes the proof of the algorithmic equivalence of the algorithms on the specified problems. □

The equivalence of the (1+1)-SMS-EMOA and the (1+1)-EA as formulated in Th. 3.12 holds for all bi-objective problems and any specification of the algorithms. For a certain class of problems and a certain variation, we show a linear convergence rate of the algorithms to an optimal point. This can now be done easily since this result for the (1+1)-EA already exists, which we briefly recall.

Algorithm 3.5: EA using mutation with self-adaptation

choose $X^{(0)} \in \mathbb{R}^n$, set $t = 0$ /* initialization */
choose $\sigma^{(0)} > 0$, $\delta > 0$, $\gamma > 1$, set $p_s = 1/5, k = 0$
repeat
  draw $Z^{(t)}$ from a multivariate standard normal distribution
  $Y^{(t)} = X^{(t)} + \sigma^{(t)} Z^{(t)}$ /* mutation */
  do selection
  if $Y^{(t)}$ survived **then** $k{+}{+}$
  if $t \mod \delta \neq 0$ **then** $\sigma^{(t+1)} = \sigma^{(t)}$
  else /* step size adaptation */
    $q_s = k/\delta, \quad k = 0$
    if $q_s \geq p_s$ **then** $\sigma^{(t+1)} = \sigma \cdot \gamma$ **else** $\sigma^{(t+1)} = \sigma/\gamma$
  $t{+}{+}$
until termination criterion fulfilled

Let the (1+1)-EA from Alg. 3.3 be further specified by Alg. 3.5. It uses mutations by addition of a vector drawn from a multivariate standard distribution scaled by step size σ. The step size is controlled via self-adaptation by the so-called 1/5-rule, with success rate $p_s = 1/5$. Is the ratio of offspring that replaced the parent lower than 1/5 within the observation interval δ, the step size is reduced by division through the adaptation factor γ, and otherwise increased by γ.
For the (1+1)-EA with self-adaptation with the 1/5-rule Jägersküpper (2006, Th.4) proved the following convergence rate:

Theorem 3.13 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top\mathbf{x} + c$ with positive definite matrix A with bounded bandwidth. The (1+1)-EA (Alg. 3.3) with self-adaptation as in Alg. 3.5 using $\delta = \Theta(n)$ and $p_s = \frac{1}{5}$ halves the distance to the optimum in expectation in $O(n)$ iterations with overwhelming probability, provided that $\sigma^{(0)} = \Theta(D/n)$ where D is the initial distance to the minimizer. (Jägersküpper (2006, Th. 4)).*

The theorem actually says that the approximation error decreases exponentially fast and so the (1+1)-EA has a linear convergence rate toward the optimum under the specified conditions. We transfer this result to the (1+1)-SMS-EMOA. Compared to our result in Beume et al. (2010)*, we changed the conditions of the objective functions to have bounded bandwidth.

Corollary 3.14 *The (1+1)-SMS-EMOA as in Alg. 3.4 with self-adaptation as in Alg. 3.5 applied to a bi-objective minimization problem $f : \mathbb{R}^n \to \mathbb{R}^2$ approaches an element of the Pareto front with linear order of convergence if both objective functions are quadratically convex of bounded bandwidth. (cf. Beume et al. (2010)*).*

Proof. We show that under the assumed properties of the objective functions, their sum $f_1(\mathbf{x}) + f_2(\mathbf{x})$ fulfills the conditions required to invoke Th. 3.13 for the (1+1)-EA. Then, thanks to the algorithmic equivalence of the (1+1)-EA and the (1+1)-SMS-EMOA (Th. 3.12), the convergence rate holds for the (1+1)-SMS-EMOA as well.
The quadratic convex functions f_1, f_2 with bounded bandwidth have matrices A, B with bounded bandwidth due to Lemma 3.9. It remains to show that the sum of the functions is quadratically convex with $A + B$ of bounded bandwidth.
Let $\nu(A) = \{\{a_1, \ldots, a_n\}\}$ denote the eigenvalues of A labeled in ascending order, and $\nu(B) = \{\{b_1, \ldots, b_n\}\}$ be defined accordingly. The so-called Rayleigh quotient is enclosed by the minimal and maximal eigenvalues as $\nu_{\min} \leq \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} \leq \nu_{\max}$. Therefore,

$$a_1 + b_1 \leq \frac{\mathbf{x}^\top(A+B)\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} = \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} + \frac{\mathbf{x}^\top B\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} \leq a_n + b_n. \tag{3.12}$$

From the bounded bandwidth of A and B follow $a_n \leq \kappa_a \cdot a_1$ and $b_n \leq \kappa_b \cdot b_1$. Thus, $a_n + b_n \leq \kappa_a \cdot a_1 + \kappa_b \cdot b_1 \leq \max\{\kappa_a + \kappa_b\}(a_1 + b_1)$. As $A + B$ is supposed to be a non zero matrix and has bounded bandwidth, we follow that $A + B$ is positive definite. Thus, the sum $f_1 + f_2$ fulfills the condition to invoke Theorem 3.13 that guarantees linear convergence rate of the (1+1)-EA with self-adaptation for $f^s(\mathbf{x})$. An optimum of the sum of the objective functions corresponds to a point on the Pareto front, so the (1+1)-SMS-EMOA approaches the Pareto front with linear convergence rate.

□

3.2.2.2 (1,2)-SMS-EMOA

The algorithmic equivalence of a (1,2)-SMS-EMOA to a (1,2)-EA can be shown analogously to the case of the (1+1) selection. In the (1,2) selection, the algorithms generate two offspring from one parent and decide which offspring replaces the parent. Thus, the selection performs a decision among two individuals just like in the (1+1) case. For the algorithmic equivalence of the (1+1) versions we did

not invoke any information how the individuals have been generated, so the proof holds for selection among any two individuals, thus also for the (1,2) selection. We deduce the following results, retrieved from Beume et al. (2011)*.

Theorem 3.15 *The (1,2)-SMS-EMOA with adaptive reference point applied to a bi-objective minimization problem* $f : \mathbb{R}^n \to \mathbb{R}^2, \min\{f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x})) \mid \mathbf{x} \in \mathbb{R}^n\}$ *is algorithmically equivalent to a (1,2)-EA applied to the minimization of the single-objective function* $f^s : \mathbb{R}^n \to \mathbb{R}$ *with* $f^s(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$*, presumed all operators apart from the selection are identical in both algorithms. (Beume et al. (2011)*).* □

For the (1,2)-EA, a convergence result by Rudolph (1997) exists, which we can now transfer directly to the (1,2)-SMS-EMOA.

Theorem 3.16 *Let the (1,2)-SMS-EMOA with adaptive reference point use variation by mutations due to a multivariate uniform distribution on the surface of a hypersphere with radius* $r > 0$ *and self-adaptation of the step-size (i.e., the radius)* r *via the 1/5-success rule (as in Alg. 3.5) with* $\delta > 0$*,* $p_s = \frac{1}{5}$*, and* $\gamma > 1$*. Applied to a bi-objective optimization problem* $f : \mathbb{R}^n \to \mathbb{R}^2, \min\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}$ *the algorithm approaches an element of the Pareto front with at most linear order of convergence if both objective functions are convex with bounded bandwidth. (cf. Beume et al. (2011)*).* □

3.2.2.3 Limitations of the Approach

The algorithmic equivalence of the (1+1)-SMS-EMOA and the (1+1)-EA optimizing the sum of the objective function does not hold for more general scenarios with more than two objectives or populations of individuals.

Higher dimensional Problems

Theorem 3.17 *The (1+1)-SMS-EMOA with adaptive reference point applied to a multi-objective minimization problem* $f : \mathbb{R}^n \to \mathbb{R}^d, \min\{f(\mathbf{x}) | \mathbf{x} \in \mathbb{R}^n\}$ *is not algorithmically equivalent to a (1+1)-EA applied to the minimization of the single-objective function* $f^s : \mathbb{R}^n \to \mathbb{R}$ *with* $f^s(\mathbf{x}) = \sum_{i=1}^{d} f_i(\mathbf{x})$*, presumed all operators apart from the selection are identical in both algorithms, when* $d \geq 3$*. (cf. Beume et al. (2010)*).*

Proof. We show the theorem by a counter example with differing selection behavior. Suppose there are two incomparable points $f(\mathbf{x}), f(\mathbf{y}) \in \mathbb{R}^d$ with $f(\mathbf{x}) = (0, 0, \ldots, 0)$ and $f(\mathbf{y}) = (-1, \ldots, -1, d-1-\epsilon)$ where $\epsilon \in (0, 1) \subset \mathbb{R}$. The adaptive

reference point for the set $A = \{f(\mathbf{x}), f(\mathbf{y})\}$ is $\mathbf{r} = (1, \ldots, 1, d - \epsilon)$. The hypervolume contribution of each point is here simply calculated by the product of the distances to the neighbor in each dimension:

$$H(f(\mathbf{x}), A, \mathbf{r}) = \left(\prod_{i=1}^{d-1}(r_i - f_i(\mathbf{x}))\right) \cdot \big(f_d(\mathbf{y}) - f_d(\mathbf{x})\big) \qquad = d - 1 - \epsilon$$
$$H(f(\mathbf{y}), A, \mathbf{r}) = \left(\prod_{i=1}^{d-1}(f_i(\mathbf{x}) - f_i(\mathbf{y}))\right) \cdot \big(r_d - f_d(\mathbf{y})\big) \qquad = 1.$$

The values of the sum of objectives are

$$f^s(\mathbf{x}) = \sum_{i=1}^{d} f_i(\mathbf{x}) \qquad = 0$$
$$f^s(\mathbf{y}) = \sum_{i=1}^{d} f_i(\mathbf{y}) = (d-1) \cdot (-1) + (d - 1 - \epsilon) \qquad = -\epsilon.$$

For $d \geq 3$, the (1+1)-SMS-EMOA accepts $\mathbf{x}$ due to the higher hypervolume but the (1+1)-EA accepts $\mathbf{y}$ due to its lower function value, so their selections differ. Note that the algorithms behave equivalent for the case of $d = 2$ with $f(\mathbf{x}) = (0, 0)$, $f(\mathbf{y}) = (-1, d - 1 - \epsilon)$, and $\mathbf{r} = (1, d - \epsilon)$: The (1+1)-SMS-EMOA accepts $\mathbf{y}$ as $H(f(\mathbf{x}), A, \mathbf{r}) = 1 - \epsilon < H(f(\mathbf{y}), A, \mathbf{r}) = 1$.
For $d \geq 3$ the selection according to the hypervolume contribution is not equivalent to the selection according to the absolute hypervolume. However, the latter approach is no back door here as an analogous counter example indicates the inequivalence to the selection of the (1+1)-EA (cf. Beume et al. (2010)*). Compared to the example above, only the sign of ϵ is changed in $f(\mathbf{y})$, leading to $f(\mathbf{y}) = (-1, \ldots, -1, d - 1 + \epsilon)$, with $\epsilon \in (0, 1) \subset \mathbb{R}$, $\mathbf{r} = (1, \ldots, 1, d + \epsilon)$, and $f^s(\mathbf{y}) = \epsilon$. The absolute dominated hypervolume of each point is

$$H(\{f(\mathbf{x})\}, \mathbf{r}) = \prod_{i=1}^{d}(r_i - f_i(\mathbf{x})) = (1 - 0)^{d-1} \cdot (d + \epsilon - 0) \qquad = d + \epsilon$$
$$H(\{f(\mathbf{y})\}, \mathbf{r}) = \prod_{i=1}^{d}(r_i - f_i(\mathbf{y})) = (1 + 1)^{d-1} \cdot (d + \epsilon - d + 1 - \epsilon) \qquad = 2^{d-1}.$$

For $d \geq 3$, the (1+1)-SMS-EMOA accepts $\mathbf{y}$ due to the higher hypervolume but the (1+1)-EA accepts $\mathbf{x}$ due to $f^s(\mathbf{x}) = 0 < f^s(\mathbf{y}) = \epsilon$, so the selections differ. Again, the algorithms behave algorithmically equivalent for $d = 2$: The (1+1)-SMS-EMOA accepts $\mathbf{x}$ as $H(\{f(\mathbf{x})\}, \mathbf{r}) = 2 + \epsilon > H(\{f(\mathbf{y})\}, \mathbf{r}) = 2$.

□

Population-based SMS-EMOA

The following part is mainly transferred literally from Beume et al. (2011)*. We show by a counter-example that the order induced by the hypervolume values of the points is not equivalent to the order induced by the unweighted sum f^s.

Theorem 3.18 *The order of individuals induced by the sum of objectives* $f^s(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ *differs from the order according to hypervolume values when the reference point is chosen adaptively depending on more than two points. (Beume et al. (2011)*).*

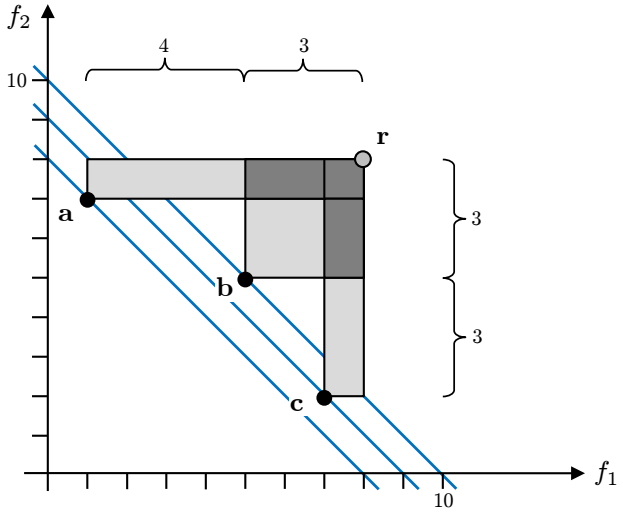


Fig. 3.3: The adaptive reference point $\mathbf{r} = \mathbf{r}(\{\mathbf{a}, \mathbf{b}, \mathbf{c}\})$ is chosen relative to the three points. The blue lines indicate isolines of the sum f^s. The maximal and minimal elements according to the sum do not correspond to the minimal and maximal elements due to the hypervolume (contributions).

Proof. In order to generate corresponding selection behavior of the SMS-EMOA and the EA, the order of hypervolume values has to be inverse to the order of the sum of objectives. We show that contradictions already occur for bi-objective problems with a set of three points regarding the minimal and maximal values. Consider the set $M = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ in Figure 3.3 with $\mathbf{a} = (1, 7), \mathbf{b} = (5, 5), \mathbf{c} = (7, 2), \mathbf{r}(M) = (8, 8)$.

The orders according to the hypervolume contribution, the absolute hypervolume, and the sum of obectives of each point are:

$$\begin{array}{rcrcl} H(\mathbf{a}, M, \mathbf{r}(M)) = 4 & = & H(\mathbf{b}, M, \mathbf{r}(M)) = 4 & > & H(\mathbf{c}, M, \mathbf{r}(M)) = 3 \\ H(\{\mathbf{b}\}, \mathbf{r}(M)) = 9 & > & H(\{\mathbf{a}\}, \mathbf{r}(M)) = 7 & > & H(\{\mathbf{c}\}, \mathbf{r}(M)) = 6 \\ f^s(\mathbf{a}) = 8 & < & f^s(\mathbf{c}) = 9 & < & f^s(\mathbf{b}) = 10. \end{array}$$

Obviously, the orders of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ differ, especially regarding the minima and maxima, so neither the hypervolume contributions nor the absolute hypervolume values correspond to the sum of objectives. □

From Theorem 3.18 it follows that the selection according to the hypervolume of a point in the SMS-EMOA and the selection of the EA according to the sum of objectives are in general not equivalent when the reference point depends on more than two points. This already holds for a set of three individuals, which can occur when the parent or the offspring population consist of more than one individual since the orders already differ in their minimal and maximal elements.

Corollary 3.19 *When the reference point of the hypervolume is chosen adaptively depending on more than two points, then for $\mu > 1$ or $\lambda > 1$ the $(\mu + \lambda)$-SMS-EMOA and the $(\mu+\lambda)$-EA with the sum of objectives as fitness function, presumed all operators apart from the selection are identical in both algorithms, are not algorithmically equivalent; as well as the (μ, λ)-SMS-EMOA and the (μ, λ)-EA with $\lambda > 2$. (Beume et al. (2011)*).* □

3.2.3 Population-based SMS-EMOA using Pairwise Comparisons

Respecting the limitations of our approach of algorithmic equivalence not being valid for the selection among a set of more than two points, we develop a new selection concept based on pairwise comparisons. Note that we thereby deviate from the standard selection concept of the SMS-EMOA. However, the new concept enables us to establish algorithmic equivalence between population-based versions of the SMS-EMOA and single-objective EA.
The selection within the SMS-EMOA is now performed based on pairwise comparisons in a tournament modus (cf. Alg. 3.6). First, the individuals (parents and/or offspring) to select among are compiled in one set. In each step, the set is divided into pairs of individuals. Among the pairs, the individuals are compared against their partner regarding the hypervolume w. r. t. an adaptive reference point depending on that pair only. Then, the best individual can be determined as follows: The winners of the comparisons, i.e., the individuals with higher hypervolume values are kept, the others discarded from the data structure. The remaining set of

winners is again divided into pairs and the process proceeds iteratively until the set of winners contains only one individual.

Algorithm 3.6: Tournament selection in set T with pairwise comparisons

```
1 while T.size > 1 do
2     divide set T into pairs
3     foreach pair (T^(i), T^(j)) do
4         r = (max{f_1(T^(i)), f_1(T^(j))} + 1, max{f_2(T^(i)), f_2(T^(j))} + 1)^⊤
5         if f(T^(i)) ⪯ f(T^(j)) or
            (f(T^(i)) ∥ f(T^(j)) and H({f(T^(i))}, r) > H({f(T^(j))}, r)) then
6             discard T^(j) from T
7         else discard T^(i) from T
```

As shown by Lemma 2.5, the absolute hypervolume values and the hypervolume contributions result in the same ordering of two points, so the hypervolume contribution could be used as well. As proved in Theorem 3.12, the order according to the hypervolume equals the inverse order created by the sum f^s, thus the tournament selection based on pairwise comparisons is equal to the selection according to the unweighted sum performed by a single-objective EA. So, the individual winning the tournament regarding the pairwisely highest hypervolume values is the individual with lowest value of the sum f^s. The selection is less efficient than the normal one but this does not effect the convergence results as these are calculated with respect to the black-box complexity and the tournament does not require additional function evaluations.

We specify different tournaments for different selection aims. For all tournaments holds that in case of equal hypervolume values offspring are preferred over parents, i.e., the parent loses the comparison as done by the selection of a single-objective EA. Remaining ties are broken randomly.

Determining one best individual suffices for a $(1, \lambda)$ or a $(1 + \lambda)$ selection, with the ranking of the remaining individuals being irrelevant. Algorithm 3.6 performs this task when called with the set of the λ offspring for the first case and called with the set of the parent and the λ offspring for the latter. Thereby, we establish algorithmic equivalence between the SMS-EMOA and a single-objective EA.

Determining one worst individual is the goal of a $(\mu + 1)$ selection. The tournament is performed as described above, whereas the losing individuals proceed the tournament and the winning individuals are discarded from the tournament.

For a $(\mu \dot{+} \lambda)$ selection, the ranking of the best μ, or alternatively the worst λ for the $\dot{+}$-selection, respectively the worst $\lambda - \mu$ for the ,-selection have to be determined. We know from the equivalence to the inverse order of the sum f^s that a total order is produced when all individuals are compared pairwisely against each

other. The best individual is the one that wins each comparison, the second best wins each comparison but one against the best individual and so on. Thereby a ranking establishes among the individuals and can be carved out as far as necessary for the selection. Our argumentation leads to the following lemma of algorithmic equivalence.

Lemma 3.20 *The $(\mu \dot{+} \lambda)$-SMS-EMOA with tournament selection using pairwise comparisons of hypervolume values w. r. t. the adaptive reference point applied to a bi-objective optimization problem $\min\{f(x) \mid x \in \mathbb{R}^n\}$ is algorithmically equivalent to a single-objective $(\mu \dot{+} \lambda)$-EA applied to the single-objective function $f^s : \mathbb{R}^n \to \mathbb{R}$ with $f^s(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$, assuming the same choice of μ, λ and $+$ or $\dot{+}$, for the selection and the remaining operators being identical. (Beume et al. (2011)*).* □

Due to the algorithmic equivalence, we are able to transfer known convergence rates from the single-objective EA to the SMS-EMOA. The following theorems are taken literally from Beume et al. (2011)*.

The convergence result in Rudolph (1997, p. 184, Th. 6.11) for the $(1, \lambda)$-EA can directly be transcribed to the $(1, \lambda)$-SMS, due to their algorithmic equivalence (Lemma 3.20).

Theorem 3.21 *Suppose the objective functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ lead to a (K, Q)-strongly convex surrogate function via $f^s(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ with $f^s : \mathbb{R}^n \to \mathbb{R}$. Let the constant M_λ denote the expectation of the maximum of λ independent samples from a Beta distribution with support $(-1, 1)$ and parameters $p = q = (n-1)/2$. Then the convergence rate of the $(1, \lambda)$-SMS-EMOA, $2 \leq \lambda < \infty$ with adaptive reference point, pairwise comparisons, and variation operations as described in (Rudolph, 1997, p. 184, Th. 6.11) is*

(a) $1 - M_\lambda^2$ if f^s is quadratic, the Hessian matrix $\nabla^2 f^s(\mathbf{x})$ is known in advance and the length of the gradient in ellipsoid norm $\|\cdot\|_{\nabla^2 f^s(\mathbf{x})}$ can be determined;

(b) $\leq 1 - M_\lambda^2/Q^{3/2}$ if f^s is quadratic, the Hessian matrix $\nabla^2 f^s(\mathbf{x})$ is known in advance and the Euclidean length of the gradient can be determined;

(c) $\leq 1 - M_\lambda^2/Q^2$ if the largest eigenvalue of the Hessian (or constant L) is known in advance and the Euclidean length of the gradient can be determined. (Beume et al. (2011)).* □

Evidently, the gradient and Hessian of the surrogate function follow from the gradients and Hessians of the objective functions:

$$\begin{aligned} \nabla f^s(\mathbf{x}) &= \nabla(f_1(\mathbf{x}) + f_2(\mathbf{x})) &&= \nabla f_1(\mathbf{x}) + \nabla f_2(\mathbf{x}) \\ \nabla^2 f^s(\mathbf{x}) &= \nabla^2(f_1(\mathbf{x}) + f_2(\mathbf{x})) &&= \nabla^2 f_1(\mathbf{x}) + \nabla^2 f_2(\mathbf{x}). \end{aligned}$$

Due to the algorithmic equivalence (Lemma 3.20), the convergence result in Rudolph (1997, p. 190, Th. 6.12) for the $(1+\lambda)$-EA can be transferred to the $(1+\lambda)$-SMS-EMOA.

Theorem 3.22 *Suppose the objective functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ lead to a (K,Q)-strongly convex surrogate function via $f^s(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ with $f^s : \mathbb{R}^n \to \mathbb{R}$. If the Euclidean length of the gradient is available for all $\mathbf{x} \in \mathbb{R}^n$ then the following property holds for the $(1+\lambda)$-SMS-EMOA with $1 \leq \lambda < \infty$, using the adaptive reference point, pairwise comparisons of hypervolume values, and mutation of a parent X_t via $X_t + \|\nabla f^s(X_t)\| \cdot Z$ where Z is a spherically distributed random vector with support $\mathbb{R}^n$: The algorithm converges almost surely and in mean to the Pareto front and the mean velocity of approach is geometrically fast (i.e., linear convergence rate) regardless of the actual choice of the distribution of Z and its parameterization. (Beume et al. (2011)*).* □

Due to the algorithmic equivalence (Lemma 3.20), the bounds for the $(\mu+1)$-EA shown by Jägersküpper and Witt (2005, Th. 2,3) result in the following theorem for the $(\mu+1)$-SMS-EMOA.

Theorem 3.23 *Suppose the objective functions $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ lead to a unimodal surrogate function $f^s(\mathbf{x}) = f_1(\mathbf{x}) + f_2(\mathbf{x})$ satisfying $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{x}^*| < |\mathbf{y} - \mathbf{x}^*| \Rightarrow f(\mathbf{x}) < f(\mathbf{y})$, where $\mathbf{x}^* \in \mathbb{R}^n$ is the unique minimizer. The $(\mu+1)$-SMS-EMOA, where $2 \leq \mu = poly(n)$, with adaptive reference point and pairwise comparisons of hypervolume values within the selection, optimizes such a bi-objective optimization problem where the number of steps until the distance to x^* has been halved is $\Omega(\mu n)$ with probability $1 - 2^{\Omega(n)}$ and also in expectation, and $O(\mu n)$ w.o.p. with $\sigma^{(0)} = \Theta(D/n)$, where D denotes the initial distance to the minimizer. (cf. Beume et al. (2011)*).* □

For a $(\mu \overset{+}{,} \lambda)$-EA no convergence rates are known for both μ and λ greater than one, to the best of our knowledge. Convergence has been proven (see e.g. Rudolph (1997, Sec. 6.3)) but that is not our focus of interest here. However, appropriate results will hold for the SMS-EMOA accordingly.

3.2.4 SMS-EMOA with Fixed Reference Point

In contrast to the previous sections where we considered the adaptive reference point, we here analyze the SMS-EMOA with a reference point that remains fixed throughout the optimization process.

Theorem 3.24 *The (1+1)-SMS-EMOA with fixed reference point $\mathbf{r} \in \mathbb{R}^2$ for minimizing $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}))^\top$ is algorithmically equivalent to the (1+1)-EA maximizing the hypervolume*

$$H(\{f(\mathbf{x})\}, \mathbf{r}) = [r_1 - f_1(\mathbf{x})]\,[r_2 - f_2(\mathbf{x})], \tag{3.13}$$

presumed all operators apart from the selection are identical in both algorithms. (Beume et al. (2011)).*

Proof. Due to the conditions of the theorem, it suffices to show the algorithmic equivalence of the selection operators, whereas the argumentation of the proof is analogous to Theorem 3.12. Before the start of the optimization process, fix a reference point $\mathbf{r} \in \mathbb{R}^2$ such that it is dominated by all possible individuals. Let $\mathbf{x} \in \mathbb{R}^n$ be the current parent and $\mathbf{y} \in \mathbb{R}^n$ the current offspring of both algorithms. We detail that whenever the SMS-EMOA accepts the offspring, so does the EA; omitting the analogous proof that whenever the SMS-EMOA rejects the offspring, so does the EA.
If $f(\mathbf{y}) \preceq f(\mathbf{x})$ holds, the SMS-EMOA accepts the offspring without hypervolume calculation. Then it holds $H(\{f(\mathbf{y})\}, \mathbf{r}) \geq H(\{f(\mathbf{x})\}, \mathbf{r})$, which is intuitively clear and proved in Theorem 2.4.2, so $\mathbf{y}$ is as well accepted by the (1+1)-EA. If $f(\mathbf{y}) \parallel f(\mathbf{x})$, the offspring is accepted if its hypervolume contributions is larger than the parent's. Lemma 2.5 tells us that in case of two points, the ordering of the hypervolume contribution and the absolute hypervolume (used as fitness function in the EA) are equal. Thus the selection among incomparable individuals is algorithmically equivalent for both algorithms which completes the proof.

□

Since the argumentation above considered the selection without further properties of the individuals, the same result holds for the selection among any two points, like the two offspring in a $(1, 2)$-selection. So, the algorithmic equivalence holds analogously for that case.

Corollary 3.25 *The $(1, 2)$-SMS-EMOA with fixed reference point $\mathbf{r} \in \mathbb{R}^2$ for minimizing $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}))^\top$ is algorithmically equivalent to the $(1, 2)$-EA that maximizes the hypervolume $H(\{f(\mathbf{x})\}, \mathbf{r}) = [r_1 - f_1(\mathbf{x})]\,[r_2 - f_2(\mathbf{x})]$, presumed all operators apart from the selection are identical in both algorithms.*

We proof a new convergence rate for the single-objective (1+1)-EA that also holds for the (1+1)-SMS-EMOA due to their algorithmic equivalence (Theorem 3.24). The following part is mainly transferred literally from Beume et al. (2011)* with adapted notation.

Theorem 3.26 *The (1+1)-SMS-EMOA with fixed reference point and variation by uniform mutations on the surface of a unit hypersphere and step sizes proportional to the length of the gradient maximizes the dominated hypervolume of the linear problem*

$$f(\mathbf{x}) = (\mathbf{a}^\top \mathbf{x} + a_0, \mathbf{b}^\top \mathbf{x} + b_0)^\top \to \min!, \quad \mathbf{x} \in \mathbb{R}^n$$

with convergence rate

$$c(\mathbf{x}) = 1 - \frac{\kappa(\mathbf{x})}{\nu_1^2} \cdot \frac{0.405}{n}$$

where $0 \leq \kappa(\mathbf{x}) \leq \nu_1^2$ *and* ν_1 *is the largest eigenvalue of matrix* $\mathbf{a}\,\mathbf{b}^\top$ *provided that* $\mathbf{a}\,\mathbf{b}^\top$ *is negative semidefinite. (Beume et al. (2011)*).*

Proof. Owing to Lemma 3.7 and the preconditions of this theorem it is guaranteed that the $[\ldots]$ hypervolume in (3.13) is a quadratic form whose Hessian matrix is negative semidefinite with rank 1. In order to invoke Theorem 3.24 for deriving a convergence result for the (1+1)-SMS-EMOA a convergence result for the (1+1)-EA for semidefinite quadratic forms is required. Apparently, such a result does not exist yet. Therefore, the remainder of the this section is devoted to establish such a result in Theorem 3.29 below. Since the maximization of a concave function is identical to minimizing a convex function the application of Theorem 3.29 completes the proof of this theorem. □

At first, some facts from matrix theory are required.

