TECHNISCHE UNIVERSITÄT DORTMUND

REIHE COMPUTATIONAL INTELLIGENCE

COLLABORATIVE RESEARCH CENTER 531

Design and Management of Complex Technical Processes and Systems by means of Computational Intelligence Methods

Benefits and Drawbacks for the Use of ε-Dominance in Evolutionary Multi-Objective Optimization

Christian Horoba and Frank Neumann

No. CI-248/08

Technical Report ISSN 1433-3325 May 2008

Secretary of the SFB 531 · Technische Universität Dortmund · Dept. of Computer Science/LS 2 · 44221 Dortmund · Germany

This work is a product of the Collaborative Research Center 531, "Computational Intelligence," at the Technische Universität Dortmund and was printed with financial support of the Deutsche Forschungsgemeinschaft.

Benefits and Drawbacks for the Use of ε-Dominance in Evolutionary Multi-Objective Optimization

Christian Horoba*
Fakultät für Informatik, LS 2
Technische Universität Dortmund
Dortmund, Germany

Frank Neumann
Algorithms and Complexity
Max-Planck-Institut für Informatik
Saarbrücken, Germany

May 14, 2008

Abstract

Using diversity mechanisms in evolutionary algorithms for multi-objective optimization problems is considered as an important issue for the design of successful algorithms. This is in particular the case for problems where the number of non-dominated feasible objective vectors is exponential with respect to the problem size. In this case the goal is to compute a good approximation of the Pareto front. We investigate how this goal can be achieved by using the diversity mechanism of ε-dominance and point out where this concept is provably helpful to obtain a good approximation of an exponentially large Pareto front in expected polynomial time. Afterwards, we consider the drawbacks of this approach and point out situations where the use of ε-dominance slows down the optimization process significantly.

1 Introduction

Evolutionary algorithms (EAs) are general problem solvers which have especially shown to be successful in the context of multi-objective optimization [2, 3, 4]. For this kind of problems these algorithms seem to be well suited as the task in multi-objective optimization is to search for a set of solutions instead of a single one. Due to this circumstance multi-objective optimization is often considered as more difficult than single-objective optimization. Common generalizations of classical approaches to single-objective optimization do rarely result in successful algorithms for multi-objective optimization. A popular approach is to transfer the original multi-objective problem into a single-objective one by considering a linear combination of the different objective functions. Using several restarts with different weighting coefficients it is possible to generate some non-dominated solutions.

*This author was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the Collaborative Research Center "Computational Intelligence" (SFB 531).

However, this approach does often lead to a poor approximation of the Pareto front, since solutions from the non-convex part of the Pareto front can not be found and an even spread of weights does typically not lead to an even spread of solutions on the Pareto front. In contrast to this, multi-objective evolutionary algorithms (MOEAs) incorporating the concept of Pareto dominance into their selection mechanism are often successful in obtaining a good approximation of the Pareto front in a single run, since their population evolves over time into an approximation of the Pareto front.

In the context of multi-objective optimization the runtime behavior of a simple MOEA, called Global SEMO, has been studied. Initial studies considered investigations of artificial pseudo-Boolean functions [15, 8, 9] as well as some classical multi-objective combinatorial optimization problems [14, 16]. Later on, this algorithm has been used to study the question whether multi-objective models for single-objective optimization problems can significantly speed up the optimization process [17, 7]. Additionally, the effects of adding objectives to a given problem have been investigated by analyzing the runtime behavior on example functions [1].

A characteristic feature of Global SEMO is that it keeps for each discovered non-dominated objective vector a corresponding individual in its population. Often the number of non-dominated feasible objective vectors grows exponentially in the problem size. This is especially the case for many NP-hard multi-objective combinatorial optimization problems [6]. The class of problems with this property includes minimum spanning tree or shortest path problems with more than one weight function on the edges. For such problems it is not possible to obtain the whole Pareto front efficiently and a common approach to deal with this circumstance is to look for good approximations. When using MOEAs such as IBEA [19], SPEA2 [20], or NSGA-II [5] to approximate a large Pareto front, specific diversity mechanisms are applied to spread the individuals of the population over the whole Pareto front.

We study the concept of ε-dominance introduced by Laumanns et al. [13] and investigate its impact with respect to the runtime behavior. In the mentioned approach diversity is ensured by partitioning the objective space into boxes of appropriate size. The applied MOEA is allowed to keep at most one individual of each box in its population. A usual scenario is to divide the objective space into boxes such that their number is logarithmic with respect to the number of objective vectors.

The aim of this paper is to show where the mentioned approach is provably beneficial. Therefore, we compare a variation of Global SEMO using the concept of ε-dominance with the original algorithm. To point out situations where this concept leads provably to a better optimization process, we present a class of instances with an exponential number of non-dominated feasible objective vectors. We show that using the concept of ε-dominance a good approximation of the Pareto front is constructed efficiently while the approach not using this concept can not achieve this goal in expected polynomial time. Later on, we present instances where the concept of ε-dominance prevents the algorithm from constructing good approximations of the Pareto front. For the efficient optimization of these instances it is essential that the population contains more than one individual per box to construct other individuals that are needed for a good approximation of the Pareto front. In contrast to this, we prove that the approach without using the diversity mechanism constructs the whole Pareto front in expected polynomial time.

The outline of the paper is as follows. In Section 2, we introduce the concept of ε-dominance and the algorithms that are subject of our analyses. Situations where the mentioned diversity concept is provably helpful are presented in Section 3 and the drawbacks of this approach are investigated in Section 4. Finally, we finish with some concluding remarks.

2 Algorithms and ε-Dominance

We start with some basic notations and definitions that will be used throughout this paper. We denote the set of all Boolean values by $\mathbb{B}$ and the set of all real numbers by $\mathbb{R}$ and investigate the maximization of functions fitting in the shape of $f\colon \mathbb{B}^n \to \mathbb{R}^m$. We call f *objective function*, $\mathbb{B}^n$ *decision space* and $\mathbb{R}^m$ *objective space*. The elements of $\mathbb{B}^n$ are called *decision vectors* and the elements of $\mathbb{R}^m$ *objective vectors*. Let $u = (u_1, \ldots, u_m) \in \mathbb{R}^m$ and $v = (v_1, \ldots, v_m) \in \mathbb{R}^m$ be two objective vectors. We define that u *weakly dominates* v, denoted by $u \succeq v$, precisely if $u_i \geq v_i$ for all $i \in \{1, \ldots, m\}$, and u *dominates* v, denoted by $u \succ v$, precisely if $u \succeq v$ and $v \not\succeq u$. Let $x \in \mathbb{B}^n$ and $y \in \mathbb{B}^n$ be two decision vectors. We are able to use the same manners of speaking and notations for decision vectors, since the definition $x \succeq y :\Leftrightarrow f(x) \succeq f(y)$ transfers the concept of dominance from the objective space to the decision space.

