

Analyzing Hypervolume Indicator Based Algorithms

Dimo Brockhoff¹, Tobias Friedrich², and Frank Neumann²

¹ Computer Engineering and Networks Lab, ETH Zurich, 8092 Zurich, Switzerland
dimo.brockhoff@tik.ee.ethz.ch

² Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany
firstname.lastname@mpi-inf.mpg.de

Abstract Indicator-based methods to tackle multiobjective problems have become popular recently, mainly because they allow to incorporate user preferences into the search explicitly. Multiobjective Evolutionary Algorithms (MOEAs) using the hypervolume indicator in particular showed better performance than classical MOEAs in experimental comparisons. In this paper, the use of indicator-based MOEAs is investigated for the first time from a theoretical point of view. We carry out running time analyses for an evolutionary algorithm with a $(\mu + 1)$ -selection scheme based on the hypervolume indicator as it is used in most of the recently proposed MOEAs. Our analyses point out two important aspects of the search process. First, we examine how such algorithms can approach the Pareto front. Later on, we point out how they can achieve a good approximation for an exponentially large Pareto front.

1 Introduction

In the last decades, there has been a growing interest in developing evolutionary algorithms for multiobjective optimization problems. Many variants proposed in the last years make use of special indicator functions that explicitly define the optimization goal—independent from the algorithm itself. That is an advantage compared to earlier algorithms where user preferences were incorporated in the algorithms implicitly.

The hypervolume indicator, first introduced by Zitzler et al. as the ‘size of the space covered’ [13] and also known as the \mathcal{S} -metric [1, 7], is used in many cases as the underlying indicator function. Up to now, it is—together with its weighted version of [11]—the only known indicator that is compliant with the concept of Pareto-dominance, i.e., whenever a set of solutions dominates another set, its hypervolume indicator value is higher than the one of the latter. This is the main reason why most of the recently proposed indicator based algorithms like IBEA [12], SMS-EMOA [1], or the multiobjective version of CMA-ES [6] use the hypervolume as the underlying indicator—although its calculation time is exponential in the number of objectives. It was shown experimentally even for a higher number of objectives that hypervolume-based algorithms outperform standard MOEAs [10]. A theoretical understanding why hypervolume-based algorithms outperform their Pareto-dominance based counterparts is still missing.

This paper is a first step towards a general explanation why hypervolume-based algorithms perform better on the known test problems than other state-of-the-art algorithms. Our aim is to gain insights into the optimization process of hypervolume-based

algorithms by carrying out rigorous running time analyses. Besides very general non-convergence results on steady-state MOEAs by Zitzler et al. [14], there are no results on the runtime behavior of indicator based evolutionary algorithms known so far. This paper achieves the first results of this kind. Comparisons to former running time analysis results of non-hypervolume-based algorithms allow first conclusions that and when hypervolume-based algorithms are preferable to other algorithms. Within this paper, we consider two important parts of the optimization process. First, we examine how hypervolume-based evolutionary algorithms may approach the Pareto optimal set (Section 3). By considering the function LOTZ, we point out how the population moves to the Pareto front. Second, we examine in Section 4 how the hypervolume indicator helps to spread the individuals of a population over a large Pareto front such that a good approximation of the Pareto optimal set can be achieved. In the following section, we provide the basis for our analyses to follow.

2 The Hypervolume Indicator and Hypervolume-based Algorithms

Our aim is to analyze hypervolume-based algorithms for multi-objective optimization problems. Without loss of generality, we assume that k objective functions $f = (f_1, \dots, f_k)$ that map solutions $x \in X$ from the decision space X to an objective vector $f(x) = (f_1(x), \dots, f_k(x)) \subseteq \mathbb{R}^k$ have to be maximized. Throughout this study, we assume that X is the set of binary strings of length n . Instead of optimizing the weak Pareto-dominance relation $\succeq := \{(x, y) \mid x, y \in X \wedge \forall 1 \leq i \leq k: f_i(x) \geq f_i(y)\}$, i.e., finding its maximal elements also called Pareto optimal solutions³, the goal for hypervolume-based algorithms is to maximize the hypervolume indicator I_H . The hypervolume indicator $I_H(A)$ of a solution set $A \subseteq X$ can be defined as the hypervolume of the space that is dominated by the set A and is bounded by a reference point $r = (r_1, \dots, r_k) \in \mathbb{R}^k$:

$$I_H(A) = \lambda \left(\bigcup_{a \in A} [f_1(a), r_1] \times [f_2(a), r_2] \times \dots \times [f_k(a), r_k] \right)$$

where $\lambda(S)$ is the Lebesgue measure of a set S and $[f_1(a), r_1] \times [f_2(a), r_2] \times \dots \times [f_k(a), r_k]$ is the k -dimensional hypercuboid consisting of all points that are weakly dominated by the point a but not weakly dominated by the reference point.