Lemma 3.27 *Let* $A : n \times n$ *be a symmetric and positive semidefinite matrix with* $\mathrm{rank}(A) < n$*.*

a) The eigenvalues $\nu_1 \geq \nu_2 \geq \ldots \geq \nu_n \geq 0$ *of* A *are nonnegative with* $\nu_1 > 0$ *and* $\nu_n = 0$*.*

b) The eigenvalues of A^2 *are* ν_i^2 *for* $i = 1, \ldots, n$*.*

c) $\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} : 0 = \nu_n \leq \frac{\mathbf{x}^\top A \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \leq \nu_1$*.*

Proof. See e.g. Magnus and Neudecker (1999). □

Next, a probabilistic result is of essential importance.

Lemma 3.28 *If* $\mathbf{u}$ *is an* n*-dimensional random vector that is uniformly distributed on the surface of a unit hypersphere then* $\|\mathbf{u}\| = 1$ *with probability* 1 *and* $\mathbf{x}^\top \mathbf{u} = \|\mathbf{x}\| \cdot B$ *where* B *is a Beta random variable with support* $[-1, 1]$ *and parameters* $p = q = (n-1)/2$*.*

Proof. See (Rudolph, 1997, p. 22). □ □

Now we are in the position to prove the desired result.

Theorem 3.29 *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a quadratic function with positive semidefinite Hessian matrix whose rank is less than* n*. The (1+1)-EA with uniform mutations*

on the surface of a unit hypersphere and step sizes proportional to the length of the gradient minimizes $f(\cdot)$ with position-dependent mean convergence rate

$$c(\mathbf{x}) = 1 - \frac{\kappa(\mathbf{x})}{\nu_1^2} \cdot \frac{0.405}{n}$$

with $0 \leq \kappa(\mathbf{x}) \leq \nu_1^2$. (Beume et al. (2011)).*

Proof. It suffices to consider the case $f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ where A is positive semidefinite with $\operatorname{rank}(A) < n$. Let $s > 0$ be the deterministic step size and $\mathbf{u}$ an n-dimensional random vector that is uniformly distributed on the surface of a hypersphere (i.e., $\mathbf{u}$ is a random direction of unit length).
Let $\mathbf{x} \in \mathbb{R}^n$ be the current position of the (1+1)-EA. The random fitness value of the offspring $\mathbf{x} + s\,\mathbf{u}$ is given by

$$\begin{aligned}
f(\mathbf{x} + s\,\mathbf{u}) &= (\mathbf{x} + s\,\mathbf{u})' A (\mathbf{x} + s\,\mathbf{u}) \\
&= \mathbf{x}^\top A\mathbf{x} + 2\,s\,\mathbf{u}' A\mathbf{x} + s^2\,\mathbf{u}' A\mathbf{u} \\
&= \mathbf{x}^\top A\mathbf{x} + 2\,s\,\|A\mathbf{x}\| \cdot B + s^2\,\mathbf{u}' A\mathbf{u} && (3.14) \\
&\leq \mathbf{x}^\top A\mathbf{x} + 2\,s\,\|A\mathbf{x}\| \cdot B + s^2 \cdot \nu_1 && (3.15) \\
&= f(\mathbf{x}) + \frac{\|A\mathbf{x}\|^2}{\nu_1} \left(2\gamma \cdot B + \gamma^2\right) && (3.16)
\end{aligned}$$

where (3.14) results from Lemma 3.28, (3.15) from Lemma 3.27(c), and (3.16) from setting $s = \gamma\,\|A\mathbf{x}\|/\nu_1$, with $\gamma > 0$. After selection in generation t (indicated by the subscript (t)), the fitness value of the selected individual is

$$f(\mathbf{x}_{(t+1)}) = f(\mathbf{x}_{(t)}) + \frac{\|A\mathbf{x}_{(t)}\|^2}{\nu_1} \min\{0, 2\gamma \cdot B + \gamma^2\}$$

since a positive value of $2\gamma \cdot B + \gamma^2$ would lead to a rejection of the offspring. This in turn leads to the expected fitness value

$$\begin{aligned}
\mathrm{E}\left(f(\mathbf{x}_{(t+1)})\right) &= f(\mathbf{x}_{(t)}) + \frac{\|A\mathbf{x}_{(t)}\|^2}{\nu_1} \mathrm{E}\left(\min\{0, 2\gamma \cdot B + \gamma^2\}\right) \\
&\leq f(\mathbf{x}_{(t)}) - \frac{\|A\mathbf{x}_{(t)}\|^2}{\nu_1}\, c_n && (3.17)
\end{aligned}$$

where

$$c_n = -\min_{\gamma > 0}\{\mathrm{E}\left(\min\{0, 2\gamma \cdot B + \gamma^2\}\right)\} \approx \frac{0.405}{n} > 0$$

for large n (see Rudolph (1997), p. 171f.). According to Lemma 3.27 inequality (3.17) can be further processed via

$$\begin{aligned} \mathrm{E}\left(f(\mathbf{x}_{(t+1)})\right) &\leq f(\mathbf{x}_{(t)}) - \frac{(A\mathbf{x}_{(t)})^\top (A\mathbf{x}_{(t)})}{\nu_1} c_n \\ &= f(\mathbf{x}_{(t)}) - \frac{\mathbf{x}_{(t)}^\top A^2 \mathbf{x}_{(t)}}{\nu_1} c_n \\ &= f(\mathbf{x}_{(t)}) - \frac{\kappa(\mathbf{x}_{(t)})\, \mathbf{x}_{(t)}^\top \mathbf{x}_{(t)}}{\nu_1} c_n \end{aligned}$$

where $\kappa(\mathbf{x}) \in [0, \nu_1^2]$. Since $f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} \leq \nu_1\, \mathbf{x}^\top \mathbf{x}$

$$\begin{aligned} \mathrm{E}\left(f(\mathbf{x}^{(t+1)})\right) &\leq f(\mathbf{x}_{(t)}) - \frac{\kappa(\mathbf{x}_{(t)})\, f(\mathbf{x}_{(t)})}{\nu_1^2} c_n \\ &= f(\mathbf{x}_{(t)}) \left(1 - \frac{\kappa(\mathbf{x}_{(t)})\, c_n}{\nu_1^2}\right) \end{aligned}$$

which proves the theorem. □

3.2.5 Other EMOA

We consider other popular EMOA to examine whether the properties of the SMS-EMOA are outstanding. It is observed that the IBEA behaves equivalent to the SMS-EMOA, whereas NSGA-II and SPEA2 do not reach a competitive convergence rate. The remainder of the subsection is partly transferred literally from Beume et al. (2010)*, whereas the result for (1+1)-IBEA$_{HD}$ is an addition.

IBEA

We show that a (1+1)-IBEA (see Sec. 3.1, Zitzler and Künzli (2004)) selecting according to the additive ϵ-indicator $I_{\epsilon+}$ (see Zitzler et al. (2003)) or according to the pairwise hypervolume I_{HD} performs equally to the (1+1)-SMS-EMOA for two objectives. IBEA prefers non-dominated individuals over dominated ones, so for the case of two comparable individuals, the behavior of acceptance and rejection is clearly equal to the one of the SMS-EMOA. For incomparable individuals, the indicators $I_{\epsilon+}$ or I_{HD} come into play, which are relative binary indicators, originally defined on two sets of points. For two points, $I_{\epsilon+}(\mathbf{a}, \mathbf{b})$ calculates the minimal distance ϵ by which $\mathbf{a}$ can be or needs to be moved in each direction in order to weakly dominate $\mathbf{b}$, given as (3.2) in Sec. 3.1 (cf. Fig. 3.4, left):

$$I_{\epsilon+}(\mathbf{a}, \mathbf{b}) = \min_{\epsilon \in \mathbb{R}} \{\forall i \in \{1, \ldots, d\} : f_i(\mathbf{a}) - \epsilon \leq f_i(\mathbf{b})\}.$$

The indicator value is negative or zero for $\mathbf{a} \prec \mathbf{b}$, so the smaller the better. For two points, the fitness function as in (3.4) reduces to $F(\mathbf{a}) = -e^{-I_{\epsilon+}(\mathbf{b},\mathbf{a})/\kappa}$ and

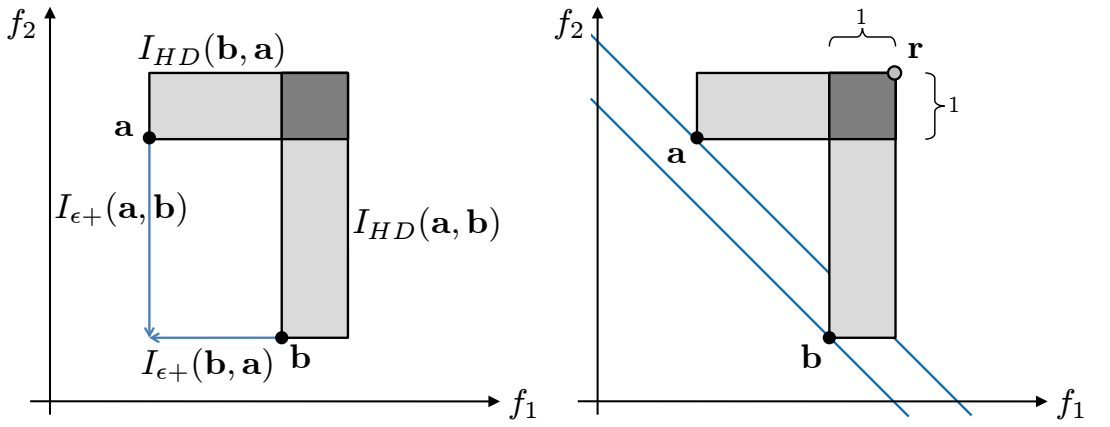


Fig. 3.4: IBEA's indicator values (left) correspond to the hypervolume contributions w. r. t. the adaptive reference point (right). The blue arrows give the $I_{\epsilon+}$ values and the light gray areas the I_{HD} values.

$F(\mathbf{b}) = -e^{-I_{\epsilon+}(\mathbf{a},\mathbf{b})/\kappa}$, where we can neglect the scaling factor κ for the ranking. So actually each point is assigned the value of the other point, allocated such that a smaller indicator value leads to a smaller fitness. Since the point with the lower fitness is discarded, we can conclude that a point $\mathbf{b}$ survives if the indicator value of point $\mathbf{a}$ is the greater one. This value corresponds to the hypervolume contribution of point $\mathbf{b}$ (cf. Fig. 3.4), which has been shown to reduce to a distance in this scenario. Thus we have an analogous behavior to the (1+1)-SMS-EMOA.
For our scenario of two points, I_{HD} reduces to the hypervolume contribution, such that $I_{HD}(\mathbf{a},\mathbf{b}) = H(\mathbf{b},\{\mathbf{a},\mathbf{b}\},\mathbf{r})$ and $I_{HD}(\mathbf{b},\mathbf{a}) = H(\mathbf{a},\{\mathbf{a},\mathbf{b}\},\mathbf{r})$ (cf. Fig. 3.4), presumed the reference point is chosen alike. With the fitness assignment analogous to $I_{\epsilon+}$, it results that a point $\mathbf{b}$ survives if it has greater fitness, thus if its hypervolume contribution is larger just like in the (1+1)-SMS-EMOA.

Theorem 3.30 *For bi-objective problems, the (1+1)-IBEA$_{\epsilon+}$ as well as the (1+1)-IBEA$_{HD}$ with equal reference point is algorithmically equivalent to the (1+1)-SMS-EMOA with adaptive reference point, presumed all operators apart from the selection being identical. (cf. Beume et al. (2010)*).* □

Corollary 3.31 *Theorem 3.12 and Corollary 3.14 hold as well for the (1+1)-IBEA$_{\epsilon+}$ as well as for the (1+1)-IBEA$_{HD}$. (cf. Beume et al. (2010)*).*

NSGA-II and SPEA-2

NSGA-II (see Sec. 3.1, Deb et al. (2002a)) has been developed with a $(\mu + \mu)$ selection, and is here considered for $\mu = 1$. The selection starts by performing non-dominated sorting on the set of parent and offspring. If the individuals are

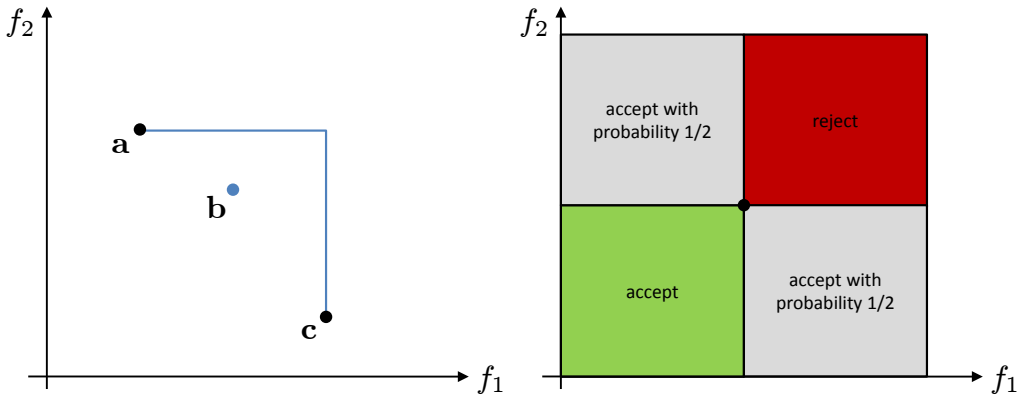


Fig. 3.5: Left: The length of the blue lines defines the crowding distance of point **b** in NSGA-II, whereas it is undefined for the boundary points **a** and **c**. Right: Regions of acceptance, rejection, or random acceptance in case of incomparable points for (1+1)-NSGA-II, (1+1)-SPEA2, and the simple (1+1)-EA ROSEA, analyzed in Rudolph (1998).

comparable, the dominating one is kept and the dominated one is discarded. In case of incomparable individuals the crowding distance is invoked. It rewards individuals with a large distance to their neighbors (cf. Fig. 3.5, left), and assigns a value of infinity to points at the boundary of the non-dominated front, i.e., those not having a worse neighbor in at least one dimension. Here, both points are boundary points with equal crowding distance values. Thus, one is chosen to be discarded uniformly at random, so in case of incomparable points, each is accepted with probability 1/2 (cf. Fig. 3.5, right).
This special handling of the boundary points is common for the SMS-EMOA as well, yet the hypervolume contribution can be performed to obtain a total order of the individuals. In contrast, the crowding distance is not defined for boundary points. Analogously to the hypervolume, a reference point could be defined as a stand-in for the missing neighbors. Yet, even with this idea there seems to be no reasonable definition that fits into the concept of the crowding distance and repairs its drawbacks. However the measure bases on Manhattan distances between the points, the values for two points will always be equal as the distance is symmetric.

The same result holds for the (1+1)-SPEA2 (see Sec. 3.1, Zitzler et al. (2002)). For incomparable individuals, there are neither dominated nor dominating ones, thus the raw fitness of both individuals is zero. So, the secondary indicator based on a k-nearest neighbor method is used. The resulting values for the individuals are equal since they both are their only neighbors and distances are symmetric.
We declare that both NSGA-II and SPEA-2 in their (1+1) version are algorithmically equivalent to the (1+1)-EMOA termed *ROSEA (Random Objective Selection EA)* in Algorithm 3.7, which has been analyzed by Rudolph (1998) and is based

Algorithm 3.7: ROSEA

1 choose $X^{(0)} \in \mathbb{R}^n$, set $t = 0$
2 **repeat**
3 draw $Z^{(t)}$ from $U[\partial S_n(1)]$ /* mutation */
4 $\sigma^{(t)} = \frac{1}{2} \min_{w \in [0,1]} \{\|(1-w)\nabla f_1(X^{(t)}) + w\nabla f_2(X^{(t)})\|\}$
5 $Y^{(t)} = X^{(t)} + \sigma^{(t)} Z^{(t)}$
6 draw i with $Pr(i = 1) = Pr(i = 2) = \frac{1}{2}$ /* choose objective i */
7 **if** $f_i(Y^{(t)}) \leq f_i(X^{(t)})$ **then** /* selection acc. to objective i */
8 $X^{(t+1)} = Y^{(t)}$
9 **else** $X^{(t+1)} = X^{(t)}$
10 increment t
11 **until** termination criterion fulfilled

on the framework proposed by Kursawe (1991). ROSEA chooses uniformly at random one fitness function for selection (line 6). The better individual regarding the function is kept, the other one discarded. Two incomparable individuals have both worst and best values in interchanged functions. So, choosing a function is equivalent to choosing the preferred individual. This results in the following theorem.

Theorem 3.32 *The (1+1)-NSGA-II and the (1+1)-SPEA2 are algorithmically equivalent to the (1+1)-ROSEA, assuming all operators apart from the selection being identical. (Beume et al. (2010)*).* □

Since Rudolph (1998) proves that for ROSEA convergence is given but only with a sub-linear rate for at least one instance from the problem class, we immediately get the following result.

Corollary 3.33 *The (1+1)-NSGA-II and the (1+1)-SPEA2 have sub-linear convergence rate under conditions for the step sizes given in Rudolph (1998), i.e., with D_t denoting the distance to Pareto set at step $t \geq 0$, it holds $D_t \to 0$ in mean and with probability 1, and even $E(D_{t+1}) \leq c_t \cdot E(D_t)$ with $c_t \to 0$ as $t \to \infty$. (Beume et al. (2010)*).* □

It is still unclear how this step size rule can be realized in practice and, thus, whether NSGA-II and SPEA2 converge at all for any other known mutation operator. Nevertheless, our result indicates that sub-linear convergence might be the best one can hope for.

3.2.6 Conclusions

This work compasses a substantial progress of the convergence analysis of state-of-the-art EMOA. Due to a new proof mechanism of showing the algorithmic equiva-

lence between single- and multi-objective EA, we are able to transfer convergence results from the single- to the multi-objective case. We gain convergence rates for state-of-the-art EMOA, mainly the SMS-EMOA, for bi-objective continuous problems.
We proved the following algorithmic equivalences of the SMS-EMOA applied to any bi-objective minimization problem and a single-objective EA, where the operators apart from the selection are arbitrary but assumed to be equal for both algorithms.

- The SMS-EMOA with adaptive reference point and pairwise selection is algorithmically equivalent to the single-objective EA minimizing the sum of objectives f^s with the same selection scheme for all cases of a $(\mu \dot{+} \lambda)$ selection. For a $(1+1)$ or $(1,2)$ selection, the pairwise selection equals the normal selection as it is performed on a set of size 2.

- The (1+1)-SMS-EMOA with fixed reference point is algorithmically equivalent to the (1+1)-EA maximizing the hypervolume.

It is remarkable that the algorithmic equivalence holds for *any* bi-objective problem. Thereby, any future insight gained for the single-objective EA on a function class containing the fitness function f^s automatically holds for the corresponding SMS-EMOA as well. It is an exceptional observation that two algorithms with different operators, different computational complexity and different behavior for other cases, behave identical in this bi-objective scenario.
The concept of pairwise tournament selection has been developed to establish algorithmic equivalence. It is not recommended to become a standard operator in the SMS-EMOA, since the behavior of the selection according to a weighted sum is equipped with the known drawbacks of simple a priori techniques, e.g. in case of concave functions only points of the boundary of the Pareto front are optimal.

The following convergence rates hold for the SMS-EMOA with a certain variation operator applied to a bi-objective minimization problem of a certain class as well as for the algorithmically equivalent single-objective EA.

- For variation by mutation on a hypersphere with self-adaptation by the 1/5-rule, and problems with both objective functions being quadratically convex of bounded bandwidth, the convergence rate of the SMS-EMOA with adaptive reference point and selection as (1+1) or (1,2) is linear.

- For mutation as specified by Rudolph (1997) and problems with (K,Q)-strongly convex functions, several bounds for the convergence rates of the $(1, \lambda)$-SMS-EMOA with adaptive reference point and pairwise selection are given.

- For mutation as specified by Rudolph (1997) and problems with (K,Q)-strongly convex functions, the $(1 + \lambda)$-SMS-EMOA with adaptive reference point and pairwise selection has linear convergence rate.

- For the $(\mu + 1)$-SMS-EMOA with adaptive reference point and pairwise selection on certain problems where f^s is unimodal with a unique optimizer, a lower bound $\Omega(\mu n)$ and an upper bound $O(\mu n)$ are proven conditionally for halving the initial distance to the optimum.

- For variation by mutation on a hypersphere and certain linear problems, a position-dependent convergence rate is shown for the (1+1)-SMS-EMOA with fixed reference point. Having proved the algorithmic equivalence to the (1+1)-EA maximizing the hypervolume, the analysis is performed for the single-objective case, where a convergence rate was not known before for the (1+1)-EA in this scenario.

For the first time a linear convergence rate was proved for EMOA that do not use an explicit weighting of objectives. As visible by the algorithmic equivalence of the hypervolume and the (unweighted or equally weighted) sum, the hypervolume realizes a kind of implicit weighting for the case of two points and the adaptive reference point.
The approach of establishing algorithmic equivalence of the hypervolume selection and the selection according to f^s has the following limitations as shown by counter-examples.

- The equivalence of the (1+1)-SMS-EMOA with adaptive reference point to a single-objective EA minimizing f^s does not hold for more than 2 objectives. Nevertheless, other fitting substitute functions may exist.

- The selection of the SMS-EMOA among a set of more than 2 points is not equivalent to the selection of a single-objective EA w. r. t. f^s. This result motivates the concept of the pairwise selection, to reestablish the selection among 2 points.

Apart from the SMS-EMOA, we proved results for NSGA-II, SPEA2 and IBEA.

- The (1+1)-IBEA with selection according to the additive ϵ-indicator $I_{\epsilon+}$ or the pairwise hypervolume I_{HD} performs equally to the (1+1)-SMS-EMOA for two objectives. So the results of algorithmic equivalence and convergence rates hold for this algorithm accordingly.

- The (1+1)-NSGA-II and the (1+1)-SPEA2 have a sub-linear convergence rate on the considered class of functions since their selection mechanism among incomparable individuals degenerates to random choice in case of 2 points.

With our presented proof techniques more results may be gained, and we do not guarantee completeness. Our results shall be considered as prototypic and more may follow alike.

For future work on convergence properties, the analysis of multi-objective problems with more than two objectives is interesting. This section demonstrated that the very simple function f^s behaves equally to the hypervolume in certain scenarios. As the hypervolume is complex, this motivates to look for other simple measures that may replace the hypervolume in specific scenarios i.e., share the positive properties of the hypervolume w. r. t. the dominance relation. Other surrogate functions may be found to establish the algorithmic equivalence of single- and multiobjective EA. Thereby, the concept of pairwise selection is a help, but it is desired to become obsolete. The results for the SMS-EMOA with a fixed reference point shall be extended. An aspirational goal is the analysis of the original SMS-EMOA with the $(\mu+1)$-selection and the adaptive reference point. The analysis could be enhanced to include other multi-objective optimization algorithms.
More general, it is desirable to transfer results of single-objective EA to the multi-objective case. More techniques like our one shall help to bridge the gap between the theory on these algorithm classes and to catch up with the lead of the single-objective case.

3.3 Optimality of Steady-State-Selection

The previous section dealt with the convergence speed of the SMS-EMOA towards the Pareto front. To take a different point of view, we here consider the optimization of the population's quality *on* the Pareto front of bi-objective problems, i.e., the situation when the population already consists of Pareto-optimal points. The results are taken from our publication Beume et al. (2009b)* and so are all blue indicated literal citations of this subsection, whereas the notation is adapted to this document.

The SMS-EMOA has been developed with a $(\mu + 1)$-section scheme, so only one individual is exchanged in the population per iteration. This selection scheme is called *steady-state selection* or *1-greedy selection* in this context. The question arises whether these small changes suffice to reach a global optimum of the population's quality and escape from locally optimal compositions. As the SMS-EMOA seeks to optimize the hypervolume, the natural question is whether it succeeds, so the population's quality is measured regarding the hypervolume. An improvement of the hypervolume is in continuous spaces always possible as long as there are points dominating a population member, assuming that these points can be sampled with positive probability. Note that regarding continuous spaces the Pareto front is practically unreachable.

In contrast to the default selection scheme using the adaptive reference point, the reference point is here fixed throughout the optimization process. This is motivated by the assumption that the reference point does not change very much—at least not in magnitudes—when the population is already situated on the Pareto front. The fixed reference point is chosen as if it was the adaptive reference point w. r. t. the boundary points of the Pareto front. Thereby, the global optimum of the population's hypervolume is well-defined for a fixed population size and we can consider the process of approaching it.

Related work on this topic took a worst-case perception by providing counter-examples indicating that certain algorithms cannot reach the global optimum in general. Knowing that counter-examples exist is of theoretical interest, whereas we are interested in *when* these counter-examples occur and whether they occur in practice. So, we take a more practical perspective by case-related analysis of certain classes of functions.

The next Section 3.3.1 details the related work and gives basic definitions. Our starting point in Section 3.3.2 is a counter-example from literature indicating that a $(2 + 1)$ selection according to the hypervolume gets stuck in a local optimum. To relativize its artificial properties, we transform the example from a discrete to a two-part continuous Pareto front while maintaining its essential properties, analyze it theoretically and study experimentally whether the global optimum is attainable for algorithms exchanging more than one individual. It is revealed that it is also difficult for these algorithms to reach the global optimum due to strong

local attractors. In Section 3.3.3 we analyze and experimentally study simpler shaped connected Pareto fronts. For linear Pareto fronts, it is proved that a 1-greedy hypervolume selection suffices to reach the global optimum from any starting configuration. Experiments on Pareto fronts of different curvature allude that the problems are solvable to optimality with a 1-greedy hypervolume selection. Our results are summarized and discussed in Section 3.3.4.

3.3.1 Preliminaries

Related work

Bringmann and Friedrich (2011) denote EMOA as *ineffective* in case there exists a start population from which a set with maximal hypervolume cannot be reached throughout the optimization process. The maximal hypervolume seems to be considered w. r. t. a fixed reference point.
Further, they distinguish among hypervolume-based *non-decreasing* EMOA which select such that the population's hypervolume w. r. t. a fixed reference point never decreases compared to the preceding generation, and *locally optimal* EMOA whereas the next population is chosen such that a somehow local hypervolume value is maximal, e.g. w. r. t. the adaptive reference point like in the SMS-EMOA.
Trivially, $(\mu + \lambda)$-EMOA with $\lambda \geq \mu$ are effective, assuming they can sample the optimal set with positive probability, as proven in (Bringmann and Friedrich, 2011, Th. 3.4) for locally optimal EMOA and in (Zitzler et al., 2010, Th. 4.4) for non-decreasing ones.
Zitzler et al. (2010, Cor. 4.6) show by an extreme counter example that non-decreasing $(\mu + 1)$-EMOA are not effective in general, respectively not *1-greedy* as they call it. Bringmann and Friedrich (2011, Th. 3.5) prove that all non-decreasing $(\mu + \lambda)$-EMOA with $\lambda < \mu$ are ineffective.
Moreover Bringmann and Friedrich (2011) give approximation results, saying that $(\mu+\lambda)$-EMOA with $\lambda < \mu$ can always find a population of at least half the optimal hypervolume, and no hypervolume-based EMOA can always achieve a population's hypervolume greater than $1/(1+0.1338(1/\lambda-1/\mu))$ times the optimal hypervolume.

Basic Definitions

Definition 3.34 *Zitzler et al. (2010) 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 examine the hypervolume maximization in the sense of such a preference relation, so that the population with the best hypervolume is preferred. We transfer the definition to hypervolume-based EMOA and multiobjective problems.

Definition 3.35 *We denote a hypervolume-based EMOA as* k-greedy *if it performs a* $(\mu + k)$*-selection, where in each generation the population is chosen such that the population's hypervolume is maximal w. r. t. a specified fixed reference point.*

So, from the set of size $\mu + k$, the subset of size μ with the highest hypervolume among all those subsets is chosen. In our scenario, the considered EMOA are *non-decreasing* (in the context of the definitions by Bringmann and Friedrich (2011)) because the reference point is fixed throughout the optimization and so is the optimization goal, in contrast to the locally optimal SMS-EMOA with adaptive reference point. Due to Bringmann and Friedrich (2011) all non-decreasing EMOA are ineffective, which means that a counter example exist for any $(\mu + \lambda)$-EMOA with $\lambda < \mu$. Yet, we analyze these EMOA on certain classes of functions to further understand when these failures occur.

Definition 3.36 *A problem is denoted as* k-greedy solvable *if for any initial population, there exists a finite number of iterations of a* k*-greedy EMOA resulting in the optimal set, i.e., the set of size* μ *with maximal hypervolume w. r. t. the specified fixed reference point.*

Note that [...] any problem is μ-greedy solvable for a $(\mu + k)$-EMOA with $k \geq \mu$ assuming that all search points are sampled with positive probability. The composition of sets maximizing the hypervolume among all sets of fixed size has been analyzed by Auger et al. (2009a, 2010) as *optimal* μ*-distributions.* To further examine the potential of small changes, we define a more restricted scenario.

Definition 3.37 *A problem is denoted as* locally k-greedy solvable *if it is* k*-greedy solvable by exchanging each individual only with a neighboring one, in the sense that their distance in the search space is small.*

For the problems we consider, a small distance in the search space (realized by a small step size) implies a small distance in the objective space. Analogously, an EMOA is *locally* k*-greedy* if its variations performs small changes as defined above. Due to the $(\mu + 1)$-selection of the SMS-EMOA we focus on 1-greediness.

Test Functions

For the experimental investigation, we compiled the following set of academic test functions. The functions T1-T4 have either a convex or a concave Pareto front. Note that T1 and T4 describe the same Pareto front. Test function T5 changes its curvature from concave to convex, while still being connected and continuously

differentiable. T6 is the continuous conversion of the counter example from Zitzler et al. (2010) described in detail in Sec. 3.3.2. It is multi-modal with respect to the hypervolume of the population. The search space is restricted to one decision variable x as $x \in [0, 1]$ for T1-T5 and T7, and $x \in [1, 7]$ for T6. For all functions but T7 the domain is restricted such that all points are Pareto-optimal. The Pareto front of T7 consist of five convex parts and is chosen as an example of a highly disconnected one. For the experiments on T1-T5 and T7, the reference point applied in the selection of the EMOA is fixed to $\mathbf{r} = (2, 2)^\top$, for T6 it is $\mathbf{r} = (10, 7)^\top$.