The set $\mathcal{PF}(f) := \{u \in f(\mathbb{B}^n) \mid \forall v \in f(\mathbb{B}^n)\colon v \not\succ u\}$ is called the *Pareto front of* f and the set $\mathcal{P}(f) := f^{-1}(\mathcal{PF}(f)) = \{x \in \mathbb{B}^n \mid \forall y \in \mathbb{B}^n\colon y \not\succ x\}$ *Pareto set of* f.[1] The set $\{(x, f(x)) \mid x \in \mathcal{P}(f)\}$ constitutes the canonical solution of an optimization problem of the considered kind. In the literature a set of the form $\{(x, f(x)) \mid x \in X\}$ with $X \subseteq \mathcal{P}(f)$ is also considered as a valid solution if $f(X) = \mathcal{PF}(f)$. This means that it is sufficient to determine for all non-dominated objective vectors $u \in \mathcal{PF}(f)$ at least one decision vector $x \in \mathbb{B}^n$ with $f(x) = u$.

The next algorithm called Global Simple Evolutionary Multi-objective Optimizer (Global SEMO) can be considered as one of the simplest population-based EAs for multi-objective optimization problems and has been analyzed with respect to its runtime behavior on pseudo-Boolean functions [8, 9, 1] as well as classical combinatorial optimization problems [16, 17, 7]. Global SEMO maintains a population of variable size which serves as an archive for the discovered non-dominated individuals as well as a pool of possible parents. The population is initialized with a single individual which is drawn uniformly at random from the decision space. In each generation an individual x is drawn uniformly at random from the current population P. An offspring y is created by applying a mutation operator to x. We resort to the global mutation operator which flips each bit of x with probability $1/n$ throughout this paper. The offspring is added to the population if it is not dominated by any other individual of P. All individuals which are weakly dominated by y are in turn deleted from the population. The last step ensures that the population stores for each discovered non-dominated objective vector u just the most recently created decision vector x with $f(x) = u$.

Algorithm 1. *Global SEMO*

1. *Choose* $x \in \mathbb{B}^n$ *uniformly at random.*
2. *Initialize* $P := \{x\}$.
3. *Repeat*
   - *Choose* $x \in P$ *uniformly at random.*
   - *Create an offspring* y *by flipping each bit of* x *with probability* $1/n$.
   - *Update* $P := (P \setminus \{z \in P \mid y \succeq z\}) \cup \{y\}$ *if* $\{z \in P \mid z \succ y\} = \emptyset$.

For theoretical investigations, we count the number of rounds until a desired goal has been achieved. The number of these rounds is called the runtime of the considered algorithm. The

[1] Note, that the preimage $f^{-1}(U)$ of a set $U \subseteq \mathbb{R}^m$ under a function $f\colon \mathbb{B}^n \to \mathbb{R}^m$ is also well defined if f is not a bijection.

expected runtime refers to the expectation of this random variable. For exact optimization often the expected optimization time is considered which equals the expected number of iterations until a decision vector for each objective vector of $\mathcal{PF}(f)$ has been included into the population. We are mainly interested in approximations. Therefore, we are interested in the number of rounds until a MOEA has achieved an approximation of the Pareto front with a certain quality.

We are considering the following model 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 $i \in \{1, \ldots, m\}$. We call a set $\mathcal{PF}_\varepsilon(f) \subseteq f(\mathbb{B}^n)$ an *ε-approximate Pareto front of* f if

$$\forall u \in f(\mathbb{B}^n)\colon \quad \exists v \in \mathcal{PF}_\varepsilon(f)\colon \quad v \succeq_\varepsilon u,$$

and a set $\mathcal{PF}^*_\varepsilon(f) \subseteq \mathcal{PF}(f)$ an *ε-Pareto front of* f if $\mathcal{PF}^*_\varepsilon(f)$ is an ε-approximate Pareto front. The corresponding Pareto sets are naturally defined, i. e., $\mathcal{P}_\varepsilon(f) := f^{-1}(\mathcal{PF}_\varepsilon(f))$ and $\mathcal{P}^*_\varepsilon(f) := f^{-1}(\mathcal{PF}^*_\varepsilon(f))$. We point out that it is possible that there are several different ε-approximate Pareto fronts or ε-Pareto fronts for a given objective function. We also emphasize that ε-Pareto fronts are of more value than ε-approximate Pareto fronts to a decision maker, since all objective vectors of an ε-Pareto front are non-dominated with respect to the classical concept of dominance. In the following sections, we limit our considerations to functions where the Pareto set contains all decision vectors and therefore the distinction between ε-approximate Pareto fronts and ε-Pareto fronts collapses.

The following algorithm, called Global Diversity Evolutionary Multi-objective Optimizer (Global DEMO$_\varepsilon$), incorporates the concept of ε-dominance [13]. The idea is to partition the objective space into boxes such that all objective vectors in a box ε-dominate each other. The algorithm maintains at each time step at most one individual per box. This approach ensures that the individuals contained in the population show some kind of diversity with respect to their objective vectors and that the size of the population can be controlled in a better way. These properties seem to be very important if we intend to approximate a large Pareto front. We formalize this idea by introducing the so-called box index vector which maps each decision vector to the index of its box. We assume a positive and normalized objective space, i. e., $f_i(x) \geq 1$ for all $i \in \{1, \ldots, m\}$ and $x \in \mathbb{B}^n$. Let

$$b_i(x) := \left\lfloor \frac{\log(f_i(x))}{\log(1+\varepsilon)} \right\rfloor$$

and denote by $b(x) := (b_1(x), \ldots, b_m(x))$ the *box index vector* of a decision vector x. Global DEMO$_\varepsilon$ works as Global SEMO with the exceptions that it does not accept an offspring with a dominated box index vector and that it deletes all individuals from the population whose box index vectors are weakly dominated by the box index vector of the offspring. This approach ensures that at most one individual per non-dominated box resides in the population.

Algorithm 2. *Global DEMO$_\varepsilon$*

1. *Choose $x \in \mathbb{B}^n$ uniformly at random.*
2. *Initialize $P := \{x\}$.*
3. *Repeat*
    - *Choose $x \in P$ uniformly at random.*

- *Create an offspring y by flipping each bit of x with probability $1/n$.*
- *Update $P := (P \setminus \{z \in P \mid b(y) \succeq b(z)\}) \cup \{y\}$ if $\{z \in P \mid b(z) \succ b(y) \vee z \succ y\} = \emptyset$.*

Global DEMO$_\varepsilon$ features two important properties. The first one is that the population contains an ε-Pareto set of the so far sampled decision vectors. This is made precise in the following lemma which shows that dominance with respect to the box index vector induces ε-dominance.