Note that the hypervolume indicator is Pareto-dominance compliant, i.e., whenever a solution set $A \subseteq X$ is strictly better than a set $B \subseteq X$ with respect to the weak Pareto-dominance relation⁴ the hypervolume of A is also strictly better than the one for B ($I_H(A) > I_H(B)$). Therefore, the objective vectors of a set $X^* \subseteq X$ that maximizes the hypervolume indicator cover the Pareto front entirely [2].

³ Usually, an approximation of the so-called Pareto front is sought: the set of objective vectors the preimages of which are Pareto optimal is usually smaller than the Pareto optimal set itself.

⁴ We say that a set of solutions A is strictly better than another set B iff (for maximization) $A \succeq B \wedge B \not\succeq A$ where A is dominating B ($A \succeq B$) iff $\forall b \in B : \exists a \in A : a \succeq b$.

Fixing the maximal number μ of solutions in an evolutionary algorithm \mathcal{A} , the goal of maximizing the hypervolume indicator changes to finding a set of μ solutions that have the maximal hypervolume indicator value among all sets of μ solutions. The time until such a solution set is found for the first time is referred to as the optimization time of \mathcal{A} ; its expectation is denoted by the term *expected optimization time*.

Several evolutionary algorithms to optimize the hypervolume have been proposed in the literature [1, 6, 11, 12]. Most of them use the same $(\mu + \lambda)$ -selection scheme which will be also investigated in the remainder of the paper. The population P of the next generation with $|P| = \mu$ is computed from the set P' of solutions that is the union of the previous population and the λ generated offsprings in the following way: after a non-dominated sorting of P' [4], the non-dominated fronts are, starting with the best front, completely inserted into the new population P until the size of P is at least μ . For the first front F the inclusion of which yields a population size larger than μ , the solutions x in this front with the smallest indicator loss $d(x) := I_H(F) - I_H(F \setminus \{x\})$ are successively removed from the new population where the indicator loss is recalculated every time a solution is removed.

The algorithm $(\mu + 1)$ -SIBEA, we investigate in the following, is based on the Simple Indicator-Based Evolutionary Algorithm (SIBEA) proposed in [11] that also uses the above mentioned selection scheme. For our theoretical investigations, we consider a simplified version of SIBEA (see Algorithm 1). It uses a population P of size μ and produces in each iteration one single offspring x . By removing the individual with the smallest hypervolume loss from $P \cup \{x\}$, the new parent population is obtained. The omission of the non-dominated sorting step is not crucial for our obtained results, i.e., all running time bounds are the same than with the sorting. Only dominated points are handled differently: with the original selection scheme, always the worst point on the worst front is deleted, whereas in our version, any dominated point is deleted with the same probability.

Algorithm 1 $(\mu + 1)$ -SIBEA

Parameters: population size μ

Step 1 (Initialization):

Generate an initial (multi)-set of decision vectors $P \subseteq \{0, 1\}^n$ of size μ uniformly at random.

Step 2 (Repeat):

- Select an element x from P uniformly at random. Flip each bit of x with probability $1/n$ to obtain an offspring x' . Set $P' := P \cup \{x'\}$.
 - For each solution $x \in P'$ determine the hypervolume loss $d(x)$ if it is removed from P' , i.e., $d(x) := I_H(P') - I_H(P' \setminus \{x\})$.
 - Choose an element $z \in P'$ with smallest loss in P' uniformly at random, i.e., $z = \operatorname{argmin}_{x \in P'} d(x)$ and set $P := P' \setminus \{z\}$.
-

The goal of the next sections is to analyze the runtime behavior of $(\mu + 1)$ -SIBEA on some example functions. These analyses point out some basic concepts how the algo-

rithm can make progress during the optimization process. Additionally, it gives insights how a good spread over the whole Pareto front can be achieved using the hypervolume indicator.

3 Exploring a Small Pareto Front

In this section, we examine the well-known bi-objective problem LOTZ with a Pareto front of size $n + 1$ and show that the expected optimization time of the $(\mu + 1)$ -SIBEA is $O(\mu n^2)$ if μ is large enough to find all optima, i.e., $\mu \geq n + 1$.