T1: $f_1(x) = x^2, \quad f_2 = (1 - x)^2$ — Schaffer, Schaffer (1985), convex

T2: $f_1(x) = x, \quad f_2 = 1 - x$ — DTLZ1, Deb et al. (2002b), convex

T3: $f_1(x) = \sin((\pi/2)x), \quad f_2 = \cos((\pi/2)x)$ — DTLZ2, Deb et al. (2002b), 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, Zitzler et al. (2000)

3.3.2 Understanding a Counter Example

Discrete Counter Example

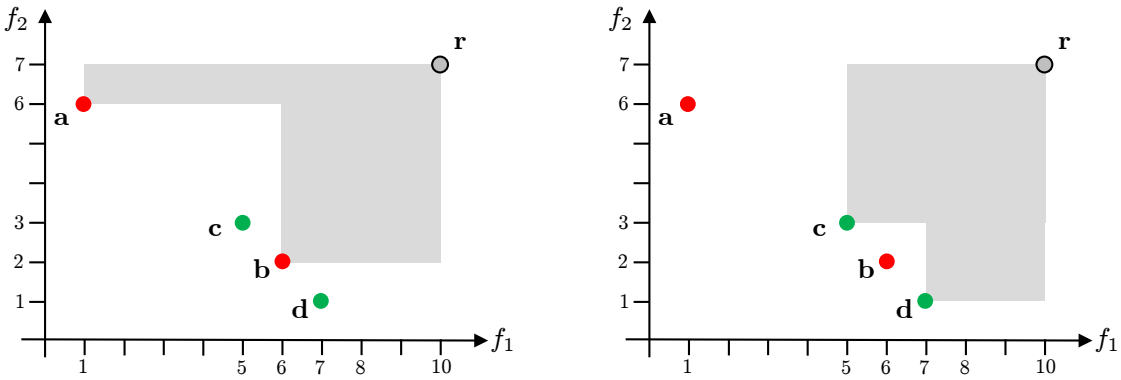


Fig. 3.6: Counter example for a non-decreasing (2+1)-EMOA with hypervolume-based selection. Points **c** and **d** are optimal but a population initialized with **a** and **b** cannot escape this local optimum.

Zitzler et al. (2010, Cor. 4.6) proved that, in general, a 1-greedy EMOA is not able to obtain the set which covers the maximal dominated hypervolume. They showed

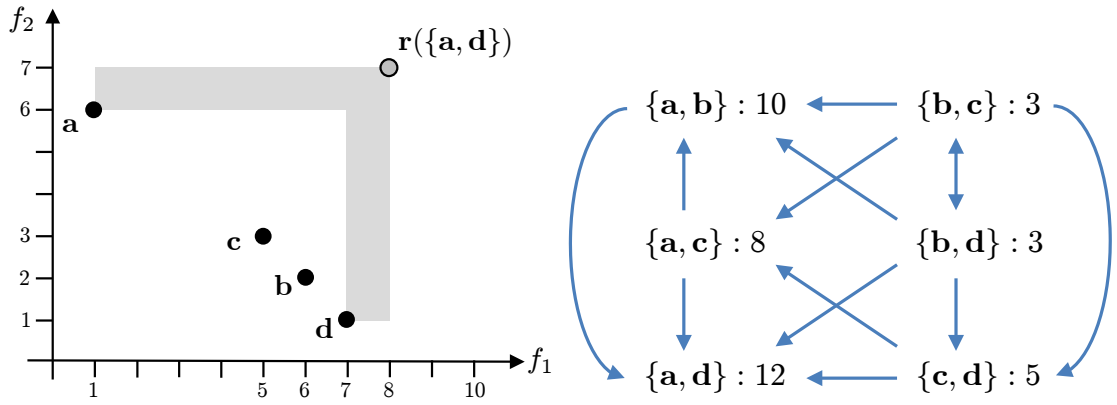


Fig. 3.7: The counter example does not hold for the adaptive reference point. Left: The maximal value is attained by $H(\{\mathbf{a},\mathbf{d}\},\mathbf{r}(\{\mathbf{a},\mathbf{d}\}))$. Right: Transition graph of the sets of size 2 towards the global optimum $\{\mathbf{a},\mathbf{d}\}$. The arrows indicate the exchange of 1 point.

this by a counter example in a two-dimensional objective space with a Pareto front consisting of the four points $\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}$ and a fixed reference point $\mathbf{r}$ as reproduced in Fig. 3.6. The algorithm shall optimize the distribution of a population of two individuals, so using a (2+1)-selection. When the population is initialized with the two points $\mathbf{a}$ and $\mathbf{b}$ (with $H(\{\mathbf{a},\mathbf{b}\},\mathbf{r}) = 25$), the global optimum formed by the points $\mathbf{c}$ and $\mathbf{d}$ (with $H(\{\mathbf{c},\mathbf{d}\},\mathbf{r}) = 26$) is unreachable for a 1-greedy EMOA. Any combination of either $\mathbf{a}$ or $\mathbf{b}$ with a different point (sets $\{\mathbf{a},\mathbf{c}\},\{\mathbf{a},\mathbf{d}\},\{\mathbf{b},\mathbf{c}\},\{\mathbf{b},\mathbf{d}\}$) leads to a worse hypervolume value and is therefore not accepted. Thus, the set $\{\mathbf{a},\mathbf{b}\}$ is a local maximum of the population's hypervolume. 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 $\mathbf{a}$ is positioned close enough to the reference point to avoid an optimal distribution which includes this point.

The counter example does not apply when selecting according the hypervolume using the adaptive reference point, since the hypervolume optimization becomes dynamic. The highest hypervolume is then achieved by $H(\{\mathbf{a},\mathbf{d}\},\mathbf{r}) = 12$ (Fig. 3.7, left). There is no local optimum and the optimal set is reachable from any starting configuration as illustrated in Fig. 3.7, right. The arrows indicate transitions from one population to a non-inferior successive set by exchanging one individual.

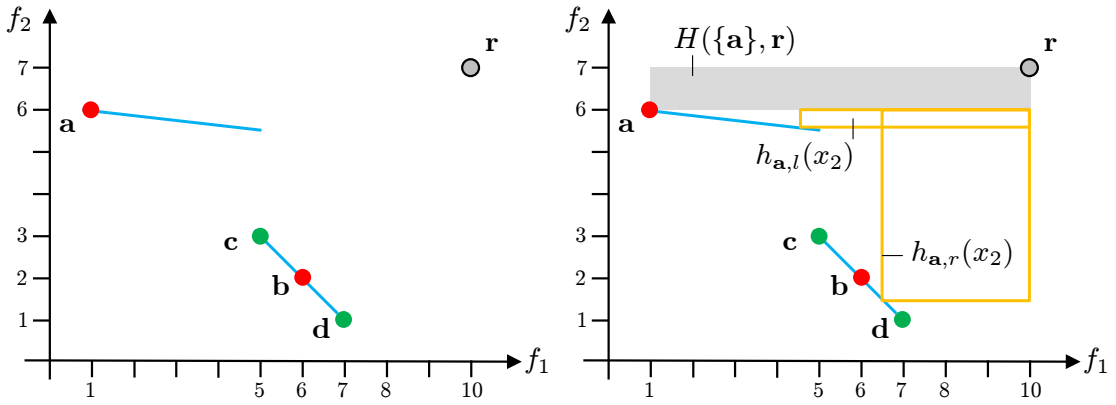


Fig. 3.8: Left: Conversion of the discrete pathological example for 1-greedy EMOA to the continuous space with a disconnected Pareto front of two linear segments (T6). Still the points **c** and **d** are optimal, whereas **a** and **b** form a local optimum. Right: Dissection of the hypervolume for one point fixed at **a** and one moving on either the left ($h_{\mathbf{a},l}$) or the right ($h_{\mathbf{a},r}$) line (case (1) in the proof of Th. 3.38).

Counter Example Transformed to Continuous Space

The discrete example described above 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. 3.8, defined as function T6 (Sec. 3.3.1). 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 in the objective space, whose parameters are the two first (or x-) coordinates of the two search points on the Pareto front. A population is then denoted as $\mathbf{x} = (x_1, x_2)$, and e.g. the global maximum as $\mathbf{x} = (5, 7)$ or $\mathbf{x} = (7, 5)$, and the local maximum as $\mathbf{x} = (1, 6)$ or $\mathbf{x} = (6, 1)$.

T6 is not 1-greedy solvable

Theorem 3.38 *The function T6 is not 1-greedy solvable for the 1-greedy (2+1)-EMOA with selection due to non-decreasing hypervolume w. r. t. the fixed reference point* $\mathbf{r} = (10, 7)$. *(Beume et al. (2009b)*).*

Proof. We prove that the set $\{\mathbf{a}, \mathbf{b}\}$ with hypervolume $H(\{\mathbf{a}, \mathbf{b}\}, \mathbf{r}) = 25$ is indeed a local optimum for the considered selection scheme so that a population $\mathbf{x} = (1, 6)$ cannot reach the global optimum $\mathbf{x} = (5, 7)$ of hypervolume $H(\{\mathbf{c}, \mathbf{d}\}, \mathbf{r}) = 26$. Exchanging x_1 and x_2 yields the same result. Let $\mathbf{x} = (1, 6)$ denote the population,

then we consider the two cases that either $x_1 = 1$ remains fixed and x_2 changes, or $x_2 = 6$ is fixed and x_1 changes.
(1) Let $x_1 = 1$. The hypervolume of the population can be calculated as the hypervolume $H(\{\mathbf{a}\}, \mathbf{r}) = 9$ dominated by $\mathbf{a}$ plus the hypervolume contribution of the point at x_2. The point at x_2 can be positioned either on the left or right line (Fig. 3.8, right). For x_2 on the left line segment, the function

$$h_{\mathbf{a},l}(x_2) = (10 - x_2)\left(6 - \left(-\frac{x_2}{8} + \frac{49}{8}\right)\right) = -\frac{x_2^2}{8} + \frac{11x_2}{8} - \frac{10}{8}$$

gives its hypervolume contribution. Standard calculus obtains an optimum of the negative parabola at $x_2 = \frac{11}{2}$, which is beyond the domain of the line. With x_2 at the right boundary of the line, the total hypervolume is still below 13, so smaller than 25. For x_2 on the right line, the hypervolume contribution is

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

The maximum is at $x_2 = 6$, so there is no better second point than $\mathbf{b}$.
(2) Let $x_2 = 6$. For the point at x_1, we distinguish the three cases that the point is on the left line, or on the right line left or right from $\mathbf{b}$. For x_1 on the left line, the total hypervolume is the hypervolume dominated by $\mathbf{b}$ ($H(\{\mathbf{b}\}, \mathbf{r})$=20) and the contribution of the point at x_1 given by

$$h_{\mathbf{b},l}(x_1) = (6 - x_1)\left(7 - \left(-\frac{x_1}{8} + \frac{49}{8}\right)\right) = -\frac{x_1^2}{8} - \frac{x_1}{8} + \frac{42}{8}.$$

with a maximum at $x_1 = -\frac{1}{2}$ outside the domain. The best valid point is $\mathbf{a}$ with a total hypervolume of 25. The two cases of x_1 on right line are handled together by a function calculating the hypervolume of two points as their bounding box through the reference point minus a small rectangle with the points as diagonal corners, which is a square due to the right line's slope of -1:

$$h_{\mathbf{b},r}(u, v) = (10 - u)\,(7 - (-v + 8)) - (v - u)^2 = -u^2 - v^2 + uv + u + 10v - 10$$

assuming $u < v$. For $x_1 < 6$, respectively $x_1 > 6$ this leads to

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

with the maximum values at

$$\arg\max_{x_1}(h_{\mathbf{b},r}(x_1, 6)) = \frac{7}{2} \quad \text{and} \quad \arg\max_{x_1}(h_{\mathbf{b},r}(6, x_1)) = 8$$

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

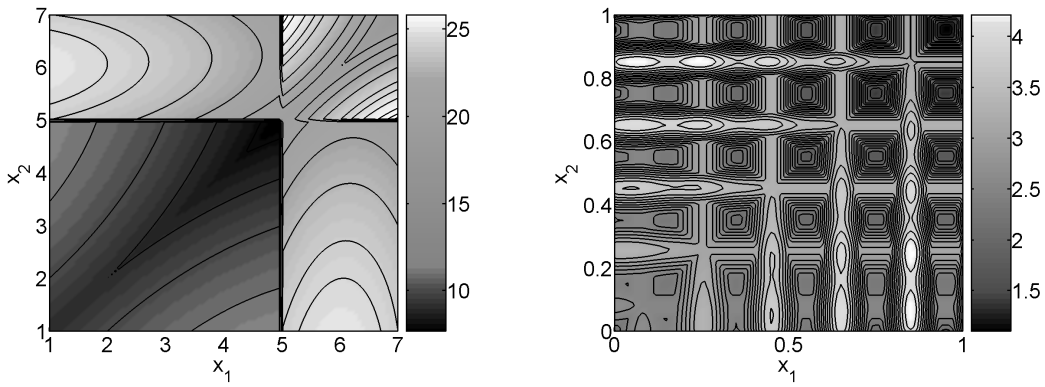


Fig. 3.9: Hypervolume landscapes for a set of two individuals on the disconnected Pareto front of T6 (left) or T7 (right). For T6, there is no improving line-search step from the local optimum at (1,6) to the global optimum at (5,7).

Hypervolume Landscape

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. 3.9, 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. 3.9. The contour lines indicate that a 1-greedy EMOA is not able to leave the local optimum. A μ-greedy scheme would not encounter this principal difficulty as it allows for steps in any direction. However, it also faces the problem of locating a good area, which may be also difficult. The function T6 is studied in the following experiment.

Experiment: How μ-greedy solvable is T6?

Pre-experimental planning We consider a 1-greedy and a μ-greedy single-objective EA moving on the Pareto front only (resembling SMS-EMOA for example). So, the search space is the Pareto set and the EA directly maximize the hypervolume 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 $\mathbf{x} = (5, 7)$, and the boundary treatment hinders coming near to it. Both have been switched off hereafter.

Task We compare the algorithms and expect that the μ-greedy EA performs significantly better than the 1-greedy EA in terms of success rates.

Setup All four algorithms are run 100 times per initial mutation step size (0.1 and 0.5) being the only free parameter, which is only adapted by the CMA-ES. Up to 5000 evaluations are allowed 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).

Results The results are given in Table 3.1 by means of success rates.

Tab. 3.1: Success rates (per 100 repeats) of different EA types for converging against the distribution of two points with globally maximal hypervolume.

Step Size	1-greedy EA	μ-greedy EA	$(1, 5)$-CMA-ES	$(5, 10)$-CMA-ES
0.1	12%	15%	55%	100%
0.5	15%	13%	61%	100%

Observations Table 3.1 documents that the 1-greedy EA indeed fails, but so also does the μ-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 μ-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 μ-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 μ-greedy solvable, μ-greedy EA without additional features like step-size adaptation have roughly the same chance of getting to the global optimum as 1-greedy EA. Note that the same applies to the original discrete example presented by Zitzler et al. (2010), 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 μ-greedy scheme, the continuous one does.

Generalizations of the Counter Example

The provided example of a non 1-greedy function is fragile: Moving the reference point from $\mathbf{r} = (10, 7)^\top$ towards symmetry makes it 1-greedy solvable again. Also note that the whole construction breaks down when increasing the population size. Experimental 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 μ-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. 3.9). However, such a problem will also become increasingly difficult for a μ-greedy algorithm.

Conjecture 3.39 *Continuously defined functions, which are not* 1*-greedy solvable for large population sizes ($\mu \gg 3$) are not generally considerably easier for μ-greedy algorithms.*

One may however pose the question if these non 1-greedy solvable functions have to be defined piecewise. From Fig. 3.9, 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 3.40 *Non 1-greedy solvable, continuously differentiable functions can be constructed for any finite population size.*

3.3.3 Continuously Differentiable Pareto Fronts

This section analyzes the convergence of 1-greedy EMOA to the distribution maximizing the hypervolume for two special cases in the first part. Afterwards, the properties of 1-greedy EMOA on differently shaped Pareto fronts are experimentally studied.

3.3.3.1 Theoretical Analysis

Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ 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

$$f(X^*) = \left\{ \begin{pmatrix} s \\ \gamma(s) \end{pmatrix} : s \in [\,0,1\,] \subset \mathbb{R} \right\}, \tag{3.18}$$

where $\gamma : [\,0,1\,] \to \mathbb{R}$ is twice continuously differentiable. Let $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(\mu)} \in f(X^*)$ be distinct objective vectors on the Pareto front. According to (3.18), we have $\mathbf{y}^{(i)} = (s_i, \gamma(s_i))^\top$ for $i = 1, \ldots, \mu$. The hypervolume of the points $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(\mu)}$ is given by (cf. Eq. (2.6))

$$H(s, \mathbf{r}) = (r_1 - s_1)\,[\,r_2 - \gamma(s_1)\,] + \sum_{i=2}^{\mu} (r_1 - s_i)\,[\,\gamma(s_{i-1}) - \gamma(s_i)\,], \tag{3.19}$$

w. r. t. the reference point $\mathbf{r} = (r_1, r_2)^\top$ with $0 \le s_1 < s_2 < \cdots < s_\mu \le 1 \le r_1$ and $r_2 \ge \gamma(0)$.
Whenever the hypervolume function is concave (cf. Def. 3.4), the 1-greedy selection scheme of the SMS-EMOA with fixed reference point and a population of μ 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 hypervolume 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 negative semidefinite (cf. Def. 3.6). Partial differentiation of (3.19) leads to

$$\begin{aligned}
\frac{\partial H(s, \mathbf{r})}{\partial s_1} &= \gamma(s_1) - r_2 + (s_1 - s_2)\,\gamma'(s_1) \\
\frac{\partial H(s, \mathbf{r})}{\partial s_i} &= \gamma(s_i) - \gamma(s_{i-1}) + (s_i - s_{i+1})\,\gamma'(s_i) \qquad (i = 2, \ldots, \mu - 1) \\
\frac{\partial H(s, \mathbf{r})}{\partial s_\mu} &= \gamma(s_\mu) - \gamma(s_{\mu-1}) + (r_1 - s_\mu)\,\gamma'(s_\mu)
\end{aligned}$$

and finally to

$$\begin{aligned}
\frac{\partial^2 H(s,\mathbf{r})}{\partial s_i \partial s_{i-1}} &= -\gamma'(s_{i-1}) && (i = 2,\ldots,\mu)\\
\frac{\partial^2 H(s,\mathbf{r})}{\partial s_i \partial s_i} &= 2\,\gamma'(s_i) + (s_i - s_{i+1})\,\gamma''(s_i) && (i = 1,\ldots,\mu)\\
\frac{\partial^2 H(s,\mathbf{r})}{\partial s_i \partial s_{i+1}} &= -\gamma'(s_i) && (i = 1,\ldots,\mu-1)
\end{aligned}$$

with $s_{\mu+1} := r_1$. Other second partial derivatives are zero. Thus, the Hessian matrix $\nabla^2 H(s,\mathbf{r})$ of the hypervolume as given in (3.19) is a tridiagonal matrix.

Linear Pareto Front

Suppose that $\gamma(s) = m\,s + b$ 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 $2\,m < 0$ 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 negative definite if all diagonal entries are negative. It is easily seen that these conditions are fulfilled. As a result, we have proven:

Theorem 3.41 *If the Pareto front of a bi-objective minimization problem is linear, then the [...] hypervolume of μ distinct points on the Pareto front is a strictly concave function. (Beume et al. (2009b)*).* □

From this result, we can deduce that it is sufficient to move a single point at a time for reaching the maximal hypervolume 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 negative 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 $\mathbf{r} = (1,1)^\top$.

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

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

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 $s = (\frac{1}{100}, \frac{1}{4}, \frac{4}{9})^\top$, 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 3.42 *If the Pareto front of a bi-objective minimization problem is convex, then the [...] hypervolume of μ distinct points on the Pareto front is not a concave function in general. (Beume et al. (2009b)*).* □

However, this result does imply neither that there are no convex fronts with concave hypervolume nor that the 1-greedy selection scheme of the SMS-EMOA gets necessarily stuck on convex fronts.

3.3.3.2 Experimental Analysis on Connected Pareto fronts

It is experimentally analyzed whether the 1-greedy SMS-EMOA can robustly obtain the hypervolume-optimal distribution of points for the approximation of piecewise continuous Pareto fronts with different curvature (convex to concave). We consider the local 1-greediness as defined in Section 3.3.1. Recall that a local 1-greedy solvable problem is also 1-greedy solvable. We perform a comprehensive study on the set of simple test functions T1-T5 (Sec. 3.3.1), whereas we restrict our analyses to small populations, due to our results in Section 3.3.2.

Pre-experimental planning To gain a deeper understanding of the hypervolume progress, we first analyze hypervolume contributions and local 1-greedy steps.
Hypervolume contributions: 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. This assumption is wrong, as demonstrated for T4 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 $\alpha = 1$. The analytically determined optimal positions of the points w. r. t. the reference point $\mathbf{r} = (1, 1)^\top$ are shown in Fig. 3.10 (left) and the corresponding hypervolume contributions on the right side. The graphs are symmetric to the bisecting line. Accordingly, the distributions are symmetric to the

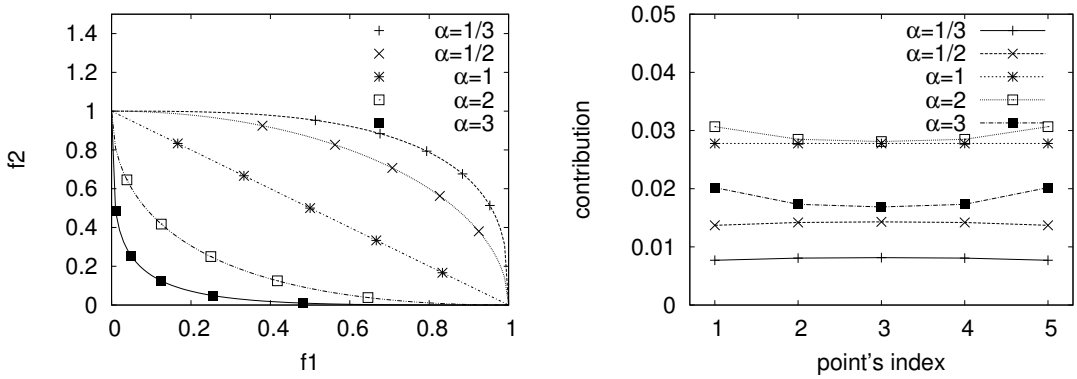


Fig. 3.10: Optimal positions of points on Pareto fronts of T4 with different curvature (left) and corresponding hypervolume contributions (right), w. r. t. the reference point $\mathbf{r} = (1, 1)^\top$. The distributions are symmetric and the contributions are not all equal.

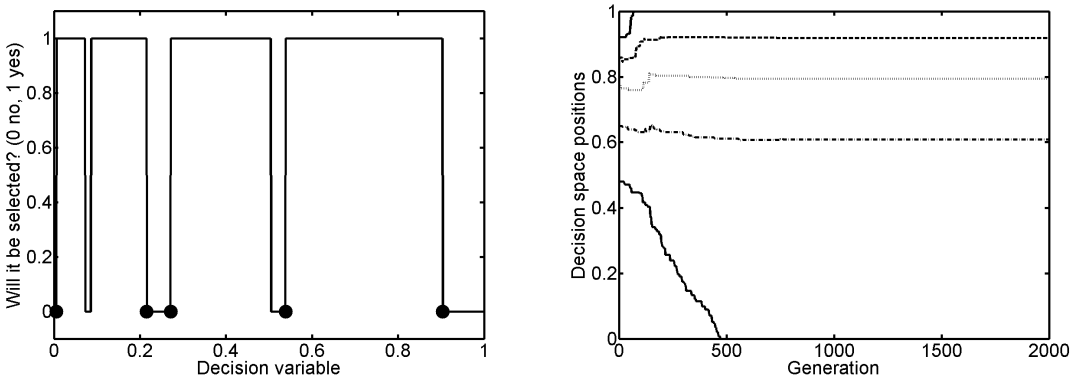


Fig. 3.11: Left: Areas of acceptance/rejection of points due to their hypervolume contribution, relative to the decision space variables of a population on the Pareto front of T1. Improving areas are adjacent to the points.
Right: Progress of moving the points from their starting positions to their optimal positions by a local search $(5 + 1)$-SMS-EMOA on T4 ($\alpha = 1/3$).

middle point on the bisecting line, and so are the hypervolume contributions. The contributions are displayed sorted according to the first objective. For $\alpha < 3$, the values tend to grow with increasing α. On concave Pareto fronts, the middle point has the highest contribution and values decrease monotonic towards the boundaries, whereas it is the other way round for convex Pareto fronts, so that the knee point has the lowest contribution. Only on the linear front, all contributions are equal, and the points are equally spaced as proved in Lemma 2.7.

Local 1-greedyness: Fig. 3.11 (left) shows the areas of acceptance and rejection of points, exemplarily for a randomly initialized population on T1 (analogous results have been observed for T2-T5). Note that the acceptance regions are directly

adjacent to the current points, so that a point can be improved by a local movement. To further investigate the effect of single local refinements, a local search SMS-EMOA is applied, with variation only by Gaussian mutation with the small step size $\sigma = 0.01$. Fig. 3.11 (right) displays the progress of the decision space variables of this local search $(5 + 1)$-SMS-EMOA on T4 with $\alpha = 1/3$ w. r. t. the fixed reference point $\mathbf{r}' = (2, 2)^\top$. As starting positions, the optimal distribution for the closer reference point $\mathbf{r} = (1, 1)^\top$ is used. The algorithm is able to guide the solutions from the old to the new optimal positions. [...] For the new reference point $\mathbf{r}'$ the optimal points move closer to the boundaries of the Pareto front as there is more hypervolume to gain due to the farther reference point. The other points follow these extremal solutions to cover the resulting distance.
The following experiment shall support the arising assumption that even a local search-based $(\mu+1)$ SMS-EMOA converges towards the optimal population.

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

Setup For each test function T1-T5 and population sizes $\mu \in \{1, \ldots, 6\}$, approximations for the hypervolume-optimal distribution are globally calculated by the MATLAB implementation of the (5,10)-CMA-ES by Hansen and Ostermeier (2001), where no limit on the function evaluations, but a lower limit on σ_i $(i = 1, \ldots, \mu)$ of 10^{-12} is specified. We trust in the quality of the results due to the experiences in Sec. 3.3.2. 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.

Results/Observations The local search SMS-EMOA detects the optimal distribution in all runs for the convex and concave test functions 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.

3.3.4 Conclusions

We investigated the process of the SMS-EMOA when optimizing the distribution of points on the Pareto front in order to maximize the hypervolume. Our posed question was, whether the steady-state (or 1-greedy) selection of exchanging only one individual per generation is able to reach a global optimum of the population's hypervolume which is defined w. r. t. a fixed reference point.
To deeper understand the optimization process, we analyzed a counter-example. It is revealed that the local optimum where the EMOA gets stuck is not a singularity but has an actual strong attractor so that also μ-greedy or non-elitist algorithms fail to detect the global optimum. This result is a step forward from theoretical possibilities to more practical probabilities. The constructed counter examples appear artificial as the Pareto fronts are disconnected and one may think that the non-optimality is achieved by forbidding points at the right positions. However, connecting the parts shall keep the negative properties while making the function continuously differentiable, so we conjecture that counter examples also exist for this class of problems where an arbitrary movement of points is possible.
We proved that the hypervolume is a concave function in case of a linear Pareto front, thus the function does not have a non-global local optimum and the continuous improvement of the hypervolume results in the global optimum. Thorough experiments showed that the optimal hypervolume is reliably detected by the 1-greedy SMS-EMOA on the chosen examples of continuously differentiable functions with connected Pareto fronts. These results strongly suggests that these functions are 1-greedy solvable, i.e., there is no configuration from where the optimum cannot be reached. Failures have only been detected for a (1+1)-SMS-EMOA, so we think the risk of not being 1-greedy solvable decreases with increasing population size of the EMOA.
Recall that we considered a model scenario with a fixed reference point, so the results do not directly apply to the SMS-EMOA with the normally used adaptive reference point. For a large population size, the scenarios shall behave similar as the reference point is expected to change only slightly after the points are spread along the whole Pareto front. For these cases, we experimentally did not observe any failure on the considered functions. For small population sizes, we witnessed some failures of getting stuck in a local optimum. These problems shall not occur with an adaptive reference point, where the optimization problem is dynamic due to the changing reference point, so that the population is unlikely to get stuck. This again advocates the concept of the adaptive reference point.
This topic is still under examination as the recent work by Bringmann and Friedrich (2011) shows. Our conjectures are still neither proved nor disproved.

Open problems are to identify other function classes where the 1-greedy concept suffices or if not—to give the lowest k so that a k-greedy selection succeeds. Even more desirable than analyzing the general possibility of solvability (if a path to the optimum exists from each configuration) is the consideration of the probabilistic solvability, i.e., analyzing how likely it is to reach the global optimum. Our experimental studies are a first step, whereas more insights are worthwhile to generalize our results.

3.4 Performance on Many-Objective Problems

First studies of the SMS-EMOA demonstrated its performance on two- and three-objective problems, see Emmerich et al. (2005)*, Naujoks et al. (2005a,b)*, Beume et al. (2007)* and Section 3.1; and the preceding sections analyzed certain subprocesses of the algorithm. Here, the quality of the SMS-EMOA's finally generated approximations of the Pareto front shall be analyzed, when optimizing problems of more than three objectives—termed *many-objective* by Farina and Amato (2002). Due to the lack of theoretical tools for performance analysis of EMOA on complex problems, we study the performance through experimental analyses as a comparative benchmark including other state-of-the-art algorithms. It is revealed that also in this scenario, the SMS-EMOA shows outstanding performance. The results are taken from Wagner et al. (2007a)* and so are the literal citations with the notation adapted to this document, except for a paragraph in Sec. 3.4.1 taken from Wagner et al. (2006)*.