Lemma 1. *If $b(x) \succeq b(y)$ then $x \succeq_\varepsilon y$.*

Proof. We have to show that $b_i(x) \geq b_i(y)$ implies $(1+\varepsilon)\cdot f_i(x) \geq f_i(y)$ for all $i \in \{1, \ldots, m\}$. Hence,

$$\begin{array}{rrcl} & b_i(x) & \succeq & b_i(y) \\ \Leftrightarrow & \lfloor \log(f_i(x))/\log(1+\varepsilon) \rfloor & \succeq & \lfloor \log(f_i(y))/\log(1+\varepsilon) \rfloor \\ \Rightarrow & \log(f_i(x))/\log(1+\varepsilon) & \succeq & \log(f_i(y))/\log(1+\varepsilon) - 1 \\ \Leftrightarrow & 1 + \log(f_i(x))/\log(1+\varepsilon) & \succeq & \log(f_i(y))/\log(1+\varepsilon) \\ \Leftrightarrow & \log(1+\varepsilon) + \log(f_i(x)) & \succeq & \log(f_i(y)) \\ \Leftrightarrow & \log((1+\varepsilon)\cdot f_i(x)) & \succeq & \log(f_i(y)) \\ \Leftrightarrow & (1+\varepsilon)\cdot f_i(x) & \succeq & f_i(y) \end{array}$$

proves the lemma. □

The second important property is that the population size can be bounded in the following way. Denote by $F_i^{\max} := \max_{x \in \mathbb{B}^n} f_i(x)$ the largest value of the ith objective function f_i and by $F^{\max} := \max_{i \in \{1,\ldots,m\}} F_i^{\max}$ the largest value of the objective function f. Laumanns et al. [13] give the following upper bound on the population size.

Lemma 2. *The size of the population of Global DEMO$_\varepsilon$ is upper bounded by*

$$\left(\frac{\log F^{\max}}{\log(1+\varepsilon)} + 1 \right)^{k-1}.$$

3 Benefits

In this section, we examine in which situations the concept of ε-dominance leads provably to a better approximation behavior. The behavior of Global DEMO$_\varepsilon$ depends on the choice of ε. Our aim is to give a class of instances that is parameterized by ε. In particular, we present for each fixed choice of $\varepsilon > 0$ an instance where Global DEMO$_\varepsilon$ significantly outperforms Global SEMO.

W. l. o. g. we assume that n is even, i. e., each decision vector consists of an even number of bits. We denote the first half of a decision vector $x = (x_1, \ldots, x_n)$ by $x' = (x_1, \ldots, x_{n/2})$ and its second half by $x'' = (x_{n/2+1}, \ldots, x_n)$. 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 $\overline{x}$. In addition, we define the function

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

which interprets a bit-string x as the encoded natural number with respect to the binary numeral system.

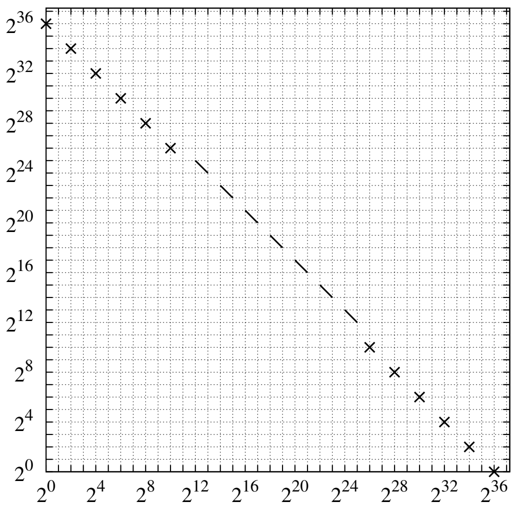

Figure 1: Outline of the Pareto front of LF_ε for $\varepsilon = 1$ and $n = 36$.

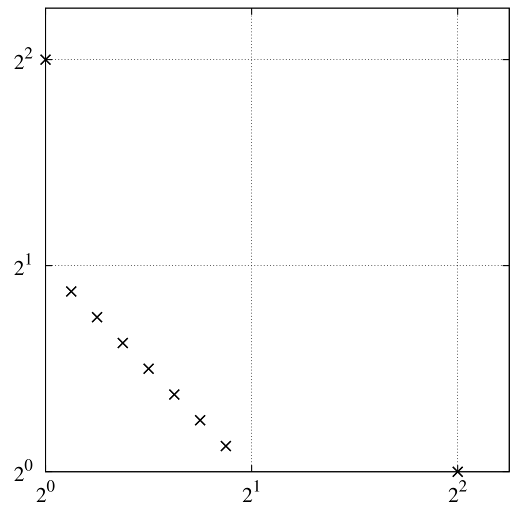

Figure 2: Pareto front of SF_ε for $\varepsilon = 1$ and $n = 8$.

To point out the benefits that can be gained when using the concept of ε-dominance we consider the function LF (large front) defined as $\mathrm{LF}_\varepsilon(x) = (\mathrm{LF}_{\varepsilon,1}(x), \mathrm{LF}_{\varepsilon,2}(x))$, where

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

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

The Pareto set includes all decision vectors, since $\mathrm{LF}_{\varepsilon,1}$ and $\mathrm{LF}_{\varepsilon,2}$ behave complementarily. An outline of the Pareto front of LF_ε for $\varepsilon = 1$ and $n = 36$ is shown in Figure 1.

Let x and y be two decision vectors with $|x'|_1 = |y'|_1$. Then $b(x) = b(y) = (2\cdot|x'|_1, 2\cdot|x'|_0)$ and therefore $x \succeq_\varepsilon y$ and $y \succeq_\varepsilon x$ due to Lemma 1. Hence, to achieve an ε-Pareto set it is sufficient to obtain for each $k \in \{0, \ldots, n/2\}$ a decision vector x with $|x'|_1 = k$. On the other hand, let x and y be two decision vectors with $|x'|_1 \neq |y'|_1$. Then either

$$(1+\varepsilon)\cdot \mathrm{LF}_{\varepsilon,1}(x) < \mathrm{LF}_{\varepsilon,1}(y) \quad \text{and} \quad (1+\varepsilon)\cdot \mathrm{LF}_{\varepsilon,2}(y) < \mathrm{LF}_{\varepsilon,2}(x)$$

or

$$(1+\varepsilon)\cdot \mathrm{LF}_{\varepsilon,2}(x) < \mathrm{LF}_{\varepsilon,2}(y) \quad \text{and} \quad (1+\varepsilon)\cdot \mathrm{LF}_{\varepsilon,1}(y) < \mathrm{LF}_{\varepsilon,1}(x)$$

and therefore $x \not\succeq_\varepsilon y$ and $y \not\succeq_\varepsilon x$. Hence, to achieve an ε-Pareto set it is also necessary to obtain a decision vector x with $|x'|_1 = k$ for each $k \in \{0, \ldots, n/2\}$.

First, we consider Global SEMO and show that this algorithm is unable to achieve an ε-Pareto front of LF_ε within a polynomial number of steps. The basic idea is to show that its population quickly grows before it obtains decision vectors with a large or small number of ones in the first half of the bit-string. To show this behavior we utilize the following results on the number of Hamming neighbors for a given set of elements of the Boolean hypercube.