LOTZ was first investigated in [9] and has been considered in several previous studies concerning the running time analysis of MOEAs. It is defined as LOTZ: $\{0, 1\}^n \rightarrow \mathbb{N}^2$ with

$$f_1(x) = \text{LO}(x) = \sum_{i=1}^n \prod_{j=1}^i x_j \quad \text{and} \quad f_2(x) = \text{TZ}(x) = \sum_{i=1}^n \prod_{j=i}^n (1 - x_j).$$

Without loss of generality, we fix the reference point for computing the hypervolume to $(-1, -1)$. All results of this section still hold as long as the reference point (r_1, r_2) is chosen such that r_1 and r_2 are negative.

Lemma 1 *The expected time until the $(\mu + 1)$ -SIBEA has obtained for the first time a Pareto optimal solution of LOTZ is $O(\mu n^2)$.*

Proof. Throughout this proof, we consider the situation where no Pareto optimal search point belongs to the current population P . Let $\{x_1, x_2, \dots, x_k\} \subseteq P$ be the set of individuals that are not dominated by any other individual in P . Denote by H the hypervolume covered by these points. Without loss of generality, we assume that $\text{LO}(x_i) \leq \text{LO}(x_{i+1})$, $1 \leq i \leq k - 1$ holds which also implies $\text{TZ}(x_i) \geq \text{TZ}(x_{i+1})$, $1 \leq i \leq k - 1$, as the k individuals do not dominate each other.

Let $X_1 = \text{LO}(x_1) + 1$, $X_i = \text{LO}(x_i) - \text{LO}(x_{i-1})$, $2 \leq i \leq k$ and denote by $X_{\max} = \sum_{i=1}^k X_i$ the maximum LO-value with respect to the reference point $(-1, -1)$. Similar, define $Y_1 = \text{TZ}(x_k) + 1$ and $Y_i = \text{TZ}(x_{k-i}) - \text{TZ}(x_{k-i+1})$, $2 \leq i \leq k$, and denote by $Y_{\max} = \sum_{i=1}^k Y_i$ the maximum TZ-value with respect to the reference point $(-1, -1)$.

Considering one single solution x_i of the k non-dominated solutions of P , we study how the hypervolume can increase. Flipping the single bit which increases its LO-value increases the hypervolume by at least Y_{k-i+1} . Flipping the single bit which increases its TZ-value increases the hypervolume by at least X_i . We call all these 1-bit flips applied to one of the k individuals good. Each of these $2k$ good operations happens with probability $\frac{1}{\mu} \cdot \frac{1}{n} \cdot (1 - 1/n)^{n-1} \geq \frac{1}{e\mu n}$ in the next step.

Note, that each good operation is accepted as it leads to a population with a larger hypervolume. The total increase of all good operations with respect to the current hypervolume H is at least $X_{\max} + Y_{\max} \geq \sqrt{X_{\max} \cdot Y_{\max}} \geq \sqrt{H}$.

Choosing one of these $2k$ good operations uniformly at random, the expected increase of the hypervolume is at least $\sqrt{H}/(2k)$. Hence, the expected number of good

operations needed to increase the hypervolume by \sqrt{H} is upper bounded by $2k$. Using Markov's inequality, the probability of having at least $4k$ operations to achieve this goal is upper bounded by $1/2$. Hence, with probability at least $1/2$ a phase containing $4k$ good operations is successful, i.e., increases the hypervolume by \sqrt{H} with probability at least $1/2$. This implies that an expected number of 2 of these phases carrying out $4k$ such good operations each is enough to increase the hypervolume by \sqrt{H} .

Considering all good 1-bit flips together, the probability of carrying out one good operation in the next step of the algorithm is at least $\frac{2k}{e\mu n}$. Hence, the expected waiting time for a good operation is $O(\mu n / (2k))$ and the expected waiting time for increasing the hypervolume by at least \sqrt{H} is therefore upper bounded by $O(\frac{\mu n}{2k} \cdot 2 \cdot 4k) = O(\mu n)$.