At the time of this study, it was common practice to test algorithms only for the two- and three-dimensional cases. Thus it was unknown how the optimizers perform on higher dimensional problems, so that the field of many-objective optimization achieved some surprising results. Our study, among others, contributed to the understanding that performing well in the popular 2- and 3-objective cases does not imply the same for more objectives.

The many-objective scenario constitutes a special challenge due to the high-dimensional objective space. As described in Section 1.2, the fraction of the space that is comparable to a certain point decreases exponentially as 2^{-d+1} with increasing dimension d (cf. Farina and Amato (2002)). So, the typical case is that points are incomparable and the Pareto dominance gives no information how to sort and select them. Therefore, a sound additional selection criterion, like the hypervolume contribution, is required to resolve these cases. As we see the main challenge in the selection, we setup the benchmark such that all operators apart from the selection are equal (or similar as far as possible) so that the selection operators shall decide on success or failure of the algorithms.

For the benchmark we consider algorithms with three classes of selection concepts: based on the Pareto dominance (NSGA-II, SPEA2, ϵ-MOEA), aggregation techniques (MSOPS, RSO by Hughes (2003, 2005)), or on quality indicators (IBEA, SMS-EMOA). The main interest of the study are the magnitudes of the results, i.e., whether the algorithms are able to approach the Pareto front or not. It is revealed that the well-established EMOA NSGA-II and SPEA2 rapidly degradate with increasing number of objectives, and we analyze why. Modern EMOA like ϵ-MOEA, MSOPS, IBEA and SMS-EMOA cope well with the high-dimensional objectives space, and the SMS-EMOA seems to perform best.

The following Section 3.4.1 gives an overview on related work and specifies our benchmark setup. Section 3.4.2 presents the results of Pareto-based EMOA and

an additional study to pursuing the cause of their failure. Section 3.4.3 describes the aggregation methods including a new methods of generating weight vectors for MSOPS, next to their performance results. The indicator-based EMOA are studied in Section 3.4.4. Finally, Section 3.4.5 gives concluding remarks and discusses open problems of many-objective optimization.

3.4.1 Preliminaries

Related Work

Farina and Amato (2002) introduced the term many-objective optimization and have been among the pioneers to sensitize the research community to this challenge. Few previous studies on many-objective optimization by Purshouse and Fleming (2003) and Hughes (2005) focus to demonstrate the bad performance of NSGA-II by Deb et al. (2002a). Hughes observed a simple single-objective restart strategy outperforming NSGA-II on a six-objective function in a two-dimensional decision space. Upon this, he implied a generalization to all Pareto-based techniques. Contrarily, we demonstrate that the modern ϵ-dominance-based ϵ-MOEA performs well.
Ishibuchi et al. (2008) give a brief overview of challenges and useful techniques in many-objective optimization, where scalarization approaches like MSOPS are considered as promising. The study of our modification of NSGA-II presented in Section 3.4.2 is continued and confirmed. Adra and Fleming (2011) propose a new diversity mechanism and demonstrate that NSGA-II using the measure is significantly improved on many-objective problems.
Recently, Bringmann et al. (2011) presented an EMOA guided by a formal notion of approximation, namely the additive approximation of the Pareto front. Its performance is demonstrated on DTLZ functions of up to 20 objectives.
Our benchmark is setup such that it extends existing studies on low-dimensional problems, i.e., it uses the same parameterization as in Deb et al. (2003), and uses well-established test problems.

Benchmark Settings

All algorithms, except RSO, have been implemented within the PISA framework (cf. Bleuler et al. (2003)) since an integrative framework simplifies comparisons. The same variation operators are used with exactly the same parameterization, which is chosen according to the studies of Deb et al. (2003). Simulated binary crossover (SBX) and polynomial mutation (PM) as described by Deb (2001) (cf. Sec. 3.1) are applied with mutation probability $p_m = 1/n$ per decision variable and recombination probability $p_c = 1$ per individual. The distribution indices $\eta_c = 15$ and $\eta_m = 20$ are used. If not otherwise stated, a $(\mu + \mu)$ strategy and a binary tournament for mating selection are applied. A number of $30,000$ function

evaluations is accomplished and the population size $\mu = 100$ is chosen. For each EMOA, besides SMS-EMOA, on each test function, 20 runs are performed. Due to the exponential runtime and the small standard deviation in the observed runs, SMS-EMOA is only repeated 5 times.

Test Functions

To benchmark the performance of the considered EMOA, the functions DTLZ1 and DTLZ2 of the DTLZ test function family by Deb et al. (2002b) are invoked (cf. Sec. 1.5). These functions are scalable in the number of objectives and thus allow for a many-objective study. The decision vector is divided into two subvectors. The first one of length $d - 1$ contains the parameters defining the position on the given surface while the second of length ν specifies the distance to the Pareto front. This results in dimension $d + \nu - 1$ of the decision space. According to Deb et al. (2002b), $\nu = 5$ is used in DTLZ1 and $\nu = 10$ is used in DTLZ2 respectively.

The Pareto front of DTLZ1 is a linear hyperplane. DTLZ2 features a Pareto front that corresponds to the positive part of the unit hypersphere. [...] Here, the interaction between objectives is nonlinear. The domain of all decision variables is $[0, 1]$. Due to different scaling constants in the distance function, the co-domain of objective values for DTLZ1 is $[0, 1 + 225\nu]$ and $[0, 1 + 0.25\nu]$ for DTLZ2, respectively. The Pareto set of both test functions corresponds to $x_d = \ldots = x_n = 0.5$ with arbitrary values for $x_1, \ldots, x_{d-1}$.

Performance Assessment

The quality of the final Pareto front approximations is evaluated via the hypervolume and a convergence measure by Deb et al. (2003). The reference points $\mathbf{r} = (0.7, \ldots, 0.7)$ for DTLZ1 and $\mathbf{r} = (1.1, \ldots, 1.1)$ for DTLZ2 were used in previous studies e.g. Deb et al. (2003), Naujoks et al. (2005a)* and are close to the Pareto front in order to emphasize on the distribution of optimal solutions. Points that do not dominate the reference point are discarded for the hypervolume calculation. The indicator values are normalized by calculating the fraction of the analytical optimal value. Note that exactly 100% are unreachable with a finite number of points.

The convergence measure describes the average distance of the approximation to the Pareto front in the objective space. In contrast to the study by Deb et al. (2003), the Euclidean distance to the nearest optimal solution is determined analytically without using a reference set. This is possible due to the special structure of the employed Pareto fronts.

The distance is calculated as follows (description taken literally from Wagner et al. (2006)*). In analytic geometry, the distance of a point to a hyperplane can be

calculated using the Hesse normal form. [...] The DTLZ1 hyperplane possesses the normal vector $(z_1, \ldots, z_d)^\top = (1, \ldots, 1)^\top$ resulting in the distance

$$d_{DTLZ1}(\mathbf{x}) = |f_1(\mathbf{x}) + \ldots + f_d(\mathbf{x}) - 0.5| \tag{3.20}$$

of a solution $\mathbf{x}$ to the Pareto front.

On DTLZ2, the shortest way to the Pareto front is the line between the position vector of the solution and the origin of the objective coordinate system. Since the distance of the origin to each point of the Pareto front is 1, the distance of a solution $\mathbf{x}$ to the Pareto front constitutes

$$d_{DTLZ2}(\mathbf{x}) = |f(\mathbf{x})| - 1. \tag{3.21}$$

3.4.2 Pareto-based EMOA

As *Pareto-based* EMOA, we classify EMOA with selection criteria that are mainly based on the qualitative information of Pareto-dominance, Pareto-based ranking, or counting.

We assign NSGA-II to this category due to its non-dominated sorting and consider it because it has clearly been the most frequently applied EMOA at the time of the study and still is the most popular one. SPEA2 belongs to this class because of the fitness assignment based on counting of dominated and dominating individuals. The ϵ-MOEA is a representative of the more modern ϵ-dominance concept. The algorithms are described in Section 3.1, so that here only their parameterization is detailed. All EMOA are executed in the PISA framework.

For the NSGA-II all parameters have already been mentioned in Section 3.4.1. For the k-nearest neighbor method of SPEA2, in PISA $k = 1$ is chosen. The ϵ-MOEA has been proposed as steady-state algorithm, whereas we seek to parameterize it similar to the other algorithms using the $(\mu + \mu)$ selection. As a compromise, we choose a $(2 + 2)$-selection to have a small offspring population size and $\mu = \lambda$. For the ϵ-MOEA, the resulting set is not the population but the archive, so ϵ is chosen such that its archive E finally contains about 100 solutions in order to assure comparability with the sets of the other EMOA. The ϵ values are determined in a pre-experimental study, where the ϵ-MOEA showed an extremely high sensitivity to this parameter, which not only shall distinguish the Pareto front appropriately but also the space on the way to it. All vector components of ϵ are chosen to be equal. For DTLZ1, it is $\epsilon = (\epsilon_i, \epsilon_i, \epsilon_i)$ with $\epsilon_i = 0.03$ in case of 3 dimensions, $\epsilon_i = 0.047$ in 4 dimensions, $\epsilon_i = 0.057$ in 5 dimensions, and $\epsilon_i = 0.066$ in 6 dimensions. For DTLZ2, it is $\epsilon_i = 0.058$ in case of 3 dimensions, $\epsilon_i = 0.125$ in 4 dimensions, $\epsilon_i = 0.18$ in 5 dimensions, and $\epsilon_i = 0.232$ in 6 dimensions.

Experimental Results

The quality indicator values of the final sets are regarding the convergence given in Table 3.2 and regarding the relative hypervolume in Table 3.3.

Tab. 3.2: Convergence measure of the Pareto dominance based EMOA.

		DTLZ1			DTLZ2		
obj.	algorithm	mean	std.dev	median	mean	std.dev	median
3	ε-MOEA	0.00614	0.00413	0.00484	0.00102	0.00022	0.00105
	NSGA-II	0.06333	0.15581	0.01002	0.01049	0.00162	0.01027
	SPEA2	0.06783	0.16435	0.00792	0.00801	0.00112	0.00806
4	ε-MOEA	0.15990	0.34073	0.01990	0.00129	0.00024	0.00126
	NSGA-II	1.70260	1.95260	0.69515	0.08522	0.02580	0.08060
	SPEA2	3.47990	4.78910	1.66910	0.08164	0.01676	0.08901
5	ε-MOEA	0.22348	0.41685	0.01941	0.02681	0.00120	0.02670
	NSGA-II	300.416	37.2461	317.506	1.06780	0.14504	1.07770
	SPEA2	358.818	25.0853	366.236	1.30970	0.15758	1.27760
6	ε-MOEA	0.97014	1.39920	0.27217	0.00272	0.00067	0.00266
	NSGA-II	393.674	17.6076	388.689	2.15610	0.09584	2.16910
	SPEA2	482.742	13.6757	479.577	2.32000	0.09617	2.36070

Tab. 3.3: Relative hypervolume of the Pareto dominance based EMOA.

		DTLZ1, $\mathbf{r} = (0.7, \ldots, 0.7)$			DTLZ2, $\mathbf{r} = (1.1, \ldots, 1.1)$		
obj.	algorithm	mean	std.dev	median	mean	std.dev	median
3	ε-MOEA	0.94560	0.01005	0.94662	0.92858	0.00118	0.92836
	NSGA-II	0.94333	0.11423	0.96923	0.86913	0.00803	0.86918
	SPEA2	0.98010	0.00152	0.98068	0.90760	0.00350	0.90782
4	ε-MOEA	0.85493	0.18655	0.92697	0.87722	0.00186	0.87766
	NSGA-II	0.45730	0.40600	0.46204	0.71644	0.01971	0.71733
	SPEA2	0.62316	0.34319	0.72224	0.78461	0.01258	0.78202
5	ε-MOEA	0.82261	0.16668	0.86933	0.83847	0.00308	0.83809
	NSGA-II	0	0	0	0.11570	0.06842	0.11734
	SPEA2	0	0	0	0.12528	0.06942	0.12864
6	ε-MOEA	0.64563	0.38344	0.81552	0.85332	0.01434	0.85497
	NSGA-II	0	0	0	0	0	0
	SPEA2	0	0	0	0	0	0

NSGA-II and SPEA2 rapidly decrease in quality with increasing dimension of the objective space. If more than four objectives are considered, these algorithms do

not converge to the Pareto front as indicated by the high distance values. With dimension greater than four, no relative hypervolume is measured because no point dominating the reference point is achieved. The results of the ϵ-MOEA are in magnitudes better than the others, and the algorithm especially achieves a positive relative hypervolume in each dimension. So, the ϵ-dominance concept seems to work also in high dimensions, provided that the size of the ϵ-boxes is chosen appropriately. However, the high-dimensional results are only for DTLZ2 close to the best considered algorithms.
All algorithms perform reasonably in case of 3 dimensions, whereas it seems that especially the concepts of NSGA-II and SPEA2 are not qualified for higher dimensional problems. Since the remaining operators are chosen equally, we attribute this failure to the selection operators. The following additional study shall reveal a deeper insight of the malfunction.

Additional Convergence Study

We increase the budget of function evaluations to $1,000,000$ in order to exhibit whether a convergence occurs maybe later. The progress is shown for the 6-objective DTLZ1 in Fig. 3.12, where the graph of ϵ-MOEA is hardly visible due to its convergence. Contrarily, NSGA-II and SPEA2 increase the distance to the Pareto front during the first generations. We assume that this behavior stems from the selection based on diversity measures which favor higher distances between solutions. As the boundary points (regarding the measure) are always kept, the population spreads without a tendency to the Pareto front.

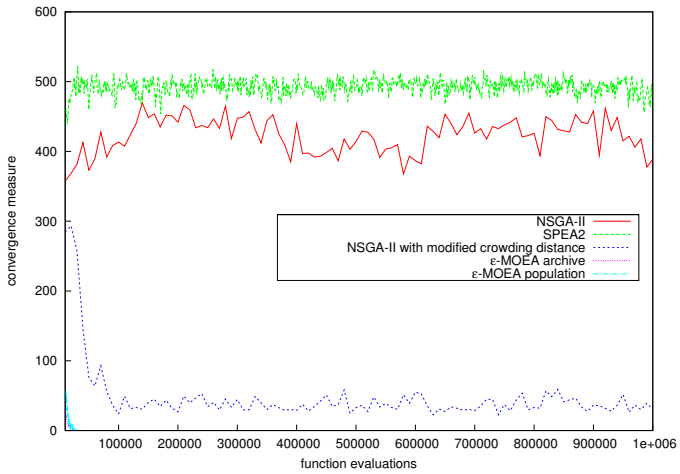


Fig. 3.12: Convergence measure during the optimization run performing the median result on six-objective DTLZ1.

To confirm this assumptions and improve NSGA-II, a slight modification of the crowding distance is studied. Originally, an individual without a neighbor regarding one dimension of the objective space is assigned an infinite crowding distance. Instead of that, a value of zero is used, causing that non-dominated solutions with extremal values are rejected. Although this variant is not able to converge to the Pareto front, an improvement of the average distance within the first $100,000$ function evaluations is obvious (Fig. 3.12). Then, most of the decision variables have reached their optimal value. Only one or two of them remain in a local optimum. This experiment shows that a diversity measure with emphasis on a spread of the population can misguide the EMOA to deterioration and the loss of promising non-dominated solutions. Ishibuchi et al. (2008) confirm that our modification improves the convergence while it decreases the diversity. We consider both as an advantage since diversity over the objective space far away from the Pareto front is not to be aspired, but the aim is diversity along the Pareto front.

3.4.3 Aggregation-based EMOA

Basic aggregation methods are single-objective optimizers, which multiply the objective values with weights and accumulate them to a scalar value. The EMOA considered here, enhance aggregation concepts in order to produce a set of solutions. In contrast to the other EMOA considered, aggregation-based approaches require the a priori definition of relations between objective functions. This results in a certain focus during the optimization. We consider several variants of MSOPS and a simple restart strategy called RSO, both developed by Hughes (2003, 2005).

MSOPS

Multiple Single Objective Pareto Sampling (MSOPS) does not feature Pareto methods, but handles all objectives in parallel. The decision maker has to choose q vectors of weights for every objective function to enable an aggregation. Hughes (2003) recommends weighted min-max (termed MSOPS 1 in the following) and a combination of this approach with *Vector-Angle-Distance-Scaling (VADS)* called *dual optimization* (MSOPS 2). Depending on the aggregation strategy, one receives a set of q or $2q$ aggregated scores per solution. The scores are held in a score matrix S, where each row belongs to a solution and each column represents an aggregated score. Each column of the matrix S is ranked, giving the best performing population member rank one. The rank values are stored in a matrix R. Each row of R is sorted ascending, resulting in a lexicographical order of the individuals. The runtime is in $O(\mu q d)$ for the computation of the aggregated scores, and in $O(\mu q \log q)$ and $O(q\mu \log \mu)$ respectively to perform the sort. Thus, the runtime of MSOPS is $O(\mu q(d + \log q + \log \mu)$ per generation.
Obviously, the choice of weight vectors determines the distribution properties of MSOPS. Each weight vector $\mathbf{w} = (w_1, \ldots, w_d)$ corresponds to a direction, given

analytically by a target vector starting in the origin. The aim of the aggregation methods is to reach the point on the corresponding direction vector which is as close as possible to the origin. To this end, weighted min-max focuses on the distance to the origin, while *VADS* favors solutions whose position vector has a small intersecting angle with the target vector.
In this study, the optimization shall not have a special focus, but an approximation of the whole Pareto front is desired and the weight vectors have to be chosen appropriately. In the benchmarking in Hughes (2005), '*50 target vectors spread uniformly across the search space*' are used. The target vectors [...] are created by calculating an initial number of steps $s = \lfloor \sqrt[d]{q} \rfloor$ and constructing each possible vector containing multiples of $1/s$ between 0 and 1. Afterward, these target vectors are normalized and doubles are removed. If the number of targets is lower than desired, s is incremented and the procedure is repeated. At the end, a next neighbor technique is used to prune the set of target vectors to the desired size. Because the PISA implementation of MSOPS uses weight vectors, a transformation of the target vectors into weights is necessary. We recommend—deviant from Hughes (2005)—the following procedure for transformation, which can also be used to transform a set of utopia or reference points into weights and avoids numerically unstable calculations in many cases.
From the aggregation methods can be referred that a weight vector for a specified target fulfills the following $d-1$ conditions, with the target vector $\mathbf{t} = (t_1, \ldots, t)$:

$$w_1 \cdot t_1 = w_2 \cdot t_2, \qquad w_2 \cdot t_2 = w_3 \cdot t_3, \qquad \ldots \qquad w_{d-1} \cdot t_{d-1} = w_d \cdot t_d$$

The normalizing condition $w_1 + \ldots + w_d = 1$ is added in order to obtain a completely defined system of equations. Thus, the components of the corresponding weight vector can be computed as follows:

$$w_i = \frac{\prod_{j \neq i} t_j}{\sum_{k=1}^{d} \prod_{j \neq k} t_j}, \quad i = 1, \ldots, d \tag{3.22}$$

To extremal solutions with value 0 in $d-1$ objectives, a small ε needs to be added to allow the above calculation. Hughes (2003) generally recommends to use a number of target vectors that is lower than the population size. Besides, he states that the number of target vectors has to be increased for more objectives. To cover both needs, three different sets of target vectors are used. The first contains 50 vectors, the second 100 vectors, and the third 200 vectors.
MSOPS-II has been developed by Hughes (2007) after this study. It is a powerful and user-friendly algorithm due to its automatic target vector generation, yet it does not perform clearly better than MSOPS.

RSO

A restart strategy of a conventional single-objective evolutionary optimizer is applied as well and abbreviated *RSO (Repeated Single Objective)* according to Hughes

Tab. 3.4: Convergence measure of aggregation algorithms. The leftmost column gives the number of objectives.

		DTLZ1			DTLZ2		
	algorithm	mean	std.dev	median	mean	std.dev	median
3	MSOPS 1 50	0.00276	0.00235	0.00185	0.00013	0.00014	$9.0{\cdot}10^{-5}$
	MSOPS 1 100	0.00278	0.00241	0.00244	0.00015	0.00010	0.00015
	MSOPS 1 200	0.00234	0.00156	0.00210	0.00080	0.00020	0.00076
	MSOPS 2 50	0.00214	0.00221	0.00161	$9.0{\cdot}10^{-5}$	$5.9{\cdot}10^{-5}$	$8.4{\cdot}10^{-5}$
	MSOPS 2 100	0.00222	0.00172	0.00191	0.00037	0.00013	0.00035
	MSOPS 2 200	0.00128	0.00074	0.00116	0.00168	0.00034	0.00168
	RSO	62.9990	15.2960	59.7140	0.26753	0.04901	0.26776
4	MSOPS 1 50	0.00392	0.00451	0.00269	0.00023	0.00023	0.00012
	MSOPS 1 100	0.00292	0.00252	0.00231	0.00024	0.00039	0.00013
	MSOPS 1 200	0.00365	0.00319	0.00264	0.00072	0.00028	0.00067
	MSOPS 2 50	0.00246	0.00216	0.00182	0.00016	0.00010	0.00012
	MSOPS 2 100	0.00849	0.02369	0.00282	0.00074	0.00024	0.00072
	MSOPS 2 200	0.00439	0.00378	0.00260	0,00203	0.00047	0.00195
	RSO	118.260	33.4420	121.190	0.56473	0.07953	0.57386
5	MSOPS 1 50	0.08016	0.31475	0.00814	0.00059	0.00027	0.00060
	MSOPS 1 100	0.05667	0.23459	0.00337	0.00017	0.00023	$7.1{\cdot}10^{-5}$
	MSOPS 1 200	0.00779	0.00556	0.00651	0.00096	0.00033	0.00092
	MSOPS 2 50	0.13676	0.26271	0.01882	0.00113	0.00038	0.00097
	MSOPS 2 100	0.03308	0.11179	0.00614	0.00138	0.00065	0.00119
	MSOPS 2 200	0.00870	0.01079	0.00535	0,00231	0.00059	0.00233
	RSO	111.960	35.1240	112.140	0.73556	0.15491	0.72211
6	MSOPS 1 50	0.02207	0.06509	0.00604	0.00044	0.00030	0.00044
	MSOPS 1 100	0.00936	0.01579	0.00406	0.00012	$8.7{\cdot}10^{-5}$	$9.7{\cdot}10^{-5}$
	MSOPS 1 200	0.00734	0.00420	0.00712	0.00048	0.00028	0.00039
	MSOPS 2 50	0.27890	0.63926	0.02603	0.00091	0.00058	0.00069
	MSOPS 2 100	0.18106	0.32499	0.02496	0.00190	0.00097	0.00180
	MSOPS 2 200	0.01344	0.01134	0.01026	0.00118	0.00056	0.00116
	RSO	110.910	42.7920	113.600	0.67628	0.13970	0.69903

(2005). Here, a single-objective run is performed for each of the 100 weight vectors. Thus, the number of function evaluations has to be divided among them, resulting in only 300 evaluations per run.

The derandomized mutation operator by Ostermeier et al. (1994) is applied in a $(1, 10)$-evolution strategy. This operator was a first step towards the popular Covariance Matrix Adaptation (CMA) operator by Hansen and Ostermeier (2001),

which is known to produce good results within limited function evaluations. To handle multiple objectives in a single-objective EA, the weighted min-max approach was chosen like in MSOPS. RSO is not part of the PISA framework.

Experimental Results

Tab. 3.5: Relative hypervolume of aggregation algorithms. The leftmost column gives the number of objectives.

		DTLZ1, $\mathbf{r} = (0.7, \ldots, 0.7)$			DTLZ2, $\mathbf{r} = (1.1, \ldots, 1.1)$		
	algorithm	mean	std.dev	median	mean	std.dev	median
3	MSOPS 1 50	0.97142	0.00127	0.97184	0.89663	0.00717	0.89817
	MSOPS 1 100	0.96484	0.00171	0.96537	0.88344	0.00208	0.88341
	MSOPS 1 200	0.96180	0.00955	0.96625	0.88752	0.02681	0.88490
	MSOPS 2 50	0.97278	0.00111	0.97317	0.89822	0.00054	0.89799
	MSOPS 2 100	0.96719	0.00623	0.96776	0.91774	0.01203	0.92105
	MSOPS 2 200	0.95744	0.00965	0.96020	0.91117	0.00775	0.91253
	RSO	0	0	0	0.67735	0.03730	0.68188
4	MSOPS 1 50	0.96590	0.00107	0.96623	0.84765	0.01438	0.85238
	MSOPS 1 100	0.94724	0.00573	0.94887	0.72575	0.03761	0.73177
	MSOPS 1 200	0.94764	0.01187	0.94968	0.81489	0.03289	0.82292
	MSOPS 2 50	0.96726	0.00062	0.96730	0.85284	0.00049	0.85273
	MSOPS 2 100	0.96908	0.00258	0.96955	0.86206	0.00609	0.86445
	MSOPS 2 200	0.95605	0.00561	0.95742	0.85938	0.01289	0.86395
	RSO	0	0	0	0.39649	0.02363	0.39435
5	MSOPS 1 50	0.97740	0.00614	0.97956	0.78971	0.05479	0.80668
	MSOPS 1 100	0.96312	0.01848	0.97160	0.48432	0.32422	0.72034
	MSOPS 1 200	0.97749	0.00584	0.97694	0.82177	0.01404	0.82490
	MSOPS 2 50	0.93235	0.16743	0.98387	0.81037	0.00915	0.80863
	MSOPS 2 100	0.98743	0.00119	0.98762	0.86497	0.00606	0.86565
	MSOPS 2 200	0.97966	0.00296	0.97987	0.84002	0.01467	0.84609
	RSO	0	0	0	0.04960	0.03184	0.05873
6	MSOPS 1 50	0.98688	0.00469	0.98770	0.70669	0.18905	0.76654
	MSOPS 1 100	0.95343	0.02840	0.96312	0.63285	0.13323	0.68515
	MSOPS 1 200	0.99046	0.00169	0.99056	0.81435	0.03071	0.81964
	MSOPS 2 50	0.92549	0.18116	0.99355	0.84659	0.00215	0.84627
	MSOPS 2 100	0.96533	0.06398	0.98592	0.79881	0.01918	0.79436
	MSOPS 2 200	0.99122	0.00160	0.99154	0.81208	0.11049	0.83925
	RSO	0	0	0	0.16333	0.03440	0.15121

The convergence and relative hypervolume results are displayed in Tab. 3.4 and 3.5. While MSOPS obtains very promising results, RSO does not succeed in reaching the Pareto front. This is due to a too small number of function evaluations per run and the loss of information with every restart. Confirming the observations of Hughes (2005), RSO outperforms NSGA-II and SPEA2 in case of five and six objectives.

Almost all variants of MSOPS attain very low average distances indicating that nearly optimal solutions have been found. Only for five or six objectives, variants using a lower number of target vectors fail to converge to the Pareto front in some of the runs. In Tab. 3.4, this behavior can be inferred from a high standard deviation and high differences between the mean and the median value.

From the obtained hypervolume can be concluded that the distribution properties can be slightly improved by the supporting use of VADS. Hughes' assumption that the number of target vectors should be increased if more objectives are concerned is confirmed. For three objectives, the variants of MSOPS using 50 target vectors obtain the maximal hypervolume among the aggregation methods. With increasing objectives, the best values can be obtained with a higher number of target vectors. In general, the results show that the method used to design the target vectors is able to generate well distributed Pareto front approximations. Even for three objectives, NSGA-II and ϵ-MOEA (DTLZ1), respectively NSGA-II and SPEA2 (DTLZ2) can be outperformed regarding the hypervolume. Note that the given method to generate the target vectors only performs well on continuous Pareto fronts. As observed by Hughes (2003), a refinement of the targets is necessary for more complicated problems.

3.4.4 Indicator-based EMOA

The term *indicator-based EA (IBEA)* was introduced by Zitzler and Künzli (2004) for EMOA guided by a general preference information. The EMOA's selection operator uses a preference function (indicator) as a single-objective substitute for the d-dimensional objective function. In contrast to the aggregation methods, this preference information describes a general aim as expressed by quality measures like the hypervolume.

We consider SMS-EMOA and two versions of IBEA, implemented in PISA. The SMS-EMOA is applied in its original version (see Sec. 3.1), i.e., it performs non-dominated sorting and selection due to the hypervolume contribution. The adaptive reference point is here chosen according to the elements of the worst ranked fronts only, instead of the whole population, in order to give even less emphasis to boundary solutions. The IBEA (cf. Sec. 3.1) is considered in the variants IBEA_{HD} using the pairwise hypervolume and $\text{IBEA}_{\epsilon+}$ with the additive ϵ-indicator. The scaling factor $\kappa = 0.05$ is chosen as recommended in Zitzler and Künzli (2004).

We apply the recommended adaptive variant of the IBEA (cf. Zitzler and Künzli (2004)), which uses a normalization of the objective values to the interval $[0, 1]$.