The Boolean hypercube of dimension $n \in \mathbb{N}$ is defined as the undirected graph $G = (V, E)$ with $V = \mathbb{B}^n$ and $E = \{(v, w) \in V^2 \mid H(v, w) = 1\}$, where $H(v, w)$ denotes the Hamming distance of $v = (v_1, \ldots, v_n) \in \mathbb{B}^n$ and $w = (w_1, \ldots, w_n) \in \mathbb{B}^n$, i. e., $H(v, w) = \sum_{i=1}^{n} |v_i - w_i|$. A cut (S, T) is a partition of the vertices V of a graph $G = (V, E)$ into two sets S and T and the size $s(S, T)$ of a cut (S, T) is defined as the total number of edges crossing the cut, i. e., $s(S, T) = |\{(s, t) \in E \mid s \in S \wedge t \in T\}|$. Furthermore, we denote the number of positive bits of the representation of a non-negative integer i according to the binary numeral system by $h(i)$. The following statements lower bound the size of particular cuts in the Boolean hypercube.

Lemma 3 (Hart [11]). *Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$ with $0 < k \leq 2^n$, then*

$$\min \left\{ s(S, \mathbb{B}^n \setminus S) \mid S \subseteq \mathbb{B}^n, |S| = k \right\} = n \cdot k - 2 \cdot \sum_{i=0}^{k-1} h(i).$$

Lemma 4 (Hart [11]). *Let $r \in \mathbb{N}$, then*

$$\sum_{i=0}^{2^r - 1} h(i) = r \cdot 2^{r-1}.$$

Lemma 3 and Lemma 4 can be used to bound the expected number of undiscovered Hamming neighbors of a randomly chosen individual $x \in P \subseteq \mathbb{B}^n$.

Corollary 1. *Let $n \in \mathbb{N}$, $r \in \mathbb{N}$ with $0 \leq r \leq n$, and $P \subseteq \mathbb{B}^n$ with $0 < |P| \leq 2^r$, choose $x \in P$ uniformly at random, and observe the random variable X measuring the number of Hamming neighbors of x not contained in P, then*

$$\mathrm{E}(X) \geq n - 2 \cdot r.$$

Proof. The expected number of Hamming neighbors not contained in P is related to the size of the cut $(P, \mathbb{B}^n \setminus P)$, since $\mathrm{E}(X) = s(P, \mathbb{B}^n \setminus P)/|P|$. Let $m \in \mathbb{N}$ with $2^{m-1} < |P| \leq 2^m$. Due to Lemma 3 and Lemma 4

$$\begin{aligned} \frac{s(P, \mathbb{B}^n \setminus P)}{|P|} &\geq \frac{n \cdot |P| - 2 \cdot \sum_{i=0}^{|P|-1} h(i)}{|P|} \\ &\geq n - \frac{2 \cdot \sum_{i=0}^{2^m - 1} h(i)}{2^{m-1}} \\ &= n - \frac{2 \cdot m \cdot 2^{m-1}}{2^{m-1}} \geq n - 2 \cdot r. \end{aligned}$$

□

Using Corollary 1 we are able to show that Global SEMO needs with high probability an exponential number of iterations to achieve an ε-Pareto set.

Theorem 1. *The time until Global SEMO has achieved an ε-Pareto set of LF_ε is lower bounded by $2^{\Omega(n^{1/4})}$ with probability $1 - 2^{-\Omega(n^{1/4})}$.*

Proof. To find an ε-Pareto set of LF_ε Global SEMO has to produce for each $k \in \{0, \ldots, n/2\}$ at least one decision vector x with $|x'|_1 = k$. In the following, we lower bound the time needed to produce a decision vector x with $|x'|_1 = n/2$. To achieve this goal we follow a typical run of the algorithm. Whenever we upper bound the probability of the occurence of an unlikely event with e_i we work in the following under the assumption that this event has not occured.

Due to Chernoff bounds $|x'|_1 > 3n/8$ holds for the initial individual x with probability

$$e_1 := 2^{-\Omega(n)}.$$

As the probability that a mutation flips at least i bits is upper bounded by

$$\binom{n}{i} \cdot \left(\frac{1}{n}\right)^i \leq \left(\frac{en}{i}\right)^i \cdot \left(\frac{1}{n}\right)^i = \left(\frac{e}{i}\right)^i,$$

the probability that it flips more than $n^{1/4}$ bits is at most $(e/n^{1/4})^{n^{1/4}} = 2^{-\Omega(n^{1/4} \log n)}$. Hence, the probability that one of the first $2^{n^{1/4}}$ mutations flips more than $n^{1/4}$ bits is upper bounded by

$$e_2 := 2^{n^{1/4}} \cdot 2^{-\Omega(n^{1/4} \log n)} = 2^{-\Omega(n^{1/4} \log n)}.$$

In the following, we limit our considerations to the first $2^{n^{1/4}}$ steps of the algorithm.

We wait until Global SEMO creates for the first time an individual x with $|x'|_1 \geq n/2 - n^{1/2} + n^{1/4}$. Since at most $n^{1/4}$ bits flip in a mutation, also $|x'|_1 < n/2 - n^{1/2} + 2n^{1/4}$. We call the creation of an individual x with $|x'|_1 > \max_{y \in P} |y'|_1$ an *improvement*. There are at least

$$\frac{n^{1/2} - 2n^{1/4} + 1}{n^{1/4}} = n^{1/4} - 2 + n^{-1/4} \geq \frac{n^{1/4}}{2}$$

improvements needed to reach an individual x with $|x'|_1 = n/2$.

We divide the following part of a run in phases, where the ith phase ends when an improvement is achieved. We show that the completion of these phases takes an exponential number of steps with high probability. An improvement requires that an individual x with $|x'|_1 + n^{1/4} > \max_{y \in P} |y'|_1$ is chosen for mutation and at least one 0-bit in its first half is flipped. Denote the size of the population at the beginning of the ith phase by $|P_i|$. As the population contains at most $n^{1/4}$ individuals x with $|x'|_1 + n^{1/4} > \max_{y \in P} |y'|_1$ and each of these individuals contains less than $n^{1/2}$ 0-bits in its first half, the probability of an improvement in a step in the ith phase is upper bounded by $n^{1/4}/|P_i| \cdot n^{1/2}/n \leq 1/|P_i|$. The probability that the ith phase takes at least $|P_i|$ steps is lower bounded by $(1-1/|P_i|)^{|P_i|-1} \geq 1/e$. Due to Chernoff bounds the probability that less than $1/2 \cdot n^{1/4}/2 \cdot 1/e = n^{1/4}/(4e)$ phases take at least $|P_i|$ steps is upper bounded by

$$e_3 := 2^{-\Omega(n^{1/4})}.$$

We call theses $\ell \geq n^{1/4}/(4e)$ phases *long phases* and denote their indices by $i_1, \ldots, i_\ell$.