It remains to show that $O(n)$ successive increases of the hypervolume by its square-root fraction suffice to reach the maximum hypervolume of $O(n^2)$. Let $h(t)$ be the hypervolume of the current solutions after t increases by at least $\sqrt{h(t)}$. Then, $h(t+1) \geq h(t) + \sqrt{h(t)}$. We want to prove by induction that $h(t) \geq t^2/5$. The induction basis case holds trivially since $h(0) \geq 0$. In general,

$$\begin{aligned} h(t) &\geq h(t-1) + \sqrt{h(t-1)} \geq \frac{(t-1)^2}{5} + \frac{t-1}{\sqrt{5}} \\ &\geq \frac{t^2}{5} + t \left(\frac{1}{\sqrt{5}} - \frac{2}{5} \right) - \left(\frac{1}{\sqrt{5}} - \frac{1}{5} \right) \geq \frac{t^2}{5}. \end{aligned}$$

Therefore, the expected number of iterations for the situation where no solution of the current population is Pareto optimal is upper bounded by $O(\mu n^2)$. \square

Theorem 2 *Choosing $\mu \geq n + 1$, the expected optimization time of the $(\mu + 1)$ -SIBEA on LOTZ is $O(\mu n^2)$.*

Proof. Using Lemma 1, the expected time until a first Pareto optimal solution has been obtained is $O(\mu n^2)$. Due to the hypervolume-based selection and the fact that at most $n + 1$ solutions are mutually non-dominated in LOTZ [9], a Pareto optimal solution that has been found with $(\mu + 1)$ -SIBEA will stay from that moment in the population or another solution mapping to the same objective vector will enter the population. Increasing the number of Pareto optimal solutions in the population increases the hypervolume indicator, i.e., the highest hypervolume value is achieved if and only if the entire Pareto front is found. Therefore, there is at least one solution in the population which has a Hamming neighbor that is Pareto optimal and not contained in the current population—unless the whole Pareto optimal set is already found. Hence, the expected waiting time for increasing the number of Pareto optimal solutions in the population is $O(\mu n)$. Having reached a Pareto optimal solution for the first time at most n additional Pareto optimal solutions have to be produced which implies that the expected time to achieve a population including all Pareto optimal solutions is $O(\mu n^2)$. \square

4 Approximating a Large Pareto Front

The goal of this section is to examine how the hypervolume indicator helps to achieve a good spread over a larger Pareto front. In the case of a large Pareto front, we are

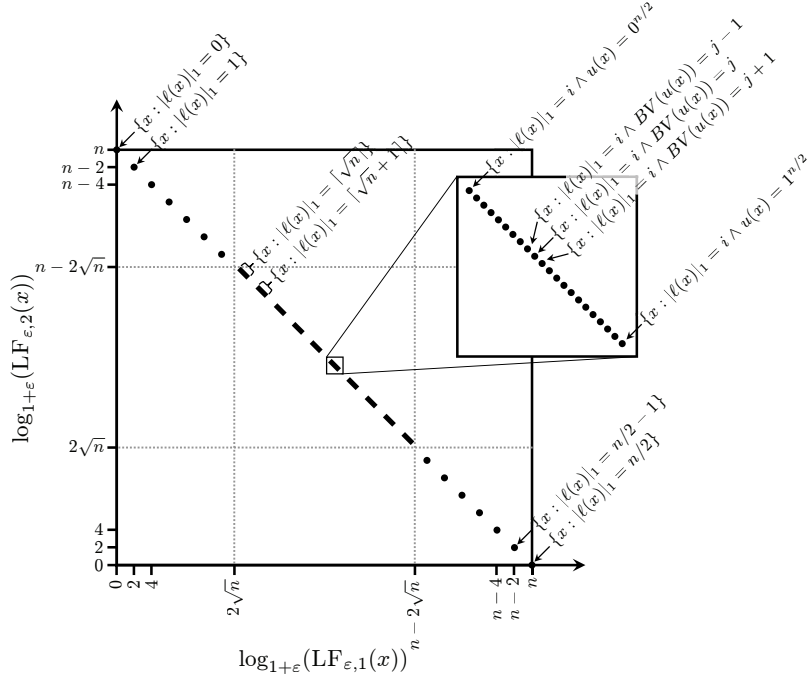


Figure 1. Illustration of the objective space of LF_ε . In addition, the corresponding solutions in decision space are annotated as well. It is important to note that instead of the original objective values the logarithms of the objective values are plotted for clarity.

interested in the time until an algorithm has achieved a good approximation of the Pareto front. We are considering the multiplicative ε -dominance relation [8] to measure the quality of an approximation. Let $\varepsilon \in \mathbb{R}^+$ be a positive real number. We define that an objective vector u ε -dominates v , denoted by $u \succeq_\varepsilon v$, precisely if $(1 + \varepsilon) \cdot u_i \geq v_i$ for all $i \in \{1, \dots, m\}$. An evolutionary algorithm has achieved an ε -approximation for a given problem if there exists for each objective vector v in the objective space a solution with objective vector u in the population such that $u \succeq_\varepsilon v$. In the following, we present for each choice of ε a function where the Pareto-dominance based algorithm Global SEMO [3] cannot obtain an ε -approximation while $(\mu + 1)$ -SIBEA is able to achieve this goal in expected polynomial time.