Experimental Results

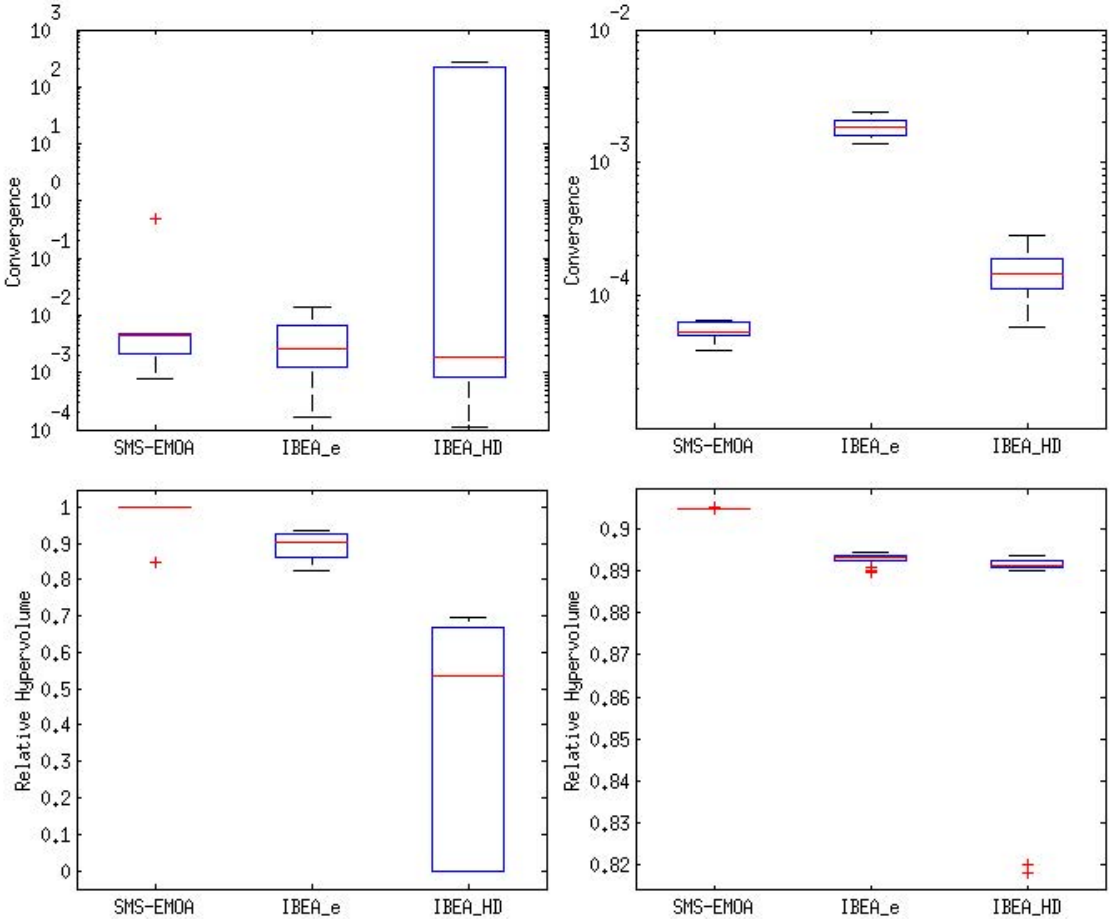


Fig. 3.13: Boxplots on the convergence (top) and the relative hypervolume (bottom) of indicator-based EMOA on DTLZ1 (left) and DTLZ2 (right) with 6 objectives.

An overview of the results is exemplarily given for the 6-dimensional case by the box plots in Fig. 3.13, the complete values of the convergence and the relative hypervolume are given in Tab. 3.6 and 3.7. As can be inferred from the convergence measure, both IBEA variants reach the Pareto front of DTLZ2. On DTLZ1, only IBEA$_{\epsilon+}$ converges towards the Pareto front for all dimensions. IBEA$_{HD}$ reaches a very good distance value on DTLZ1 with three dimensions but fails in case of more objectives. This is due to the normalization of objective values to $[0, 1]$, tending the hypervolume indicator to favor extremal solutions, which hinder the progression. Surprisingly, the IBEA$_{\epsilon+}$ using the additive ϵ-indicator reaches better hypervolume values than the IBEA$_{HD}$ invoking the hypervolume indicator. The consideration

Tab. 3.6: Convergence measure of indicator-based EMOA. The leftmost column gives the number of objectives.

		DTLZ1			DTLZ2		
	algorithm	mean	std.dev.	median	mean	std.dev.	median
3	$IBEA_{\epsilon+}$	0.04399	0.17481	0.00057	0.00015	$5.0{\cdot}10^{-5}$	0.00014
	$IBEA_{HD}$	0.00137	0.00337	0.00029	$1.3{\cdot}10^{-5}$	$5.3{\cdot}10^{-6}$	$1.2{\cdot}10^{-5}$
	SMS-EMOA	0.00110	0.00148	0.00039	$3.4{\cdot}10^{-6}$	$1.2{\cdot}10^{-6}$	$2.8{\cdot}10^{-6}$
4	$IBEA_{\epsilon+}$	0.01790	0.02940	0.00096	0.00071	0.00012	0.00069
	$IBEA_{HD}$	76.1230	119.550	0.00136	$4.5{\cdot}10^{-5}$	$1.3{\cdot}10^{-5}$	$4.2{\cdot}10^{-5}$
	SMS-EMOA	0.00193	0.00176	0.00100	$1.4{\cdot}10^{-5}$	$5.0{\cdot}10^{-6}$	$1.2{\cdot}10^{-5}$
5	$IBEA_{\epsilon+}$	0.02056	0.06678	0.00129	0.00115	0.00019	0.00112
	$IBEA_{HD}$	151.310	131.820	215.000	0.00013	0.00014	0.00010
	SMS-EMOA	0.00333	0.00215	0.00351	$3.7{\cdot}10^{-5}$	$9.2{\cdot}10^{-6}$	$3.8{\cdot}10^{-5}$
6	$IBEA_{\epsilon+}$	0.00467	0.00450	0.00256	0.00187	0.00031	0.00184
	$IBEA_{HD}$	82.1580	116.410	0.00182	0.00015	$5.6{\cdot}10^{-5}$	0.00014
	SMS-EMOA	0.10278	0.22310	0.00444	$5.4{\cdot}10^{-5}$	$1.1{\cdot}10^{-5}$	$5.2{\cdot}10^{-5}$

Tab. 3.7: Relative hypervolume of indicator-based EMOA. The leftmost column gives the number of objectives.

		DTLZ1, $\mathbf{r} = (0.7, \ldots, 0.7)$			DTLZ2, $\mathbf{r} = (1.1, \ldots, 1.1)$		
	algorithm	mean	std.dev.	median	mean	std.dev	median
3	$IBEA_{\epsilon+}$	0.77693	0.03182	0.78033	0.92991	0.00075	0.93002
	$IBEA_{HD}$	0.73929	0.03144	0.74208	0.92023	0.00071	0.92008
	SMS-EMOA	0.98352	0.00071	0.98387	0.93870	$6.3{\cdot}10^{-5}$	0.93873
4	$IBEA_{\epsilon+}$	0.82920	0.02445	0.83425	0.89477	0.00059	0.89484
	$IBEA_{HD}$	0.51417	0.35620	0.70647	0.88633	0.00090	0.88619
	SMS-EMOA	0.97612	0.00034	0.97627	0.90370	$6.4{\cdot}10^{-5}$	0.90368
5	$IBEA_{\epsilon+}$	0.87018	0.02777	0.86961	0.88571	0.00097	0.88584
	$IBEA_{HD}$	0.26292	0.33673	0	0.88250	0.00122	0.88259
	SMS-EMOA	0.99182	0.00019	0.99182	0.89619	$9.5{\cdot}10^{-5}$	0.89624
6	$IBEA_{\epsilon+}$	0.89146	0.03569	0.90029	0.89283	0.00130	0.89322
	$IBEA_{HD}$	0.40153	0.30853	0.53634	0.88431	0.02231	0.89124
	SMS-EMOA	0.96688	0.06741	0.99698	0.90483	0.00014	0.90481

of translation lengths in the additive ϵ-indicator causes a good distribution of solutions. Contrarily, the approximation of the hypervolume contribution through the binary hypervolume indicator tends to spiral downward with increasing dimension of objective space. Both adaptive IBEA fail to produce a good distribution on

DTLZ1, which we ascribe to the high-scaled co-domain and the resulting difficulties in the scaling of the fitness values.

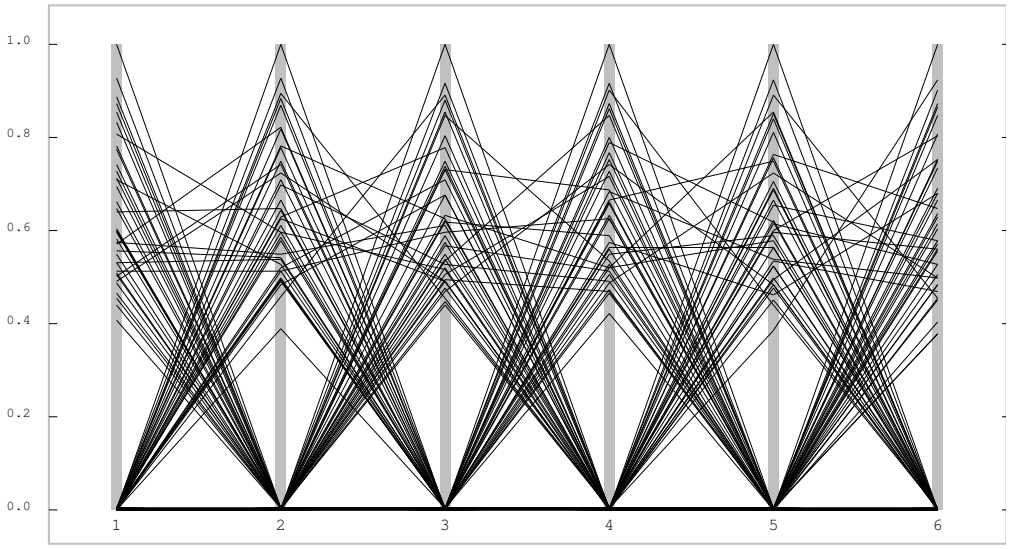


Fig. 3.14: Results of one run of SMS-EMOA on 6-objective DTLZ2. In the parallel plot, each column corresponds to one objective.

SMS-EMOA reaches the best hypervolume values of all considered algorithms. The distance values are very good as well and all runs except one reached the Pareto front. This run on six-objective DTLZ1 stagnated since one decision variable—which defines the distance—remains static at a non-optimal value due to an unusual loss of diversity in decision space in the beginning of the optimization process. Since the selector modules in PISA only decide regarding the objective values, this effect cannot be blamed to the selection properties of SMS-EMOA. Figure 3.14 exemplarily pictures the distribution of a typical six-objective result set of SMS-EMOA in a parallel plot. Every objective is covered and the structure of the set is almost symmetric, indicating a uniformly spread distribution of solutions over the whole Pareto front.

3.4.5 Conclusions

Our study demonstrated that concepts performing well in case of two or even three objective may not be qualified for higher dimensional problems. The bad performance of early Pareto-based methods like NSGA-II and SPEA2 observed by Hughes (2005) and Purshouse and Fleming (2003) is confirmed. They show a rapid degradation with increasing number of objectives. Some additional studies show that they do not converge to the Pareto front at all and stagnate far away from it. The performance of ϵ-MOEA refutes the hypothesis of Hughes that a Pareto-based

approach cannot succeed on many-objective problem instances. Instead, favoring extremal solutions has been shown to hinder the progression in many-objective spaces, which is also obviously for IBEA.
It is shown that more recent EMOA using indicators, which feature more than just distribution aspects, perform very well in many-objective optimization. Especially, SMS-EMOA [...] outperforms the other algorithms on all considered test functions. Moreover, an aggregation-based EMOA, namely MSOPS, performs well with respect to convergence aspects. A sophisticated scheme for the generation of weight vectors is introduced and also produces well distributed solution sets. In comparison to the simple restart strategy RSO, MSOPS benefits from structural equalities of good solutions by optimizing all weight vectors in parallel. [...]
Our results give valuable insights for practitioners which algorithm to choose depending on the number of objectives. While all established EMOA are suitable for bi-objective problems, optimizers for high-dimensional problems have to be chosen carefully. The SMS-EMOA shows outstanding effectiveness. The drawback of its exponential runtime can in the future be avoided by invoking an approximation of the hypervolume instead of the exact calculation, so it shall additionally become efficient.
Future research may develop new concepts for handling large numbers of objectives or optimizers combining ideas from indicator-based EMOA and aggregation algorithms like MSOPS.
The study emphasizes the necessity that new optimizers which are not dedicated to low-dimensional problems but shall be able to handle an arbitrary number of objectives have to be tested explicitly on many-objective problems.
An open problem is how optimizers shall be benchmarked. First attempts on professional benchmarks are made with the CEC competition organized by Huang et al. (2007). An issue is the parameterization of optimizers that influences their performance and may make a benchmarking unfair. The following sections give insights on the effects of parameter tuning.

3.5 Parameter Tuning

The previous Section 3.4 compared optimizers regarding the magnitudes of their performance with the main interest whether they come close to the Pareto front or not. This section shall analyze the performance of the SMS-EMOA more detailedly when using tuned parameterizations. Our focus is the influence of variation operators on the performance and the potential of parameter tuning. The presentation is based on Wessing et al. (2010) including literally adopted parts with adapted notation.

For single-objective EA, the variation as done in the CMA-ES gained general acceptance. The CMA-variation has been transferred to the multiobjective case by the MO-CMA-ES but the situation of operators is yet more multifarious due to different requirements. In order to approximate the Pareto front instead of just one optimum as in the single-objective case, diversity in the population is an important demand which shall also be supported by the variation operators. Most EMOA apply the operators SBX and PM by Deb and Agrawal (1995) or variation by differential evolution (DE, e.g. by Storn and Price (1997)). While the first is considered as a standard since years, EMOA using the latter gained more attention in recent years, emphasized by their good performance in the EMOA benchmark (with 8 participants) organized by Huang et al. (2007) for the CEC 2007 competition. Since their success has been attributed to the DE variation, the question arises whether the SMS-EMOA with that variation performs better than with the usually applied SBX and PM combination.

So, we experimentally analyze the performance of the DE variation compared to the combination of SBX and PM within the SMS-EMOA. We aim at matching the latest standards of experimental analyses so that this study may serve as a commendable example. According to e.g. Bartz-Beielstein et al. (2010a, Ch. 2.6), a comparison shall be performed with parameterizations optimized to the considered problems so that the optimizers can max out their potential. This contrasts the common practice in benchmarks to compare optimizers based on their default or on equal parameterization which may be more suitable for one algorithm than for another. Here, we perform the recommended parameter tuning using SPO (Sequential Parameter Optimization by Bartz-Beielstein et al. (2010b)). Parameter optimization itself is an optimization problem (cf. Sec. 1.5) with unknown optima. So we cannot guarantee to compare optimal parameterizations but only 'optimized' ones. The variants of the SMS-EMOA are then analyzed using the finally found parameterizations.

The following Section 3.5.1 describes related work and our chosen tuning method. Section 3.5.2 specifies the experimental setup. The experiment on the academic test problems is described in Section 3.5.3, and the analogous study on aerodynamic real-world problems in Section 3.5.4. Section 3.5.5 concludes this study.

3.5.1 Preliminaries

Related Work

By considering parameter tuning as an optimization problem, in principle any direct search optimization algorithm could be applied. Yet, there are sophisticated methodologies dedicated to this challenging task. Methods for parameter tuning and analyses are an active research area, a survey of the latest state-of-the-art methods is given by Bartz-Beielstein et al. (2010a) or Eiben and Smit (2011). We briefly mention related techniques before detailing our chosen method SPO.
Nannen and Eiben developed a method called *Relevance Estimation and Value Calibration (REVAC)* and combined it with other techniques, namely racing and sharpening, in Smit and Eiben (2009). The name stands for the assessment of the importance of a parameter on the algorithm's performance and the parameter tuning towards good values. The REVAC algorithm is an EA, respectively an estimation of distribution algorithm, which works on a population of parameter vectors that undergo an evolutionary optimization process. The limitations of REVAC are that it can only be applied for numeric parameters and does not give insights (like joint distributions for multiple parameters) into parameter interactions.
Several studies by the authors of REVAC and others analyzed the parameterizations of single-objective EA. However, this topic seemed to be neglected for EMOA. On the other hand, parameter tuning of EA has been modeled as a multiobjective problem due to conflicting demands, approaches are e.g. M-FETA by Smit et al. (2010) or performance fronts by Dréo (2009).
REVAC and SPO seem to be the most sophisticated tuning methodologies, whereas SPO appeared to be more suitable for our study of multiobjective performances. Performance assessment, analyses and visualization in case of bi-objective problems may be done by attainment surfaces (see Grunert da Fonseca and Fonseca (2010)). Yet the measuring via a quality indicator seems to be more suitable in combination with SPO and three-objective test problems. SPO has been applied to EMOA for the first time by Naujoks et al. (2006). They optimized the SMS-EMOA regarding its population size and the choice of its variation operator, i.e., had a more course-grained perspective compared to our study that includes the optimization of variation operators' parameterizations. Wessing and Naujoks (2010) describe a preliminary study to this work, focusing on the stability of SPO and the suitability of indicators.

Sequential Parameter Optimization (SPO)

For parameter tuning, we apply the method *Sequential Parameter Optimization (SPO)* by Bartz-Beielstein et al. (2005) realized in the SPO Toolbox (SPOT) by Bartz-Beielstein et al. (2010b) briefly described here.
The main idea of SPO is to treat optimizer runs as experiments, using methods from Design of Experiments (DoE), cf. Montgomery (1997), and Design and Analysis of

Computer Experiments (DACE), cf. Sacks et al. (1989). The optimizer's exogenous parameters, which are set before the start of the optimization are considered as the experiment's design variables that are sought to be optimized. SPO expects a region of interest (ROI) in the search space specified for each parameter. Then SPO starts by initialization in the ROIs according to a latin hypercube sampling (LHS) (see Bartz-Beielstein et al. (2010b)). A scalar value is required to describe the optimizer's performance, for which a quality indicator is chosen analogously to the selection within an EMOA. Based on these evaluations a surrogate model is built, in our case DACE Kriging by Lophaven et al. (2002). As the optimizer's results are stochastic, each point is sampled several times and the results are averaged. In an optimization loop, the model is then used to predict promising parameter configurations. The new candidates are evaluated and the data is fed back into the model. If no new best configuration is found in a step, the number of repetitions is increased.

3.5.2 Setup of Experimental Analysis

We describe the studied variation operators, academic test and real-world problems, and the performance measures to evaluate the optimizers' results. The experiments are then documented using the reporting scheme by Preuss (2007).

Variation Parameters to be Tuned

We consider the variation operators SBX and PM compared to the variation according to differential evolution (DE).
SBX and PM are described in Section 3.1. Recall that the parameters $\eta_c, \eta_m \in \mathbb{R}^+$ control the variance of the distributions. The application of SBX to create an offspring is determined by p_c. The application of altering a variable of the offspring's genome via PM is given by p_m, and decided independently for each variable. The combination of SBX and PM is referred to as *SBX* in the following for convenience.
For variation via differential evolution (DE) the variant by Storn and Price (1997) is considered with an alteration. In the original, the offspring only compete with their parents in the selection to keep changes local and maintain diversity in the population. However, we want to consider DE as a variation operator only. Three parents are chosen uniformly at random and a fourth parent is copied to generate an offspring. Then each variable is proceeded as follows (cf. Algorithm 3.8): The difference of the first two parents, weighted by the parameter $F \in [0, 2]$, plus the third parent replaces the variable of the offspring, in case that a random number is smaller than the parameter CR, or the index of the variable equals a random index chosen before, assuring that this procedure is executed at least once. Otherwise the offspring keeps the variable of the fourth parent.

Algorithm 3.8: DE variation

input : parents $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{3)}, \mathbf{x}^{(4)}$ from population $P^{(t)}$, parameters F, CR
output : offspring $\mathbf{y}$

```
index ← choose uniformly at random from {1, ..., n}
for i = 1 to n do
    if (i == index) or (rand() < CR) then        /* rand() ~ U(0,1) */
        y_i ← x_i^(3) + F(x_i^(1) − x_i^(2))
    else
        y_i ← x_i^(4)
```

Performance Measures

In multiobjective optimization, it is recommended to use more than one performance measure, yet SPO can handle only one. In the study by Wessing and Naujoks (2010) three quality measures, namely the hypervolume, the additive epsilon indicator $I_{\epsilon+}$, and the $R2$-indicator I_{R2} (see Zitzler et al. (2008a) for detailed definitions) have been considered each for the performance assessment within SPO. Thereby, the SPO optimization guided by the hypervolume achieves the best results. So, we here focus on the hypervolume only as it seems to be the most suitable measure for SPO.

Test Problems

From the CEC 2007 testbed by Huang et al. (2007) we have chosen the two-objective problems OKA2, SYM-PART, the shifted as well as rotated variants of ZDT problems, and three-objective DTLZ and WFG problems, so 13 functions in total. Note that the results are not directly comparable to others obtained before in the environment of the CEC 2007 contest, since a number of bugs in the implementation of the benchmark have been fixed.
The second experiment studies two aerodynamic problems. The design of an airfoil represented as Bezier splines is optimized w. r. t. different objective functions. NACA is a two-objective redesign problem, see Naujoks et al. (2002) for a detailed description. Two target airfoils with one being nearly optimal for the take off phase (with high lift) and the other for the cruising phase (with low drag) are given. An airfoil shall be designed such that the differences of the pressure distributions to the two target airfoils are minimized as the two objectives. Compromises between the two target airfoils describe the Pareto front with the targets as its extremes.
The problem RAE 2822 (referred to as *RAE*) is the drag minimization of an airfoil for three flow conditions as objective functions (cf. Emmerich et al. (2006)). Several

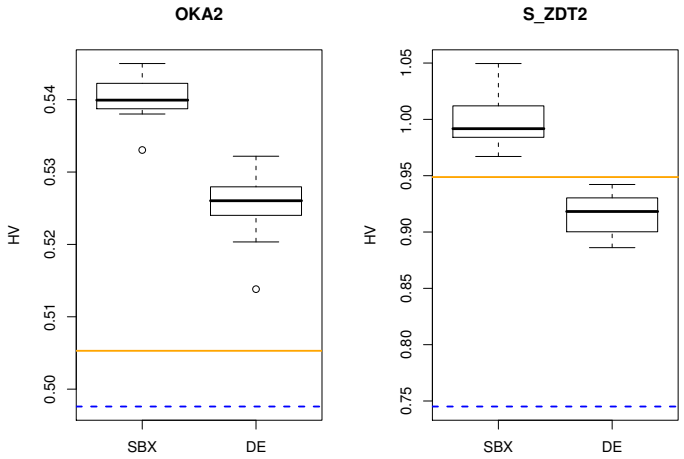


Fig. 3.15: Boxplots show the performance distributions in terms of dominated hypervolume (HV) of 20 SPO runs. Additional lines mark the default SBX (solid) and DE (dashed) configurations' mean performance.

constraints shall guarantee that the differences to a reference airfoil are not too strong. Geometric constraints can be checked without flow simulation, whereas aerodynamic performance constraints like lift and pitching moment are checked after the completed evaluation via computational fluid dynamics simulations.

3.5.3 DE vs. SBX on CEC 2007 Problems

Research Question How does DE compare to SBX variation on the CEC 2007 test case collection by Huang et al. (2007)?

Preexperimental planning For the experiment's preparation, some SPO runs were carried out to determine the parameters' regions of interest (see Tab. 3.8). Additionally, the optimization of SBX configurations on OKA2 and S_ZDT2 with 1,000 problem evaluations was repeated 20 times, to get an estimate of SPO's reliability. It is not necessary that SPO always delivers the same parameterization as the optimized one, because not all parameters have influence on the performance. But it is desired that an algorithm set up with the final parameterization achieves Pareto front approximations of similar quality. Fig. 3.15 shows that the performance could be increased in all cases and we regard the variance as small enough for meaningful comparisons, even when only one SPO run is performed per problem. The tuned parameter configurations will be called DE* and SBX* in the remainder.

Task After SPO has finished, the new configurations are run 50 times and evaluated via the hypervolume. These samples are compared to same-sized samples of the default configurations and each other. For each comparison, a two-sided U-Test

(see Hollander and Wolfe (1973)) is employed. The null hypothesis is that there is no difference in the medians and we require a significance level of 5% to reject it.

Setup SPO is applied to all 2-objective and 3-objective test problems in the CEC 2007 suite. The contained 5-objective problems are excluded, because of the SMS-EMOA's high runtime on these. Two different run lengths, namely $500d$ and $5000d$ function evaluations with d denoting the number of objectives, of the SMS-EMOA are examined to detect possible *floor* or *ceiling effects*. [...] Tables 3.8 and 3.9 show the regions of interest and the setup for the experiments. The default parameters for DE variation are chosen according to Kukkonen and Lampinen (2007). DE's lower bound for μ is higher than that for SBX, because it uses more parents for variation. The performance evaluation is generally done according to the CEC 2007 contest rules by Huang et al. (2007), i.e., the whole objective space of each problem is approximately normalized to $[1, 2]^d$. The reference point of the hypervolume is then set to $(2.1, \ldots, 2.1)^\top$, although Wessing and Naujoks (2010) show that the whole approach can have drawbacks on some problems.

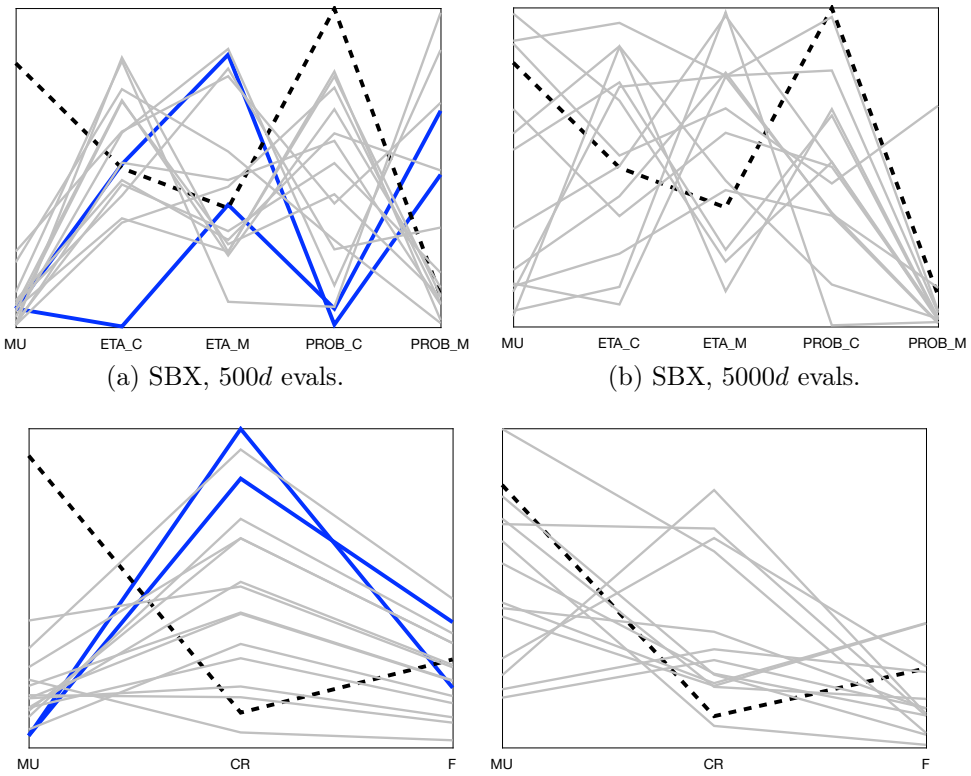
Tab. 3.8: The default values and region of interest (ROI) of parameters. The ROI is the range on which the search is conducted.

	DE			SBX				
Param.	μ_{DE}	CR	F	μ_{SBX}	η_c	η_m	p_c	p_m
Default	100	0.1	0.5	100	20.0	15.0	1.0	0.1
ROI	$\{6, \ldots, 120\}$	$[0, 1]$	$[0, 2]$	$\{3, \ldots, 120\}$	$[0, 40]$	$[0, 40]$	$[0, 1]$	$[0, 1]$

Tab. 3.9: Setup of the experiment on CEC 2007 test problems.

Problems	Two- and three-objective CEC 2007 problems
SPO budget	500 algorithm runs
Algorithm initialization	Uniform random
Stopping criterion	$500d$ and $5000d$ problem evaluations
Algorithm	SMS-EMOA
Parameters	DE: μ, CR, F; SBX: μ, η_c, η_m, p_c, p_m
Initial experimental design	Latin Hypercube (50 points, 3 repeats per point)
Performance measure	Hypervolume

Results/Visualization Tables 3.10 and 3.11 show the performance results of DE and SBX variation. Optimized configurations that are significantly better than the competing optimized configuration are highlighted in bold face. Figure 3.16 shows parallel plots of the configurations. More details on the parameter configurations are provided by Wessing (2009).



(a) SBX, 500d evals. (b) SBX, 5000d evals.

(c) DE, 500d evals. (d) DE, 5000d evals.

Fig. 3.16: Parallel plots of best parameter configurations found by SPO. Parameters for all 13 test functions are shown in light gray. Default configurations are marked as dashed lines. Results from the experiment on aerodynamic problems in Sec. 3.5.4 are shown as blue bold lines.

Observations SBX reaches [...] better mean values than DE for all test cases. For 500d problem evaluations, SBX* is better than DE* on ten problems, while the opposite is true on only two problems (there is one tie). For 5000d evaluations, SBX* wins seven times and DE* four times (there are two ties). Except for SBX* on R_ZDT4, both operators can always improve significantly compared to their default configurations. Figure 3.16 shows that small population sizes should be used on the short runs. Especially for long runs, low values of p_m are a good choice. It also seems to be promising to choose $CR > F$. The rest of the parameters does not follow any general trend.

Discussion The experiment shows that the decision which variation is chosen is less important than the decision to tune the chosen variation operator, because the differences between the default and optimized configurations are much bigger than between different optimized configurations. It is also obvious that the default setting is completely opposing the optimal configuration on some problems. SBX*

Tab. 3.10: Mean hypervolume and standard deviation with $500d$ evaluations.