The population grows if (1) an individual x with $n^{1/2} \leq |x'|_1 \leq n/2 - n^{1/2}$ is selected and (2) afterwards exactly one bit in its second half is flipped such that a new individual $y \notin P$ emerges. Since the population contains at least

$$\frac{n/2 - n^{1/2} - 3n/8}{n^{1/4}} = \frac{n^{3/4}}{8} - n^{1/4} \geq \frac{n^{3/4}}{16}$$

individuals x with $3n/8 < |x'|_1 \leq n/2 - n^{1/2}$ and at most $2n^{1/2}$ individuals x with $|x'|_1 <$

$n^{1/2}$ or $|x'|_1 > n/2 - n^{1/2}$, the first probability is lower bounded by

$$1 - \frac{2n^{1/2}}{n^{3/4}/16} = 1 - 32n^{-1/4} \geq 1/2.$$

To lower bound the second probability we utilize the following reverse version of Markov's inequality which applies to bounded random variables. This version can be easily derived from Markov's inequality. Let X be a non-negative random variable. If there are $a, b \in \mathbb{R}_0^+$ such that $\text{Prob}(X \leq a) = 1$ and $b < \text{E}(X)$, then $\text{Prob}(X > b) \geq (\text{E}(X) - b)/(a - b)$. Let X be the random variable measuring the number of individuals $y \notin P$ which differ from an uniformly at random chosen individual $x \in P$ in exactly one bit in their second half. As the population size is upper bounded by $2^{n/8}$, we can use Corollary 1 with $r = n/8$ to deduce that $\text{E}(X) \geq n/2 - 2 \cdot n/8 = n/4$. Using the inequality from above with $a = n/2$ and $b = \text{E}(X)/2$, we conclude that

$$\text{Prob}\left(X \geq \frac{n}{8}\right) \geq \frac{\text{E}(X) - \text{E}(X)/2}{n/2 - \text{E}(X)/2} \geq \frac{1}{3}.$$

Therefore, the second probability is lower bounded by $1/3 \cdot n/8 \cdot 1/n \cdot (1 - 1/n)^{n-1} \geq 1/(24e)$. Altogether, the probability that the population grows is lower bounded by $1/(48e)$. Due to Chernoff bounds the probability that in a long phase the population grows by less than $|P_i|/(96e) \geq |P_i|/261$ individuals is upper bounded by $2^{-\Omega(|P_i|)} = 2^{-\Omega(n^{3/4})}$. The probability that this event happens in at least one long phase is upper bounded by

$$e_4 := \ell \cdot 2^{-\Omega(n^{3/4})} \leq n^{1/2} \cdot 2^{-\Omega(n^{3/4})} = 2^{-\Omega(n^{3/4})}.$$

The inequalities

$$|P_{i_1}| \geq \frac{1}{16} \cdot n^{3/4} \quad \text{and} \quad |P_{i_{j+1}}| \geq |P_{i_j}| + \frac{1}{261} \cdot |P_{i_j}|.$$

lower bound the growth of the population in the long phases. By solving the recursive inequalities we get

$$|P_{i_j}| \geq \left(\frac{262}{261}\right)^{j-1} \cdot \frac{1}{16} \cdot n^{3/4}.$$

Since the error probabilities $e_1, \ldots, e_4$ sum up to $2^{-\Omega(n^{1/4})}$, we conclude that the last long phase takes a least

$$\left(\frac{262}{261}\right)^{n^{1/4}/(4e)-1} \cdot \frac{1}{16} \cdot n^{3/4} = 2^{\Omega(n^{1/4})}$$

steps with probability $1 - 2^{-\Omega(n^{1/4})}$ which shows the theorem. □

We emphasize that the exponentially small success probability of Global SEMO on LF_ε implies that not even sequential or parallel runs of Global SEMO help to find an ε-Pareto set of LF_ε in polynomial time. If we observe at most $2^{cn^{1/4}}$ runs of Global SEMO on LF_ε and grant each run at most $2^{c'n^{1/4}}$ steps, where $c > 0$ and $c' > 0$ are two sufficiently small constants, then the probability that at least one run finds an ε-Pareto set of LF_ε is still at most $2^{-\Omega(n^{1/4})}$.

In the following, we show that Global DEMO$_\varepsilon$ is efficient on LF$_\varepsilon$. The advantage of Global DEMO$_\varepsilon$ is that it always works with a population that is much smaller than the population of Global SEMO.

Theorem 2. *The expected time until Global DEMO$_\varepsilon$ has achieved an ε-Pareto set of* LF$_\varepsilon$ *is $\mathcal{O}(n^2 \log n)$.*

Proof. Denote the set of covered $|\cdot|_1$ values by $A := \{|x'|_1 \mid x \in P\}$ and the set of uncovered $|\cdot|_1$ values by $B := \{0, \dots, n/2\} \setminus A$. The population P includes for each non-dominated feasible objective vector a corresponding decision vector precisely if $A = \{0, \dots, n/2\}$.

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 $b = a+1$. In addition, $|P|$ is upper bounded by

$$\left(\frac{\log F^{\max}}{\log(1+\varepsilon)} + 1\right)^{k-1} < \left(\frac{\log(1+\varepsilon)^{2n+1}}{\log(1+\varepsilon)} + 1\right)^{2-1} = 2n+2$$

due to Lemma 2. Let $x \in P$ be the individual with $|x'|_1 = a$. The probability to choose x in the next step and flip exactly one proper bit to obtain a decision vector y with $|y'|_1 = b$ is at least

$$\frac{1}{|P|} \cdot \frac{\min\{b+1, n/2-b+1\}}{en} \geq \frac{\min\{b, n/2-b\}+1}{3en^2}.$$

To obtain the upper bound, we sum up over the different values that b can attain. Therefore, the expected optimization time is upper bounded by

$$3en^2 \cdot \sum_{b=0}^{n/2} \frac{1}{\min\{b, n/2-b\}+1} \leq 6en^2 \cdot \sum_{b=1}^{n/4+1} \frac{1}{b} = \mathcal{O}(n^2 \log n)$$

which completes the proof. □

4 Drawbacks

In the previous section, we have shown that the concept of ε-dominance can help to achieve a good approximation of the Pareto set. A drawback of the mentioned approach is that it can prevent an algorithm from proceeding from a non-dominated feasible objective vector to another if there is a large number of such objective vectors within a single box. We present a class of problems parameterized by ε where Global DEMO$_\varepsilon$ is not efficient. In contrast to this Global SEMO can obtain the whole Pareto front in expected polynomial time independently of the choice of ε. Again, all decision vectors are non-dominated and it is easy to obtain a new objective vector by flipping a single bit in at least one decision vector of the current population. In contrast to the example class analyzed in the previous section the number of objective vectors is polynomially bounded.

To point out the drawbacks that the use of ε-dominance might have, we consider the two-objective function SF (small front) defined as $\mathrm{SF}_\varepsilon(x) = (\mathrm{SF}_{\varepsilon,1}(x), \mathrm{SF}_{\varepsilon,2}(x))$, where

$$\mathrm{SF}_{\varepsilon,1}(x) := (1+\varepsilon)^{|x|_1/n + \lfloor |x|_1/n \rfloor},$$
$$\mathrm{SF}_{\varepsilon,2}(x) := (1+\varepsilon)^{|x|_0/n + \lfloor |x|_0/n \rfloor}.$$

It is obvious that the Pareto set includes all decision vectors. The Pareto front of SF_ε is shown in Figure 2 for $\varepsilon = 1$ and $n = 8$.