We consider the bi-objective problem LF_ε (large front) introduced in [5] which is parametrized by the value ε coming from the definition of ε -dominance. Without loss of generality, we assume that n is even, i.e., each decision vector consists of an even number of bits. We denote the lower half of a decision vector $x = (x_1, \dots, x_n)$ by $\ell(x) = (x_1, \dots, x_{n/2})$ and its upper half by $u(x) = (x_{n/2+1}, \dots, x_n)$. Furthermore, we denote the length of a bit-string x by $|x|$, the number of its 1-bits by $|x|_1$, the number

of its 0-bits by $|x|_0$, and its complement by \bar{x} . In addition, we define the function

$$\text{BV}(x) := \sum_{i=1}^{|x|} 2^{|x|-i} \cdot x_i$$

which interprets a bit-string x as the encoded natural number with respect to the binary numeral system. We consider the function $\text{LF}_\varepsilon : \{0, 1\}^n \rightarrow \mathbb{R}^2$ (large front), for a given $\varepsilon \in \mathbb{R}^+$, defined as

$$f_1(x) = \text{LF}_{\varepsilon,1}(x) := \begin{cases} (1 + \varepsilon)^{2 \cdot |\ell(x)|_1 + 2^{-n/2} \cdot \text{BV}(u(x))} & \min\{|\ell(x)|_0, |\ell(x)|_1\} \geq \sqrt{n} \\ (1 + \varepsilon)^{2 \cdot |\ell(x)|_1} & \text{otherwise,} \end{cases}$$

$$f_2(x) = \text{LF}_{\varepsilon,2}(x) := \begin{cases} (1 + \varepsilon)^{2 \cdot |\ell(x)|_0 + 2^{-n/2} \cdot \text{BV}(\overline{u(x)})} & \min\{|\ell(x)|_0, |\ell(x)|_1\} \geq \sqrt{n} \\ (1 + \varepsilon)^{2 \cdot |\ell(x)|_0} & \text{otherwise.} \end{cases}$$

The function $\text{LF}_\varepsilon(x)$ is illustrated in Figure 1. In the following proofs it will sometimes be helpful to use the following equivalent formulation of $f_2(x)$:

$$\text{LF}_{\varepsilon,2}(x) = \begin{cases} (1 + \varepsilon)^{n - 2|\ell(x)|_1 + 1 - 2^{-n/2} \text{BV}(u(x)) - 2^{-n/2}} & \min\{|\ell(x)|_0, |\ell(x)|_1\} \geq \sqrt{n} \\ (1 + \varepsilon)^{n - 2|\ell(x)|_1} & \text{otherwise.} \end{cases}$$

It has been shown in [5] that Global SEMO needs with probability exponentially close to 1 a number of steps that is exponential in the number of bits to achieve an ε -approximation of LF. The reason for this negative result is that the population of Global SEMO becomes exponentially large before obtaining an ε -approximation. On the other hand it has been pointed out in this paper that the use of ε -dominance with the choice of ε as used for the definition of LF achieves an ε -approximation in expected polynomial time.

In the following, we show that this goal can also be achieved by using the $(\mu + 1)$ -SIBEA with a population of reasonable size. Our result holds for each $\varepsilon \in \mathbb{R}^+$ —in contrast to the usage of ε -dominance based algorithms examined in [5] where the exact knowledge of ε is necessary to achieve a good approximation.

Let the reference point for computing the hypervolume be $((1 + \varepsilon)^{-1}, (1 + \varepsilon)^{-1})$ corresponding to the point $(-1, -1)$ in the plot of Figure 1. Note that the following results also hold for any reference point (r_1, r_2) with $r_1, r_2 \leq (1 + \varepsilon)^{-1}$.