Problem	SBX		SBX*		DE		DE*	
OKA2	0.5053	±0,013	**0.5450**	±0,011	0.4976	±0,013	0.5322	±0,021
SYM-PART	1.1640	±0,009	**1.2063**	±0,001	1.0335	±0,023	1.1935	±0,011
S_ZDT1	1.0189	±0,020	**1.1024**	±0,015	0.8706	±0,019	1.0403	±0,026
S_ZDT2	0.9488	±0,023	**1.0496**	±0,042	0.7451	±0,028	0.9422	±0,033
S_ZDT4	0.9505	±0,027	1.0407	±0,043	0.8489	±0,025	1.0546	±0,027
R_ZDT4	1.0994	±0,018	1.1214	±0,039	1.0605	±0,022	**1.1350**	±0,024
S_ZDT6	0.7340	±0,011	**0.7592**	±0,016	0.6573	±0,007	0.7013	±0,011
S_DTLZ2	1.3270	±0,001	**1.3291**	±0,002	1.3189	±0,003	1.3290	±0,001
R_DTLZ2	1.3100	±0,009	1.3204	±0,008	1.2531	±0,024	**1.3243**	±0,003
S_DTLZ3	1.3165	±0,003	**1.3304**	±0,001	1.3054	±0,005	1.3257	±0,003
WFG1	0.9051	±0,005	**0.9936**	±0,005	0.8677	±0,012	0.9084	±0,008
WFG8	1.1101	±0,014	**1.2231**	±0,009	1.1062	±0,015	1.1985	±0,009
WFG9	1.1651	±0,020	**1.2147**	±0,013	1.1474	±0,024	1.2012	±0,019

Tab. 3.11: Mean hypervolume and standard deviation with $5000d$ evaluations.

Problem	SBX		SBX*		DE		DE*	
OKA2	0.5610	±0,010	**0.5725**	±0,011	0.5480	±0,008	0.5676	±0,010
SYM-PART	1.2074	±3.0e-4	1.2095	±1.4e-4	1.1194	±0,012	**1.2098**	±4.0e-5
S_ZDT1	1.1508	±0,004	1.1684	±0,005	0.8755	±0,012	**1.1686**	±0,002
S_ZDT2	1.0668	±0,006	**1.1271**	±0,009	0.7394	±0,023	1.0928	±0,018
S_ZDT4	1.1241	±0,015	**1.2029**	±0,003	0.8441	±0,020	1.1654	±0,011
R_ZDT4	1.1908	±0,010	1.1933	±0,008	1.0845	±0,015	1.1923	±0,006
S_ZDT6	0.8658	±0,007	0.9293	±0,024	0.6584	±0,004	**0.9574**	±0,016
S_DTLZ2	1.3301	±1.1e-5	1.3302	±4.0e-6	1.3224	±5.7e-4	1.3302	±1.6e-5
R_DTLZ2	1.3296	±3.2e-5	**1.33000**	±3.2e-5	1.2638	±1.3e-2	1.32997	±4.2e-5
S_DTLZ3	1.3306	±2.3e-4	**1.33099**	±1.4e-5	1.3028	±2.2e-3	1.33097	±2.1e-5
WFG1	0.9684	±0,003	**1.0803**	±0,017	0.8929	±0,005	1.0633	±0,025
WFG8	1.2197	±0,005	**1.2702**	±0,003	1.2003	±0,046	1.2615	±0,003
WFG9	1.2459	±0,007	1.2592	±0,008	1.1894	±0,007	**1.2607**	±0,010

winning more often might be due to a biased set of problems. The result is more balanced on the longer runs, so it would be interesting to test if DE* performance increases for run lengths extended even further.

3.5.4 DE vs. SBX on Aerodynamic Problems

From the academic test cases we have learned that optimal parameterizations differ considerably. As a consequence, there is no general near-optimal default parameterization. However, this could be due to artificial structures of the academic test

cases. Therefore, the next experiment shall reveal whether our insights gained so far are transferable to real-world problems.

Research Question How does DE compare to SBX variation on examples of real-world problems? How do the results compare to the ones from the previous experiment?

Preexperimental planning As real-world problems from the aerodynamic field, the bi-objective NACA and the 3-objective RAE are chosen (cf. Section 3.5.2). The problems have been subject of previous studies, see Beume et al. (2007)*, Emmerich et al. (2006), Naujoks et al. (2002). Due to the large calculation times for the computational fluid dynamics simulations, a restricted number of 1000 objective function evaluations is allowed and the SPO budget is slightly decreased to 300 algorithm runs (see Tab. 3.12).

Task See the previous experiment in Section 3.5.3.

Setup The airfoils are represented as Bezier points with 6 degrees of freedom for NACA, and 18 for RAE. For RAE, the initial population always includes the baseline design. The default configurations and regions of interest are identical to the first experiment (see Tab. 3.8). Table 3.12 shows the differences in the experimental setup compared to the experiment in Section 3.5.3. The reference point is set to $(0.4, 0.4)^\top$ for the NACA problem and $(10, 10, 10)^\top$ for the RAE problem.

Tab. 3.12: Settings for experiment on aerodynamic problems that differ from Tab. 3.9.

Problems	NACA, RAE
SPO budget	300 algorithm runs
Stopping criterion	1000 problem evaluations
Initial experimental design	Latin Hypercube (25 points, 4 repeats per point)

Results/Visualization Table 3.13 shows the found optimized configurations, which are also included in Fig. 3.16 as bold lines. Table 3.14 shows the performance results.

Observations All optimized configurations are significant improvements over their default configurations. The difference between SBX* and DE* is not significant on NACA, but on RAE. The possible improvements by parameter tuning can be gleaned from Tab. 3.14: SPO is able to improve the NACA values by 2.7% using

Tab. 3.13: Parameter results on the aerodynamic problems

	DE Configuration			SBX Configuration				
Problem	μ_{DE}	CR	F	μ_{SBX}	η_c	η_m	p_c	p_m
NACA	21	0.90	0.34	10	0.16	15.43	0.06	0.68
RAE	14	0.76	0.71	10	20.50	34.24	0.01	0.48

Tab. 3.14: Mean hypervolume and standard deviation on aerodynamic problems

Problem	SBX	SBX*	DE	DE*
NACA	0.1462 ± 0.0012	0.1501 ± 0.0007	0.1467 ± 0.0009	0.1502 ± 0.0007
RAE	993.663 ± 0.005	993.844 ± 0.022	993.672 ± 0.033	**993.869** ± 0.041

SBX* and 2.4% featuring DE*. However, the results on RAE cannot be improved accordingly, here the improvements are about 0.02%.
Interestingly, the same population size is identified for SBX variation on both test cases. For DE*, a roughly similar population size was identified for the RAE case as well, while the best value for the NACA case is about twice as big. Concerning the operators' probabilities, SBX* variation focuses on mutation. The application probabilities for the recombination operator are very small, which means that η_c cannot have much influence. Generally, it is remarkable that SBX* and DE* are completely opposed to the default configurations.

Discussion The improvement seems so low on RAE, because the initial population always contains the mentioned near-optimal baseline solution, which already dominates a hypervolume of 993.662. But in fact, the default SBX configuration fails to find any other feasible solution in 49 of the 50 runs. The default DE configuration 'only' fails in 39 runs. DE* and SBX*, on the other hand, achieve success rates of 100% for this measure.

3.5.5 Conclusions

Responding the recent popularity of DE variation, we studied the performance of the SMS-EMOA with this variation compared to its usual operator combination SBX and PM. Thorough experimental analyses have been performed as a parameter tuning study using SPO, on academic test problems as well as on aerodynamic real-world problems.
The main results are: (1) The performance of the tuned operators improved significantly compared to the default parameterizations. (2) The performance of the two tuned variation operators is very similar. (3) The optimized parameter configurations for the considered problems are very different.

So, our experimental analyses could not verify any advantage of the DE variation over the combination of SBX and PM, at least not within the SMS-EMOA. More general than that, our study puts the common practice of benchmarking and tuning studies into question.
The study shows exemplarily that parameter tuning has more potential of performance improvement than the choice of a supposedly better operator. So, tuning shall always be performed before a conclusive rating of methods; it shall become standard. For academic benchmarks, this means that comparing EMOA without considering parameterizations is unfair and may give misleading hints like in the CEC 2007 competition by Huang et al. (2007) regarding DE variation. As described by Bartz-Beielstein et al. (2010a, Ch. 2.6), a comparison based on equal or untuned parameterizations is inappropriate. Instead the tuning shall be part of the benchmarking. For contests, we recommend to dedicate a budget of function evaluations for parameter tuning and maybe specifying the tuning method. In publications presenting a new optimizer, the new method is typically parameterized with expert knowledge but tested against standard optimizers using default parameters. It is desirable to quantify the effort put in finding good parameterizations and to study ranges of parameters and their sensitivity.
For practitioners our message is analogously: Instead of investing time on a comprehensive study of methods, quickly choose one optimizer and invest the time to adapt it to your problem. Do not expect default parameterizations to perform well. When parameter tuning is performed, it is still common practice to do this with a chosen optimizer on a test problem which is efficiently evaluable instead of the actual problem. Our study demonstrates that this procedure is not promising, since the optimized parameter configurations for the academic test suite differ a lot from those for the real-world problems. It cannot be expected that parameterization performing well on one problem do so on another.
Although parameter tuning is elaborate, this tool shall not be neglected due to the significant potential. When evaluations of the optimization problem are too time-consuming, an alternative is to perform the tuning on a surrogate model instead. Preuss et al. (2010) achieved successes with this approach and demonstrate that the EA using the gained tuned parameterization outperforms the EA using the default parameterization on the original problem.
Professional tools for parameter tuning are available with SPOT and REVAC (cf. Nannen and Eiben (2007)) but parameter tuning requires considerable resources. To reduce the time spent on expensive function evaluations, further investigations on the suitability of surrogate models instead of original problems for parameter tuning are desirable. Moreover, the interaction of parameters is interesting and dependencies and sensitivity analyses shall give valuable insights.

3.6 Hybrid Hypervolume-Gradient Metaheuristic

While the previous sections dealt with the SMS-EMOA or its components, this section presents a hybridization of the SMS-EMOA with an additional hypervolume-based technique as a new multiobjective optimizer. The new technique performs a steepest descent according to the gradient of the hypervolume of a population of points, thereby optimizing the positions such that their hypervolume is maximized. The presentation is based on Emmerich et al. (2007)* and so are the literally adopted parts with notation adapted to this document.

Hybridizations shall join the advantages of several methodologies. The SMS-EMOA performs satisfactorily, yet it might be improvable with additional techniques. We see potential in gradient-based methods, which are highly recommended when gradient information is available, but even promising when the gradient has to be estimated. The methods are local optimizers that are applied for convergence to a local optimum in order to refine or improve a starting solution to make the best possible use of its potential.

The gradient of the hypervolume of a set of points is introduced, and the optimization guided by this gradient increases the hypervolume of the set by determining an improved position for all points at once. Special care has to be taken for dominated points when no local movement can make them non-dominated so that the population's hypervolume remains unchanged. A penalty approach is introduced to make the gradients non-zero. Analytical calculation as well as empirical approximation of the gradient are discussed. A steepest descent method guided by the hypervolume gradient is introduced as a multiobjective optimizer. This method is then hybridized with the SMS-EMOA in a high-level relay hybrid (cf. Talbi (2002)). The SMS-EMOA is executed first to roughly approach the population along the Pareto front. The gradient method takes the final population of the SMS-EMOA as its input and performs a fine-tuning of the points' positions according to the hypervolume maximization. The hybrid optimizer is studied on bi-objective academic test problems of the ZDT suite, amongst others comparing different switching times of the optimization methods. The proposed method is especially tested with a low budget of function evaluations that shall investigate its suitability for real-world applications which cannot afford tens of thousands time-consuming function evaluations.

The following Section 3.6.1 gives an overview of related work including other hybridizations with the SMS-EMOA, and the used mathematical notation. In Section 3.6.2 the hypervolume gradient of a population of points is defined and its analytical as well as empirical calculation are described. A steepest descent gradient method using the hypervolume gradient is introduced in Section 3.6.3. Section 3.6.4 presents experiments of a relay hybrid of the SMS-EMOA and the gradient-based optimizer. Section 3.6.5 concludes this topic.

3.6.1 Preliminaries

This section gives an overview of related work, especially referring to other hybridizations of the SMS-EMOA. Afterwards, the mathematical notation required for the gradient is introduced.

Related Work

Other hybridizations related to the hypervolume have been considered and are briefly reviewed.
Koch et al. (2009)* couple the SMS-EMOA respectively with the following three local search methods. The direct search methods by Hooke and Jeeves (1961) searches along coordinate axes and minimizes in this multiobjective scenario the weighted Tchebycheff distance to a utopian point. The first order method of multi-objective steepest descent according to Fliege and Svaiter (2000) performs a steepest descent search in the cone of dominating solutions, where the direction is calculated based on the Jacobian matrix. The second order Newton method was transferred to the multiobjective case by Fliege et al. (2009). The SMS-EMOA is coupled with the first two above methods in a relay, where the final population of the SMS-EMOA is then refined by the local search method. A concurrent hybrid of the SMS-EMOA and the multiobjective Newton method is presented where it is probabilistically decided whether the local search is performed for offspring. Experimental analyses are performed on the ZDT suite, where different switching schedules between the methods are studied.
Voß et al. (2008)* compare the MO-CMA-ES (cf. Sec. 3.1) against hybridizations of the single-objective CMA-ES with scalarization approaches. The scalarization is performed via a weighted sum or the Tchebycheff method which is as opposed to the former able to approach any Pareto-optimal point. Experimental analyses are performed on established bi-objective function, including the ZDT suite. The Tchebycheff-based optimizer outperforms the one using the weighted sum, whereas both hybrid optimizers are outperformed by the MO-CMA-ES on most considered functions.
Besides the methods above, several hybridization for multiobjective optimization have been developed, whereas most aim at improving single points instead of a whole set like in our approach. Similar to Fliege and Svaiter (2000), Bosman and de Jong (2006) aim at searching the cone of dominating points by local mutants. Similar to our approach Schütze et al. (2005) use a gradient-based search to approximate a well-distributed set by means of Karush-Kuhn-Tucker points, which is only conditionally possible. Shukla (2007) discusses the hybridization of NSGA-II with two gradient based local search operator, thereby estimating the gradient with the help of a perturbation technique. Sindhya et al. (2008) as well combine NSGA-II with a local search method with the help of an achievement scalarization function. Lately, Hughes (2011) presented an algorithm called MODELS which combines

the weight vector aggregation techniques from MSOPS with a golden section line search algorithm, yielding competitive results. Moreover, the calculation of the hypervolume gradient relates to the work on optimal μ-distributions by Auger et al. (2009a) which as well seek to position all points optimally.

Mathematical Notation

To compute the hypervolume of a population, we use the following representation as a *population vector*.

Definition 3.43 *The representation of μ population members $\mathbf{x}^{(i)} \in \mathbb{R}^n$, $i = 1, \ldots, \mu$, in the search space is combined to one vector of length μn called* population vector*:*

$$\mathbf{p} = (x_1^{(1)}, \ldots, x_n^{(1)}, \ldots, x_1^{(\mu)}, \ldots, x_n^{(\mu)})^\top = (p_1, \ldots, p_{\mu \cdot n})^\top.$$

The mapping from population vectors back to populations can be realized as

$$\Psi(\mathbf{p}) = \{\{ \ (x_1^{(i)}, \ldots, x_n^{(i)})^\top \mid i = 1, \ldots, \mu \ \}\}. \tag{3.23}$$

$\mathbf{F}$ denotes the function evaluation of points according the objective function $f(\mathbf{x})$, while maintaining their structure as

$$\mathbf{F}(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(\mu)}) = (f(\mathbf{x}^{(1)}), f(\mathbf{x}^{(2)}), \ldots, f(\mathbf{x}^{(\mu)}))^\top \tag{3.24}$$

or respectively

$$\mathbf{F}(\ \{\{ \ \mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(\mu)} \ \}\} \ ) = \{\{ \ f(\mathbf{x}^{(1)}), f(\mathbf{x}^{(2)}), \ldots, f(\mathbf{x}^{(\mu)}) \ \}\}. \tag{3.25}$$

The optimization of the hypervolume is then performed on evaluated population vectors, as

$$H(\mathbf{p}) := H(F(\Psi(\mathbf{p}))), \tag{3.26}$$

which is a valid representation as the global optima of this problem can be mapped back to the global optima of the hypervolume defined on sets. The representation with population vectors usually introduces additional local optima as permuting the positions of points in the vector results in the same population.
For notational convenience, we write the hypervolume briefly without a reference point and apply it directly to an argument list of vectors instead of a set of vectors, using the above mappings implicitly.
In summary, the following mappings are applied to express the gradient of the hypervolume:

$$\mathbb{R}^{\mu \cdot n} \underbrace{\xrightarrow{\mathbf{F}}}_{\text{decision to objective space}} \mathbb{R}^{\mu \cdot d} \underbrace{\xrightarrow{H}}_{\text{objective space to hypervolume}} \mathbb{R}^+. \tag{3.27}$$

3.6.2 Hypervolume Gradient

A general definition of the hypervolume gradient for the space of population vectors is

$$\nabla H(\mathbf{p}) = \left(\frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_{\mu n}} \right)^\top \tag{3.28}$$

[...] Using the chain rule the gradient can be rewritten as

$$\nabla H(F(\mathbf{p})) \;=\; (H \circ F)'(\mathbf{p}) \;=\; H'(F(\mathbf{p})) \cdot F'(\mathbf{p}) \tag{3.29}$$

as detailed in the following. Let $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(\mu)}$ be μ points in the decision space, then $\nabla H(\mathbf{p})$ can be written as:

$$H' \text{ at } \begin{pmatrix} f(\mathbf{x^{(1)}}) \\ f(\mathbf{x^{(2)}}) \\ \cdots \\ f(\mathbf{x}^{(\mu)}) \end{pmatrix} \circ \begin{pmatrix} f' \text{ at } \mathbf{x}^{(1)} & 0 & 0 & \cdots & 0 \\ 0 & f' \text{ at } \mathbf{x}^{(2)} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & f' \text{ at } \mathbf{x}^{(\mu)} \end{pmatrix} \tag{3.30}$$

The top level structure of the matrix associated to the linear mapping $\mathbf{F}'$ is a diagonal matrix of size μ whose diagonal elements are matrices of size $d \times n$ associated to the linear maps f' at $\mathbf{x}^{(j)}$, where $j = 1, 2, \ldots, \mu$ and each of the off-diagonal elements is the zero matrix of size $d \times n$ as well.
A more detailed description of this matrix is given as:

$$\underbrace{\begin{pmatrix} \frac{\partial H}{\partial y_1^{(1)}} \\ \vdots \\ \frac{\partial H}{\partial y_d^{(1)}} \\ \vdots \\ \frac{\partial H}{\partial y_1^{(\mu)}} \\ \vdots \\ \frac{\partial H}{\partial y_d^{(\mu)}} \end{pmatrix}^\top}_{\nabla H(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(\mu)})} \cdot \underbrace{\begin{pmatrix} \frac{\partial f_1(\mathbf{x}^{(1)})}{\partial x_1^{(1)}} & \cdots & \frac{\partial f_1(\mathbf{x}^{(1)})}{\partial x_n^{(1)}} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_d(\mathbf{x}^{(1)})}{\partial x_1^{(1)}} & \cdots & \frac{\partial f_d(\mathbf{x}^{(1)})}{\partial x_n^{(1)}} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \vdots & \cdots & \vdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \vdots & \cdots & \vdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{\partial f_1(\mathbf{x}^{(\mu)})}{\partial x_1^{(\mu)}} & \cdots & \frac{\partial f_1(\mathbf{x}^{(\mu)})}{\partial x_n^{(\mu)}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \frac{\partial f_d(\mathbf{x}^{(\mu)})}{\partial x_1^{(\mu)}} & \cdots & \frac{\partial f_d(\mathbf{x}^{(\mu)})}{\partial x_n^{(\mu)}} \end{pmatrix}}_{\mathbf{F}'(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\mu)})} \tag{3.31}$$

with $\mathbf{y}^{(i)} = f(\mathbf{x}^{(i)})$. Note that $\mathbf{F}'(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\mu)})$ depends solely on the gradient functions ∇f_i at the sites $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\mu)}$. Hence, if these $d \cdot \mu$ local gradients are known, the desired gradient $\nabla H(\mathbf{p})$ can be computed.

Distinguishing Types of Sets

The computation of $\nabla H(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(\mu)})$ is discussed next. Three cases of the set $\{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(\mu)}\}$ need to be considered: (1) mutually non-dominated sets, (2) sets with strictly dominated points, and (3) sets with edge points.

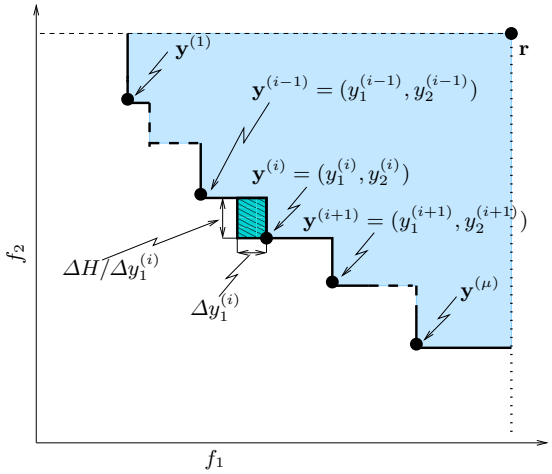


Fig. 3.17: Partial derivative of the hypervolume for $d = 2$ and non-dominated sets. The lengths of the line-segments of the attainment curve correspond to the values of the partial derivatives of H. Only for boundary points do the values of the partial derivatives depend on the reference point.

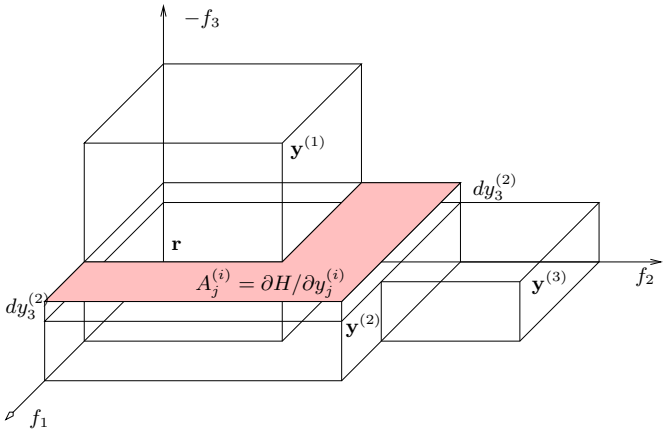


Fig. 3.18: Partial derivative for $d = 3$. By changing a point $\mathbf{y}^{(i)}$ differentially in the j-th coordinate direction, the hypervolume grows with the area $A_j^{(i)}$ of the 'visible' face of the exclusively contributed hypervolume of that point in the direction of the movement. Hence $A_j^{(i)}$ is the partial derivative $\partial H / \partial y_j^{(i)}$.

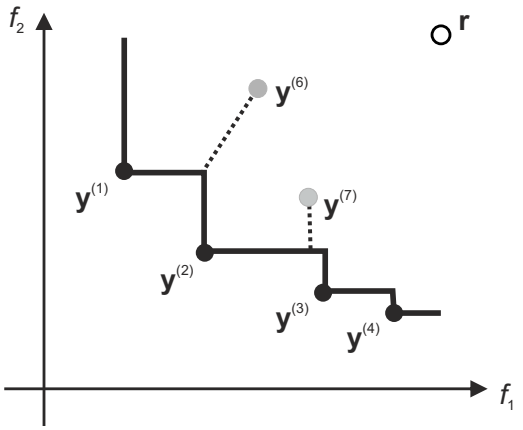


Fig. 3.19: The penalty function is defined as the sum of the Euclidean distance (dashed lines) of the dominated points (gray) to the attainment curve (solid line) shaped by the non-dominated points (black) and bounded by the reference point $\mathbf{r}$. The penalty is subtracted from the hypervolume value to give an influence to the dominated points.

Mutually non-dominated sets. For $d = 1$ holds $\frac{\partial H}{\partial y_1^{(i)}} = 1$, and for $d = 2$ holds (assuming vectors $\mathbf{y}^{(i)}$ are sorted in descending order of f):

$$\frac{\partial H}{\partial y_1^{(i)}} = y_2^{(i-1)} - y_2^{(i)} \quad \text{and} \quad \frac{\partial H}{\partial y_2^{(i)}} = y_1^{(i-1)} - y_1^{(i)}, \quad i = 1, \ldots, \mu \tag{3.32}$$

as illustrated in Fig. 3.17. Note that boundary points need special treatment, as their contribution to the gradient is influenced by the reference point. In three dimensions ($d = 3$), the computation of the partial derivative gets more tedious. The general principle is sketched in Fig. 3.18.

Sets with strictly dominated points. The gradient equals zero in case of dominated points—provided that a slight perturbation does not make them non-dominated—since no improvement of the hypervolume can be observed for any movement. Therefore, dominated points do not move during a search with gradient methods but just remain in their position. To enable an improvement of dominated points, a *penalty value* can be subtracted from the hypervolume value, that is negative if and only if points are dominated and otherwise zero. For each dominated point, the minimal Euclidean distance to the attainment surface shaped by the non-dominated points is calculated (Fig. 3.19). The sum of these values is subtracted from the hypervolume value of the whole set of points. This way, the movement of dominated points influences the improvement of the penalized hypervolume and a local gradient of the dominated points is computed that points in the direction of the nearest point on the attainment curve. In a gradient descent

method the movement of the non-dominated points is delayed by the dominated ones. Anyway, this drawback is a smaller deficit than completely losing the dominated points. Since any non-dominated point contributes to the hypervolume value, the primary aim is to make all points non-dominated.

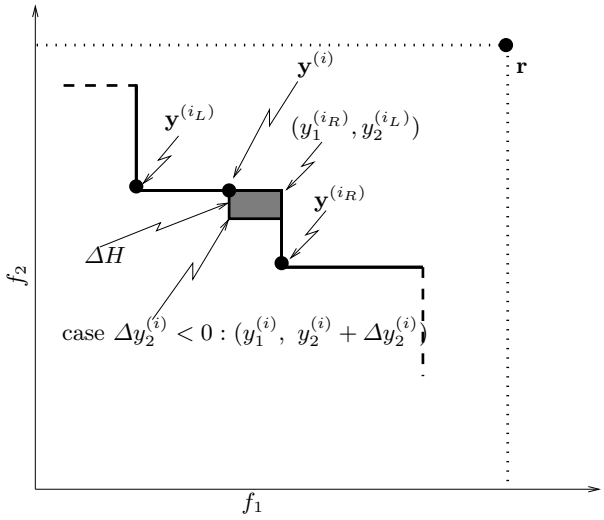


Fig. 3.20: Partial derivative for edge points in a 2-dimensional objective space. These points are dominated but not strictly dominated.

Sets with edge points. Points that are dominated but not strictly dominated (called edge points) lie on the attainment surface of the non-dominated points. Slight movements can make the points either remain weakly dominated, become strictly dominated or non-dominated. Thus, the gradient at these points is not continuous. The left-sided derivative $\frac{\partial^- H}{y_j^{(i)}}$ may be positive, while the right-sided derivative $\frac{\partial^+ H}{y_j^{(i)}}$ is always zero. For $d = 2$ positive one-sided derivatives can be determined as the length of the segment of the attainment curve. Let $\mathbf{y}^{(i_L)}$ determine the neighbor of the edge point $\mathbf{y}^{(i)}$ on the upper left corner of the attainment curve, and $\mathbf{y}^{(i_R)}$ the neighbor on the lower right corner (see Fig. 3.20). If the point $\mathbf{y}^{(i)}$ lies on the segment $\mathbf{y}^{(i_L)}$ to $(y_1^{(i_R)}, y_2^{(i_L)})^\top$, then $\frac{\partial^+ H}{\partial y_1^{(i)}} = 0$ and $\frac{\partial^- H}{\partial y_2^{(i)}} = y_1^{(i_R)} - y_1^{(i)}$ (see also Fig. 3.20); else if the point lies on the segment $\mathbf{y}^{(i_R)}$ to $(y_1^{(i_R)}, y_2^{(i_L)})^\top$, then $\frac{\partial^- H}{\partial y_1^{(i)}} = y_2^{(i_L)} - y_2^{(i)}$ and $\frac{\partial^+ H}{\partial y_2^{(i)}} = 0$. The fact that $H(\mathbf{p})$ is in general not continuously differentiable at edge points makes it problematic to work with gradient-based methods that make use of second order derivatives.

Edge points can also cause non-dominated points to have discontinuous local derivatives, which is comprehensible by arguments similar to the ones above. Besides

degenerated points in the search space can cause discontinuous derivatives. These are, loosely defined, search points [...] with the same image.

Empirical Gradient Determination

In practice the computation of the gradient can be approximated for example by using numerical differentiation. Since there may be points which are not continuously differentiable, we need to take one-sided derivatives in both directions into account. For a small positive ϵ we compute them via:

$$\frac{\partial H}{\partial p_i} \approx \frac{H((p_1, \ldots, p_i \pm \epsilon, \ldots, p_{\mu n})^\top) \pm H((p_1, \ldots, p_i, \ldots, p_{\mu n})^\top)}{\epsilon} \tag{3.33}$$

The algebraic signs we need to use depend on the gradients of the objective function. In case of continuously differentiable objective functions, it is numerically safer to compute the derivatives of the objective functions first, and then use the chain rule to compute the derivatives of the hypervolume taking special care of the different types of sets as described above. Both the computation of Equation (3.33) and the computation of the gradients of all objective functions at all points (that can be used to compute the gradient via the chain rule) requires μn evaluations of the objective function vectors.

Analytical Solution of Hypervolume Maximization

To demonstrate the exact calculation of the maximal hypervolume by means of partial derivatives, an example is given in Lemma 2.7 for a linear Pareto front. The hypervolume is maximal when the points are equally spaced over the Pareto front. The Pareto front is due to the definition of a scalable generalized Schaffer problem by Emmerich and Deutz (2007), which is also considered for proof of concept experiments in Section 3.6.4.