For each decision vector $x \notin \{0^n, 1^n\}$ holds

$$(1+\varepsilon) \cdot \mathrm{SF}_{\varepsilon,1}(0^n) \geq \mathrm{SF}_{\varepsilon,1}(x) \quad \text{and} \quad \mathrm{SF}_{\varepsilon,2}(0^n) \geq \mathrm{SF}_{\varepsilon,2}(x)$$

and

$$\mathrm{SF}_{\varepsilon,1}(1^n) \geq \mathrm{SF}_{\varepsilon,1}(x) \quad \text{and} \quad (1+\varepsilon) \cdot \mathrm{SF}_{\varepsilon,2}(1^n) \geq \mathrm{SF}_{\varepsilon,2}(x).$$

To compute an ε-Pareto set it therefore suffices to reach a population that includes the decision vectors 0^n and 1^n. These decision vectors are in turn included in each ε-Pareto set as $(1+\varepsilon) \cdot \mathrm{SF}_{\varepsilon,1}(x) < \mathrm{SF}_{\varepsilon,1}(1^n)$ for each $x \neq 1^n$ and $(1+\varepsilon) \cdot \mathrm{SF}_{\varepsilon,2}(x) < \mathrm{SF}_{\varepsilon,2}(0^n)$ for each $x \neq 0^n$.

We show that Global $\mathrm{DEMO}_\varepsilon$ is not able to efficiently find an ε-Pareto set of SF_ε. To achieve our goal we resort to a powerful theory called drift analysis which is able to deduce properties of a sequence of random variables from its drift [10]. Drift analysis has been introduced into the runtime analysis of evolutionary algorithms by [12]. The following theorem is stated in [18].

Theorem 3 (Drift Theorem)**.** *Let X_t, $t \geq 0$, be the random variables describing a Markov process over a state space S and $g\colon S \to \mathbb{R}_0^+$ a function mapping each state to a non-negative real number. Pick two real numbers $a(n)$ and $b(n)$ depending on a parameter n such that $0 \leq a(n) < b(n)$. Let $T(n) := \min\{t \geq 0 \mid g(X_t) \leq a\}$. If there are $\lambda(n) > 0$, $D(n) \geq 1$, and $p(n) > 0$ such that for all $t \geq 0$ the conditions*

1. $\mathrm{E}\left(e^{-\lambda(n)\cdot(g(X_{t+1})-g(X_t))} \mid a(n) < g(X_t) < b(n)\right) \leq 1 - 1/p(n)$,
2. $\mathrm{E}\left(e^{-\lambda(n)\cdot(g(X_{t+1})-b(n))} \mid b(n) \leq g(X_t)\right) \leq D(n)$

hold, then for all time bounds $B(n) \geq 0$

$$\mathrm{Prob}\big(T(n) \leq B(n) \mid g(X_0) \geq b(n)\big) \leq e^{\lambda(n)\cdot(a(n)-b(n))} \cdot B(n) \cdot D(n) \cdot p(n).$$

Using Theorem 3 we are able to show that Global $\mathrm{DEMO}_\varepsilon$ needs with high probability an exponential number of steps to achieve an ε-Pareto set.

Theorem 4. *The time until Global $\mathrm{DEMO}_\varepsilon$ has achieved an ε-Pareto set of SF_ε is lower bounded by $2^{\Omega(n)}$ with probability $1 - 2^{-\Omega(n)}$.*

Proof. We interpret a run of Global $\mathrm{DEMO}_\varepsilon$ on SF_ε as a Markov process, i. e., the state space corresponds to the set of all possible populations. Our aim is to derive a lower bound on the runtime until a decision vector of $\{0^n, 1^n\}$ has been obtained. To utilize Theorem 3, we define $g(X_t) := \min\{\min\{|x|_1, |x|_0\} \mid x \in X_t\}$ and choose $a(n) := 0$, $b(n) := n/128$, $\lambda(n) := \lambda := \ln 4$, $D(n) := D := 2^{14}$, and $p(n) := p := 25$. Note, that X_t contains a single decision vector as long as $g(X_t) > 0$. Observe the population $X_t = \{x\}$ and set $i := g(X_t) = \min\{|x|_1, |x|_0\}$. The probability p_{-j} to decrease i by j, $1 \leq j \leq i$, is at most

$$p_{-j} \leq \binom{i}{j} \cdot \left(\frac{1}{n}\right)^j + \binom{n-i}{n-2i+j} \cdot \left(\frac{1}{n}\right)^{n-2i+j} \leq 2 \cdot \binom{i}{j} \cdot \left(\frac{1}{n}\right)^j \leq 2 \cdot \left(\frac{ei}{j}\right)^j \cdot \left(\frac{1}{n}\right)^j,$$

since it is required that at least j ones or $n-2i+j$ zeros flip if $|x|_1 \leq n/2$ or at least j zeros or $n-2i+j$ ones flip if $|x|_1 \geq n/2$. The probability p_j to increase i by j, $1 \leq j \leq n/2 - i$, is at least

$$p_j \geq \binom{n-i}{j} \cdot \left(\frac{1}{n}\right)^j \cdot \left(1-\frac{1}{n}\right)^{n-j} \geq \left(\frac{n-i}{j}\right)^j \cdot \left(\frac{1}{n}\right)^j \cdot \frac{1}{e},$$

since it is sufficient that exactly j zeros flip if $|x|_1 \leq n/2$ or exactly j ones flip if $|x|_1 \geq n/2$. We show that the two statements given in Theorem 3 are satisfied.

1. If $0 < i < n/128$, then $p_{-j} \leq 2 \cdot (e/(128j))^j$ and $p_j \geq 1/e \cdot (127/(128j))^j$. Hence, for all $t \geq 0$

$$\begin{aligned}
&\mathrm{E}\left(e^{-\lambda \cdot (g(X_{t+1})-g(X_t))} \mid a(n) < g(X_t) < b(n)\right) \\
&= \sum_{j=1}^{i} p_{-j} \cdot e^{\lambda \cdot j} + \left(1 - \sum_{j=1}^{i} p_{-j} - \sum_{j=1}^{n/2-i} p_j\right) + \sum_{j=1}^{n/2-i} p_j \cdot e^{-\lambda \cdot j} \\
&\leq \sum_{j=1}^{i} 2 \cdot \left(\frac{e}{128j}\right)^j \cdot e^{\ln 4 \cdot j} + 1 - \sum_{j=1}^{i} 2 \cdot \left(\frac{e}{128j}\right)^j - \sum_{j=1}^{n/2-i} \frac{1}{e} \cdot \left(\frac{127}{128j}\right)^j \\
&\quad + \sum_{j=1}^{n/2-i} \frac{1}{e} \cdot \left(\frac{127}{128j}\right)^j \cdot e^{-\ln 4 \cdot j},
\end{aligned}$$

since $e^{\lambda \cdot j} > 1 > e^{-\lambda \cdot j}$ for all $j \geq 1$. The last expression can be upper bounded by

$$\begin{aligned}
&2 \cdot \sum_{j=1}^{\infty} \left(\frac{e}{32}\right)^j + 1 - \frac{127}{128e} + \frac{1}{e} \cdot \sum_{j=1}^{\infty} \left(\frac{127}{512}\right)^j \\
&= 2 \cdot \left(\frac{1}{1-e/32} - 1\right) + 1 - \frac{127}{128e} + \frac{1}{e} \cdot \left(\frac{1}{1-127/512} - 1\right) \\
&= \frac{2e}{32-e} + 1 - \frac{127}{128e} + \frac{127}{385e} \leq \frac{19}{100} + 1 - \frac{36}{100} + \frac{13}{100} = 1 - \frac{1}{p}.
\end{aligned}$$