Theorem 3 *Choosing $\mu \geq n/2 + 3$, the expected time until $(\mu + 1)$ -SIBEA has achieved an ε -approximation of LF_ε is $O(\mu n \log n)$.*

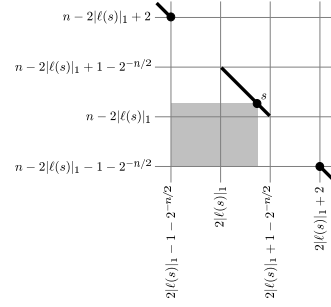
To prove Theorem 3 we have to show that if no other solution x with $|\ell(x)|_1 = |\ell(s)|_1$ is contained in the population, the $(\mu + 1)$ -SIBEA will not remove the solution s from the population:

Lemma 4 *When optimizing LF_ε , the $(\mu + 1)$ -SIBEA with $\mu \geq n/2 + 3$ will not remove a solution s with $\{x \in P : |\ell(x)|_1 = |\ell(s)|_1\} = \{s\}$ from the population P .*

Proof. If $\{x \in P : |\ell(x)|_1 = k\} = \{s\}$ for some k , such a solution s will be called sole. To show the lemma, it suffices to prove that sole solutions are not removed from the

population. Let $d(x) = I_H(P') - I_H(P' \setminus \{x\})$ be the hypervolume loss of a solution x in the population P' of $(\mu + 1)$ -SIBEA and let s be a sole solution. We will show that there is always another solution z with $d(z) < d(s)$ which proves the lemma due to the selection step of $(\mu + 1)$ -SIBEA. To this end, we first calculate a lower bound for $d(s)$ and then upper bound $d(z)$. The small sketches to the right of the volume calculations in this proof use the same double-logarithmic axes as Figure 1. If $\min\{|\ell(s)|_0, |\ell(s)|_1\} \geq \sqrt{n}$, then (we can ignore the $-2^{-n/2}$ in the exponent of the first subtrahend)

$$\begin{aligned}
d(s) &> [(1 + \varepsilon)^{2|\ell(s)|_1 + 2^{-n/2}} \text{BV}(u(s)) - (1 + \varepsilon)^{2|\ell(s)|_1 - 1}] \\
&\quad \cdot [(1 + \varepsilon)^{n - 2|\ell(s)|_1 + 1 - 2^{-n/2}} \text{BV}(u(s)) - 2^{-n/2} \\
&\quad - (1 + \varepsilon)^{n - 2|\ell(s)|_1 - 1 - 2^{-n/2}}] \\
&= (1 + \varepsilon)^{n + 1 - 2^{-n/2}} + (1 + \varepsilon)^{n - 2 - 2^{-n/2}} \\
&\quad - (1 + \varepsilon)^{n - 2^{-n/2}} \text{BV}(u(s)) - 2^{-n/2} \\
&\quad - (1 + \varepsilon)^{n - 1 + 2^{-n/2}} \text{BV}(u(s)) - 2^{-n/2} \\
&\geq (1 + \varepsilon)^{n + 1 - 2^{-n/2}} - (1 + \varepsilon)^{n - 2^{-n/2}} \\
&\quad - (1 + \varepsilon)^{n - 1 - 2^{-n/2}} + (1 + \varepsilon)^{n - 2 - 2^{-n/2}}.
\end{aligned}$$

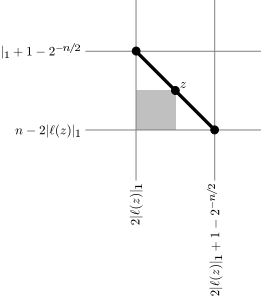


where the last inequality stems from the fact that

$$\max_{0 \leq \Delta \leq 1} (1 + \varepsilon)^\Delta + (1 + \varepsilon)^{1 - \Delta} = (1 + \varepsilon)^0 + (1 + \varepsilon)^1.$$

It remains to prove the existence of a solution z with $d(z) < d(s)$. If there is a solution z with $\min\{|\ell(z)|_0, |\ell(z)|_1\} < \sqrt{n}$ and $|\{x \in P : |\ell(x)|_1 = |\ell(z)|_1\}| \geq 2$, then $d(z) = 0$ and the lemma is proven. If there is a k with $|\{x \in P : |\ell(x)|_1 = k\}| > 2$, then there is a solution z with $|\ell(z)|_1 = k$ and

$$\begin{aligned}
d(z) &\leq [(1 + \varepsilon)^{2|\ell(z)|_1 + 2^{-n/2}} \text{BV}(u(z)) - (1 + \varepsilon)^{2|\ell(z)|_1}] \\
&\quad \cdot [(1 + \varepsilon)^{n - 2|\ell(z)|_1 + 1 - 2^{-n/2}} \text{BV}(u(z)) - 2^{-n/2} \\
&\quad - (1 + \varepsilon)^{n - 2|\ell(z)|_1}] \\
&< (1 + \varepsilon)^{n + 1} - (1 + \varepsilon)^{n + 2^{-n/2}} \text{BV}(u(z)) \\
&\quad + (1 + \varepsilon)^n - (1 + \varepsilon)^{n + 1 - 2^{-n/2}} \text{BV}(u(z)) \\
&\leq (1 + \varepsilon)^{n + 1} + (1 + \varepsilon)^n - 2(1 + \varepsilon)^{n + 1/2}.
\end{aligned}$$