3.6.3 Gradient-based Pareto Optimization

Due to the known problems with second-order gradient methods, which require twice continuous differentiability, a first-order gradient method, namely the steepest descent/ascent method with backtracking line search has been implemented (cf. Boyd and Vandenberghe (2006)). The pseudo-code of our implementation is provided in Algorithm 3.9. The line-search algorithm has been kept simple to maintain transparency of the search process. It will however converge to a local maximizer relative to the line search direction. Note, that the line search may move to the same point in two subsequent iterations. In this case the evaluation of the objective function vectors of the population can be omitted. The convergence speed and accuracy of the line search can be controlled with the parameters τ and $s_{\min}$, respectively. The stepsize s is decreased via the reduction parameter τ and

Algorithm 3.9: Gradient-based hypervolume maximization (cf. Emmerich et al. (2007)*)

input : initial population as population vector $\mathbf{p}$
parameters: accuracy of line search $s_{\min}$, step size reduction rate $\tau \in (0,1)$

$s \leftarrow 1.0$ /* initialize step size s */
$i \leftarrow 0; \mathbf{p}^{best} \leftarrow \mathbf{p}^0$
$\mathbf{v}^{(0)} \leftarrow \nabla H(\mathbf{p}^{best})$ /* initialize search direction */
while $|\mathbf{v}^{(i)}| > \epsilon$ **do**
  $s \leftarrow 1.0$
  while $s > s_{\min}$ **do** /* line search in gradient direction */
    $\mathbf{p}^{new} \leftarrow \mathbf{p}^{best} + s\mathbf{v}^{(i)}$ /* try positive direction */
    if $H(F(\Psi(\mathbf{p}^{best}))) \geq H(F(\Psi(\mathbf{p}^{new})))$ **then**
      $\mathbf{p}^{new} \leftarrow \mathbf{p}^{best} - s\mathbf{v}^{(i)}$ /* try negative direction */
      if $H(F(\Psi(\mathbf{p}^{best}))) \geq H(F(\Psi(\mathbf{p}^{new})))$ **then**
        $s \leftarrow s \cdot \tau$ /* no success: reduce step size s */
        $\mathbf{p}^{new} \leftarrow \mathbf{p}^{best}$ /* new best point is old best point */
    $\mathbf{p}^{best} \leftarrow \mathbf{p}^{new}$
  $\mathbf{v}^{(i+1)} \leftarrow \nabla H(\mathbf{p}^{new}), i \leftarrow i+1$ /* compute new gradient direction */
return $\mathbf{p}^{best}$

the line search is performed as long as $s > s_{\min}$ holds. Since the length of the gradient decreases when the algorithm converges to the optimum of a differentiable function, $s_{\min}$ does not have to be very low, because the length of the gradient influences the step-size as well. In case of unknown exact gradients, the gradients are approximated via (3.33) at the cost of μn function evaluations. The hypervolume is thereby computed w. r. t. a fixed reference point.

3.6.4 Experiments on SMS-EMOA-Gradient-Hybrid

Since the gradient optimizer has been conceived as a fine-tuning of a Pareto front approximation, it does not seem to be promising as a stand-alone optimizer. Therefore, it is hybridized with an EMOA, and we choose the SMS-EMOA for this task. The SMS-EMOA generates a rough Pareto front approximation, and the gradient optimizer take the SMS-EMOA's final population as its input. This hybridization scheme is called high-level relay hybrid in the taxonomy by Talbi (2002). Relay denotes the sequential execution of the algorithms and high-level means that the algorithms are applied in their original versions, not just taking parts of them.
Note that the SMS-EMOA is applied in a version without internal reference point (cf. Sec. 3.1). Only bi-objective problems are considered here, and for these holds

that only the hypervolume contributions of the boundary points, i.e., the points without worse neighbors in one objective, depend on the reference point. These points are selected to be always kept in the population, which works in fine in case of $d = 2$ but not in higher dimensional spaces as discussed in Section 3.4.
In the following experiments, a budget of function evaluation is partitioned among the algorithms, so that the time of switching over is a parameter of the hybrid algorithm.

Experiments on Generalized Schaffer Problem

We conducted two experiments to analyze the limit behavior of the hybrid algorithm on the generalized Schaffer problem (cf. Section 3.6.2) defined as follows.

Definition 3.44 *The generalized Schaffer problem (cf. Emmerich and Deutz (2007) is defined as $f = (f_1, f_2)$ with*

$$f_1(\mathbf{x}) = \frac{1}{n^\alpha}\left(\sum_{i=1}^{n} x_i^2\right)^\alpha \quad \text{and} \quad f_2(\mathbf{x}) = \frac{1}{n^\alpha}\left(\sum_{i=1}^{n}(1-x_i)^2\right)^\alpha$$

with $\mathbf{x} \in [0,1]^n, \alpha \in \mathbb{R}^+$ and both objectives to be minimized.

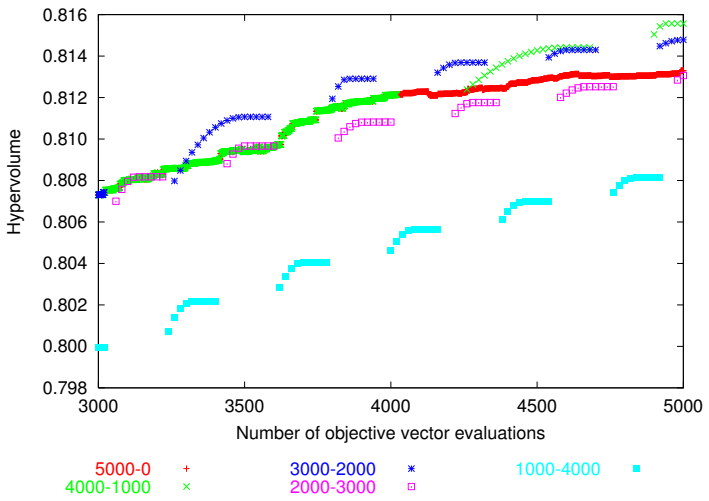


Fig. 3.21: Hypervolume achieved by the hybrid optimizer using different switching times. In the legend, the first number denotes the number of evaluations of the SMS part and the second of the gradient part.

A simple version has also been considered in Section 3.3 as T1. In a preliminary study, we consider the generalized Schaffer problem with $n = 10$ and $\alpha = \frac{1}{2}$, hence the Pareto front equaling $\{(y_1, y_2) \mid y_2 = 1 - y_1 \text{ and } y_1 \in [0, 1]\}$ is linear (cf. Section 3.6.2). Recall that, due to Theorem 3.41, we know that the hypervolume of this problem is a strictly concave function, so that the gradient method converges to the global optimum. Experiments with the gradient method, initialized with a population evolved over 1000 iterations by SMS-EMOA, show that the convergence is indeed linear for small population sizes of e.g. $\mu = 5$, and almost linear for higher population sizes. The dimension of the problem's search space seems to have less influence on the convergence rate.
Fig. 3.21 shows the results for the generalized Schaffer problem with $\alpha = 1$, the dimension of the search space $n = 10$, and $\mu = 10$. Typical runs are shown, where the SMS-EMOA is always started using the same random seed.. The Pareto front is equal to $\{(y_1, y_2) \mid y_2 = 1 - 2\sqrt{y_1} + y_1 \text{ and } 0 \leq y_1 \leq 1\}$ and the maximally attainable hypervolume is $1 - \frac{1}{6} \approx 0.833333$. The limit behavior of the hybrid optimizer using different switching times are compared. The discontinuities in the progress correspond to the end of a line search, and a gap indicates that function evaluations are spend on the $\mu n = 100$ gradient calculation. The picture shows that once the gradient part of the hybrid method is supplied with a reasonably good approximation set to the Pareto front, the gradient part of the method outperforms the pure SMS-EMOA.

Studies on ZDT Test Suite

To compare the hybrid optimizer with or without using the penalty function for dominated points, experiments on ZDT6 of the ZDT test suite (Zitzler et al. (2000), cf. Sec. 1.5) are performed. Parameters are chosen as search space dimension $n = 10$, population size $\mu = 20$, $s_{\min} = 0.1$, $\tau = 0.5$, and the reference point $\mathbf{r} = (10, 10)$ for the internal hypervolume computation as well as for the external performance evaluation. The total number of function evaluations in each run was 2000 in order to test the ability to work with a low budget of function evaluations (contrarily to the 20,000 function evaluations in previous benchmarks, e.g. Beume et al. (2007)*) which is desirable for real-world applications. Five different strategies were performed, listed with increasing number of function evaluations dedicated to the SMS-EMOA part: 20, 200, 400, 1000, and 2000, respectively. The remainder of the 2000 function evaluations was used for the gradient part.
Fig. 3.22 reveals that it pays off to apply the gradient part of the algorithm as soon as a rough approximation set has been found. The speed-up occurs especially at the beginning and thus the hybrid approach is useful in order to obtain very good results with few function evaluations. Secondly the picture also shows that giving a penalty to points in the population which are dominated gives far better approximation sets w. r. t. the hypervolume. [...] Clearly, the hybrid algorithm converges in each case to a population with maximal hypervolume. Also the pure

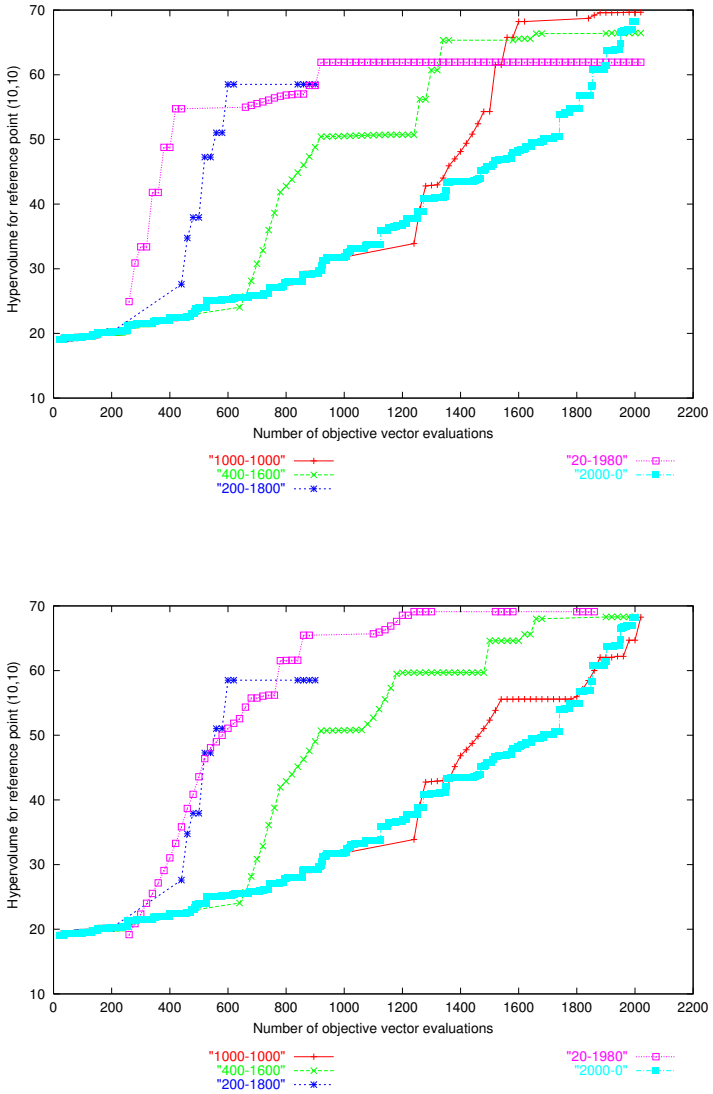


Fig. 3.22: Progress of the hybrid optimizer on ZDT6, with $n = 10, \mu = 20$ and 2000 function evaluations. Different switching points are used, denoted as the number of function evaluations of the SMS-EMOA and for the gradient. The penalty function for dominated points is only used in the lower figure.

SMS-EMO eventually catches up with the hybrid algorithm and converges to the maximum.

Table 3.15 shows the results of running the hybrid algorithm on the ZDT test suite (ZDT1-ZDT4, and ZDT6). Due to the good performance, the hybrid optimizer is applied using the penalty function for dominated points. It is parameterized using

Tab. 3.15: Results for different switching time (ST) strategies of the hybrid optimizer on ZDT functions, given as averaged hypervolume values over 5 repetitions each.

		1000 evaluations			2000 evaluations		
ZDT	ST	min	avg	max	min	avg	max
1	20	21.88	23.28	24.20	21.88	23.90	24.51
	500	21.12	23.39	24.34	23.60	24.12	24.46
	1000	17.20	20.02	21.85	24.18	24.38	24.48
	1500	17.20	20.02	21.85	24.11	24.26	24.40
	2000	17.20	20.02	21.85	23.06	23.73	24.37
2	20	19.16	21.25	24.06	19.41	22.81	24.13
	500	18.13	20.58	23.10	19.70	22.11	23.96
	1000	14.66	18.10	19.87	14.66	20.65	23.69
	1500	14.66	18.10	19.87	20.00	22.15	23.64
	2000	14.66	18.10	19.87	20.00	20.82	22.21
3	20	22.40	25.89	27.15	22.40	26.24	27.37
	500	24.46	26.16	27.22	24.54	26.65	27.49
	1000	19.14	21.62	23.35	23.80	25.97	27.40
	1500	19.14	21.62	23.35	23.56	26.10	27.33
	2000	19.14	21.62	23.35	24.72	25.78	27.26
4	20	0.00	0.00	0.00	0.00	0.00	0.00
	500	1.99	8.74	12.30	1.99	10.07	17.98
	1000	6.07	10.53	14.04	10.53	13.01	16.24
	1500	6.07	10.53	14.04	8.05	12.31	16.31
	2000	6.07	10.53	14.04	8.59	12.61	16.12
6	20	57.83	70.86	78.10	60.36	73.88	83.30
	500	36.83	61.74	72.66	51.55	72.63	79.14
	1000	38.01	51.38	63.31	71.33	78.21	85.35
	1500	38.01	51.38	63.31	75.96	80.44	85.03
	2000	38.01	51.38	63.31	81.74	88.17	91.65

five different switching times of 2000 function evaluations in total, listed by the number of function evaluations for the SMS-EMOA part: 20, 500, 1000, 1500, and 2000. Each version of the hybrid algorithm is repeated five times with different random seeds. The performance is evaluated by the hypervolume w. r. t. the reference point $\mathbf{r} = (5, 5)$ for ZDT1-4, and $\mathbf{r} = (10, 10)$ for ZDT6, and the same reference point is used for the internal hypervolume calculation. Other parameters are chosen as $n = 10$, $s_{\min} = 0.01$, $\tau = 0.8$. The minimal, average, and maximal hypervolume values calculated concerning the five repetitions of a strategy, are recorded after 1000 and 2000 evaluations. The results are not tested for statistical significance due to the low number of runs. The study shall just give rough tendencies. Re-

garding the averaged values, the hybrid optimizer using 20 or 500 as switching time outperforms the pure SMS-EMOA (corresponding to the line of ST= 2000) after 1000 function evaluations on all functions expect ZDT4. These differences become marginal after 2000 function evaluations, where the pure SMS-EMOA is occasionally achieves the best results. On ZDT4, it seems that SMS-EMOA with only 20 evaluations did not run long enough to generate a good starting population, so that the gradient method was not able to generate any point dominating the reference point. The more function evaluations are spend for the SMS-EMOA, the better the results on the multimodal functions ZDT4 and ZDT6. Concluding, we see that the gradient method may gain a speed-up, especially in the beginning of the optimization.

3.6.5 Conclusions

The gradient of the hypervolume of a set has been introduced and integrated into a hybrid metaheuristic. Using the chain rule, the gradient of the hypervolume can be computed from the gradients of the objective functions. It is important to distinguish between non-dominated, strictly dominated, and so called edge point, i.e., dominated points on the boundary of the dominated surface. While for non-dominated sets differentiability is inherited from the objective functions, in the presence of edge points one-sided derivatives occur. For strictly dominated points sub-gradients with value zero occur. They make it impossible to improve these points by means of gradient methods. This problem can be partly circumvented by introducing a penalty approach.

A steepest descent method guided by the experimental approximation of the hypervolume gradient is presented. Yet, it is not dedicated as a stand-alone optimizer as it requires a good starting set. So, a high-level relay hybridization with the SMS-EMOA is developed. The SMS-EMOA generates a rough approximation of the Pareto front which is then used as the initialization of the gradient optimizer which fine-tunes the positions of all points at once to a high precision approximation.

Proof of concept studies confirmed a fast convergence of the gradient method and demonstrated that the hybrid optimizer outperforms the original SMS-EMOA on the considered bi-objective functions. A study with a low budget of function evaluations revealed that very fast progress is made with sparse resources so that the hybrid method can especially be recommended for real-world applications where only few function evaluations are possible.

Moreover, the gradient method resolves the limitations of the SMS-EMOA regarding the selection scheme: Although its steady-state approach is highly effective, it is shown in Section 3.3 that the maximal hypervolume value is not always reachable by exchanging only one population member. This drawback is overcome by the gradient method that optimizes all positions of points at once.

An open problem is the gradient computation in case of more than two objectives which has so far only been sketched. Then, studies on higher dimensional problems are desirable. Besides, the switching time from the EMOA to the gradient is a parameter of the hybrid optimizer which deserves further investigation. Furthermore a sophisticated stopping criterion for the local optimizer is of interest.

4 Summary

This thesis concludes with the achieved core results and open problems. More detailed statements can be found in the respective conclusion sections above. The two main research topics are properties of the hypervolume calculation and characteristics of hypervolume-based evolutionary multiobjective optimization algorithms. Thereby, this thesis takes the challenging perspective of gaining mainly positive results and constructive solutions in complex scenarios that are relevant for practitioners.

The problem of calculating the hypervolume is known to belong to the complexity class $\#P$, yet this gives very abstract limits on the runtime of algorithms. We provide the first concrete lower bound of $\Omega(m \log m)$ for the hypervolume of m points in arbitrary dimensions, shown by reduction from a geometric problem. The lower bound is sharp for the hypervolume in two or three dimensions as known from matching upper bounds, resulting in the solved problem complexity of $\Theta(m \log m)$ for the dimension $d \in \{2, 3\}$. For the hypervolume in higher dimensions it is assumed that the problem complexity is actually higher, yet it is still the best known lower bound. Classifying the hypervolume as a special case of Klee's measure problem (KMP) grounded a new theoretical bases and allowed to transfer the knowledge from computational geometry. We here describe an adaption of the complex canonical KMP algorithm to the hypervolume calculation with significant simplifications while the runtime of $O(m^{d/2} \log m)$ is maintained. Further results built upon this algorithm. Our contributions and further known properties of the hypervolume are presented in simple and exact notation.

The main open challenges regarding the hypervolume are the further advancement of approximation algorithms, faster exact algorithms, and dimension-specific results of the problem complexity.

The presented study on characteristics of hypervolume-based evolutionary multiobjective optimization algorithms (EMOA) starts with convergence rates on bi-objective problems. The main tool is to establish algorithmic equivalence between EMOA and single-objective EA, namely for the SMS-EMOA on arbitrary bi-objective problems and an EA optimizing the sum of these objectives. These finding allows to transfer known convergence results from the single-objective EA to the SMS-EMOA or to analyze the structurally easier single-objective EA in order to gain new results. Several convergence rates are proved w. r. t. different mutation operators. The linear convergence rate proved for the SMS-EMOA is the first

known one for an EMOA without explicit weighting. The new proof techniques establish a solid framework for the development of further results.

Complementary to the speed of approaching the Pareto front discussed above, we separately considered the optimization process when the population is positioned on a continuous Pareto front. It is shown that the steady-state selection of exchanging one individual per generation may end in a local optimum of the hypervolume w. r. t. to a fixed reference point, where from the maximal value is unreachable. This local optimum may be a strong attractor also drawing in optimizers using other selection schemes. Experimental analyses on popular test problems with connected Pareto fronts showed that the optimal hypervolume is always reached by population-based versions of the SMS-EMOA. The hypervolume of linear Pareto fronts only has a global optimum that is therefore always achievable by steady-state selection. More analyses on the probabilistic solvability, i.e., the probability of reaching the global optimum, of other function classes are desired to explore the limits of this effective selection scheme. Note that the population does not get stuck when the hypervolume is calculated w. r. t. the adaptive reference point which supports this concept.

In the study on established test problems with many objectives, SMS-EMOA outperforms all other considered EMOA, emphasizing that the hypervolume is an expressive quality measure and suitable selection criterion also in high dimensional objective spaces. Besides, the ϵ-MOEA, MSOPS with a new weight vector generation method, and IBEA$_{\epsilon+}$ perform well. The older algorithms NSGA-II and SPEA2 completely fail to approach the Pareto front and even diverge. Exemplarily for NSGA-II it is demonstrated that the divergence can be avoided by rejecting boundary solutions instead of favoring them, which we consider as a general guideline for EMOA on many-objective problems.

A parameter tuning study reveals that the SMS-EMOA's performance can be improved a lot by adapting its parameters to the application problem. Contrarily to benchmarks in the literature, the choice of the variation operator plays a minor role, but the tuning causes the main improvements, resulting in more or less equal performance of the considered variation operators, while the tuned parameter values are quite different. We postulate that parameter tuning shall become standard in both real-world applications and academic benchmarks. We strongly put into question the common practice to tune an optimizer on test problems and to apply it with these tuned parameters to a real-world problem, due to the experiences that the optimal parameterizations differ significantly depending on the problem. Instead, it is recommended to perform the parameter tuning on a metamodel developed from the expensive application problem, which shall result in performance improvements compared to default parameterizations.

A gradient of the hypervolume regarding a population of points has been defined and calculated in order to allude to optimal positions of the points maximizing the hypervolume. Based on this technique, an optimizer is specified. The SMS-EMOA

is hybridized with this gradient approach by first executing the SMS-EMOA and then the gradient optimizer in relay. This way, the SMS-EMOA roughly approaches the Pareto front and the gradient technique performs a high precision approximation of the Pareto front. First studies on bi-objective problems give promising results. The method can especially be recommended in case of a low budget of function evaluations. The concept as yet needs to be worked out for more than two objectives.

Only continuous problems have been considered here but the algorithms—hypervolume algorithms as well as EMOA—are also applicable to discrete spaces.

In addition to proceeding the research above on hypervolume-based methods, multiobjective optimization faces the following major challenges.
The MO-CMA-ES provides a solid variation concept, but dedicated variation operators for the multiobjective case are to date still missing. It shall be promising to consider the relation of parents in the search space and objective space to estimate whether the resulting offspring are supposed to be dominated or not. Moreover, further methods for step size adaptation are of interest.
The integration of preferences deserves further effort in order to flexibly incorporate the user in the optimization process. To this end, EMOA can be hybridized with various techniques from multi-criteria decision making (MCDM).
Theoretical analyses of EMOA lack behind the knowledge on single-objective EA due to the more complex processes. It seems to be promising to transfer tools and insights from the single-objective case to the multiobjective optimization to overcome this gap. Some achievements have been accomplished, including our tools for convergence analyses, yet more are desirable.
The parameterized complexity seems to bear good respects for a different perspective of theoretical analyses. Traditionally, the runtime of EA is formulated w. r. t. the dimension of the search space. In many real-world problems, this input size is fixed and the user is interested in how the runtime of optimizers scales regarding other parameters. Some work has been done in this field but more is worthwhile.
Real-world problems in both single- or multiobjective optimization may be too demanding for exact evaluations of each candidate solution. The research field of metamodelling is a key to handle these problems and deserves further attention.

Hypervolume-based EMOA are currently the favored multiobjective optimizers. The question arises whether this is the end of the road or other concepts may fulfill the practitioner's needs even better.

Nomenclature

Notation

$\mathcal{M}(\mathbb{R}^d)$	set of all multisets with elements from $\mathbb{R}^d$
$\mathbf{v}$	bold, roman letter denotes a point or column vector
$\mathbf{v}^\top$	column vector transposed to row vector
$\mathbf{v}^{(i)}$	superscript denotes that $\mathbf{v}$ is the ith component within a sorting
$\{\{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(m)}\}\}$	double curly brackets denote multiset
v_i	subscript denotes the ith component of vector $\mathbf{v}$
set	transformation from multiset to set by removing copies

Multiobjective Optimization

$\mathbf{a} \parallel \mathbf{b}$	$\mathbf{a}$ and $\mathbf{b}$ are incomparable
$\mathbf{a} \prec \mathbf{b}$	$\mathbf{a}$ dominates $\mathbf{b}$
$\mathbf{a} \preceq \mathbf{b}$	$\mathbf{a}$ weakly dominates $\mathbf{b}$
$\mathbf{a} \prec\!\prec \mathbf{b}$	$\mathbf{a}$ strongly dominates $\mathbf{b}$
$\mathbf{nad}(M)$	nadir point of multiset M
d	dimension of objective space of multiobjective problem
n	dimension of search space of multiobjective problem
$ndms(M)$	non-dominated multiset of multiset M
$nds(M)$	non-dominated set of multiset M

Quality Assessment

$\mathbf{r}$	reference point bounding the hypervolume $H(M, \mathbf{r})$
$H(\mathbf{y}, M, \mathbf{r})$	hypervolume contribution of $\mathbf{v} \in M$ to the hypervolume of M
$H(M, \mathbf{r})$	hypervolume of multiset M bounded by reference point $\mathbf{r}$
Leb	Lebesgue measure

Parameters of Evolutionary Algorithms

η_c	distribution index of simulated binary crossover (SBX)
η_m	distribution index of polynomial mutation (PM)
λ	offpring population size
μ	population size
CR	crossover rate of DE
F	scaling parameter of DE
N	normal distribution
n	number of decision variables, size of search space
$P^{(t)}$	population in generation t
p_c	application probability of crossover operator
p_m	application probability of mutation operator
U	uniform distribution

Abbreviations

DE	Differential Evolution
DTLZ	Test problems by Deb, Thiele, Laumanns, and Zitzler
EA	Evolutionary Algorithm
EMOA	Evolutionary Multiobjective Optimization Algorithm, MOEA

KMP	Klee's Measure Problem
MOEA	Multi-Objective Evolutionary Optimization Algorithm, EMOA
PM	Polynomial Mutation
SBX	Simulated Binary Crossover
SMS-EMOA	S-metric Selection EMOA
UG	Uniform Gap
ZDT	Test problems by Zitzler, Deb, and Thiele

List of Figures

List of Tables

List of Algorithms

Bibliography

A. Abraham and R. Goldberg, editors. *Evolutionary Multiobjective Optimization: Theoretical Advances and Applications*. Springer, Berlin, 2005.

S. F. Adra and P. J. Fleming. Diversity management in evolutionary many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 15 (2):183–195, 2011.

A. Auger and B. Doerr, editors. *Theory of Randomized Search Heuristics: Foundations and Recent Developments*. World Scientific Publishing, London, 2011.

A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the hypervolume indicator: Optimal μ-distributions and the choice of the reference point. In I. Garibay et al., editors, *Proc. of the 10th ACM SIGEVO workshop on Foundations of Genetic Algorithms (FOGA 2009)*, pages 87–102. ACM, New York, NY, 2009a.

A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Articulating user preferences in many-objective problems by sampling the weighted hypervolume. In G. Raidl et al., editors, *Proc. of the 11th Annual conference on Genetic and Evolutionary Computation (GECCO 2009)*, pages 555–562. ACM, New York, NY, 2009b.

A. Auger, J. Bader, and D. Brockhoff. Theoretically investigating optimal μ-distributions for the hypervolume indicator: First results for three objectives. In R. Schaefer et al., editors, *Proc. of the 11th International Conference on Parallel Problem Solving from Nature – PPSN XI: Part I*, volume 6238 of *Lecture Notes in Computer Science*, pages 586–596. Springer, Berlin, 2010.

T. Bäck, D. Fogel, and Z. Michalewicz, editors. *Handbook of Evolutionary Computation*. IOP Publishing and Oxford University Press, Bristol/New York, 1997.

J. Bader and E. Zitzler. HypE: an algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 19(1):45–76, 2011.

T. Bartz-Beielstein, C. Lasarczyk, and M. Preuß. Sequential parameter optimization. In B. McKay et al., editors, *Proc. of the 2005 IEEE Congress on Evolutionary Computation (CEC 2005)*, volume 1, pages 773–780. IEEE Press, Piscataway, NJ, 2005.

T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, editors. *Experimental Methods for the Analysis of Optimization Algorithms*. Springer, Berlin, 2010a.

T. Bartz-Beielstein, C. Lasarczyk, and M. Preuss. The sequential parameter optimization toolbox. In T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, editors, *Experimental Methods for the Analysis of Optimization Algorithms*, pages 337–362. Springer, Berlin, 2010b.

J. L. Bentley. Algorithms for Klee's rectangle problem. Unpublished notes, Department of Computer Science, CMU, 1977.

N. Beume. *Hypervolumen-basierte Selektion in einem evolutionären Algorithmus zur Mehrzieloptimierung*. Diploma thesis, University of Dortmund, Department of Computer Science, March 2006.

N. Beume. S-metric calculation by considering dominated hypervolume as Klee's measure problem. *Evolutionary Computation*, 17(4):477–492, 2009.

N. Beume and G. Rudolph. Faster S-metric calculation by considering dominated hypervolume as Klee's measure problem. In B. Kovalerchuk, editor, *Proc. of the IASTED International Conference on Computational Intelligence (CI 2006)*, pages 231–236. ACTA Press, Anaheim, CA, 2006a.

N. Beume and G. Rudolph. Faster S-metric calculation by considering dominated hypervolume as Klee's measure problem. Technical Report of the Collaborative Research Center 531 *Computational Intelligence* CI–216/06, Universität Dortmund, 2006b.

N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.

N. Beume, B. Naujoks, and G. Rudolph. SMS-EMOA: Effektive evolutionäre Mehrzieloptimierung. *at-Automatisierungstechnik*, 56(7):357–364, 2008.

N. Beume, C. M. Fonseca, M. López-Ibáñez, L. Paquete, and J. Vahrenhold. On the complexity of computing the hypervolume indicator. *IEEE Transactions on Evolutionary Computation*, 13(5):1075–1082, 2009a.

N. Beume, B. Naujoks, M. Preuss, G. Rudolph, and T. Wagner. Effects of 1-greedy S-metric-selection on innumerably large pareto fronts. In M. Ehrgott et al., editors, *Proc. of the 5th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2009)*, volume 5467 of *Lecture Notes in Computer Science*, pages 21–35. Springer, Berlin, 2009b.

N. Beume, M. Laumanns, and G. Rudolph. Convergence rates of (1+1) evolutionary multiobjective algorithms. In R. Schaefer et al., editors, *Proc. of the 11th International Conference on Parallel Problem Solving from Nature – PPSN XI: Part I*, volume 6238 of *Lecture Notes in Computer Science*, pages 597–606. Springer, Berlin, 2010.