2. If $n/128 \leq i \leq n/2$, then $p_{-j} \leq 2 \cdot (e/(2j))^j$ and $p_j \geq 0$. Hence, for all $t \geq 0$

$$\begin{aligned}
&\mathrm{E}\left(e^{-\lambda \cdot (g(X_{t+1})-b(n))} \mid b(n) \leq g(X_t)\right) \\
&\leq \mathrm{E}\left(e^{-\lambda \cdot (g(X_{t+1})-g(X_t))} \mid b(n) \leq g(X_t)\right) \\
&= \sum_{j=1}^{i} p_{-j} \cdot e^{\lambda \cdot j} + \left(1 - \sum_{j=1}^{i} p_{-j} - \sum_{j=1}^{n/2-i} p_j\right) + \sum_{j=1}^{n/2-i} p_j \cdot e^{-\lambda \cdot j} \\
&\leq \sum_{j=1}^{i} 2 \cdot \left(\frac{e}{2j}\right)^j \cdot e^{\ln 4 \cdot j} + 1 - \sum_{j=1}^{i} 2 \cdot \left(\frac{e}{2j}\right)^j,
\end{aligned}$$

since $e^{\lambda \cdot j} > 1 > e^{-\lambda \cdot j}$ for all $j \geq 1$. The last expression can be upper bounded by

$$2 \cdot \left(\sum_{j=1}^{5} \left(\frac{2e}{j} \right)^j + \sum_{j=6}^{i} \left(\frac{2e}{j} \right)^j \right) + 1 \leq 2 \cdot \left(\sum_{j=0}^{5} (2e)^j + \sum_{j=0}^{\infty} \left(\frac{e}{3} \right)^j \right) + 1$$
$$= 2 \cdot \left(\frac{1-(2e)^6}{1-2e} + \frac{1}{1-e/3} \right) + 1 \leq 2^{14} = D.$$

Since $n/128 \leq |x|_1 \leq 127n/128$ holds for the initial individual $x \in \mathbb{B}^n$ with probability $1-2^{-\Omega(n)}$ due to Chernoff bounds, an application of Theorem 3 yields that a decision vector of $\{0^n, 1^n\}$ has been found within $B(n) := 2^{n/2^7} = 2^{\Omega(n)}$ steps with probability at most

$$e^{\lambda \cdot (a(n)-b(n))} \cdot B(n) \cdot D \cdot p = e^{\ln 4 \cdot (-n/2^7)} \cdot 2^{n/2^7} \cdot 2^{14} \cdot 25 = 2^{-\Omega(n)}. \qquad \square$$

Just recently a simplified version of Theorem 3 has been proposed [18] that leads to an easier proof of Theorem 4.

Theorem 5 (Simplified Drift Theorem)**.** *Let X_t, $t \geq 0$, be the random variables describing a Markov process over a state space S and $g\colon S \to \{0, \ldots, N\}$, $N \in \mathbb{N}$, a function mapping each state to a natural number between 0 and N. Pick two real numbers $a(n)$ and $b(n)$ depending on a parameter n such that $0 \leq a(n) < b(n)$. Let $T(n) := \min\{t \geq 0 \mid g(X_t) \leq a(n)\}$ and $\Delta_t(i) := (g(X_{t+1}) - g(X_t) \mid g(X_t) = i)$ for all $t \geq 0$ and $0 \leq i \leq N$. If there are constants $\delta, \varepsilon, r > 0$ such that for all $t \geq 0$ the conditions*

1. $\mathrm{E}\left(\Delta_t(i)\right) \geq \varepsilon$ *for all* $a(n) < i < b(n)$
2. $\mathrm{Prob}\big(\Delta_t(i) = -j\big) \leq 1/(1+\delta)^{j-r}$ *for all* $i > a(n)$ *and* $j \geq 1$

hold, then there is a constant $c^* > 0$ such that

$$\mathrm{Prob}\big(T(n) \leq 2^{c^*(b(n)-a(n))} \mid g(X_0) \geq b(n)\big) = 2^{-\Omega(b(n)-a(n))}.$$

Alternate proof of Theorem 4. We interpret a run of Global DEMO$_\varepsilon$ on SF$_\varepsilon$ as a Markov process, i. e., the state space corresponds to the set of all possible populations. To utilize Theorem 5, we define $g(X_t) := \min\{\min\{|x|_1, |x|_0\} \mid x \in X_t\}$ and choose $a(n) := 0$, $b(n) := n/20$, $\varepsilon := 3/100$, $\delta := 1$, and $r := 2$. We show that the two statements given in Theorem 5 are satisfied.

1. For all $t \geq 0$ and $a(n) < i < b(n)$ holds

$$\begin{aligned}
\mathrm{E}(\Delta_t(i)) &= \sum_{j=1}^{i} \mathrm{Prob}(\Delta_t(i) = -j) \cdot (-j) + \sum_{j=1}^{n/2-i} \mathrm{Prob}(\Delta_t(i) = j) \cdot j \\
&\geq -\sum_{j=1}^{i} 2 \cdot \left(\frac{e}{20j}\right)^j \cdot j + \sum_{j=1}^{n/2-i} \frac{1}{e} \cdot \left(\frac{19}{20j}\right)^j \cdot j \\
&= -\frac{e}{10} \cdot \sum_{j=1}^{i} \left(\frac{e}{20j}\right)^{j-1} + \frac{19}{20e} \cdot \sum_{j=1}^{n/2-i} \left(\frac{19}{20j}\right)^{j-1} \\
&\geq -\frac{e}{10} \cdot \sum_{j=0}^{\infty} \left(\frac{e}{20}\right)^j + \frac{19}{20e} \\
&= -\frac{e/10}{1-e/20} + \frac{19}{20e} \geq \frac{3}{100} = \varepsilon.
\end{aligned}$$

2. For all $t \geq 0$, $i > a(n)$, and $j \geq 1$ holds

$$\mathrm{Prob}(\Delta_t(i) = -j) \leq 2 \cdot \binom{n}{j} \cdot \left(\frac{1}{n}\right)^j \leq \frac{2}{j!} \leq \left(\frac{1}{2}\right)^{j-2} = \left(\frac{1}{1+\delta}\right)^{j-r}.$$

Since $n/20 \leq |x|_1 \leq 19n/20$ holds for the initial individual $x \in \mathbb{B}^n$ with probability $1 - 2^{-\Omega(n)}$ due to Chernoff bounds, an application of Theorem 5 yields that a decision vector of $\{0^n, 1^n\}$ has been found within $2^{c^* \cdot (b(n)-a(n))} = 2^{\Omega(n)}$ steps with probability at most $2^{-\Omega(b(n)-a(n))} = 2^{-\Omega(n)}$. □

We note that the exponentially small success probability of Global DEMO$_\varepsilon$ on SF$_\varepsilon$ implies that restarts of the algorithm are not useful to significantly reduce the required time to find an ε-Pareto set of SF$_\varepsilon$.