where the last inequality holds since $\operatorname{argmin}_{0 \leq \Delta \leq 1} (1 + \varepsilon)^\Delta + (1 + \varepsilon)^{1 - \Delta} = \frac{1}{2}$. Comparing this upper bound for $d(z)$ and the above lower bound for $d(s)$ yields $d(z) < d(s)$ for all $\varepsilon > 0$ and n .

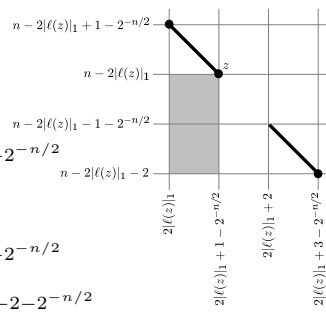
It remains to examine the case where there is neither a k with $\min\{n - k, k\} < \sqrt{n}$ and $|\{x \in P : |\ell(x)|_1 = k\}| \geq 2$ nor a k with $|\{x \in P : |\ell(x)|_1 = k\}| > 2$. As there are only $n/2 + 1$ possible values for k , but at least $\mu + 1 \geq n/2 + 4$ solutions in P' , by

the pigeonhole principle there must be a k with

$$\begin{aligned} \min\{n - k, k\} &> \lceil \sqrt{n} \rceil, \\ |\{x \in P : |\ell(x)|_1 = k\}| &\geq 2, \\ |\{x \in P : |\ell(x)|_1 = k + 1\}| &\geq 1. \end{aligned}$$

Let z be a solution with $|\ell(z)|_1 = k$, then

$$\begin{aligned} d(z) &\leq [(1 + \varepsilon)^{2|\ell(z)|_1 + 2^{-n/2}} \text{BV}(u(z)) - (1 + \varepsilon)^{2|\ell(z)|_1}] \\ &\quad \cdot [(1 + \varepsilon)^{n - 2|\ell(z)|_1 + 1 - 2^{-n/2}} \text{BV}(u(z)) - 2^{-n/2} \\ &\quad - (1 + \varepsilon)^{n - 2|\ell(z)|_1 - 2}] \\ &= (1 + \varepsilon)^{n + 1 - 2^{-n/2}} - (1 + \varepsilon)^{n + 1 - 2^{-n/2}} \text{BV}(u(z)) - 2^{-n/2} \\ &\quad - (1 + \varepsilon)^{n - 2 + 2^{-n/2}} \text{BV}(u(z)) + (1 + \varepsilon)^{n - 2} \\ &< (1 + \varepsilon)^{n + 1 - 2^{-n/2}} - (1 + \varepsilon)^{n + 1 - 2^{-n/2}} \text{BV}(u(z)) - 2^{-n/2} \\ &\quad - (1 + \varepsilon)^{n - 2 + 2^{-n/2}} \text{BV}(u(z)) - 2^{-n/2} + (1 + \varepsilon)^{n - 2 - 2^{-n/2}} \\ &\leq (1 + \varepsilon)^{n + 1 - 2^{-n/2}} - (1 + \varepsilon)^{n - 2 - 2^{-n/2}} - (1 + \varepsilon)^{n - 1 - 2^{-n/2}} + (1 + \varepsilon)^{n - 2 - 2^{-n/2}}. \end{aligned}$$



where the last inequality comes from $\operatorname{argmin}_{0 \leq \Delta \leq 1} (1 + \varepsilon)^{3 - \Delta} + (1 + \varepsilon)^\Delta = 1$. This shows $d(z) < d(s)$ and finally proves the lemma. \square

Proof of Theorem 3. An ε -approximation of the Pareto front has been achieved if and only if the population includes for each $k \in \{0, \dots, n/2\}$ a solution x with $|\ell(x)|_1 = k$ (see [5]). Denote the set of covered $|\cdot|_1$ values by $A := \{|\ell(x)|_1 \mid x \in P\}$ and the set of uncovered $|\cdot|_1$ values by $B := \{0, \dots, n/2\} \setminus A$. Due to the previous lemma, we know that during the optimization process elements that are added to A are never removed and follow the ideas given in [5].