N. Beume, M. Laumanns, and G. Rudolph. Convergence rates of SMS-EMOA on continuous bi-objective problem classes. In H.-G. Beyer and W. B. Langdon, editors, *Proc. of the 11th ACM SIGEVO workshop on Foundations of Genetic Algorithms (FOGA 2011)*, pages 243–252. ACM, New York, NY, 2011.

S. Bleuler, M. Laumanns, L. Thiele, and E. Zitzler. PISA — a platform and programming language independent interface for search algorithms. In C. M. Fonseca et al., editors, *Proc. of the 2nd International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, volume 2632 of *Lecture Notes in Computer Science*, pages 494–508. Springer, Berlin, 2003. URL `http://www.tik.ee.ethz.ch/pisa/`.

P. A. Bosman and E. D. de Jong. Combining gradient techniques for numerical multi-objective evolutionary optimization. In M. Keijzer et al., editors, *Proc. of the 8th Annual conference on Genetic and Evolutionary Computation (GECCO 2006)*, volume 1, pages 627–634. ACM, New York, NY, 2006.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2006.

J. Branke, K. Deb, K. Miettinen, and R. Slowinski, editors. *Multiobjective Optimization: Interactive and Evolutionary Approaches*, volume 5252 of *Lecture Notes in Computer Science*. Springer, Berlin, 2008.

K. Bringmann and T. Friedrich. Approximating the volume of unions and intersections of high-dimensional geometric objects. In S.-H. Hong et al., editors, *Proc. of the 19th International Symposium on Algorithms and Computation (ISAAC 2008)*, volume 5369 of *Lecture Notes in Computer Science*, pages 436–447. Springer, Berlin, 2008.

K. Bringmann and T. Friedrich. Approximating the least hypervolume contributor: NP-hard in general, but fast in practice. In M. Ehrgott et al., editors, *Proc. of the 5th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2009)*, volume 5467 of *Lecture Notes in Computer Science*, pages 6–20. Springer, Berlin, 2009a.

K. Bringmann and T. Friedrich. Don't be greedy when calculating hypervolume contributions. In I. Garibay et al., editors, *Proc. of the 10th ACM SIGEVO workshop on Foundations of Genetic Algorithms (FOGA 2009)*, pages 103–112. ACM, New York, NY, 2009b.

K. Bringmann and T. Friedrich. The maximum hypervolume set yields near-optimal approximation. In M. Pelikan and J. Branke, editors, *Proc. of the 12th Annual conference on Genetic and Evolutionary Computation (GECCO 2010)*, pages 511–518. ACM, New York, NY, 2010a.

K. Bringmann and T. Friedrich. Tight bounds for the approximation ratio of the hypervolume indicator. In R. Schaefer et al., editors, *Proc. of the 11th International Conference on Parallel Problem Solving from Nature – PPSN XI: Part I*, volume 6238 of *Lecture Notes in Computer Science*, pages 607–616. Springer, Berlin, 2010b.

K. Bringmann and T. Friedrich. Convergence of hypervolume-base archiving algorithms I: Effectiveness. In N. Krasnogor and P. L. Lanzi, editors, *Proc. of the 13th Annual conference on Genetic and Evolutionary Computation (GECCO 2011)*, pages 745–752. ACM, New York, NY, 2011.

K. Bringmann, T. Friedrich, F. Neumann, and M. Wagner. Approximation-guided evolutionary multi-objective optimization. In T. Walsh, editor, *Proc. of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, pages 1198–1203, 2011.

D. Brockhoff. Optimal μ-distributions for the hypervolume indicator for problems with linear bi-objective fronts: Exact and exhaustive results. In K. Deb et al., editors, *Simulated Evolution and Learning (SEAL 2010)*, volume 6457 of *Lecture Notes in Computer Science*, pages 24–34. Springer, Berlin, 2010a.

D. Brockhoff. Theoretical aspects of evolutionary multiobjective optimization. In A. Auger and B. Doerr, editors, *Theory of Randomized Search Heuristics: Foundations and Recent Developments*, pages 101–139. World Scientific Publishing, London, 2010b.

D. Brockhoff, T. Friedrich, and F. Neumann. Analyzing hypervolume indicator based algorithms. In G. Rudolph et al., editors, *Proc. of the 10th International Conference on Parallel Problem Solving from Nature – PPSN X*, Lecture Notes in Computer Science, pages 651–660. Springer, Berlin, 2008.

T. M. Chan. A (slightly) faster algorithm for klee's measure problem. *Computational Geometry*, 43:243–250, 2010.

C. A. Coello Coello, D. A. Van Veldhuizen, and G. B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York, 2002.

K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. Wiley, Chichester, UK, 2001.

K. Deb and R. B. Agrawal. Simulated binary crossover for continuous search space. *Complex Systems*, 9:115–148, 1995.

K. Deb and M. Goyal. A combined genetic adaptive search (GeneAS) for engineering design. *Computer Science and Informatics*, 26(4):30–45, 1996.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002a.

K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable Multi-objective Optimization Test Problems. In D. B. Fogel et al., editors, *Proc. of the 2002 IEEE Congress on Evolutionary Computation (CEC 2002)*, volume 1, pages 825–830. IEEE Press, Piscataway, NJ, 2002b.

K. Deb, M. Mohan, and S. Mishra. A Fast Multi-objective Evolutionary Algorithm for Finding Well-Spread Pareto-Optimal Solutions. KanGAL report 2003002, Indian Institute of Technology, Kanpur, India, 2003.

K. Deb, M. Mohan, and S. Mishra. Evaluating the ϵ-domination based multi-objective evolutionary algorithm for a quick computation of pareto-optimal solutions. *Evolutionary Computation*, 13(4):501–525, 2005a.

K. Deb et al. *NSGA-II. Implementation in C.* Kanpur Genetic Algorithms Laboratory (KanGAL), Indian Institute of Technology, Kanpur, 2005b. URL `http://www.iitk.ac.in/kangal/codes/nsga2/`.

R. P. Dilworth. A decomposition theorem for partially ordered sets. *Annals of Mathematics*, 51:161–165, 1950.

J. Dréo. Using performance fronts for parameter setting of stochastic metaheuristics. In G. Raidl et al., editors, *Proc. of the 11th Annual conference on Genetic and Evolutionary Computation (GECCO 2009)*, pages 2197–2200. ACM, New York, NY, 2009.

H. Edelsbrunner and M. H. Overmars. Batched dynamic solutions to decomposable searching problems. *Journal of Algorithms*, 6(4):515–542, 1985.

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

A. Eiben and J. Smith. *Introduction to Evolutionary Algorithms.* Natural Computing Series. Springer, Berlin, 2nd edition, 2007.

A. E. Eiben and S. K. Smit. Parameter tuning for configuring and analyzing evolutionary algorithms. *Swarm and Evolutionary Computation*, 1(1):19–31, 2011.

M. Emmerich and A. Deutz. Test Problems based on Lamé Superspheres. In S. Obayashi et al., editors, *EMO07*, LNCS 4403, pages 922–936. Springer, Berlin, 2007.

M. Emmerich, N. Beume, and B. Naujoks. An EMO algorithm using the hypervolume measure as selection criterion. In C. A. Coello Coello et al., editors, *Proc. of the 3rd International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, volume 3410 of *Lecture Notes in Computer Science*, pages 62–76. Springer, Berlin, 2005.

M. Emmerich, K. Giannakoglou, and B. Naujoks. Single- and multi-objective evolutionary optimization assisted by gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10(4):421–439, 2006.

M. Emmerich, A. Deutz, and N. Beume. Gradient-based/evolutionary relay hybrid for computing pareto front approximations maximizing the S-Metric. In T. Bartz-Beielstein et al., editors, *Proc. of the 4th International Workshop on Hybrid Metaheuristics (HM 2007)*, volume 4771 of *Lecture Notes in Computer Science*, pages 140–157. Springer, Berlin, 2007.

A. Engelbrecht. *Computational Intelligence: An Introduction*. John Wiley & Sons, NJ, 2nd edition, 2007.

R. Everson, J. Fieldsend, and S. Singh. Full elite-sets for multiobjective optimisation. In I. Parmee, editor, *Conference on adaptive computing in design and manufacture (ADCM 2002)*, pages 343–354. Springer, Berlin, 2002.

M. Farina and P. Amato. On the optimal solution definition for many-criteria optimization problems. In J. Keller and O. Nasraoui, editors, *Proc. of the International Conference NAFIPS-FLINT 2002*, pages 233–238. IEEE Press, Piscataway, NJ, 2002.

M. Fleischer. The measure of pareto optima: Applications to multi-objective metaheuristics. In C. M. Fonseca et al., editors, *Proc. of the 2nd International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, volume 2632 of *Lecture Notes in Computer Science*, 2003.

J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.

J. Fliege, L. M. G. Drummond, and B. Svaiter. Newton's method for multiobjective optimization. *SIAM Journal on Optimization*, 20:602–626, 2009.

C. M. Fonseca, L. Paquete, and M. López-Ibáñez. An improved dimension-sweep algorithm for the hypervolume indicator. In G. G. Yen et al., editors, *Proc. of the 2006 IEEE Congress on Evolutionary Computation (CEC 2006) within the 2006*

IEEE World Congress on Computational Intelligence, pages 1157–1163. IEEE Press, Piscataway, NJ, 2006.

M. L. Fredman and B. Weide. The complexity of computing the measure of $\bigcup[a_i, b_i]$. In *Communications of the ACM*, volume 21, pages 540–544, 1978.

T. Friedrich, C. Horoba, and F. Neumann. Multiplicative approximations and the hypervolume indicator. In F. Rothlauf, editor, *Proc. of the 11th Annual conference on Genetic and Evolutionary Computation (GECCO 2009)*, pages 571–578. ACM, New York, NY, 2009.

T. Friedrich, K. Bringmann, T. Voß, and C. Igel. The logarithmic hypervolume indicator. In H.-G. Beyer and W. B. Langdon, editors, *Proc. of the 11th ACM SIGEVO workshop on Foundations of Genetic Algorithms (FOGA 2011)*, pages 81–92. ACM, New York, NY, 2011.

D. E. Goldberg. *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley, Reading, MA, 1989.

Google Scholar. Queried by scHolar index, 2011. URL `http://interaction.lille.inria.fr/~roussel/projects/scholarindex`.

V. Grunert da Fonseca and C. M. Fonseca. The attainment-function approach to stochastic multiobjective optimizer assessment and comparison. In T. Bartz-Beielstein et al., editors, *Experimental Methods for the Analysis of Optimization Algorithms*, pages 103–130. Springer, Berlin, 2010.

T. Hanne. On the convergence of multiobjective evolutionary algorithms. *European Journal of Operational Research*, 117(3):553–564, 1999.

N. Hansen. The CMA evolution strategy: A comparing review. In J. Lozano et al., editors, *Towards a new evolutionary computation. Advances in estimation of distribution algorithms*, pages 75–102. Springer, Berlin, 2006.

N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

N. Hansen, S. D. Müller, and P. Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Springer, Berlin, 2001.

M. Hollander and D. A. Wolfe. *Nonparametric Statistical Methods*. John Wiley & Sons, New York, 1973.

R. Hooke and T. Jeeves. Direct search: Solution of numerical and statistical problems. *Journal of the ACM*, 8(2):212–229, 1961.

V. L. Huang, A. K. Qin, K. Deb, E. Zitzler, P. N. Suganthan, J. J. Liang, M. Preuss, and S. Huband. Problem definitions for performance assessment of multi-objective optimization algorithms. Technical report, Nanyang Technological University, Singapore, 2007. `http://www3.ntu.edu.sg/home/epnsugan/index_files/CEC-07/CEC07.htm`.

S. Huband, P. Hingston, L. While, and L. Barone. An evolution strategy with probabilistic mutation for multi-objective optimisation. In R. Sarker et al., editors, *Proc. of the 2003 IEEE Congress on Evolutionary Computation (CEC 2003)*, volume 4, pages 2284–2291. IEEE Press, Piscataway, NJ, 2003.

E. J. Hughes. Multiple single objective pareto sampling. In R. Sarker et al., editors, *Proc. of the 2003 IEEE Congress on Evolutionary Computation (CEC 2003)*, volume 4, pages 2678–2684. IEEE Press, Piscataway, NJ, 2003.

E. J. Hughes. Evolutionary many-objective optimisation: Many once or one many? In B. McKay et al., editors, *Proc. of the 2005 IEEE Congress on Evolutionary Computation (CEC 2005)*, volume 1, pages 222–227. IEEE Press, Piscataway, NJ, 2005.

E. J. Hughes. MSOPS-II: A general-purpose many-objective optimiser. In D. Srinivasan and L. Wang, editors, *Proc. of the 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, pages 3944–3951. IEEE Press, Piscataway, NJ, 2007.

E. J. Hughes. Many-objective directed evolutionary line search. In N. Krasnogor and P. L. Lanzi, editors, *Proc. of the 13th Annual conference on Genetic and Evolutionary Computation (GECCO 2011)*, pages 761–768. ACM, New York, NY, 2011.

C. Igel, N. Hansen, and S. Roth. Covariance matrix adaptation for multi-objective optimization. *Evolutionary Computation*, 15(1):1–28, 2007.

H. Ishibuchi, N. Tsukamoto, and Y. Nojima. Evolutionary manyobjective optimization: A short review. In J. Wang et al., editors, *Proc. of the 2008 IEEE Congress on Evolutionary Computation (CEC 2008) within the 2008 IEEE World Congress on Computational Intelligence*, pages 2424–2431. IEEE Press, Piscataway, NJ, 2008.

J. Jägersküpper. How the (1+1) ES using isotropic mutations minimizes positive definite quadratic forms. *Theoretical Computer Science*, 361(1):38–56, 2006.

J. Jägersküpper and C. Witt. Rigorous runtime analysis of a $(\mu + 1)$ es for the sphere function. In H.-G. Beyer and U.-M. O'Reilly, editors, *Proc. of the 7th*

Annual conference on Genetic and Evolutionary Computation (GECCO 2005), pages 849–856, 2005.

M. T. Jensen. Reducing the run-time complexity of multiobjective EAs: The NSGA-II and other algorithms. *IEEE Transactions on Evolutionary Computation*, 7(5):503–515, 2003.

V. Klee. Can the measure of $\bigcup_1^n [a_i, b_i]$ be computed in less than $O(n \log n)$ steps? *American Mathematics Monthly*, 84(4):284–285, 1977.

J. Klinkenberg, T. Emmerich, A. Deutz, O. Shir, and T. Bäck. A reduced-cost SMS-EMOA using kriging, self-adaptation, and parallelization. In M. Ehrgott et al., editors, *Proc. of the Conference on Multicriteria Decision Making for Sustainable Energy and Transportation Systems (MCDM 2008), part 4*, volume 634 of *Lecture Notes on Economy and Mathematical Systems*, pages 301–311. Springer, Berlin, 2010.

J. Knowles and D. Corne. Properties of an adaptive archiving algorithm for storing nondominated vectors. *IEEE Transactions on Evolutionary Computation*, 7(2): 100–116, 2003.

J. Knowles, L. Thiele, and E. Zitzler. A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers. TIK Report 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Feb. 2006.

J. Knowles, D. Corne, and K. Deb, editors. *Multiobjective Problem Solving from Nature: From Concepts to Applications*. Natural Computing Series. Springer, Berlin, 2007.

J. D. Knowles. *Local Search and Hybrid Evolutionary Algorithms for Pareto Optimization*. PhD thesis, University of Reading, Reading, UK, January 2002.

P. Koch, O. Kramer, G. Rudolph, and N. Beume. On the hybridization of SMS-EMOA and local search for continuous multiobjective optimization. In F. Rothlauf, editor, *Proc. of the 11th Annual conference on Genetic and Evolutionary Computation (GECCO 2009)*, pages 603–610. ACM, New York, NY, 2009.

A. Konar. *Computational Intelligence: Principles, Techniques and Applications*. Springer, Berlin, 2005.

D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. Wiley Series in Probability and Statistics. Wiley, Chichester, UK, 2011.

S. Kukkonen and J. Lampinen. Performance assessment of generalized differential evolution 3 (GDE3) with a given set of problems. In D. Srinivasan and L. Wang, editors, *Proc. of the 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, pages 3593–3600. IEEE Press, Piscataway, NJ, 2007.

H. T. Kung, F. Luccio, and F. P. Preparata. On finding the maxima of a set of vectors. *Journal of the ACM*, 22(4):469–476, Oct. 1975.

F. Kursawe. A variant of evolution strategies for vector optimization. In H.-P. Schwefel and R. Männer, editors, *Proc. of the 1st International Conference on Parallel Problem Solving from Nature – PPSN I*, volume 496 of *Lecture Notes in Computer Science*, pages 193–197. Springer, Berlin, 1991.

M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multi-objective optimization. *Evolutionary Computation*, 10(3):263–282, 2002.

R. Li, R. Etemaadi, M. T. M. Emmerich, and M. R. V. Chaudron. An evolutionary multiobjective optimization approach to component-based software architecture design. In A. Smith, editor, *Proc. of the 2011 IEEE Congress on Evolutionary Computation (CEC 2011)*, pages 432–439. IEEE Press, Piscataway, NJ, 2011.

S. N. Lophaven, H. B. Nielsen, and J. Søndergaard. DACE – A Matlab kriging toolbox. Technical Report IMM-REP-2002-12, Informatics and Mathematical Modelling, Technical University of Denmark, Copenhagen, Denmark, 2002. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.17.3530`.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus*. Wiley, Chichester, UK, revised edition, 1999.

Z. Michalewicz, C. F. Lima, and F. G. Lobo, editors. *Parameter Setting in Evolutionary Algorithms*, volume 54 of *Studies in Computational Intelligence*. Springer, Berlin, 2007.

K. Miettinen. *Nonlinear Multiobjective Optimization*, volume 12 of *Kluwer's International Series in Operations Research & Management Science*. Kluwer Academic Publishers, 1999.

D. C. Montgomery. *Design and Analysis of Experiments*. Wiley, New York, 4th edition, 1997.

S. Mostaghim, J. Branke, and H. Schmeck. Multi-objective particle swarm optimization on computer grids. In D. Thierens et al., editors, *Proceedings of the 2007 Genetic and Evolutionary Computation Conference*, pages 869–874. ACM, New York, NY, 2007.

V. Nannen and A. Eiben. Efficient relevance estimation and value calibration of evolutionary algorithm parameters. In D. Srinivasan and L. Wang, editors, *Proc. of the 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, pages 103–110. IEEE Press, Piscataway, NJ, 2007.

B. Naujoks, L. Willmes, T. Bäck, and W. Haase. Evaluating multi-criteria evolutionary algorithms for airfoil optimisation. In J. J. M. Guervós et al., editors, *Proc. of the 7th International Conference on Parallel Problem Solving from Nature – PPSN VII*, pages 841–850. Springer, Berlin, 2002.

B. Naujoks, N. Beume, and M. Emmerich. Multi-objective optimisation using S-metric selection: Application to three-dimensional solution spaces. In B. McKay et al., editors, *Proc. of the 2005 IEEE Congress on Evolutionary Computation (CEC 2005)*, volume 2, pages 1282–1289. IEEE Press, Piscataway, NJ, 2005a.

B. Naujoks, N. Beume, and M. Emmerich. Metamodel-assisted SMS-EMOA applied to airfoil optimization tasks. In R. Schilling et al., editors, *Proc. EUROGEN'05* (CD-ROM). FLM, TU München, 2005b.

B. Naujoks, D. Quagliarella, and T. Bartz-Beielstein. Sequential parameter optimisation of evolutionary algorithms for airfoil design. In *ERCOFTAC 2006 Design and Optimization: Methods & Applications*, pages 231–235. European Research Community on Flow, Turbulence and Combustion (ERCOFTAC), 2006.

A. Ostermeier, A. Gawelczyk, and N. Hansen. Step-size adaptation based on non-local use of selection information. In Y. Davidor et al., editors, *Proc. of the 3rd International Conference on Parallel Problem Solving from Nature – PPSN III*, volume 866 of *Lecture Notes in Computer Science*, pages 189–198. Springer, Berlin, 1994.

M. H. Overmars and C.-K. Yap. New upper bounds in Klee's measure problem. *SIAM Journal on Computing*, 20(6):1034–1045, 1991.

T. Pierrard. *Multi-Objective Artificial Immune System Algorithm based on Hypervolume*. Master thesis, CINVESTAV, Departamiento de computación, Mexico, Col. San Pedro Zacatenco, Mexico, 2011.

F. P. Preparata and M. I. Shamos. *Computational Geometry. An Introduction*. Springer, Berlin, 2nd edition, 1988.

M. Preuss. Reporting on experiments in evolutionary computation. Technical Report of the Collaborative Research Center 531 *Computational Intelligence* CI–221/07, Universität Dortmund, 2007.

M. Preuss, G. Rudolph, and S. Wessing. Tuning optimization algorithms for real-world problems by means of surrogate modeling. In M. Pelikan and J. Branke, editors, *Proc. of the 12th Annual conference on Genetic and Evolutionary Computation (GECCO 2010)*, pages 401–408. ACM, New York, NY, 2010.

R. C. Purshouse and P. J. Fleming. Evolutionary multi-objective optimisation: An exploratory analysis. In R. Sarker et al., editors, *Proc. of the 2003 IEEE Congress*

on Evolutionary Computation (CEC 2003), volume 3, pages 2066–2073. IEEE Press, Piscataway, NJ, 2003.

E. Reehuis, J. Kruisselbrink, A. Deutz, T. Baeck, and M. Emmerich. Multiobjective optimization of water distribution networks using SMS-EMOA. In C. Poloni et al., editors, *Evolutionary and Deterministic Methods for Design, Optimisation and Control*, pages 269–279, 2011.

G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Kovač, Hamburg, 1997.

G. Rudolph. On a multi–objective evolutionary algorithm and its convergence to the Pareto set. In *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation*, pages 511–516. IEEE Press, Piscataway, NJ, 1998.

J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989.

J. D. Schaffer. Multiple objective optimization with vector evaluated genetic algorithms. In J. J. Grefenstette, editor, *Proc. of the 1st International Conference on Genetic Algorithms (ICGA)*, pages 93–100. Lawrence Erlbaum, 1985.

O. Schütze, A. Dell'Aere, and M. Dellnitz. On continuation methods for the numerical treatment of multi-objective optimization problems. In *Practical Approaches to Multi-Objective Optimization*, volume 04461 of *Dagstuhl Seminar Proceedings*. IBFI, Schloss Dagstuhl, Wadern, 2005.

H. Schwefel. *Evolution and Optimum Seeking: The Sixth Generation*. John Wiley & Sons, Inc., 1995.

P. K. Shukla. Gradient based stochastic mutation operators in evolutionary multi-objective optimization. In B. Beliczynski et al., editors, *Proc. of the 8th international conference on Adaptive and Natural Computing Algorithms, Part I (ICANNGA 2007)*, pages 58–66. Springer, Berlin, 2007.

B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.

K. Sindhya, K. Deb, and K. Miettinen. A local search based evolutionary multi-objective optimization approach for fast and accurate convergence. In G. Rudolph et al., editors, *Parallel Problem Solving from Nature - PPSN X*, volume 5199 of *Lecture Notes in Computer Science*, pages 815–824. Springer, Berlin, 2008.

S. Smit and A. Eiben. Comparing parameter tuning methods for evolutionary algorithms. In P. Haddow et al., editors, *Proc. of the 2009 IEEE Congress on*

Evolutionary Computation (CEC 2009), pages 399–406. IEEE Press, Piscataway, NJ, 2009.

S. Smit, A. Eiben, and Z. Szlávik. An moea-based method to tune ea parameters on multiple objective functions. In J. Filipe and J. Kacprzyk, editors, *Proc. of the International Joint Conference on Computational Intelligence (IJCCI 2010)*, pages 261–268. SciTePress, Valencia, 2010.

R. Storn and K. Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11 (4):341–359, 1997.

T. Suttorp, N. Hansen, and C. Igel. Efficient covariance matrix update for variable metric evolution strategies. *Machine Learning*, 75(2):167–197, 2009.

E.-G. Talbi. A Taxonomy of Hybrid Metaheuristics. *Journal of Heuristics*, 8(5): 541–564, 2002.

O. Teytaud. On the hardness of offline multi-objective optimization. *Evolutionary Computation*, 15(4):475–491, 2007.

A. v. d. Kuijl, M. T. M. Emmerich, and H. Li. A robust multi-objective resource allocation scheme incorporating uncertainty and service differentiation. *Concurrency and Computation: Practice and Experience*, 22(3):314–328, 2010.

V. V. Vazirani. *Approximation Algorithms*. Springer, Berlin, 2004.

VDI-Fachbereich Bionik. *VDI guideline: VDI 6224 Blatt 1. Bionic optimization - Application of evolutionary algorithms*. VDI e.V., Düsseldorf, 2011.

T. Voß, N. Beume, G. Rudolph, and C. Igel. Scalarization versus indicator-based selection in multi-objective CMA evolution strategies. In J. Wang et al., editors, *Proc. of the 2008 IEEE Congress on Evolutionary Computation (CEC 2008) within the 2008 IEEE World Congress on Computational Intelligence*, pages 3041–3048. IEEE Press, Piscataway, NJ, 2008.

T. Voß, N. Hansen, and C. Igel. Improved step size adaptation for the MO-CMA-ES. In M. Pelikan and J. Branke, editors, *Proc. of the 12th Annual conference on Genetic and Evolutionary Computation (GECCO 2010)*, pages 487–494. ACM, New York, NY, 2010.

T. Wagner, N. Beume, and B. Naujoks. Pareto-, aggregation-, and indicator-based methods in many-objective optimization. Technical Report of the Collaborative Research Center 531 *Computational Intelligence* CI–217/06, Universität Dortmund, 2006.

T. Wagner, N. Beume, and B. Naujoks. Pareto-, aggregation-, and indicator-based methods in many-objective optimization. In S. Obayashi et al., editors, *Proc. of the 4th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, volume 4403 of *Lecture Notes in Computer Science*, pages 742–756. Springer, Berlin, 2007a.

T. Wagner, T. Michelitsch, and A. Sacharow. On the design of optimisers in surface reconstruction. In D. Thierens et al., editors, *Proc. of the 9th Annual conference on Genetic and Evolutionary Computation (GECCO 2007)*, pages 2195–2002. ACM, New York, NY, 2007b.

I. Wegener. *Complexity theory: exploring the limits of efficient algorithms*. Springer, Berlin, 2005.

S. Wessing. Towards optimal parameterizations of the S-metric selection evolutionary multi-objective algorithm. Algorithm Engineering Report TR09-2-006, Technische Universität Dortmund, 2009.

S. Wessing and B. Naujoks. Sequential Parameter Optimization for Multi-Objective Problems. In G. Fogel and H. Ishibuchi, editors, *Congress on Evolutionary Computation (CEC) within the World Congress on Computational Intelligence (WCCI)*, pages 1–8. IEEE Press, Piscataway, NJ, 2010.

S. Wessing, N. Beume, G. Rudolph, and B. Naujoks. Parameter tuning boosts performance of variation operators in multiobjective optimization. In R. Schaefer et al., editors, *Proc. of the 11th International Conference on Parallel Problem Solving from Nature – PPSN XI: Part I*, volume 6238 of *Lecture Notes in Computer Science*, pages 728–737. Springer, Berlin, 2010.

L. While. A new analysis of the LebMeasure algorithm for calculating hypervolume. In C. A. Coello Coello et al., editors, *Proc. of the 3rd International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, volume 3410 of *Lecture Notes in Computer Science*, pages 326–340. Springer, Berlin, 2005.

L. While, L. Bradstreet, L. Barone, and P. Hingston. Heuristics for optimising the calculation of hypervolume for multi-objective optimisation problems. In B. McKay et al., editors, *Proc. of the 2005 IEEE Congress on Evolutionary Computation (CEC 2005)*, volume 3, pages 2225–2232. IEEE Press, Piscataway, NJ, 2005.

L. While, P. Hingston, L. Barone, and S. Huband. A faster algorithm for calculating hypervolume. *IEEE Transactions on Evolutionary Computation*, 10(1):29–38, 2006.

E. Zitzler. *Hypervolume Metric Calculation.* Computer Engineering and Networks Laboratory (TIK), Zürich, 2001. ftp://ftp.tik.ee.ethz.ch/pub/people/zitzler/hypervol.c.

E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In X. Yao et al., editors, *Proc. of the 8th International Conference on Parallel Problem Solving from Nature – PPSN VIII*, LNCS 1498, pages 832–842. Springer, Berlin, 2004.

E. Zitzler and L. Thiele. Multiobjective Optimization Using Evolutionary Algorithms—A Comparative Case Study. In A. E. Eiben et al., editors, *Proc. of the 5th International Conference on Parallel Problem Solving from Nature – PPSN V*, volume 1498 of *Lecture Notes in Computer Science*, pages 292–301. Springer, Berlin, 1998.

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

E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. Technical Report 103, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zürich, 2001.

E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. In K. Giannakoglou et al., editors, *Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems (EUROGEN 2001)*. CIMNE, 2002.

E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.

E. Zitzler, D. Brockhoff, and L. Thiele. The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. In S. Obayashi et al., editors, *Proc. of the 4th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, volume 4403 of *Lecture Notes in Computer Science*, pages 862–876. Springer, Berlin, 2007.

E. Zitzler, J. Knowles, and L. Thiele. Quality assessment of pareto set approximations. In J. Branke et al., editors, *Multiobjective Optimization: Interactive and Evolutionary Approaches*, pages 373–404. Springer, Berlin, 2008a.

E. Zitzler, L. Thiele, and J. Bader. SPAM: Set preference algorithm for multiobjective optimization. In G. Rudolph et al., editors, *Proc. of the 10th International Conference on Parallel Problem Solving from Nature – PPSN X*, pages 847–858. Springer, Berlin, 2008b.

E. Zitzler, L. Thiele, and J. Bader. On set-based multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 14(1):58–79, 2010.