In contrast to the previous result, Global SEMO is able to compute the whole Pareto front in expected polynomial time independently of the choice of ε as shown in the following theorem.

Theorem 6. *The expected time until Global SEMO has achieved an ε-Pareto set of* SF$_\varepsilon$ *is* $\mathcal{O}(n^2 \log n)$.

Proof. Denote the set of covered $|\cdot|_1$ values by $A := \{|x|_1 \mid x \in P\}$ and the set of uncovered $|\cdot|_1$ values by $B := \{0, \ldots, n\} \setminus A$. The population P includes for each non-dominated feasible objective vector a corresponding decision vector precisely if $A = \{0, \ldots, n\}$.

As long as $A \neq \{0, \ldots, n\}$, there exists an $a \in A$ and a $b \in B$ with $b = a - 1$ or $b = a + 1$. In addition, $|P|$ is upper bounded by n. Let $x \in P$ be the individual with $|x|_1 = a$. The probability to choose x in the next step and flip exactly one proper bit to obtain a decision vector y with $|y|_1 = b$ is at least

$$\frac{1}{|P|} \cdot \frac{\min\{b+1, n-b+1\}}{en} \geq \frac{\min\{b, n-b\}+1}{en^2}.$$

To obtain the upper bound, we sum up over the different values that b can attain. Therefore, the expected optimization time is upper bounded by

$$en^2 \cdot \sum_{b=0}^{n} \frac{1}{\min\{b, n-b\}+1} \leq 2en^2 \cdot \sum_{b=1}^{n/2+1} \frac{1}{b} = \mathcal{O}(n^2 \log n). \qquad \square$$

The properties of the class of functions analyzed in this section also hold for the bi-objective model of the minimum spanning tree (MST) problem examined in [17]. In the mentioned bi-objective model the task is to minimize the weight and the number of connected components of a set of edges simultaneously. For Global SEMO a polynomial upper bound has been given in [17]. In fact, it is not to hard to come up with a class of instances for the MST problem where Global DEMO$_\varepsilon$ needs with high probability an exponential number of steps to find a minimum spanning tree.

5 Conclusions

Diversity plays an important role when using evolutionary algorithms for multi-objective optimization. This is in particular the case when the number of feasible objectives vectors grows exponentially in the problem size. We have investigated the concept of ε-dominance within simple MOEAs. Our theoretical investigations point out situations where this concept significantly helps to reduce the runtime of a MOEA until a good approximation of the Pareto front has been achieved. We have also pointed out that the concept of ε-dominance can be destructive for problems where the number of feasible objective vectors is small. In this case, the investigated mechanism of diversity can slow down the optimization process drastically.

In the future, it would be interesting to examine how the use of ε-dominance can help to achieve good approximations for classical multi-objective combinatorial optimization problems. Many of these problems can have a Pareto front of exponential size and using the mechanism of ε-dominance might help to reduce the runtime until a good approximation of the Pareto set has been achieved.

References

[1] D. Brockhoff, T. Friedrich, N. Hebbinghaus, C. Klein, F. Neumann, and E. Zitzler. Do additional objectives make a problem harder? In D. Thierens, editor, *Proc. of GECCO 2007*, pages 765–772, London, UK, 2007. ACM.

[2] C. A. Coello Coello. Recent Trends in Evolutionary Multiobjective Optimization. In *Evolutionary Multiobjective Optimization: Theoretical Advances And Applications*, pages 7–32. Springer, London, 2005.

[3] 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.

[4] K. Deb. *Multi-objective optimization using evolutionary algorithms.* Wiley, Chichester, UK, 2001.

[5] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. In *Proc. of PPSN VI*, volume 1917 of *LNCS*, pages 849–858. Springer, 2000.

[6] M. Ehrgott. *Multicriteria Optimization.* Springer, Berlin, Germany, 2nd edition, 2005.

[7] T. Friedrich, J. He, N. Hebbinghaus, F. Neumann, and C. Witt. Approximating covering problems by randomized search heuristics using multi-objective models. In *Proc. of GECCO 2007*, pages 797–804, London, UK, 2007. ACM.

[8] O. Giel. Expected runtimes of a simple multi-objective evolutionary algorithm. In *Proc. of CEC 2003*, IEEE Press, pages 1918–1925, 2003.

[9] O. Giel and P. K. Lehre. On the effect of populations in evolutionary multi-objective optimization. In *Proc. of GECCO 2006*, pages 651–658. ACM Press, 2006.

[10] B. Hajek. Hitting-time and occupation-time bounds implied by drift analysis with applications. *Advances in Applied Probability*, 14:502–525, 1982.

[11] S. Hart. A note on the edges of the n-cube. *Discrete Mathematics*, 14:157–163, 1976.

[12] J. He and X. Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 127:57–85, 2001.

[13] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10(3):263–282, 2003.

[14] M. Laumanns, L. Thiele, and E. Zitzler. Running time analysis of evolutionary algorithms on a simplified multiobjective knapsack problem. *Natural Computing*, 3(1):37–51, 2004.

[15] M. Laumanns, L. Thiele, and E. Zitzler. Running time analysis of multiobjective evolutionary algorithms on pseudo-boolean functions. *IEEE Trans. Evolutionary Computation*, 8(2):170–182, 2004.

[16] F. Neumann. Expected runtimes of a simple evolutionary algorithm for the multi-objective minimum spanning tree problem. *European Journal of Operational Research*, 181(3):1620–1629, 2007.

[17] F. Neumann and I. Wegener. Minimum spanning trees made easier via multi-objective optimization. *Natural Computing*, 5(3):305–319, 2006.

[18] P. S. Oliveto and C. Witt. Simplified drift analysis for proving lower bounds in evolutionary computation. Technical Report CI-247/08, Technische Universität Dortmund, 2008.

[19] E. Zitzler and S. Künzli. Indicator-based selection in multiobjective search. In *Proc. of PPSN VIII*, volume 3242 of *LNCS*, pages 832–842. Springer, 2004.

[20] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multiobjective Optimization. In *Proc. of EUROGEN 2001*, pages 95–100. International Center for Numerical Methods in Engineering (CIMNE), 2002.