As long as $A \neq \{0, \dots, n/2\}$, there exists an $a \in A$ and a $b \in B$ with $b = a - 1$ or $b = a + 1$. Let $x \in P$ be the individual with $|\ell(x)|_1 = a$. The probability to choose x in the next step and flip exactly one proper bit to obtain a decision vector y with $|\ell(y)|_1 = b$ is at least $\frac{1}{\mu} \cdot \frac{\min\{b+1, n/2-b+1\}}{\varepsilon n} \geq \frac{\min\{b, n/2-b\}+1}{\mu \varepsilon n}$.

Summing up over the different values that b can attain, we get a maximum waiting time of $\mu \varepsilon n \cdot \sum_{b=0}^{n/2} \frac{1}{\min\{b, n/2-b\}+1} \leq 2\mu \varepsilon n \cdot \sum_{b=1}^{n/4+1} \frac{1}{b} = \mathcal{O}(\mu n \log n)$ until solutions with all possible values of b are contained in the population which completes the proof. \square

5 Conclusions

Indicator-based evolutionary algorithms have been shown to be very successful for dealing with multiobjective optimization. With this paper, we have taken a first step in understanding these algorithms using the hypervolume indicator by rigorous running time analysis. Considering the function LOTZ, we have pointed out how the progress of such algorithms towards the Pareto front can be analyzed. Later on, we have shown that the hypervolume indicator is provable helpful for approximating large Pareto fronts.

Acknowledgments

The authors would like to thank Johannes Bader, Tim Hohm, and Tamara Ulrich for valuable discussions. Dimo Brockhoff has been supported by the Swiss National Science Foundation (SNF) under grant 112079.

Bibliography

- [1] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal on Operational Research*, 181:1653–1669, 2007.
- [2] M. Fleischer. The measure of Pareto optima. Applications to multi-objective metaheuristics. In C. M. Fonseca et al., editors, *Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, volume 2632 of *LNCS*, pages 519–533, Faro, Portugal, 2003. Springer.
- [3] O. Giel. Expected Runtimes of a Simple Multi-objective Evolutionary Algorithm. In *Congress on Evolutionary Computation (CEC 2003)*, IEEE Press, pages 1918–1925, 2003.
- [4] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Massachusetts, 1989.
- [5] C. Horoba and F. Neumann. Benefits and drawbacks for the use of ε -dominance in evolutionary multi-objective optimization. In *Genetic and Evolutionary Computation Conference (GECCO 2008)*, 2008. To appear. Also available as Technical Report CI-248/08, SFB 531, TU Dortmund.
- [6] C. Igel, N. Hansen, and S. Roth. Covariance Matrix Adaptation for Multi-objective Optimization. *Evolutionary Computation*, 15(1):1–28, 2007.
- [7] J. Knowles and D. Corne. Properties of an Adaptive Archiving Algorithm for Storing Non-dominated Vectors. *IEEE Transactions on Evolutionary Computation*, 7(2):100–116, 2003.
- [8] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining Convergence and Diversity in Evolutionary Multiobjective Optimization. *Evolutionary Computation*, 10(3):263–282, 2002.
- [9] M. Laumanns, L. Thiele, and E. Zitzler. Running Time Analysis of Multiobjective Evolutionary Algorithms on Pseudo-Boolean Functions. *IEEE Transactions on Evolutionary Computation*, 8(2):170–182, 2004.
- [10] T. Wagner, N. Beume, and B. Naujoks. Pareto-, Aggregation-, and Indicator-based Methods in Many-objective Optimization. In S. Obayashi et al., editors, *Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, volume 4403 of *LNCS*, pages 742–756, Berlin Heidelberg, Germany, 2007. Springer.
- [11] 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, *Conference on Evolutionary Multi-Criterion Optimization (EMO 2007)*, volume 4403 of *LNCS*, pages 862–876, Berlin, 2007. Springer.
- [12] E. Zitzler and S. Künzli. Indicator-Based Selection in Multiobjective Search. In *Conference on Parallel Problem Solving from Nature (PPSN VIII)*, volume 3242 of *LNCS*, pages 832–842. Springer, 2004.
- [13] E. Zitzler and L. Thiele. Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In *Conference on Parallel Problem Solving from Nature (PPSN V)*, pages 292–301, Amsterdam, 1998.
- [14] E. Zitzler, L. Thiele, and J. Bader. On Set-Based Multiobjective Optimization. Technical Report 300, Computer Engineering and Networks Laboratory, ETH Zurich, February 2008.